Biomarker discovery for EBV-associated cancers and tuberculosis

Seekkuge Yomani Dilukshi Sarathkumara

BSc (Hons), MPhil

A thesis submitted for the degree of Doctor of Philosophy
at James Cook University

December 2022

Student Number: 13670138

Australian Institute of Tropical Health and Medicine (AITHM)
College of Public Health, Medical & Veterinary Science (CPHMVS)
James Cook University (JCU), Australia

Acknowledgements

This doctoral journey has been truly a roller coaster ride for me over the last few years. It was one of the most challenging things I have done in my life so far which taught me resilience and perseverance. Undertaking this PhD has been a truly life-changing experience and for me, and it would not have been possible to do without the support and guidance that I received from many people. I have been fortunate to be surrounded by several amazing people who made this journey a memorable one.

Firstly, I would like to express my sincere gratitude to my supervisory team for their continuous support, guidance, and mentorship on my study. First and foremost, to my primary advisor Prof. Denise Doolan, for her constant support, guidance, and encouragement throughout this journey. I could not have imagined having a better advisor and mentor for my PhD study. I would be always grateful to Denise for trusting my skills and abilities to drive this project from day one. I also want to thank my secondary advisors Dr. Carla Proietti and Dr. David Pattinson for their support throughout this study you have been always kind, supportive and understanding of my capabilities. I am grateful to Carla for introducing the intricacies of analyzing protein microarray data, helping with complex data analysis, and coding. Thank you, Dave and Carla, for simply being moral supporters and ensuring that I always had support both inside and outside the lab.

I thank all the present and past members of the Doolan lab. No one could ask for a better lab group in Cairns. Particularly Dr. Martha Cooper, Mr. Daniel Browne, Ms. Ashton Kelly, Ms. Jamie Brady, and Ms. Maggie Veitch for their assistance with performing experiments. I'm very grateful to Martha for introducing me to R when I had zero knowledge in coding using computational machine learning techniques analysis when I first started my PhD. Ashton and Dan for helping me with qPCR assays.

I'm indeed very thankful to our international collaborators Prof. Allan E. Hildesheim and Dr. Zhiwei Liu, from National Cancer Institute, USA and all others involved for their immense support for my studies on EBV providing me access to their precious biological samples. I am grateful to Dr. Anna E. Coghill and Dr. Nathan W. Van Bibber from Moffitt Cancer Center & Research Institute, USA and Dr. Cliona M Rooney from Baylor College of Medicine, USA for the opportunity for the collaboration and provide access to Phase I clinical trial samples on EBVST immunotherapy study.

I extend my gratitude to A/Prof. Jeffry Warner, A/Prof. Catherine Rush, Ms. Jessica Scott from CPHMVS, AITHM, Townsville and the staff of the Balimo District Hospital, PNG for their enormous efforts in collecting samples from TB patients and the community in a very challenging field trip in Balimo. I am extremely grateful to all those who participated in these studies for their invaluable contribution to this research.

I got through every day with the support, understanding and hugs of my lab mates who are also my closest friends in this. Ashton, Max, Camila, Maggie, Judith, and Julia thank you for your genuine friendships and for simply being there for me whenever I needed some help. Thank you for always encouraging me to push myself further than I would have alone.

Thank you AITHM, Cairns for the friendly and supportive work environment where I spent most of my last four years. I would like to acknowledge JCU and the AITHM for the scholarships and funding to enable me to complete this thesis. I am grateful for the financial support from the NHMRC grants awarded to my supervisor Prof. Denise Doolan.

To my best friend, Virajie akki for being one of the biggest supporters in my research carrier since my undergraduate days at Northumbria. Chanika akki, soon-to-be Dr., thank you for being a good friend and I am happy to share our PhD journeys together.

I wouldn’t be here without the support of our loving families in Sri Lanka. I am deeply grateful to my Ammi and Thaththi for their continuous support, encouragement at every step of the way and for their unconditional love even from so far away. Thank you for Nisal’s Ammi and Thaththi for being there for us always.

Having to do a full-time PhD during a global pandemic amidst the country's worst economic and political crisis in Sri Lanka wasn’t easy for me. So, this journey wouldn’t be possible without my husband Nisal who kept me going and picked me up when I couldn’t go on anymore and kept me going. Thank you for being so patient and for all your sacrifices for me, for so long. Without him and his constant support, love and care I would not be here.

I dedicate this doctoral thesis to
my loving husband for his constant support
and unconditional love.

මෙම ආචාර්ය උපාධි නිබන්ධනය
මගේ නිසල්ට ආදරයෙන් පිලිගන්වමි.

Statement of Contribution of Others

I declare that all persons who contributed to this thesis have been included as co-authors for published papers or are acknowledged below.

Nature of Assistance	Contribution	Names, Titles, and Affiliations
Supervision	Primary advisor	Prof. Denise Doolan, AITHM, JCU
	Secondary advisors	Dr. Carla Proietti, AITHM, JCU Dr. David Pattinson, AITHM, JCU
Intellectual Support	Project plan and development	Prof. Denise Doolan, AITHM, JCU Dr. Carla Proietti, AITHM, JCU
	Experimental Design	Prof. Denise Doolan, AITHM, JCU Dr. Carla Proietti, AITHM, JCU Dr. David Pattinson, AITHM, JCU A/Prof. Jeffry Warner, CPHMVS, AITHM, JCU A/Prof. Catherine Rush, CPHMVS, AITHM, JCU Mr. Daniel Browne, AITHM, JCU
	Data analysis	Dr. Carla Proietti, AITHM, JCU Dr. Martha Cooper, AITHM, JCU Mr. Daniel Browne, AITHM, JCU Dr. Zhiwei Liu, NCI, NIH Dr. Nathan W. Van Bibber (Moffitt Cancer Center & Research Institute)
	Editing of chapters and manuscripts	Prof. Denise Doolan, AITHM, JCU Dr. Carla Proietti, AITHM, JCU Prof. Allan E. Hildesheim NCI, NIH Dr. Zhiwei Liu, NCI, NIH
	Ethical approvals	Human Research Ethics Committee (JCU) Institutional Review Board for United States National Institutes of Health (NIH) National Cancer Institute (NCI) Institutional Review Board for Baylor College of Medicine and Affiliated Hospitals

		Medical Research Advisory Committee (MRAC) of Papua New Guinea
Financial Support	Scholarship and International student fee waiver	James Cook University International Postgraduate Research Scholarships (IPRS)
	Top up scholarship	Australian Institute of Tropical Health and Medicine (AITHM, JCU)
	Publication support grant	United States National Institutes of Health (NIH) Higher Degree Research Enhancement Scheme (HDRES) (CPHMVS, JCU)
	Write-up support	Doctoral Completion Grant (CPHMVS, JCU)
	Funding for scientific studies	Grants awarded to Prof. Denise Doolan by US government / NIH/NCI (contracts) and Moffitt Cancer Center (contract) and NHMRC (PRF #1137285 and Program Grant #1132975)
Data Collection	Natural Killer/ T-cell Lymphoma and classical Hodgkin Lymphoma study samples and clinical data	Prof. Allan E. Hildesheim NCI, NIH Dr. Zhiwei Liu, NCI, NIH Dr. Kelly J. Yu, NCI, NIH Dr. Nathaniel Rothman, NCI, NIH Dr. Qing Lan, NCI, NIH Prof. Richard F. Ambinder Dr. Rena R. Xian
	EBV-specific T-cells (EBVSTs) immunotherapy samples and clinical data	Dr. Cliona M Rooney (Baylor College of Medicine) Dr. Anna E. Coghill (Moffitt Cancer Center & Research Institute)
	Tuberculosis study samples and clinical data from Balimo, PNG	A/Prof. Jeffry Warner, CPHMVS, AITHM, JCU A/Prof. Catherine Rush, CPHMVS, AITHM, JCU Ms. Jessica Scott, CPHMVS, AITHM, JCU
Technical and administrative support	Experimental assistance	Ms. Ashton Kelly, AITHM, JCU Ms. Jessica Scott, CPHMVS, AITHM, JCU
	RNA-sequencing	Australian Genome Research Facility (AGRF)

	PC2 laboratory operations and administrative tasks	Ms. Jamie Brady AITHM, JCU

Research outputs

Direct publications from PhD thesis (published, submitted, or pending submission)

1. Zhiwei Liu (ZL)[*], **Yomani D. Sarathkumara**[*] (YS), John K.C. Chan, Yok-Lam Kwong, Tai Hing Lam, Dennis Kai Ming Ip, Brian C.-H. Chiu , Jun Xu, Yu-Chieh Su, Carla Proietti (CP), Martha M. Cooper (MC), Kelly J. Yu1, Bryan Bassi, Raymond Liang, Wei Hu, Bu-Tian Ji, Anna E. Coghill, Ruth M. Pfeiffer, Allan Hildesheim (AH), Nathaniel Rothman (NR), Denise L. Doolan (DD), and Qing Lan (QL), Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma. *Scientific Reports*, 2021;11(1):23664. Incorporated in **Chapter 2.**

   [*] co-first authors

Chapter No.	Nature and extent of the intellectual input of each author
2	DD, ZL, AH, NR, and QL designed and conceived the study. Co-authors from collaborative multi-centre studies collected and provided the samples. YS performed all experiments. ZL, CP, YS and MC performed data analysis. ZL wrote the initial draft. YS, AH, DD and CP extensively edited and commented on the manuscript. All collaborators, co-authors read and approved the manuscript.

2. **Yomani D. Sarathkumara**, Rena R. Xian (RX), Zhiwei Liu, Kelly J. Yu, John K. C. Chan, Yok-Lam Kwong, Anna E. Coghill, Ruth M. Pfeiffer, Nathaniel Rothman, Richard F. Ambinder (RA), Allan Hildesheim, Carla Proietti, Qing Lan, and Denise L. Doolan, A generalized proteome-wide Epstein-Barr Virus Antibody Signature predicts classical Hodgkin's Lymphoma in geographically and ethnically distinct populations, *pending submission.* Incorporated in **Chapter 3.**

Chapter No.	Nature and extent of the intellectual input of each author
3	DD, ZL, AH, NR, QL and RA designed and conceived the study. Co-authors from collaborative multi-centre studies collected and provided the samples. YS and RX performed experiments. YS, CP, and ZL performed data analysis. YS wrote the initial draft. ZL, AH, DD and CP extensively edited and commented on the manuscript. All collaborators, co-authors read and approved the manuscript.

3. **Yomani D. Sarathkumara**, Nathan W. Van Bibber (NV), Zhiwei Liu, Sweta K. Sinha, Cliona M Rooney (CMR), Anna E. Coghill (AC), Carla Proietti, Denise L. Doolan, Characterization of the proteome-wide Epstein-Barr virus antibody responses after T-cell immunotherapy in patients with EBV-associated lymphomas, *pending submission.* Incorporated in **Chapter 4.**

Chapter No.	Nature and extent of the intellectual input of each author
4	DD, CMR, and AC designed and conceived the study. Co-authors from collaborative multi-centre studies collected and provided the samples. YS performed all experiments. YS, NV, and CP performed data analysis. YS wrote the initial draft. AC, DD, and CP extensively edited and commented on the manuscript. All collaborators, co-authors read and approved the manuscript.

4. **Yomani D. Sarathkumara**, Jessica Scott (JS), Daniel Pelowa, Robert Dowi, William Pomat, Catherine M. Rush (CR), Jeffrey Warner (JW), Carla Proietti, Denise L. Doolan Molecular biomarkers of latent and active tuberculosis in Papua New Guinea, *pending submission.* Incorporated in **Chapter 5.**

Chapter No.	Nature and extent of the intellectual input of each author
5	DD, CR, JW, and CP designed and conceived the study. JW, CR and JS collected the samples. YS, JW, and CR performed experiments. YS and CP performed analysis. YDS wrote the initial manuscript draft. DD, and CP revised and edited the manuscript extensively. All authors contributed editing and approved the final version.

5. **Yomani D. Sarathkumara**, Daniel J. Browne (DB), Ashton M. Kelly (AK), David J. Pattinson (DP), Catherine M. Rush, Jeffrey Warner, Carla Proietti, and Denise L. Doolan. The effect of tropical temperatures on the quality of RNA extracted from stabilized whole blood samples, *International Journal of Molecular Sciences*, 2022, 23(18), p. 10609. Incorporated in **Chapter 6.**

Chapter No.	Nature and extent of the intellectual input of each author
6	DD, YS, and DB designed and conceived the study. YS, and AK performed experiments. YS and CP performed analysis. YDS wrote the initial manuscript draft. DD, CP, CR, revised and edited the manuscript extensively. All authors contributed editing and approved the final version.

Abstract

Epstein-Barr Virus (EBV) and *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis (TB), are both associated with latent infections that confound diagnosis and disease control. EBV is a human oncogenic virus linked to the pathogenesis of variety of lymphomas and epithelial tumours and establishes a lifelong persistent latent infection in B-cells limiting expression of specific viral proteins in order to escape the host immunity. MTB bacteria remains dormant or inactive in infected individuals, but this pathogen reservoir can reactive months or years following infection so that persons with latent TB have a lifetime risk of developing active clinical TB. For both EBV and MTB, latent infections pose a critical challenge to disease diagnosis, prevention, and eradication. Despite considerable advances in molecular immunology research, effective diagnostics for these latent infections do not exist. There is an imperative need for biomarkers to identify populations for targeted intervention to improve early diagnosis and limit disease burden. To address this, an improved understanding of immune responses associated with latent infections is required. The studies presented in this doctoral thesis employed a proteome-wide multiplex approach and a high dimensional molecular profiling technique in combination with sophisticated computational analyses to human samples from various diseases models including EBV-associated cancers (namely, natural killer/T-cell lymphoma (NKTCL) and classical Hodgkin lymphoma (cHL)) and tuberculosis to inform immune signatures of disease and aetiology.

A custom protein microarray representing the complete EBV proteome was applied as a high throughput multiplex screening tool to comprehensively evaluate IgG and IgA antibody responses and define signatures of EBV-associated lymphomas and latency, in combination with an advanced analytical analysis pipeline. Two studies identified novel antibody signatures for NKTCL and cHL in adults in Asia. Additionally, generalizability of the antibody signature for cHL previously defined in a European population was assessed in the Asian population. A third study explored, for the first time, application of EBV proteome-wide microarrays in the context of immunotherapy, treatment option for using providing important insights into understanding immune responses and discovering effective target antigens for EBV-specific T-cell (EBVST) immunotherapy against EBV-lymphomas.

Other work describes the first profiling of transcriptional responses in active and latent TB infection in Papua New Guinea (PNG) using RNA-seq as a high-throughput approach to identify distinguishing molecular features. Signatures specific for both active TB and latent infection were defined, as well as overlapping sets of genes between active and latent TB.

To facilitate future transcriptomic studies in remote and resource-limited settings such as PNG, the final chapter assessed collection and storage of whole-blood RNA collection systems in suboptimal tropical conditions where electricity and storage facilities are compromised.

In summary, this doctoral research employed two synergistic platforms of protein expression (protein array) and gene expression (transcriptomics) to extend our understanding of immune responses at the humoral and molecular level in multiple disease models associated with latent infections. The work highlights the potential for these multiplex platforms together with advanced analytical approaches including machine learning techniques for the discovery of biomarkers of disease and latency. This information would have translational impact for early diagnosis of individuals at risk of developing disease and for disease control, to improve global public health.

Table of Contents

List of figures

List of tables

List of Abbreviations

Abbreviation	Definition
˚C	Celsius
18s	18S ribosomal RNA
28s	28S ribosomal RNA
Ab	Antibody
ADCC	Antibody-dependent cell cytotoxicity
Adj	Adjusted
AFB	Acid-fast bacilli
AGRF	Australian Genomics Research Facility
AITHM	Australian Institute of Tropical Health and Medicine
AITHM	Australian Institute of Tropical Health and Medicine
AUC	Area under the curve
BCG	Bacillus Calmette–Guérin
BDH	Balimo District Hospital
bp	base pair
BP	Biological process
cDNA	complementary DNA
CFP-10	Culture filtrate protein 10
cHL	Classical Hodgkin's Lymphoma
CIs	Confidence intervals
Ct	Cycle threshold
CTLs	Cytotoxic T lymphocytes
CV	coefficient of variation
DE	differential expression
DEG	Differentially expressed gene
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
DO	Disease Ontology
DOTS	Directly Observed Treatment Short course
DR-TB	Drug-resistant tuberculosis

EA	Early antigen
EBERs	EBV-encoded small RNAs
EBNA-LP	EBV nuclear antigen leader protein
EBNA1	EBV nuclear antigen 1
EBV	Epstein-Barr virus
EBVST	EBV-specific T-cells
EBVSTs	EBV-specific T-cells
ELISA	Enzyme-linked immunosorbent assay
EP-TB	Extrapulmonary TB
ESAT-6	Early secreted antigenic target 6
FC	Fold change
FDR	False discovery rate
GO	Gene Ontology
GP	Glycoprotein
HIV	Human immunodeficiency virus
HRS	Hodgkin Reed-Sternberg
HSCT	Hematopoietic stem cell transplant
IFN	Interferon
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IGRA	Interferon gamma-release assays
IM	Infectious mononucleosis
IRB	Institutional Review Boards
JCU	James Cook University
LJ	Löwenstein-Jensen
LMP	Latent membrane protein
LOD	limit of detection
LTB	Latent tuberculosis
MDA	MeanDecreaseAccuracy
MDG	MeanDecreaseGini
MDR-TB	Multi-drug resistant TB
MGIT	Mycobacteria growth indicator tube

MHC	Major histocompatibility complex
MRAC	Medical Research Advisory Committee
mRNA	messenger ribonucleic acid
MTB	*Mycobacterium tuberculosis*
MTBC	*M. tuberculosis* complex
NCI	National Cancer Institute
NK	Natural killer
NKTCL	Natural killer T-cell lymphoma
NPC	Nasopharyngeal carcinoma
NS	Non-significant
NTM	Nontuberculous Mycobacterium
OD	Optical density
ORs	Odds ratios
PBMCs	Peripheral blood mononuclear cells
PCA	Principal component analysis
PCR	Polymerase chain reaction
PNG	Papua New Guniea
PTB	Pulmonary TB
PTLD	Post-transplant lymphoproliferative disease
QC	Quality control
qPCR	quantitative-PCR
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROC	Receiver operating characteristic curve
RR-TB	Rifampicin-resistant tuberculosis
rRNA	ribosomal RNA
RT	Room temperature
RT-qPCR	Reverse transcription quantitative PCR
SD	Standard deviation
SDHA	Succinate dehydrogenase complex, subunit A
SOT	Solid organ transplant

sPLS-DA	sparse Partial Least Squares Discriminant Analysis
SSI	Standardized signal intensity
TB	Tuberculosis
TBP	TATA-box-binding protein
TK	Thymidine kinase
TST	Tuberculin skin test
USA	United States of America
VCA	Viral capsid antigen
VSN	Variance stabilizing normalization
VST	Virus-specific T-cells
WHO	World Health Organization
ZN	Ziehl-Neelsen

Note: All gene and protein symbols follow HGNC and MGI convention where human gene symbols are italicised, and protein symbols are in standard typeface.

Chapter 1

General Introduction

This chapter provides the outline for the thesis, detailing the background, research gaps and importance of identifying antibody and transcriptomic biomarker signatures of latent diseases. It describes the potential of Epstein-Barr Virus (EBV)-associated lymphomas and tuberculosis (TB) as disease models to address the research questions.

1 General Introduction

1.1 Epstein-Barr Virus (EBV) infection

Epstein-Barr Virus (EBV) is a ubiquitous lymphotropic virus prevalent in all human populations (1). Also known as Human Herpes Virus type 4 (HHV-4), EBV belongs to the family of the *Herpesviridae*, a subfamily of the *Gammaherpesvirinae*, Genus *Lymphocryptoviru* (2). Viruses of herpes subfamilies have existed as ancestral forms and co-evolved with different host lineages over millions of years (3). The EBV genome is comprised of linear double-stranded DNA (dsDNA) with a size of around 172 kilobase pairs (kbp) that encodes more than 80 genes, surrounded by a protein capsid (4). EBV-1 and EBV-2 (also known as types A and B) are the two major types of EBV infecting humans, and these subtypes vary in the organization of the genes that encode for EBV nuclear antigen (EBNA) but are responsible for the same pathologies (5, 6).

Over 90% of the human population worldwide is infected with EBV (7). However, EBV genotypes and several diseases associated with EBV have noticeable geographical distributions (6). The primary EBV infection usually remains asymptomatic in most individuals during childhood (8). However, it can be delayed until adolescence or young adulthood, and the primary infection manifests as infectious mononucleosis (IM), characterized by fever, tonsillar pharyngitis, and lymphadenopathy (9). EBV transmission occurs through body fluids, commonly by salivary contact (10).

The primary site of EBV infection is considered the oral compartment, whereas epithelial cells and B lymphocytes are the host cells of the virus (11). EBV has a complex life cycle. As illustrated in **Figure 1-1,** EBV primarily infects and replicates in oropharyngeal epithelial cells during acute infection (10, 12). It can further establish a lifelong persistent latent infection in memory B-cells (13, 14). EBV infects naïve B lymphocytes via the interaction of the viral envelope glycoprotein gp350 with the CD21 (also known as CR2) complement receptor molecule expressed on the surface of naïve host B-cells, facilitating virus entry and triggering endocytosis (15, 16). In addition, when the second envelope glycoprotein, gp42, interacts with B-cell HLA class II molecules triggers the virus core fusion mechanism (17). Through these interactions, viral genetic material is released into the cell following the fusion of the viral membrane with the endosomal membrane. Most of the epithelial cells lack the expression of the HLA II and CD21 molecules. Therefore, β1 integrin mediates virion

attachment via BMRF-2 EBV protein (18, 19). Virions that contain glycoproteins, gH and gL directly interact with αvβ6 and αvβ8 integrins to trigger the fusion of EBV and epithelial cells (20).

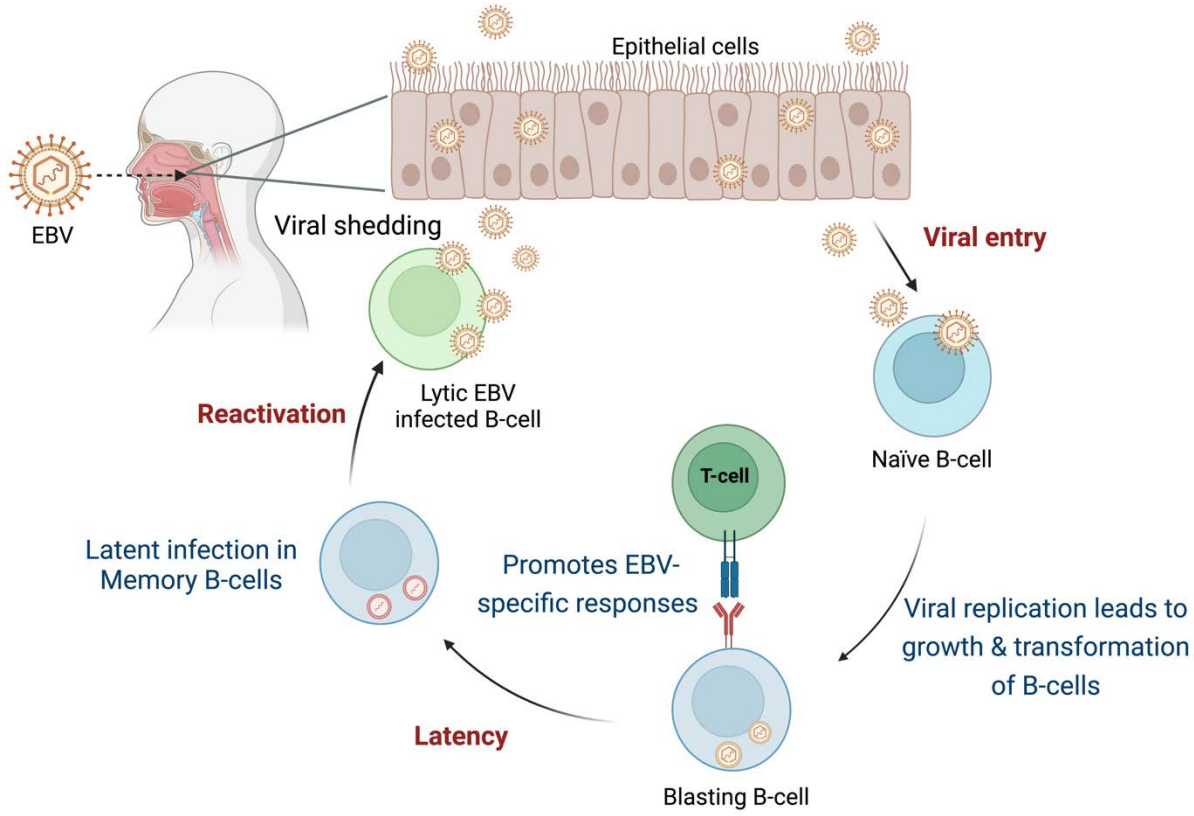


Figure 1-1 Schematic representation of Epstein–Barr Virus (EBV) life cycle.
The primary site of infection is the oro-pharynx, in which the virus initially infects epithelial cells that permit the shedding of the virus into saliva for transmission to new hosts. In the mucosal lymphoid tissue, EBV membrane glycoproteins interact with the CD21 receptor on naïve B-cells, serving as the primary entry receptor and HLA II molecules as co-receptor. Usually, these blasting B-cells are destroyed by cytotoxic T lymphocytes. Infected memory B-cells persist in lifelong latent infection with limited gene expression. Resting B-cells may be activated, resulting in periodic viral reactivation and shedding. (Created with BioRender.com)

The interactions between the human host and EBV are paradoxical. EBV has two life cycle forms in the host; latent and lytic (21). EBV can transform the resting B-cells into indefinitely proliferating cells, which can turn into immunoblasts that produce immortal lymphoblastoid cell lines (LCLs) *in vitro* (22). Most infected B-cells remain latent for life, although few may undergo a lytic cycle. However, memory B-cells are the main reservoir for

EBV reactivation, persistence, and EBV-associated lymphomas (23). The transition from latency to the lytic stage is known as EBV reactivation (11). This transition is triggered by the expression of BZLF1 and BRLF1, transcribed to encode the transactivators Zta and Rta, respectively (24). However, memory B-cells are the main reservoir for EBV reactivation, persistence, and EBV-associated lymphomas (23).

In 1956, a malignant tumour of the jaws in African children was recognized by a British surgeon, Denis Burkitt (25), and it was identified as a lymphoma based on histo-pathological investigations (26, 27). Epstein, Barr, and Achong discovered the causative virus-like particles in lymphoblasts by electron microscopy from Burkitt's lymphoma biopsy tissue in 1964 (28, 29). Hence, EBV was recognized as the first human oncogenic virus associated with several malignancies and its link to primary clinical manifestation was confirmed four years later (30, 31).

EBV can infect a broad spectrum of cells, including T lymphocytes, follicular dendritic cells, Natural Killer (NK) cells, smooth muscle cells and glandular epithelium of the thyroid, stomach, and salivary glands (32). The long-term persistence of EBV in humans is not usually a serious condition, although it increases the risk of some cancer types in certain individuals. Approximately 15-20% of human cancers are estimated to be associated with EBV infections (33). EBV is associated with a wide range of malignancies originating from B-cells and epithelial cells (33). EBV-associated lymphomas include Burkitt lymphoma (BL), Hodgkin lymphoma (HL), Natural Killer /T-cell lymphoma (NKTCL) and post-transplant lymphoproliferative disorders (PTLDs) (34, 35). Tumours of epithelial origin include nasopharyngeal carcinoma (NPC) and gastric adenocarcinoma (35, 36).

1.2 EBV latency stages, gene expression and associated cancers

There are three types of latency programs; Latency I, Latency II, and Latency III, in which EBV establishes viral latency in infected B-cells. Each latency type is characterized by a limited, distinct set of viral gene expressions (37, 38). The latency types are differentiated by their immunogenicity and the viral antigens expressed in EBV-associated lymphomas (**Figure 1-2**), as described below. Latency I is limited to transiently expressed EBNA1 in memory B-cells associated with BL (39), EBNA1, latent membrane proteins (LMPs) LMP1 and LMP2 are expressed in latency II expressing germinal centre B-cells whilst the most immunogenic type III latency pattern express EBNA1, EBNA2, EBNA3A–EBNA3C, EBNA-LP, LMP1 and LMP2 in PTLD (28, 40).

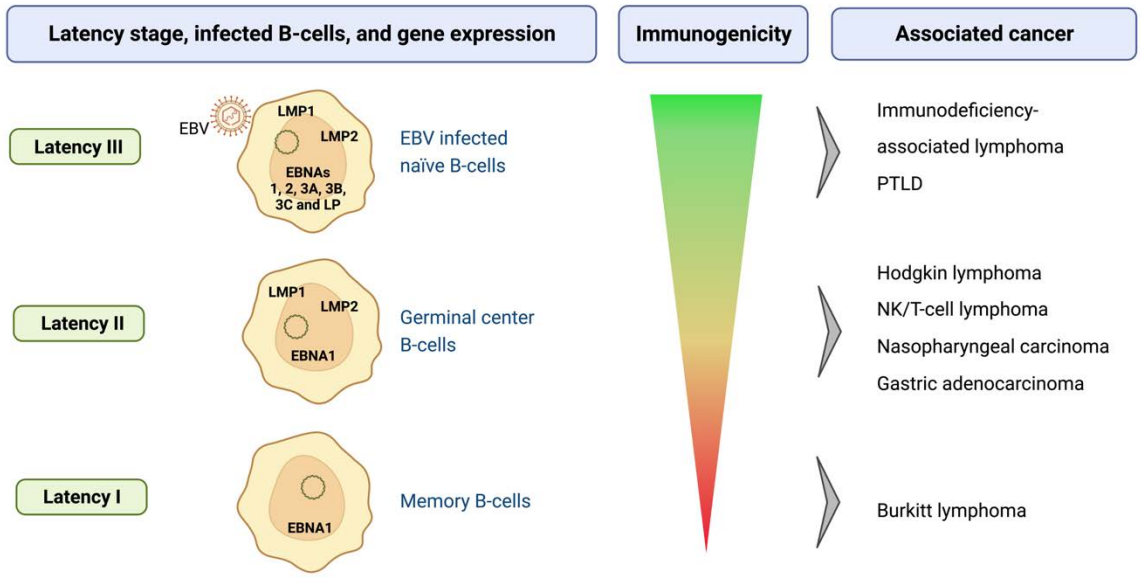


Figure 1-2 Epstein-Barr virus (EBV) latency stages characterized by limited gene expression in each latency stage, their immunogenicity, and associated malignancies. (Created with BioRender.com).

Although the EBV in cancer cells is mostly in the latent state, the viral lytic cycle also contributes to oncogenesis, and it plays a key role in the development and maintenance of cancers (41). Several studies have demonstrated that EBV lytic cycle contributes to B-cell transformation efficiency in cell cultures (42, 43) and the development of B-cell lymphoma in a humanized mouse model (44). Therefore, EBV-associated cancer cohort studies presented in this thesis are focused on recognition of lytic and latent EBV proteins.

1.3 EBV-Specific T-cells (EBVSTs) immunotherapy for EBV-lymphomas

EBV-Specific T-cells (EBVSTs) have been clinically explored as prophylaxis and therapy option for EBV-associated malignancies over the past two decades (45, 46). EBVST immunotherapies have proved successful in treating PTLD, the most immunogenic type III latency lymphoma that occurs in patients who are mostly immunosuppressed or have immunodeficiencies after transplantation (45, 47, 48). Type III latency-associated B-cells expresses the most immunogenic broader spectrum of EBV latent proteins and are more responsive to immunotherapy with EBVSTs than other latency types associated with lymphomas (46, 47). Hence, type III latency-associated lymphomas have been the most successfully treated by T-cell immunotherapy compared to other latency types (14).

Type I and II express only a handful of EBV proteins which is poorly immunogenic and more challenging to treat with EBVSTs (47). However, T-cell immunotherapy directed against Hodgkin lymphoma and extranodal NKTCL, which typically expresses type II latency, is a durable, safe, and an effective approach without significant toxicity (49-51).

Despite recent advances in this rapidly growing T-cell immunotherapy research field, subsets of patients have no responses or partial responses to immunotherapy treatments. These patients are relapsing, challenging the potential of this approach for successful treatment outcomes in the long term (52). The necessity for biomarkers that could distinguish individuals who responded (responders) from those who did not respond (non-responders) to EBVST immunotherapy would be insightful. Immuno-oncology research has focused exclusively on T-cell response, and the role of B-cells and humoral responses in cancer immunotherapies is poorly understood (53). A crucial knowledge gap is the lack of understanding of the most effective EBV antigen targets for immunotherapy, and advances in related fields could help address this.

1.4 Humoral immune responses in EBV-associated lymphomas and EBV-Specific T-cells (EBVSTs) immunotherapy

EBV has evolved a plethora of strategies to evade immune system recognition and establish latent infection in B-cells. Both cellular and humoral immune responses are critical in controlling the primary and persistent EBV infection (54). The humoral immune defence

system plays an important role in persistent EBV infection by producing antigen-specific antibodies.

The generation and selection of B cells secreting high-affinity antibodies occur in the germinal centre (GC) (55). Naïve B cells recognize and encounter cognate antigens via their B cell receptor (BCR) in lymphoid tissues (56). Upon receiving signals from activated antigen-specific CD4+ helper T-cells, activated B-cells migrate to GCs, divide rapidly, and undergo clonal expansion, transforming into rapidly proliferating centroblasts, which undergo somatic hypermutation of immunoglobulin (Ig) variable gene sequences resulting isotype-switching (57). Most of these have switched isotypes from IgM to IgG or IgA (58). Together, these processes develop long-lived antibody-secreting plasma cells and memory B-cells expressing many variants of Ig specificities binding to antigens with improved high affinity (56-59).

The role of B-cells and antibody-mediated responses is relatively understudied in the context of T-cell immunotherapy (53, 60). In tumour immunology, B-cells have known functions such as antigen presentation and producing tumour-specific antibodies that induce destroying tumour cells through antibody-dependent cell cytotoxicity (ADCC) and activation of the complement cascade (61-63). Compared to our understanding of anti-tumour responses of T-cells, current knowledge on humoral responses for T-cell immunotherapy are in its infancy, with very little work to date investigating antibody responses in EBVST immunotherapy.

The humoral response includes producing antigen-specific antibodies for both latent and lytic EBV infections. Only a few of these antigens, such as the viral capsid antigen (VCA), early antigen (EA) and EBNA1, have been widely studied and are currently used for diagnostic tests (64, 65). However, difficulties in distinguishing between EBV past infections and reactivation, defining the risk of developing EBV-associated malignancies while determining the infectious status, especially in immunocompromised patients, are considered limitations of currently used serological assays for EBV.

IgG antibodies remain detectable within weeks after the onset of primary infection and persist for many years (66). IgG antibodies against EBV proteins, VCA, EA and EBNA1, are typically associated with EBV-associated cancers (65, 67, 68). Particularly mucosal level humoral immune responses are mediated predominantly by secretory IgA antibodies (69). IgA antibodies against latent and lytic proteins have been demonstrated as a serological marker of periodic EBV reactivation at the mucosal sites (10, 70) and suggested as predictors of the

development of NPC (68, 71). Prior studies on humoral responses to EBV-associated tumours have mainly focused on measuring IgG and IgA antibodies to only a handful of EBV proteins, as reviewed in the literature (68). Thus, there is a pressing need to limit the burden of these diseases by discovering biomarkers for EBV-associated lymphomas.

1.4.1 Custom EBV protein microarrays

With the advent of modern technologies, novel proteome-wide approaches such as proteome microarrays have been developed as powerful high-throughput platforms for basic biological and clinical research (68). Protein microarrays enable the examination of a pathogen's complete or partial proteome, allowing the identification of potential targets for therapeutic or diagnostic applications (69). Microarrays assembled from pathogen proteins can play a significant role in discovering candidate biomarkers, vaccine antigens, or therapeutic targets in a wide range of infectious diseases (70, 71). Several studies have been conducted to discover sets of immunodominant antigens as potential diagnostic markers in various pathogens in infectious diseases using protein microarrays (72-74).

Briefly, a large number of proteins from the pathogen of interest (representing the complete or partial proteome) are printed at high density on a solid surface as a protein dot matrix (75). Automated high-throughput protein expression/purification using the 6xHis tag using *Escherichia coli* expression system in 96-well microplates has been developed to produce hundreds of proteins quickly and cost-effectively (76). Peptides or domains highly purified through cell-free expression systems facilitating rapid, *in situ* synthesis of proteins from their corresponding DNA templates are commonly used in protein microarrays (77). Robotic contact printing tools consisting of metal pins with solid or quill tips deliver sub-nanolitre of protein samples on the slide surface, printing hundreds of spots simultaneously (75). The printed proteins on the microarray can be then assayed for recognition by antibodies present in biological samples (sera, plasma, urine) and detected using fluorescently labelled secondary antibodies (*i.e.*, anti-human immunoglobulin antibodies) via a confocal laser scanner (such as Genepix, Molecular Devices, CA, USA).

A custom EBV protein microarray consisting of 199 sequences representing all EBV proteins (n=86) and known splice variants across five strains (AG876, Akata, B95-8, Mutu, and Raji) and including three synthetic EBV peptides commonly used as putative cancer biomarkers (VCAp18, EBNA1, and EAd p47) was developed to measure both IgA and IgG

antibodies in latent EBV infection related tumours (78). This custom EBV proteome microarray was employed together with advanced analytical data analysis pipelines to explore antibody responses against the full spectrum of EBV proteome in different populations to identify antibody signatures as biomarkers for EBV-associated cancers, including NPC, BL, and HL (78-81).

1.5 Tuberculosis

Tuberculosis (TB) poses a significant burden on global health, responsible for millions of deaths and clinical cases worldwide. Indeed, TB caused by the bacterium Mycobacterium tuberculosis (MTB) remains the second leading infectious killer after SARS-CoV2 (86). In 2021, there were an estimated 10.6 million TB cases and 1.6 million deaths (87). An increase of 3.6% in TB incidence rate has been reported in many resource-limited countries between 2020 and 2021, intensified by the COVID-19 pandemic (88). TB is also the leading cause of death among individuals infected with HIV (86, 87).

TB is a contagious, airborne chronic infectious disease that spreads from person to person. The contagious nature of TB was first postulated in the 16th century by an Italian scientist, Girolamo Fracastoro (89). In 1882, a German scientist, Dr. Robert Koch, discovered the Tubercle Bacillus, M. tuberculosis (MTB), as the causative agent of TB (90). The Mycobacterium tuberculosis complex (MTBC) consists of closely related species, including *M. tuberculosis*, *M. africanum*, and *M. bovis,* predominantly causing human disease (91). However, it is important to note that this thesis only focuses on MTB infection in humans.

The lungs are the predominant site of infection by tuberculosis, referred to as pulmonary TB (PTB). Sylvius de la Boë of Amsterdam first described pathological details of tuberculosis, characterized by lesions, tubercles presented in the lungs and their progression into abscesses, ulcers, cavities, or empyema (92). PTB is often infectious and can spread the disease via acts that propel droplets containing tubercle bacilli into the air by coughing, sneezing, spitting or simply talking (93). Sputum is also produced more often by symptomatic individuals and used as the key biological sample for most commonly used diagnostic tests. Transmission occurs when a person inhales these infected droplets containing the bacteria traverse via the upper respiratory tract (nasal passage) to the bronchi reaching the alveoli in the lungs of the lower respiratory tract (94). Although most cases are PTB, the disease can be

disseminated to other organs and sites of the body, which is then referred to as extrapulmonary TB (EP-TB) (95). EP-TB accounts for 20 to 25% of reported TB cases (96).

Tubercle bacilli are ingested by the resident alveolar macrophages and/or tissue dendritic cells in the lung epithelium (97, 98). Some of the ingested bacteria are destroyed by phagocytosis as an innate immune response by macrophages. However, some bacteria can evade phagocytosis, and a fraction of infected macrophages/dendritic cells are drained into the lymphatics or bloodstream to activate the adaptive immunity (99, 100). Granulomas, a hallmark structure of TB, consist of various aggregated immune cells (*i.e.*, fibroblasts, natural killer (NK) cells, neutrophils, T and B cells) surrounding a caseous necrotic core of MTB-infected alveolar macrophages, which are formed during primary infection, and these can trigger the host's immune mechanisms (Figure 1-3). If the host's immune system cannot control the formation of granulomas, then mycobacterial proliferation can dramatically progress into active disease, spreading to distant organs and tissues and influencing the clinical outcome (101).

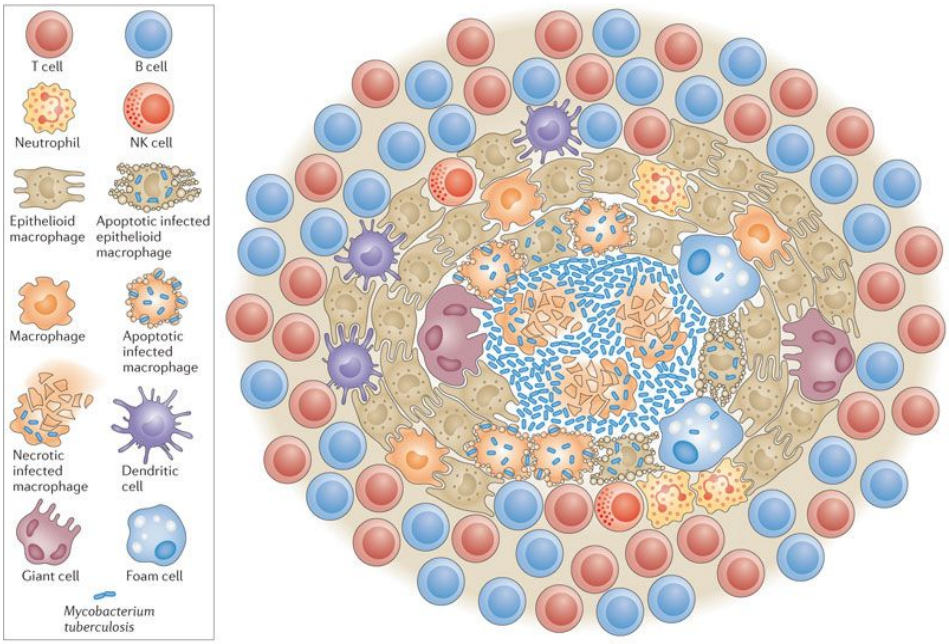


Figure 1-3 Schematic structure of tuberculous granuloma.

The granuloma is a compact structure comprised of cellular constituents and debris. Proliferated tubercle bacilli are concentrated in the central necrotic area, and some are ingested by macrophages, whilst foam cells and multi-nucleated giant cells are frequently located at the border of the necrotic centre. The granuloma is also surrounded by neutrophils, dendritic cells, B and T cells, NK cells, and fibroblasts that secrete extracellular matrix are also surrounded by the granuloma. Adopted from Ramakrishnan, 2012 (102).

1.6 Latent tuberculosis infection and diagnosis

Latent tuberculosis (LTB) infection is defined as having persistent anti-mycobacterial immune responses when individuals are infected with viable MTB without evidence of clinically manifested active TB (103). The infection remains asymptomatic in these latently infected persons with MTB without showing any active TB-related symptoms (104). Approximately, one-third of the world's population is estimated to have latent TB infection (104, 105). The risk of a latent TB individual developing active TB is estimated to be 5–10%. This process is referred to as "TB reactivation" (106).

The tuberculin skin test (TST), also known as the Mantoux test, is one of the widely available tests to diagnose latent TB infection (107). TST is an intradermal injection in the forearm of purified protein derivative (PPD) consisting of a crude antigenic mixture containing tuberculin that can elicit a delayed-type hypersensitivity reaction via a cell-mediated immune response in exposed individuals (108). Sensitized T-cells by prior TB infection migrate to the site of injection, release lymphokines, and induce other inflammatory responses through local oedema, vasodilation, erythema, fibrin deposits and migration of other inflammatory cells towards indurated skin (109, 110). The test result is usually produced in 48-72 hrs after test administration by measuring the diameter of the indurated area of the skin in millimetres and interpreted carefully to report any positive reaction observed. However, false-positive results may occur in patients vaccinated with BCG (111). Consequently, TST results have limitations regarding its sensitivity and specificity for immunocompromised conditions and previous or current exposure to Mycobacteria other than Tuberculosis or non-tuberculous Mycobacterium (NTM) (112, 113).

Interferon-gamma release assays (IGRAs) are blood tests that diagnose latent TB infections. IGRAs measure the release of IFN-γ by T-cell mediated immune response when stimulated by antigens specific for the M. tuberculosis complex, *i.e.*, early secreted antigenic target 6 (ESAT-6) and culture filtrate protein 10 (CFP-10) (114). Since these genes are absent in BCG vaccine strains or most NTM species and thus are more specific than PPD in TST, the test is recommended for BCG-vaccinated individuals (115). The QuantiFERON-TB Gold In-Tube (QFT) assay (Cellestis/Qiagen, Carnegie, Australia) and the T-SPOT.TB assay (Oxford Immunotec, Abingdon, United Kingdom) are the two commercially available IGRAs used in many countries approved by the FDA (116). The QFT assay is based on ELISA performed using whole blood whilst the T-SPOT.TB is based on an enzyme-linked immunosorbent spot (ELISPOT) assay performed on peripheral blood mononuclear cells (PBMCs) (117). Although

IGRAs cannot accurately predict the risk of infected persons developing active TB, it offers a 96-98% specificity in several studies, with >90% of the populations being BCG vaccinated in contrast with TST (118, 119). IGRA offers several operational advantages over TST since the test is read within 24 hrs and reduces the reader bias on results interpretation. However, it still requires a basic equipped laboratory setup with some technical skills to perform the test, and the cost is comparatively higher than TST, so it may not be affordable in many developing countries (107). However, a recent study reveals that Ugandan individuals who were highly exposed to MTB tested negative by both IGRA and TST and showed that these 'resisters' possess IgM, class-switched IgG antibody responses and non-IFN-γ T cell responses to the MTB-specific ESAT6 and CFP10 (120). IGRAs have also shown positive in adults with active TB infection, including EP-TB (121-123).

Available data suggest that TST and IGRA are acceptable tests but imperfect tests to identify latent TB infections accurately. However, neither test differentiates latent from active TB disease (117, 124). Therefore, the current major challenge in TB-prevalent countries is identifying asymptomatic, latently infected individuals.

1.6.1 Host transcriptomic for tuberculosis diagnosis

Host transcriptomics is a comprehensive high-throughput approach to identifying diagnostic and mechanistic immune signatures of infectious diseases based on gene expression. High-throughput RNA sequencing (RNAseq) is a platform that provides a broader dynamic range to quantify gene expression levels that measure all gene transcripts and expression of non-coding RNAs (125, 126). In contrast, targeted gene studies measure a panel of genes selected depending on a specific category (*i.e.*, pathways, function, cell subsets) or from preliminary or published data traditionally using real-time qPCR (122). Transcriptomic analysis with advanced statistical approaches enables comprehensive information on gene expression and regulation in biological samples in different conditions (128). It provides important biological insights through differential expressions of genes on either the host or the pathogen during an infectious disease (129). Understanding patterns of the host response to a specific pathogen can provide novel insights into host-pathogen interactions, which can inform a disease diagnosis.

As an example of the potential of host transcriptomics for diagnostics on TB, a whole blood RNA expression signature study conducted in South African adolescents was able to

identify a prospective six gene signature that predicted tuberculosis progression in the 12 months preceding TB diagnosis with 66% sensitivity and 81% specificity (130). Another TB-related whole-blood transcriptional study identified an interferon driven gene-signature to distinguish subjects with active TB from uninfected controls or latent TB (131). However, a follow-up study showed that this interferon signature for active TB had highly similar patterns sarcoidosis patients (132). Kaforou et al. identified a 27-transcript signature that distinguished TB from latent TB cases in the African population (133). Another study proposed a combination of four differentially expressed genes, NEMF, ASUN, DHX29, and PTPRC, as potential biomarkers to discriminate between latent and active TB (134). A recent study has identified a three-transcript signature (FCGR1A, ZNF296, C1QB) that differentiated TB from latent TB as potential biomarkers for TB (135). Although several gene signatures that measure host response to M. tuberculosis in blood samples have been proposed, none has been pursued clinical implementation (136). Further, validating any diagnostic signature in several independent study populations with clearly defined disease states is important. In fact, BCG vaccination status, variations in exposure to MTB and other NTM species, and TB endemic vs low TB prevalent areas should be considered as factors that potentially could influence TB-associated host immune responses (127).

Transcriptomic approaches can be a valuable tool for identifying biomarkers associated with latent infections. The underlying mechanisms which determining whether a latently infected person will control the Mycobacterial infection or will develop active TB or PTB are unknown. Molecular approaches offer the potential to provide key insights and transcriptomic approaches can therefore facilitate the development of biomarkers for the diagnosis and prognosis of TB essential for effective and early treatment.

1.7 Thesis hypothesis and aims

The overall hypothesis of my doctoral work was that sensitive and specific biomarker signatures for latent infections can be identified using protein microarrays, RNA-seq and computational biology. The overall aim of my doctoral studies was to discover antibody and transcriptomic biomarker signatures of latent infections using EBV-associated lymphomas and tuberculosis as disease models. Specific aims and objectives have been addressed within individual chapter, each written in the format of a scientific manuscript already published or pending submission for peer-reviewed publication.

Table 1-1 Brief outline of thesis chapters

Tittle	Chapter No.
Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma	Chapter 2
A generalized proteome-wide Epstein-Barr Virus Antibody Signature predicts classical Hodgkin's Lymphoma in geographically and ethnically distinct populations	Chapter 3
Characterization of the proteome-wide Epstein-Barr antibody responses after T-cell immunotherapy in patients with EBV-associated lymphomas	Chapter 4
Molecular biomarkers of latent and active tuberculosis in Papua New Guinea	Chapter 5
The effect of tropical temperatures on the quality of RNA extracted from stabilized whole-blood samples	Chapter 6

1.8 References

1. Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al. Human herpesviruses: biology, therapy, and immunoprophylaxis: Cambridge University Press; 2007.
2. Davison AJ. Herpesvirus systematics. Veterinary Microbiology. 2010;143(1):52-69.
3. McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EA. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. Journal of Molecular Biology 1995;247(3):443-58.
4. Baer R, Bankier A, Biggin M, Deininger P, Farrell P, Gibson T, et al. DNA sequence and expression of the B95-8 Epstein—Barr virus genome. Nature. 1984;310(5974):207.
5. Sample J, Young L, Martin B, Chatman T, Kieff E, Rickinson A. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. Journal of Virology. 1990;64(9):4084-92.
6. Tzellos S, Farrell P. Epstein-Barr virus sequence variation—biology and disease. Pathogens. 2012;1(2):156-74.
7. Niedobitek G, Meru N, Delecluse HJ. Epstein-Barr virus infection and human malignancies. International Journal of Clinical and Experimental Pathology. 2001;82(3):149-70.
8. Crawford DH, Macsween KF, Higgins CD, Thomas R, McAulay K, Williams H, et al. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. Clinical Infectious Diseases. 2006;43(3):276-82.
9. Abbott RJ, Pachnio A, Pedroza-Pacheco I, Leese AM, Begum J, Long HM, et al. Asymptomatic primary infection with Epstein-Barr virus: observations on young adult cases. Journal of Virology. 2017;91(21):e00382-17.
10. Sixbey JW, Nedrud JG, Raab-Traub N, Hanes RA, Pagano JS. Epstein–Barr virus replication in oropharyngeal epithelial cells. New England Journal of Medicine. 1984;310(19):1225-30.
11. Odumade OA, Hogquist KA, Balfour HH. Progress and problems in understanding and managing primary Epstein-Barr virus infections. Clinical Microbiology Reviews. 2011;24(1):193-209.

12. Gerber P, Lucas S, Nonoyama M, Perlin E, Goldstein L. Oral excretion of Epstein-Barr virus by healthy subjects and patients with infectious mononucleosis. The Lancet. 1972;300(7785):988-9.
13. Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA. EBV persistence in memory B cells in vivo. Immunity. 1998;9(3):395-404.
14. Laichalk LL, Hochberg D, Babcock GJ, Freeman RB, Thorley-Lawson DA. The dispersal of mucosal memory B cells: evidence from persistent EBV infection. Immunity. 2002;16(5):745-54.
15. Nemerow GR, Wolfert R, McNaughton M, Cooper NR. Identification and characterization of the Epstein-Barr virus receptor on human B lymphocytes and its relationship to the C3d complement receptor (CR2). Journal of Virology. 1985;55(2):347-51.
16. Fingeroth JD, Weis JJ, Tedder TF, Strominger JL, Biro PA, Fearon DT. Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. Proceedings of the National Academy of Sciences. 1984;81(14):4510-4.
17. Li Q, Spriggs MK, Kovats S, Turk SM, Comeau MR, Nepom B, et al. Epstein-Barr virus uses HLA class II as a cofactor for infection of B lymphocytes. Journal of Virology. 1997;71(6):4657-62.
18. Tugizov SM, Berline JW, Palefsky JM. Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. Nature Medicine. 2003;9(3):307.
19. Xiao J, Palefsky JM, Herrera R, Berline J, Tugizov SM. EBV BMRF-2 facilitates cell-to-cell spread of virus within polarized oral epithelial cells. Virology. 2009;388(2):335-43.
20. Chesnokova LS, Nishimura SL, Hutt-Fletcher LM. Fusion of epithelial cells by Epstein–Barr virus proteins is triggered by binding of viral glycoproteins gHgL to integrins αvβ6 or αvβ8. Proceedings of the National Academy of Sciences. 2009;106(48):20464-9.
21. Li H, Liu S, Hu J, Luo X, Li N, Bode AM, et al. Epstein-Barr virus lytic reactivation regulation and its pathogenic role in carcinogenesis. International Journal of Biological Sciences. 2016;12(11):1309.
22. Nakayama T, Fujisawa R, Izawa D, Hieshima K, Takada K, Yoshie O. Human B cells immortalized with Epstein-Barr virus upregulate CCR6 and CCR10 and downregulate CXCR4 and CXCR5. Journal of Virology. 2002;76(6):3072-7.
23. Korsman SNJ, van Zyl GU, Nutt L, Andersson MI, Preiser W. Epstein-Barr virus. In: Korsman SNJ, van Zyl GU, Nutt L, Andersson MI, Preiser W, editors. Virology. Edinburgh: Churchill Livingstone; 2012. p. 58-9.

24. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L. Lytic Cycle Switches of Oncogenic Human Gammaherpesviruses1. Advances in Cancer Research 2007;97:81-109.
25. Burkitt D. A sarcoma involving the jaws in African children. British Journal of Surgery. 1958;46(197):218-23.
26. Burkitt D, O'Conor GT. Malignant lymphoma in African children. I. A clinical syndrome. Cancer. 1961;14(2):258-69.
27. O'Conor GT. Malignant lymphoma in African children. II. A pathological entity. Cancer. 1961;14(2):270-83.
28. Epstein M, Barr Y, Achong B. A Second Virus-Carrying Tissue Culture Strain (EB2) of Lymphoblasts From Burkitt's Lymphoma. Pathologie et Biologie. 1964;12:1233.
29. Epstein MA. Virus particles in cultured lymphoblasts from Burkitt's lymphoma. The Lancet. 1964;1:702-3.
30. Lieberman PM. Epstein-Barr virus turns 50. Science. 2014;343(6177):1323-5.
31. Henle G, Henle W, Diehl V. Relation of Burkitt's tumor-associated herpes-ytpe virus to infectious mononucleosis. Proceedings of the National Academy of Sciences. 1968;59(1):94-101.
32. Neparidze N, Lacy J. Malignancies associated with epstein-barr virus: pathobiology, clinical features, and evolving treatments. Clinical Advances in Hematology & Oncology. 2014;12(6):358-71.
33. McLaughlin-Drubin ME, Munger K. Viruses associated with human cancer. Biochimica et Biophysica Acta. 2008;1782(3):127-50.
34. Hsu JL, Glaser SL. Epstein–Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. Critical Reviews in Oncology/Hematology. 2000;34(1):27-53.
35. Münz C. Latency and lytic replication in Epstein–Barr virus-associated oncogenesis. Nature Reviews Microbiology. 2019;17(11):691-700.
36. Kutok J, Wang F. Spectrum of Epstein-Barr virus–associated diseases. Annual Review of Pathology: Mechanisms of Disease. 2006;1:375-404.
37. Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. Frontiers in Oncology. 2019;9.
38. Delecluse HJ, Feederle R, O'Sullivan B, Taniere P. Epstein Barr virus-associated tumours: an update for the attention of the working pathologist. Journal of Clinical Pathology. 2007;60(12):1358-64.

39. Hochberg D, Middeldorp JM, Catalina M, Sullivan JL, Luzuriaga K, Thorley-Lawson DA. Demonstration of the Burkitt's lymphoma Epstein-Barr virus phenotype in dividing latently infected memory cells in vivo. Proceedings of the National Academy of Sciences. 2004;101(1):239-44.
40. Young LS, Murray PG. Epstein–Barr virus and oncogenesis: from latent genes to tumours. Oncogene. 2003;22(33):5108-21.
41. Murata T, Sato Y, Kimura H. Modes of infection and oncogenesis by the Epstein–Barr virus. Reviews in Medical Virology. 2014;24(4):242-53.
42. Katsumura KR, Maruo S, Takada K. EBV lytic infection enhances transformation of B-lymphocytes infected with EBV in the presence of T-lymphocytes. Journal of Medical Virology. 2012;84(3):504-10.
43. Kalla M, Schmeinck A, Bergbauer M, Pich D, Hammerschmidt W. AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. Proceedings of the National Academy of Sciences. 2010;107(2):850-5.
44. Ma SD, Hegde S, Young KH, Sullivan R, Rajesh D, Zhou Y, et al. A new model of Epstein-Barr virus infection reveals an important role for early lytic viral protein expression in the development of lymphomas. Journal of Virology. 2011;85(1):165-77.
45. Gottschalk S, Rooney CM. Adoptive T-Cell Immunotherapy. Current Topics in Microbiology and Immunology. 2015;391:427-54.
46. Wistinghausen B, Dave H, Bollard CM. Adoptive cellular immunotherapy for Epstein-Barr virus-associated lymphoproliferative disease. Annals of Lymphoma. 2022;6.
47. Heslop HE, Sharma S, Rooney CM. Adoptive T-Cell Therapy for Epstein-Barr Virus-Related Lymphomas. Journal of Clinical Oncology. 2021;39(5):514-24.
48. Bollard CM, Rooney CM, Heslop HE. T-cell therapy in the treatment of post-transplant lymphoproliferative disease. Nature Reviews Clinical Oncology. 2012;9(9):510-9.
49. Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. Journal of Clinical Oncology. 2014;32(8):798-808.
50. Cho SG, Kim N, Sohn HJ, Lee SK, Oh ST, Lee HJ, et al. Long-term Outcome of Extranodal NK/T Cell Lymphoma Patients Treated With Postremission Therapy Using EBV LMP1 and LMP2a-specific CTLs. Molecular Therapy. 2015;23(8):1401-9.

51. Gottschalk S, Edwards OL, Sili U, Huls MH, Goltsova T, Davis AR, et al. Generating CTLs against the subdominant Epstein-Barr virus LMP1 antigen for the adoptive immunotherapy of EBV-associated malignancies. Blood. 2003;101(5):1905-12.
52. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nature Reviews Gastroenterology & Hepatology. 2019;16(6):361-75.
53. Conejo-Garcia JR, Biswas S, Chaurio R. Humoral immune responses: Unsung heroes of the war on cancer. Seminars in Immunology. 2020;49:101419.
54. Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop AD. The immunology of Epstein-Barr virus–induced disease. Annual review of immunology. 2015;33:787-821.
55. Basso K, Dalla-Favera R. Germinal centres and B cell lymphomagenesis. Nature Reviews Immunology. 2015;15(3):172-84.
56. Stebegg M, Kumar SD, Silva-Cayetano A, Fonseca VR, Linterman MA, Graca L. Regulation of the Germinal Center Response. Frontiers in Immunology. 2018;9.
57. Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. Frontiers in Oncology. 2019;9:713.
58. Heath E, Begue-Pastor N, Chaganti S, Croom-Carter D, Shannon-Lowe C, Kube D, et al. Epstein-Barr virus infection of naïve B cells in vitro frequently selects clones with mutated immunoglobulin genotypes: implications for virus biology. PLoS Pathog. 2012;8(5):e1002697.
59. Inoue T, Moran I, Shinnakasu R, Phan TG, Kurosaki T. Generation of memory B cells and their reactivation. Immunological Reviews. 2018;283(1):138-49.
60. Kim SS, Sumner WA, Miyauchi S, Cohen EEW, Califano JA, Sharabi AB. Role of B Cells in Responses to Checkpoint Blockade Immunotherapy and Overall Survival of Cancer Patients. Clinical Cancer Research. 2021;27(22):6075-82.
61. DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B, et al. Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. Clinical Immunology. 2018;187:37-45.
62. Bruno TC, Ebner PJ, Moore BL, Squalls OG, Waugh KA, Eruslanov EB, et al. Antigen-Presenting Intratumoral B Cells Affect CD4+ TIL Phenotypes in Non–Small Cell Lung Cancer PatientsTIL-Bs Present Antigen to CD4 TILs in NSCLC. Cancer Immunology Research. 2017;5(10):898-907.

63. Kinker GS, Vitiello GAF, Ferreira WAS, Chaves AS, Cordeiro de Lima VC, Medina TDS. B Cell Orchestration of Anti-tumor Immune Responses: A Matter of Cell Localization and Communication. Frontiers in Cell and Developmental Biology. 2021;9:678127.
64. De Paschale M, Clerici P. Serological diagnosis of Epstein-Barr virus infection: Problems and solutions. World Journal of Virology. 2012;1(1):31-43.
65. Middeldorp JM. Epstein-Barr virus-specific humoral immune responses in health and disease. Epstein Barr Virus Volume 2. 2015:289-323.
66. Rea TD, Ashley RL, Russo JE, Buchwald DS. A systematic study of Epstein-Barr virus serologic assays following acute infection. American Journal of Clinical Pathology. 2002;117(1):156-61.
67. Gulley ML, Tang W. Laboratory Assays for Epstein-Barr Virus-Related Disease. The Journal of Molecular Diagnostics. 2008;10(4):279-92.
68. Coghill AE, Hildesheim A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. American Journal of Epidemiology. 2014;180(7):687-95.
69. Mantis NJ, Rol N, Corthésy B. Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. Mucosal Immunology. 2011;4:603.
70. Cohen JI. Epstein–Barr Virus Infection. New England Journal of Medicine. 2000;343(7):481-92.
71. Coghill AE, Hsu WL, Yang Q, Wang CP, Lou PJ, Yu KJ, et al. Elevated antibodies against Epstein-Barr virus among individuals predicted to carry nasopharyngeal carcinoma susceptibility variants. Journal of General Virology. 2018;99(9):1268-73.
72. Qi H, Wang F, Tao SC. Proteome microarray technology and application: higher, wider, and deeper. Expert Review of Proteomics. 2019;16(10):815-27.
73. Zhu H, Snyder M. Protein arrays and microarrays. Current opinion in chemical biology. 2001;5(1):40-5.
74. Natesan M, Ulrich RG. Protein microarrays and biomarkers of infectious disease. International Journal of Molecular Sciences. 2010;11(12):5165-83.
75. Ramachandran N, Srivastava S, Labaer J. Applications of protein microarrays for biomarker discovery. Proteomics - Clinical Applications. 2008;2(10-11):1444-59.
76. Steller S, Angenendt P, Cahill DJ, Heuberger S, Lehrach H, Kreutzberger J. Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections. Proteomics. 2005;5(8):2048-55.

77. Sundaresh S, Randall A, Unal B, Petersen JM, Belisle JT, Hartley MG, et al. From protein microarrays to diagnostic antigen discovery: a study of the pathogen Francisella tularensis. Bioinformatics. 2007;23(13):i508-18.
78. Barbour AG, Jasinskas A, Kayala MA, Davies DH, Steere AC, Baldi P, et al. A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with Borrelia burgdorferi. Infection and Immunity. 2008;76(8):3374-89.
79. Hu S, Xie Z, Qian J, Blackshaw S, Zhu H. Functional protein microarray technology. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2011;3(3):255-68.
80. Scheich C, Sievert V, Büssow K. An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. BMC Biotechnology. 2003;3(1):12.
81. Chandra H, Srivastava S. Cell-free synthesis-based protein microarrays and their applications. Proteomics. 2010;10(4):717-30.
82. Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clinical Cancer Research. 2018:clincanres. 1929.2017.
83. Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, et al. The association between the comprehensive Epstein–Barr virus serologic profile and endemic Burkitt lymphoma. Cancer Epidemiology and Prevention Biomarkers. 2020;29(1):57-62.
84. Liu Z, Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, et al. Patterns of Interindividual Variability in the Antibody Repertoire Targeting Proteins Across the Epstein-Barr Virus Proteome. The Journal of Infectious Diseases. 2018;217(12):1923-31.
85. Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. International Journal of Cancer. 2020;147(3):608-18.
86. Global tuberculosis report 2021 2021 [cited 2022 October 31]. Available from: https://www.who.int/publications/i/item/9789240037021.
87. Global tuberculosis report 2022 2022 [cited 2022 October 31]. Available from: https://www.who.int/publications/i/item/9789240061729.
88. Stop TB Partnership sounds the alarm on the TB funding crisis as TB deaths increase for a second year in a row and TB incidence rises for the first time in 20 years 2022 [cited 2022 25 Nov]. Available from: https://www.stoptb.org/news/stop-tb-partnership-sounds-

alarm-tb-funding-crisis-tb-deaths-increase-second-year-row-and-tb#:~:text=The%20TB%20incidence%20rate%20(new,of%20the%20past%202%20decades.
89. Sabbatani S. Historical insights into tuberculosis. Girolamo Fracastoro's intuition on the transmission of tuberculosis and his opponents. History of an idea. Le infezioni in medicina. 2004;12(4):284-91.
90. Koch R. The etiology of tuberculosis. Berl Klin Wochenschr. 1882;15:221-30.
91. Wlodarska M, Johnston JC, Gardy JL, Tang P. A Microbiological Revolution Meets an Ancient Disease: Improving the Management of Tuberculosis with Genomics. Clinical Microbiology Reviews. 2015;28(2):523-39.
92. Basel H. History of tuberculosis. Respiration. 1998;65(1):5-15.
93. Banuls A-L, Sanou A, Van Anh NT, Godreuil S. *Mycobacterium tuberculosis*: ecology and evolution of a human bacterium. Journal of Medical Microbiology. 2015;64(11):1261-9.
94. Torrelles JB, Schlesinger LS. Integrating lung physiology, immunology, and tuberculosis. Trends in Microbiology. 2017;25(8):688-97.
95. Transmission and Pathogenesis of Tuberculosis. Core curriculum on tuberculosis. 6 th ed. Atlanta: Centers for Disease Control and Prevention National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of Tuberculosis Elimination; 2013.
96. Ramirez-Lapausa M, Menendez-Saldana A, Noguerado-Asensio A. Extrapulmonary tuberculosis: an overview. Revista Española de Sanidad Penitenciaria. 2015;17:3-11.
97. Henderson HJ, Dannenberg AM, Lurie MB. Phagocytosis of tubercle bacilli by rabbit pulmonary alveolar macrophages and its relation to native resistance to tuberculosis. The Journal of Immunology. 1963;91(4):553-6.
98. Mohagheghpour N, van Vollenhoven A, Goodman J, Bermudez LE. Interaction of *Mycobacterium avium* with human monocyte-derived dendritic cells. Infection and Immunity. 2000;68(10):5824-9.
99. Sullivan JT, Young EF, McCann JR, Braunstein M. The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. Infection and Immunity. 2012;80(3):996-1006.
100. Shiloh MU. Mechanisms of mycobacterial transmission: how does *Mycobacterium tuberculosis* enter and escape from the human host. Future Microbiology. 2016.
101. Ravimohan S, Kornfeld H, Weissman D, Bisson GP. Tuberculosis and lung damage: from epidemiology to pathophysiology. European Respiratory Review. 2018;27(147):170077.

102. Ramakrishnan L. Revisiting the role of the granuloma in tuberculosis. Nature Reviews Immunology. 2012;12(5):352.
103. Mack U, Migliori G, Sester M, Rieder H, Ehlers S, Goletti D, et al. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. European Respiratory Journal. 2009;33(5):956-73.
104. Tuberculosis Fact sheets 2018: World Health Organisation; 2018 [cited 2019 08 May]. Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
105. Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent *Mycobacterium tuberculosis* infection. New England Journal of Medicine. 2015;372(22):2127-35.
106. Latent tuberculosis infection (LTBI) Fact sheet 2015: World Health Organization; 2016 [cited 2019 10 May]. Available from: https://www.who.int/tb/areas-of-work/preventive-care/ltbi/faqs/en/.
107. Al-Orainey IO. Diagnosis of latent tuberculosis: Can we do better? Annals of thoracic medicine. 2009;4(1):5.
108. Nayak S, Acharjya B. Mantoux test and its interpretation. Indian Dermatology Online Journal. 2012;3(1):2.
109. Loh K. Role of Mantoux test in the diagnosis of tuberculosis. Malaysia Family Physician. 2011;6(2-3):85.
110. Vukmanovic-Stejic M, Reed JR, Lacy KE, Rustin MH, Akbar AN. Mantoux test as a model for a secondary immune response in humans. Immunology Letters. 2006;107(2):93-101.
111. Wang L, Turner M, Elwood R, Schulzer M, FitzGerald J. A meta-analysis of the effect of Bacille Calmette Guerin vaccination on tuberculin skin test measurements. Thorax. 2002;57(9):804-9.
112. Regatieri A, Abdelwahed Y, Perez MT, Bush LM. Testing for tuberculosis: the roles of tuberculin skin tests and interferon gamma release assays. Laboratory Medicine. 2011;42(1):11-6.
113. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? International Journal of Tuberculosis and Lung Disease. 2006;10(11):1192-204.
114. Castro KG, Goldberg S, Jereb JA, LoBue P, Mazurek GH, Vernon A. Updated guidelines for using interferon gamma release assays to detect *Mycobacterium tuberculosis* infection--United States, 2010. Atlanta; 2010.

115. Andersen P, Munk M, Pollock J, Doherty T. Specific immune-based diagnosis of tuberculosis. The Lancet. 2000;356(9235):1099-104.
116. Abubakar I, Stagg HR, Whitworth H, Lalvani A. How should I interpret an interferon gamma release assay result for tuberculosis infection? Thorax. 2013;68(3):298-301.
117. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma interferon release assays for detection of *Mycobacterium tuberculosis* infection. Clinical Microbiology Reviews. 2014;27(1):3-20.
118. Mori T, Sakatani M, Yamagishi F, Takashima T, Kawabe Y, Nagao K, et al. Specific detection of tuberculosis infection: an interferon-γ–based assay using new antigens. American Journal of Respiratory and Critical Care Medicine. 2004;170(1):59-64.
119. Kang YA, Lee HW, Yoon HI, Cho B, Han SK, Shim Y-S, et al. Discrepancy between the tuberculin skin test and the whole-blood interferon γ assay for the diagnosis of latent tuberculosis infection in an intermediate tuberculosis-burden country. The Journal of the American Medical Association. 2005;293(22):2756-61.
120. Lu LL, Smith MT, Yu KKQ, Luedemann C, Suscovich TJ, Grace PS, et al. IFN-γ-independent immune markers of *Mycobacterium tuberculosis* exposure. Nature Medicine. 2019;25(6):977-87.
121. Fan L, Chen Z, Hao X-H, Hu Z-Y, Xiao H-P. Interferon-gamma release assays for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. FEMS Immunology & Medical Microbiology. 2012;65(3):456-66.
122. Sester M, Sotgiu G, Lange C, Giehl C, Girardi E, Migliori GB, et al. Interferon-γ release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis. European Respiratory Journal. 2011;37(1):100-11.
123. Metcalfe JZ, Everett CK, Steingart KR, Cattamanchi A, Huang L, Hopewell PC, et al. Interferon-γ Release Assays for Active Pulmonary Tuberculosis Diagnosis in Adults in Low- and Middle-Income Countries: Systematic Review and Meta-analysis. The Journal of Infectious Diseases. 2011;204(suppl_4):S1120-S9.
124. Mazurek GH, Weis SE, Moonan PK, Daley CL, Bernardo J, Lardizabal AA, et al. Prospective Comparison of the Tuberculin Skin Test and 2 Whole-Blood Interferon-γ Release Assays in Persons with Suspected Tuberculosis. Clinical Infectious Diseases. 2007;45(7):837-45.
125. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research. 2008;18(9):1509-17.

126. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods. 2008;5(7):621-8.
127. Burel JG, Babor M, Pomaznoy M, Lindestam Arlehamn CS, Khan N, Sette A, et al. Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. Frontiers in Immunology. 2019;10(221).
128. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. PLoS Computational Biology. 2017;13(5):e1005457-e.
129. Ravasi T, Mavromatis CH, Bokil NJ, Schembri MA, Sweet MJ. Co-transcriptomic analysis by RNA sequencing to simultaneously measure regulated gene expression in host and bacterial pathogen. Toll-Like Receptors: Springer; 2016. p. 145-58.
130. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
131. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973.
132. Maertzdorf J, Weiner J, Mollenkopf H-J, Network T, Bauer T, Prasse A, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proceedings of the National Academy of Sciences. 2012;109(20):7853-8.
133. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Medicine. 2013;10(10):e1001538.
134. Lee S-W, Wu LS-H, Huang G-M, Huang K-Y, Lee T-Y, Weng JT-Y. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. BMC bioinformatics. 2016;17 Suppl 1(Suppl 1):3-.
135. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, et al. Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification. bioRxiv. 2019:583674.
136. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Medicine. 2019;16(4):e1002786.

Chapter 2

Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma

This chapter titled work 'Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma' has been published in *Scientific Reports* on 8th December 2021 (DOI: https://doi.org/10.1038/s41598-021-02788-w).

2 Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma

Zhiwei Liu[1 *], Yomani D. Sarathkumara[2 *], John K.C. Chan[3 *], Yok-Lam Kwong[4 *], Tai Hing Lam[5 *], Dennis Kai Ming Ip[5 *], Brian C.-H. Chiu [6 *], Jun Xu[5], Yu-Chieh Su[7,8], Carla Proietti[2], Martha M. Cooper[2], Kelly J. Yu[1], Bryan Bassig[1], Raymond Liang[9], Wei Hu[1], Bu-Tian Ji[1], Anna E. Coghill[10], Ruth M. Pfeiffer[1], Allan Hildesheim[1 †], Nathaniel Rothman[1 †], Denise L. Doolan[2 †], and Qing Lan[1 †]

[1]. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA

[2]. Centre for Molecular Therapeutics, Australian Institute of Tropical Health of Medicine, James Cook University, Cairns, Australia

[3]. Department of Pathology, Queen Elizabeth Hospital, Hong Kong

[4]. Queen Mary Hospital, The University of Hong Kong, Hong Kong

[5]. School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The University of Hong Kong, Hong Kong

[6]. Department of Public Health Sciences, University of Chicago, Chicago, USA

[7]. Department of Medicine, School of Medicine, I-Shou University, Kaohsiung, Taiwan

[8]. Division of Hematology-Oncology, Department of Internal Medicine, E-Da Hospital, Kaohsiung, Taiwan

[9]. Hong Kong Sanatorium & Hospital, Hong Kong

10. Cancer Epidemiology Program, Division of Population Sciences, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

***** These authors contributed equally as first authors

† These authors contributed equally as last authors

2.1 Abstract

Extranodal natural killer/T-cell lymphoma (NKTCL) is an aggressive malignancy that has been etiologically linked to Epstein-Barr virus (EBV) infection, with EBV gene transcripts identified in almost all cases. However, the humoral immune response to EBV in NKTCL patients has not been well characterized. We examined the antibody response to EBV in plasma samples from 51 NKTCL cases and 154 controls from Hong Kong and Taiwan who were part of the multi-center, hospital-based AsiaLymph case-control study. The EBV-directed serological response was characterized using a protein microarray that measured IgG and IgA antibodies against 202 protein sequences representing the entire EBV proteome. We analyzed 157 IgG antibodies and 127 IgA antibodies that fulfilled quality control requirements. Associations between EBV serology and NKTCL status were disproportionately observed for IgG rather than IgA antibodies. Nine anti-EBV IgG responses were significantly elevated in NKTCL cases compared with controls and had ORs $_{\text{highest vs. lowest tertile}} > 6.0$ (Bonferroni-corrected p-values<0.05). Among these nine elevated IgG responses in NKTCL patients, three IgG antibodies (all targeting EBNA3A) are novel and have not been observed for other EBV-associated tumours of B-cell or epithelial origin. IgG antibodies against EBNA1, which have consistently been elevated in other EBV-associated tumours, were not elevated in NKTCL cases. We characterize the antibody response against EBV for patients with NKTCL and identify IgG antibody responses against six distinct EBV proteins. Our findings suggest distinct serologic patterns of this NK/T-cell lymphoma compared with other EBV-associated tumours of B-cell or epithelial origin.

Key Words: Extranodal natural killer/T-cell lymphoma, Epstein-Barr Virus, antibody, microarray

2.2 Introduction

Extranodal natural killer T-cell lymphoma (NKTCL; nasal type) is an aggressive malignancy that has been closely linked to infection with Epstein-Barr virus (EBV) (1). Nearly all NKTCL is EBV positive, with EBV gene transcripts identified in almost 100% of NKTCL tumours (2, 3). EBV establishes lifelong latency in B cells in over 90% of adults worldwide but causes cancer in only a small fraction of infected individuals (4). EBV-associated tumours include a subset of Hodgkin and non-Hodgkin lymphoma, as well as epithelial carcinomas of the nasopharynx and stomach (4). Like EBV-positive Hodgkin lymphoma (HL) and nasopharyngeal carcinoma (NPC), EBV-infected cells in patients with NKTCL have been observed to express genes of latency I (EBNA1 and EBER1/2) or latency II (LMP1/2A/2B, EBNA1, and EBER1/2) (5, 6). However, the specific role of EBV in the pathogenesis of NKTCL is still poorly understood.

NK and T cells are typically not permissive of EBV infection and, consequently, EBV is not detected in NK or T cells in the blood of healthy carriers, and is only detected at low frequency in tonsillar NK or T cells (7). A recent study suggested that EBV can infect mature peripheral T cells via binding of EBV glycoprotein gp350 to the cellular membrane protein CD21 (8), an established receptor for EBV infection of B-cells. However, EBV's role in NKTCL compared to B-cell lymphomas may differ following initial infection. It is possible that viral protein production is distinct following infection of T-cells, leading to different immune targets against which infected persons mount an antibody response. Study of the humoral (antibody) responses against EBV in patients with NKTCL, and comparison of these patterns to those observed in other EBV-associated cancers, could shed light on pathogenic mechanisms.

The humoral immune response to EBV in NKTCL patients is poorly characterized, with three case-only studies inclusive of a total of 155 patients reported to date (9-11). Those studies found suggestive elevations in antibody levels against viral capsid antigen (VCA) and early antigen (EA) but not EBV nuclear antigen (EBNA). That pattern is distinct from that observed in other EBV-related cancers including nasopharyngeal carcinoma (NPC), Burkitt lymphoma (BL), and Hodgkin lymphoma (HL) (12-15). In-depth, comprehensive characterization of serologic profiles that associate with NKTCL, and noting those that are distinct from other EBV-related cancers, could provide insight into the role of specific EBV proteins in the etiology of NKTCL. To investigate this, we utilized a multiplex technology targeting antibody responses to 202 peptide sequences representing the entire EBV proteome to comprehensively

evaluate patterns of anti-EBV antibody responses in 205 adults from Hong Kong and Taiwan, including 51 NKTCL cases and 154 matched controls.

2.3 Materials and Methods

2.3.1 Study Population

Plasma samples from 51 NKTCL cases and 154 control adults collected as part of the AsiaLymph, a multi-center hospital-based case-control study in Hong Kong and Taiwan conducted between 2012 and 2017, were selected for study. Eligible cases were aged between 18 and 79 years at diagnosis and living in the geographic region served by the partnering hospital at the time of cancer diagnosis. Cases with a prior history of lymphoma were ineligible. Blood and buccal cell collection were performed at the time of diagnosis and before receiving cancer therapy. Controls were apparently healthy individuals with no symptoms of disease (disease-free/NKTCL-free controls) who were a subset of patients seen at the same partnering hospital for diseases/conditions that were not associated with risk factors under study, including injuries and selected diseases of the circulatory, digestive, genitourinary, and central nervous system. Patients with a history of any lymphoma were not eligible for controls. Of all controls recruited in the two regions (N=1496; 1119 from Hong Kong and 377 from Taiwan), we randomly selected 154 subjects who were frequency-matched to cases on sex, age (+/- 5 years), date of enrolment (within 3 months), and region (Hong Kong/Taiwan).

The study was approved by the institutional review boards at each participating site, and the US National Institutes of Health and US National Cancer Institute. Written informed consent was obtained from all participants. All laboratory testing was conducted under a protocol approved by James Cook University Human Research Ethics Committee. All methods were performed in accordance with the Declarations of Helsinki.

2.3.2 EBV protein microarray

The comprehensive EBV protein microarray chip used in this study has been described in detail previously (12, 16). Briefly, this microarray contains 202 protein sequences representing almost the entire EBV proteome, including 199 EBV protein sequences generated from five different EBV strains (AG876, Akata, B95-8, Mutu, and Raji) and three synthetic EBV peptides for which circulating antibodies are putative cancer biomarkers (VCAp18,

EBNA1, and EAd p47). The 202 sequences represent each of the known open reading frames for EBV, as well as predicted splice variants of those open reading frames. Each of the protein sequences were cloned into the pXT7 expression vector, expressed using the E. coli cell-free protein system, and printed onto the microarray. Sequences include N-terminal 10x histidine (His) and C-terminal hemagglutinin (HA) tags for quality control and to confirm expression on the microarray. High coverage was achieved across the five prototypical EBV strains and ten Chinese strains, with >97% of the predicted sequences from each strain represented on the microarray at >99% homology. Four "noDNA" (no translated protein) spots were included to assess person-specific background.

Plasma samples from each of the study participants were tested on this EBV protein microarray as described previously (17). Slides were scanned on an Axon GenePix 4300B (Molecular Devices, Australia); raw fluorescence intensities were corrected for spot-specific background; corrected data were transformed using variance stabilizing normalization (vsn) in Gmine (18); and output was standardized to person-specific background (mean ±1.5 SD of the four "no DNA" spots). Positivity was defined as a standardized signal intensity >1.0. The standardized signal intensity for each spot was further grouped into three categories, with cut-offs for the categories defined using tertiles of the antibody distribution among the 154 controls.

Thirty-five samples were tested in duplicate, blinded to laboratory personnel, in order to assess assay reproducibility specific to this study population. The average coefficient of variation (CV) across the 202 EBV sequences was 16% [interquartile range (IQR), 14%–20%] for IgG antibody response and 19% (IQR, 16%–22%) for IgA antibody response, demonstrating a good reproducibly of our assay. We excluded 45 IgG and 75 IgA that had CVs>20%, leaving a total of 157 IgG and 127 IgA antibodies for further analysis.

2.3.3 Antibody testing using ELISA kits

To internally validate the serological findings from the EBV microarray for putative cancer biomarkers, we utilized commercial ELISA assays to test for IgG and IgA antibodies against recombinant VCA and EBNA1; these two antigens have been extensively investigated in other EBV-related cancers (19). ELISA assays were purchased from EUROIMMUN, Lübeck, Germany (IgG/IgA antibodies against VCA and IgG antibodies against EBNA1) and Zhongshan Biotech, Zhongshan, China (IgA antibodies against EBNA1) (20, 21). All samples were tested according to the manufacturers' instructions. Levels of antibodies were assessed by optical density (OD) values. Reference ODs (rODs) were obtained according to the

manufacturers' instructions by dividing OD values by a reference control. The same thirty-five blinded duplicates tested by microarray were also tested by ELISA to assess assay reproducibility. The CVs for IgG antibodies against VCA and EBNA1 were 6.9% and 7.7%, respectively; for IgA antibodies, CVs against VCA and EBNA1 were 19.1% and 25.1%, respectively.

2.3.4 Statistical analysis

Differences in the mean standardized signal intensity between NKTCL patients and controls were assessed using an unpaired Student t test. Case–control differences were considered statistically significant at the $P<0.0002$ threshold (equivalent to Bonferroni-corrected $P<0.05$) to account for the number of comparisons. Odds ratios (ORs) quantifying the association between the three-level categorical variable for each antibody and NKTCL status were estimated using logistic regression models adjusted for sex, age group (18-39, 40-49, 50-59, 60-80 years), and region. In previous work, no sociodemographic or environmental factors were found to strongly and consistently correlate with elevated anti-EBV antibody responses other than smoking (16, 22, 23); however, smoking was not associated with NKTCL in a previous study (24) and therefore was not included in our regression models. *P*-trend values were calculated from a model with each three-level antibody marker treated as an ordinal variable. Antibodies with *P*-trend<0.0002 threshold (equivalent to Bonferroni-corrected $P<0.05$) were considered as statistical significance. For results from the ELISA assays, differences in the mean rOD between NKTCL patients and controls were assessed using an unpaired Student t test.

To identify the anti-EBV IgG antibodies that are most informative for distinguishing NKTL cases from controls, we employed sparse Partial Least Squares Discriminant Analysis (sPLS-DA), which was implemented using the splda function in MixOmics R package (25, 26). The sPLS-DA is a method for identifying the key variables of complex and sparse omics datasets that are associated with a biological outcome of interest and it has been shown to be successful with applications where the number of features far outnumber the number of samples (27). This procedure involves dimension reduction using Partial Least Squares regression (PLS) for discriminant analysis in combination with a Lasso penalization for feature selection. The number of features selected per component was optimized using 10-fold cross validation repeated 5 times and the number associated with the lowest classification error rate

was chosen for the final model. The final model was then applied to the entire dataset to obtain the most important anti-EBV IgG antibodies in distinguishing NKTCL cases from controls.

Amongst controls, we estimated the correlation between antibodies using Spearman correlation coefficients. We also evaluated whether previously reported NKTCL-associated genetic variants (*i.e.*, rs13015714, mapped to IL18RAP, and rs9271588, mapped to HLA-DRB1) were associated with the level of anti-EBV antibody response using linear regression models adjusted for sex, age group (18-39, 40-49, 50-59, 60-80 years), and region.

2.4 Results

Table 2-1 shows the distributions of demographic characteristic in 51 NKTCL cases and 154 matched controls from Hong Kong and Taiwan. Cases and controls had a similar sex, age, and study region distribution, reflective of the matched study design. Approximately two thirds of adults recruited were male, and 78.4% of cases (40/51) were recruited in Hong Kong.

Table 2-1 Characteristics of study population, by NK-T cell lymphoma (NKTCL) status in Hong Kong and Taiwan.

Characteristics	NKTCL cases (N=51)	Controls (N=154)
Sex		
Male	34 (66.7)	102 (66.2)
Female	17 (33.3)	52 (33.8)
Age at diagnosis/selection (years)		
18-39	12 (23.5)	34 (22.1)
40-49	13 (25.5)	42 (27.3)
50-59	11 (21.6)	33 (21.4)
60-80	15 (29.4)	45 (29.2)
Region		
Hong Kong	40 (78.4)	123 (79.9)
Taiwan	11 (21.6)	31 (20.1)

NKTCL associations were disproportionately observed for IgG rather than IgA antibodies. Case–control comparisons of the mean standardized signal intensity for the 157 IgG

and 127 IgA antibodies on the array revealed nominal ($P < 0.05$) elevations in 52 IgG antibodies but only six IgA antibodies (**Figure 2-1**). Six anti-EBV IgG antibodies were significantly elevated in NKTCL cases compared to controls after adjustment for multiple testing ($P < 0.0002$; **Figure 2-1**). Results from the remaining 46 anti-EBV IgG and six anti-EBV IgA antibodies that were nominally significantly elevated in NKTCL cases compared to controls ($P<0.05$) are shown in **Supplementary Table 1**. Of note, we did not observe differences in anti-EBV EBNA1 IgG responses between NKTCL cases and controls (**Supplementary Table 2**).

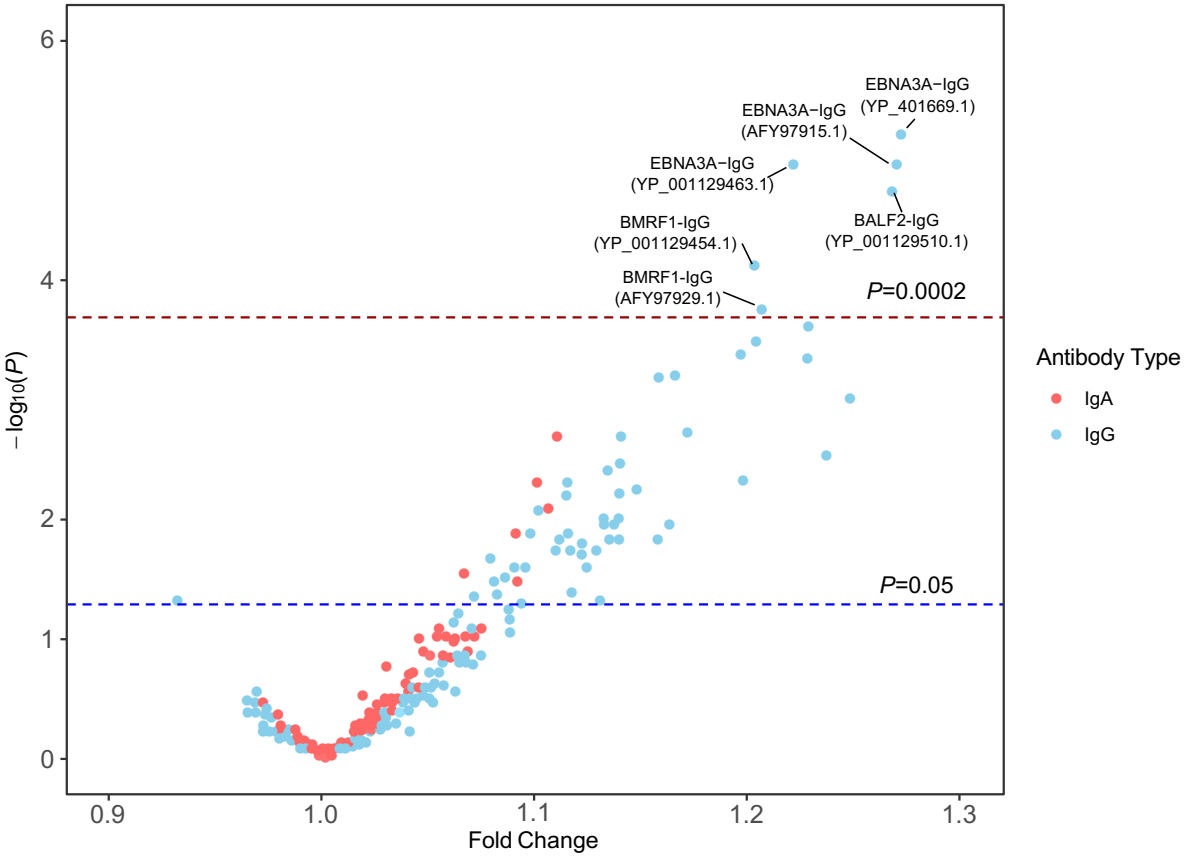


Figure 2-1 Case–control differences in the mean antibody response for 51 NK/T-cell lymphoma (NKTCL) cases versus 154 controls collected in Hong Kong and Taiwan.
The x-axis displays the fold change (case vs. control ratio of standardized signal intensity) for all antibodies with CV ≤ 20%. The y-axis illustrates the p value corresponding to the t-test for a difference in standardized signal intensity between cases and controls. Six IgG-antibodies but no IgA antibodies were significantly elevated in NKTCL cases compared to controls at the $p < 0.0002$ (Bonferroni-corrected $p < 0.05$) threshold.

Table 2-2 OR and 95% CI for the association between anti-EBV antibody level and NK-T cell lymphoma (NKTCL) in Hong Kong and Taiwan. [a]

EBV protein and array sequence	Antibody type	t test *P*	NKTCL mean (SD)	Control mean (SD)	Fold change	NKTCL positivity	Control positivity	OR tertile 2 (95% CI) [b]	OR tertile 3 (95% CI) [b]	*P*-trend [c]
EBNA3A (YP_401669.1-80382-82877)	IgG	5.99×10^{-6}	1.76 (0.47)	1.38 (0.54)	1.27	96.1%	75.3%	2.44 (0.80-7.45)	6.59 (2.38-18.22)	6.51×10^{-5}
EBNA3A (AFY97915.1-80252-82747)	IgG	1.06×10^{-5}	1.68 (0.46)	1.32 (0.52)	1.27	94.1%	64.3%	4.79 (1.29-17.73)	11.14 (3.21-38.72)	1.84×10^{-5}
EBNA3A (YP_001129463.1-80447-82888)	IgG	1.08×10^{-5}	1.85 (0.44)	1.51 (0.48)	1.22	98.0%	88.3%	8.48 (1.83-39.22)	16.33 (3.71-71.91)	1.63×10^{-5}
BALF2 [EA(D)_p138] (YP_001129510.1-165796-162410-1)	IgG	1.79×10^{-5}	1.37 (0.39)	1.08 (0.41)	1.27	80.4%	52.6%	2.34 (0.75-7.28)	7.29 (2.60-20.43)	3.03×10^{-5}
BMRF1 [EA(D)_p47] (YP_001129454.1-67745-68959)	IgG	7.64×10^{-5}	1.78 (0.44)	1.48 (0.49)	1.20	96.1%	92.9%	2.52 (0.82-7.76)	6.83 (2.45-19.08)	5.70×10^{-5}
BMRF1 [EA(D)_p47] (AFY97929.1-67486-68700)	IgG	1.81×10^{-4}	1.67 (0.45)	1.38 (0.47)	1.21	94.1%	84.4%	2.88 (0.96-8.62)	6.32 (2.27-17.61)	1.60×10^{-4}
BZLF1 [Zebra (Zta)] (YP_001129467.1-91697-91197)	IgG	4.19×10^{-4}	1.49 (0.42)	1.24 (0.39)	1.20	96.1%	74.7%	4.85 (1.3-18.09)	11.13 (3.19-38.78)	1.99×10^{-5}
BVRF2 [**VCAp40]** (YP_001129501.1-136465-138282)	IgG	6.64×10^{-4}	1.74 (0.40)	1.50 (0.46)	1.17	100.0%	95.5%	2.92 (0.97-8.79)	6.75 (2.39-19.03)	1.19×10^{-4}
BPLF1 [**Tegument protein**] (CAA24839.1-71527-62078-2)	IgG	5.82×10^{-3}	1.93 (0.40)	1.73 (0.53)	1.11	98.0%	98.7%	2.20 (0.75-6.42)	6.00 (2.25-16.01)	1.17×10^{-4}

Note: Bold text is used to highlight the canonical EBV protein name. The remaining (non-bolded) text describes the sequence details of the array probe.

Abbreviation: CI, confidence interval. SD, standard deviation

a. The table is ordered by t test *P* value (lowest to highest).

b. The odds of being a NKTCL case were calculated from a logistic regression model that included age group (18-39, 40-49, 50-59, 60-80 years), sex, region, and a three-level variable (tertiles) for anti-EBV antibody level. The tertiles were calculated using the underlying antibody distribution among disease-free controls. All ORs are expressed relative to the referent group of tertile 1 (lowest third of antibody distribution).

c. Two-sided *P* values for trend across marker categories were assessed with the Wald test using categorical values of the proteins with 1 degree of freedom.

Using logistic regression models with adjustment for sex, age, and study region, in addition to the six significant anti-EBV IgG antibodies mentioned above, we identified ab additional three IgG antibodies that were significantly elevated in NKTCL cases compared with controls. These elevations had at least a 6-fold risk ($OR_{\text{highest vs. lowest tertile}} \geq 6.0$, **Table 2-2**). The most significant *P* value was observed for IgG antibody against latent protein EBNA3A (one of three variants shown in **Figure 2-2A**). Accordingly, the strongest OR effect was observed for antibody against sequences representing EBNA3A (adjusted $OR_{\text{highest vs. lowest tertile}}$ = 16.33, 95% confidence interval [CI]: 3.71 to 71.91, *P*-trend=1.6×10^{-5}), a protein expressed in latency IIb and III phases that has not been found to be strongly associated with other EBV-associated tumors of B-cell or epithelial origin (12-14). Pronounced *p* values were also observed for the early lytic proteins BALF2 [EA(D)p138] (one representative variant shown in **Figure 2-2B)** and BMRF1 [EA(D)p47 (one of two variants shown in **Figure 2-2C**). Other IgG antibodies significantly and markedly elevated in NKTCL patients included those targeting antigens representing immediate early and late lytic proteins, BZLF2 [Zebra (Zta)], BVRF2 [VCAp40] and BPLF1 [Tegument protein] (**Table 2-2 and Figure 2-2D-F**).

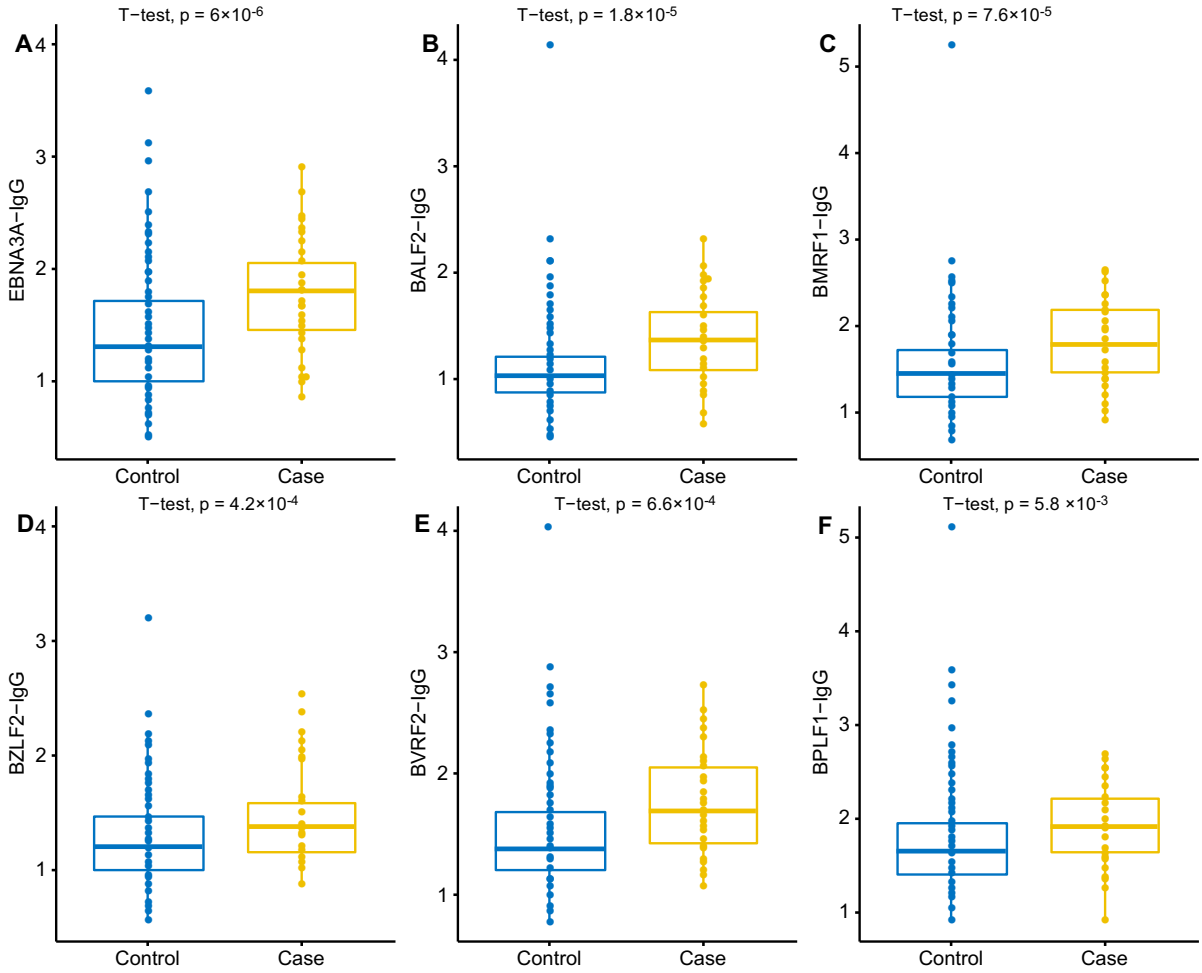


Figure 2-2 Signal intensity for the six significant anti-EBV IgG antibodies between NK/T-cell lymphoma (NKTCL) and controls.

A) EBNA3A-IgG, B) BALF2-IgG, C) BMRF1-IgG, D) BZLF2-IgG, E) BVRF2-IgG, and F) BPLF1-IgG. P-values from the t-test are listed.

We next examined the correlations between these nine highly differentially expressed anti-EBV IgG antibodies. Strong correlations were observed for antibodies targeting the same antigens (*i.e.*, three variants for EBNA3A and two variants for BMRF1), with correlations ranging from 0.903 to 0.966. More modest correlation was observed between antibodies targeting different antigens, with correlations ranging from 0.313 to 0.747 (**Supplementary Figure 1**). In a logistic regression model excluding 3 IgG antibodies (two antibodies against EBNA3A and one against BMRF1) that were highly correlated with antibodies targeting the same antigens, IgG antibodies against EBNA3A, BALF2, and BPLF1 retained statistical significance ($P < 0.05$).

In the sPLS-DA analysis, the top 10 anti-EBV IgG antibodies that were most informative for classifying NKTCL status were those targeting EBNA3A, BALF2, BRLF1, thymidine kinase (TK), BMRF1, and BZLF1 (**Supplementary Figure 2**), largely consistent with the most significant antibodies defined using the t-test.

Results from ELISA assays confirmed our array-based findings. For example, we observed that VCA-IgG was significantly elevated among NKTCL cases compared with controls (P=5.2×10^{-8}, **Supplementary Figure 3A**). There was suggestive evidence that VCA-IgA was also elevated (P=0.003) but that association was not statistically significant after adjustment for multiple testing (**Supplementary Figure 3B**). Antibodies against EBNA1 (both IgG and IgA) measured by ELISA were not elevated among NKTCL cases compared with controls (**Supplementary Figures 3C-3D**).

Finally, as an exploratory analysis, we leveraged genotyping data (28) from 94 controls included in the present study and observed that SNP rs9271588 (which maps to *HLA-DRB1*) was suggestively correlated with the most differentially expressed EBV-antibody EBNA3A-IgG (P=0.06).

2.5 Discussion

This is, to our knowledge, the first study to comprehensively evaluate EBV-directed immunity in adults diagnosed with NKTCL in Asia. We investigated both IgG and IgA responses to each protein expressed in the EBV proteome. Profound differences in the anti-EBV antibody profile between NKTCL patients and matched controls were demonstrated, with significantly elevated IgG antibody responses against six distinct EBV proteins. Notably, the strongest NKTCL–EBV associations mapped to sequences representing EBNA3A (but not EBNA1), suggesting a possible role of this latent protein in disease pathogenesis.

In addition to IgG, we examined anti-EBV IgA antibodies in the context of NKTCL. IgA reflects recent exposure along mucosal surfaces such as the oral epithelium and has proven to be an informative biomarker for EBV-associated epithelial tumours (*i.e.*, nasopharyngeal carcinoma) (12, 29). However, IgA responses did not significantly differ between NKTCL patients and controls after correction for multiple testing. Although false negative findings cannot be entirely ruled out due to a modest sample size and relatively low activity of IgA antibodies, our findings may indicate that chronic reactivation or recent exposure to the virus at a mucosal site is less important in the pathogenesis of NKTCL.

The unique association of NKTCL with IgG antibodies against EBNA3A has not been previously reported (9, 11, 30). Coghill et al. have reported an association between IgG antibodies against EBNA3A and BL in Africa, but the magnitude of association is smaller than the present study ($OR_{\text{highest vs. lowest tertile}}$=1.99) (13). Although EBV-encoded transcripts and proteins have been detected in patients with NKTCL (1, 6, 31-33), that expression pattern has generally been consistent with latency I or II infection, which is characterized by expression of EBNA-1, LMP-1, and LMP-2 genes but no other EBNA genes (1, 5, 6, 32, 33). Therefore, the higher IgG antibody levels against EBNA3A observed in the current study might not be explained by high expression of EBNA3A gene in the tumour tissue. Instead, this observation could reflect a long-term systematic exposure to the upregulation of EBNA3A gene within circulating B cells infected with EBV, which could be an early event during the development of NKTCL.

In agreement with previously reported case-only studies that included 155 patients from the U.S. and China (9, 11, 30), we confirm elevations in NKTCL patients for IgG antibodies against sequences representing EBV EA, including EAD-p47 and -p138, EBV viral capsid BVRF2 (VCAp40), as well as virion production BPLF1 (tegument protein). Furthermore, in agreement with other epidemiological research (9-11) we report here elevated IgG antibodies

against VCA and EA in patients with NKTCL, but no NKTCL associations with EBNA1. We also expand findings to the switch protein Zta (BZLF1), which has been associated with other EBV-associated malignancies (12-14) but not previously studied in the context of NKTCL. It is plausible that, again, systematic exposure to EBV, as indicated by elevations in IgG antibodies against EBV lytic proteins, potentially reflects an impaired T-cell response that allows virus to continue replication and spread from the B-cell compartment to NK/T-cells.

GWAS studies have implicated genetic susceptibility in NKTCL aetiology, with signals consistently mapped to the HLA genes in the class II region (28, 34). In our study, we observed indicative evidence that genetic variation within the HLA class II region affected anti-EBV serologic immunity in controls. It is plausible that people with susceptible HLA variants might mount altered responses to EBV infection that predisposes to NKTCL development (35-38) Future consortia-based efforts focusing on host genetic variants and anti-EBV antibodies would be required to explore the potential synergistic effects of HLA and EBV in the aetiology of NKTCL.

Our results should be interpreted in light of certain methodologic limitations. First, our observations are based on data obtained from a case-control designed study so we are unable to determine whether alterations in anti-EBV antibody responses occurred prior to disease onset; *i.e.*, predisposition to disease, as we have previously reported for other EBV-related tumours using the same EBV antibody array (12-14). However, the difficulty of conducting an adequately powered prospective study for this rare disease makes it unlikely that this limitation will be easily overcome in the future (3). Second, this is the only study to date examining the association between the proteome-wide anti-EBV antibody response and NKTCL, and we therefore lack an independent, external dataset for replication. Finally, this array was not designed to detect antibodies to conformational epitopes, which precluded us from examining NKTCL associations for selected transcripts that require glycosylation or other post-transcriptional modifications.

In conclusion, we characterize the antibody response against EBV for patients with NKTCL. Our findings suggest distinct serologic patterns of this NK/T-cell lymphoma compared with other EBV-associated tumours of B-cell or epithelial origin. This NKTCL–specific signature included pronounced differences in the immune response against six viral proteins involved in both latency and replication.

2.6 References

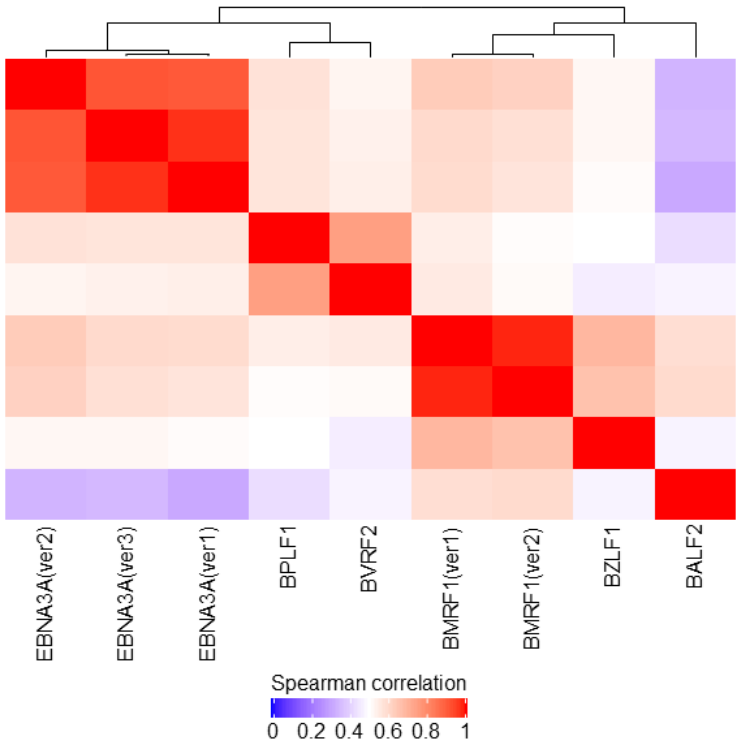
1. Asano N, Kato S, Nakamura S. Epstein–Barr virus-associated natural killer/T-cell lymphomas. Best practice & research Clinical haematology. 2013;26(1):15-21.
2. Kimura H. EBV in T-/NK-cell tumorigenesis. Human Herpesviruses. 2018:459-75.
3. Aozasa K, Takakuwa T, Hongyo T, Yang W-I. Nasal NK/T-cell lymphoma: epidemiology and pathogenesis. International Journal of Hematology. 2008;87(2):110-7.
4. Young LS, Yap LF, Murray PG. Epstein–Barr virus: more than 50 years old and still providing surprises. Nature Reviews Cancer. 2016;16(12):789-802.
5. Houldcroft CJ, Kellam P. Host genetics of Epstein–Barr virus infection, latency and disease. Reviews in medical virology. 2015;25(2):71-84.
6. Xiong J, Cui B-W, Wang N, Dai Y-T, Zhang H, Wang C-F, et al. Genomic and transcriptomic characterization of natural killer T cell lymphoma. Cancer Cell. 2020;37(3):403-19. e6.
7. Hudnall SD, Ge Y, Wei L, Yang N-P, Wang H-Q, Chen T. Distribution and phenotype of Epstein–Barr virus-infected cells in human pharyngeal tonsils. Modern Pathology. 2005;18(4):519-27.
8. Smith NA, Coleman CB, Gewurz BE, Rochford R. CD21 (Complement Receptor 2) is the receptor for Epstein-Barr virus entry into T cells. Journal of virology. 2020;94(11):e00428-20.
9. Jones JF, Shurin S, Abramowsky C, Tubbs RR, Sciotto CG, Wahl R, et al. T-cell lymphomas containing Epstein–Barr viral DNA in patients with chronic Epstein–Barr virus infections. New England Journal of Medicine. 1988;318(12):733-41.
10. Cao S-M, Liu Z, Jia W-H, Huang Q-H, Liu Q, Guo X, et al. Fluctuations of epstein-barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. PLoS One. 2011;6(4):e19100.
11. Huang Y, Rao H, Yan S, Wang F, Wu Q, Feng Y, et al. Serum EBV EA-IgA and VCA-IgA antibodies can be used for risk group stratification and prognostic prediction in extranodal NK/T cell lymphoma: 24-year experience at a single institution. Annals of hematology. 2017;96(8):1331-42.
12. Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clinical Cancer Research. 2018;24(6):1305-14.

13. Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, et al. The association between the comprehensive Epstein–Barr virus serologic profile and endemic Burkitt lymphoma. Cancer Epidemiology and Prevention Biomarkers. 2020;29(1):57-62.
14. Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. International Journal of Cancer. 2020;147(3):608-18.
15. Song L, Song M, Camargo MC, Van Duine J, Williams S, Chung Y, et al. Identification of anti-Epstein-Barr virus (EBV) antibody signature in EBV-associated gastric carcinoma. Gastric Cancer. 2021;24(4):858-67.
16. Liu Z, Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, et al. Patterns of interindividual variability in the antibody repertoire targeting proteins across the Epstein-Barr virus proteome. The Journal of infectious diseases. 2018;217(12):1923-31.
17. Lee M-H, Huang Y-H, Coghill AE, Liu Z, Yu KJ, Hsu W-L, et al. Epstein-Barr virus-based nasopharyngeal carcinoma (NPC) risk prediction scores are elevated in NPC multiplex family members in Taiwan. The Journal of infectious diseases. 2021;223(3):441-4.
18. Proietti C, Zakrzewski M, Watkins TS, Berger B, Hasan S, Ratnatunga CN, et al. Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) Web-Server. Scientific Reports. 2016;6(1):1-15.
19. Coghill AE, Hildesheim A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. American journal of epidemiology. 2014;180(7):687-95.
20. Gan Y, Fones-Tan A, Chan S, Gan L. Epstein-Barr viral antigens used in the diagnosis of nasopharyngeal carcinoma. Journal of Biomedical Science. 1996;3(3):159-69.
21. Chen MR, Liu MY, Hsu SM, Fong CC, Chen CJ, Chen IH, et al. Use of bacterially expressed EBNA-1 protein cloned from a nasopharyngeal carcinoma (NPC) biopsy as a screening test for NPC patients. Journal of medical virology. 2001;64(1):51-7.
22. Xu F-H, Xiong D, Xu Y-F, Cao S-M, Xue W-Q, Qin H-D, et al. An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein–Barr virus activation. Journal of the National Cancer Institute. 2012;104(18):1396-410.
23. Pickard A, Chen CJ, Diehl SR, Liu MY, Cheng YJ, Hsu WL, et al. Epstein-Barr virus seroreactivity among unaffected individuals within high-risk nasopharyngeal carcinoma families in Taiwan. International Journal of Cancer. 2004;111(1):117-23.

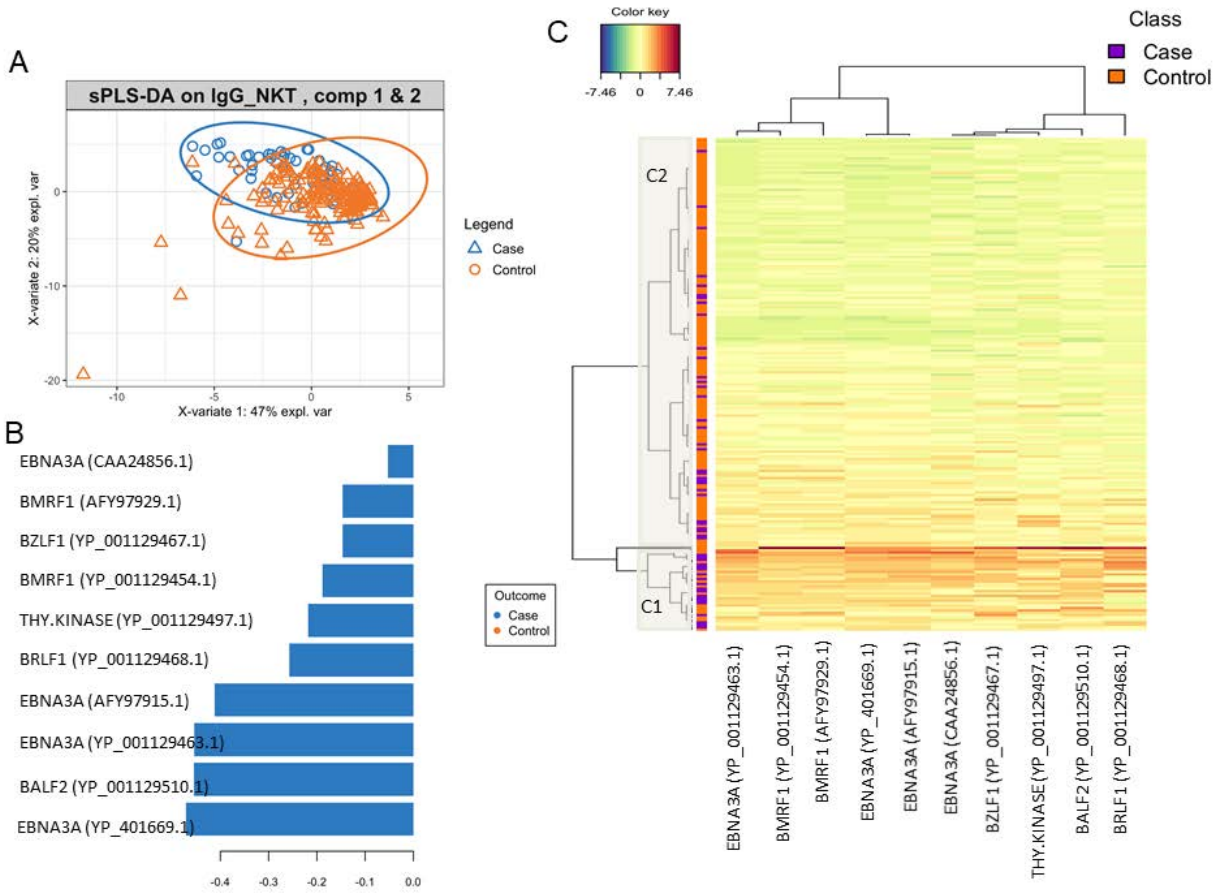
24. Xu JX, Hoshida Y, Yang WI, Inohara H, Kubo T, Kim GE, et al. Life-style and environmental factors in the development of nasal NK/T-cell lymphoma: A case–control study in East Asia. International Journal of Cancer. 2007;120(2):406-10.
25. Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC bioinformatics. 2011;12(1):1-17.
26. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS computational biology. 2017;13(11):e1005752.
27. Chung D, Keles S. Sparse partial least squares classification for high dimensional data. Statistical applications in genetics and molecular biology. 2010;9(1).
28. Lin G-W, Xu C, Chen K, Huang H-Q, Chen J, Song B, et al. Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study in multiple populations. The Lancet Oncology. 2020;21(2):306-16.
29. Bhaduri-McIntosh S, Landry ML, Nikiforow S, Rotenberg M, El-Guindy A, Miller G. Serum IgA antibodies to Epstein-Barr virus (EBV) early lytic antigens are present in primary EBV infection. The Journal of infectious diseases. 2007;195(4):483-92.
30. Huang Y-H, Wu Q-L, Zong Y-S, Feng Y-F, Hou J-H. Nasopharyngeal extranodal NK/T-cell lymphoma, nasal type: retrospective study of 18 consecutive cases in Guangzhou, China. International Journal of Surgical Pathology. 2011;19(1):51-61.
31. Peng R-J, Han B-W, Cai Q-Q, Zuo X-Y, Xia T, Chen J-R, et al. Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. Leukemia. 2019;33(6):1451-62.
32. Minarovits J, Hu L-F, Imai S, Harabuchi Y, Kataura A, Minarovits-Kormuta S, et al. Clonality, expression and methylation patterns of the Epstein-Barr virus genomes in lethal midline granulomas classified as peripheral angiocentric T cell lymphomas. Journal of general virology. 1994;75(1):77-84.
33. Tao Q, Ho FC, Loke SL, Srivastava G. Epstein-Barr virus is localized in the tumour cells of nasal lymphomas of NK, T or B cell type. International Journal of Cancer. 1995;60(3):315-20.
34. Li Z, Xia Y, Feng L-N, Chen J-R, Li H-M, Cui J, et al. Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study. The Lancet Oncology. 2016;17(9):1240-7.

35. Rubicz R, Yolken R, Drigalenko E, Carless MA, Dyer TD, Bauman L, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). PLoS genetics. 2013;9(1):e1003147.
36. Hammer C, Begemann M, McLaren PJ, Bartha I, Michel A, Klose B, et al. Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. The American Journal of Human Genetics. 2015;97(5):738-43.
37. Scepanovic P, Alanio C, Hammer C, Hodel F, Bergstedt J, Patin E, et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. Genome medicine. 2018;10(1):1-13.
38. Sallah N, Miley W, Labo N, Carstensen T, Fatumo S, Gurdasani D, et al. Distinct genetic architectures and environmental factors associate with host response to the γ2-herpesvirus infections. Nature communications. 2020;11(1):1-13.

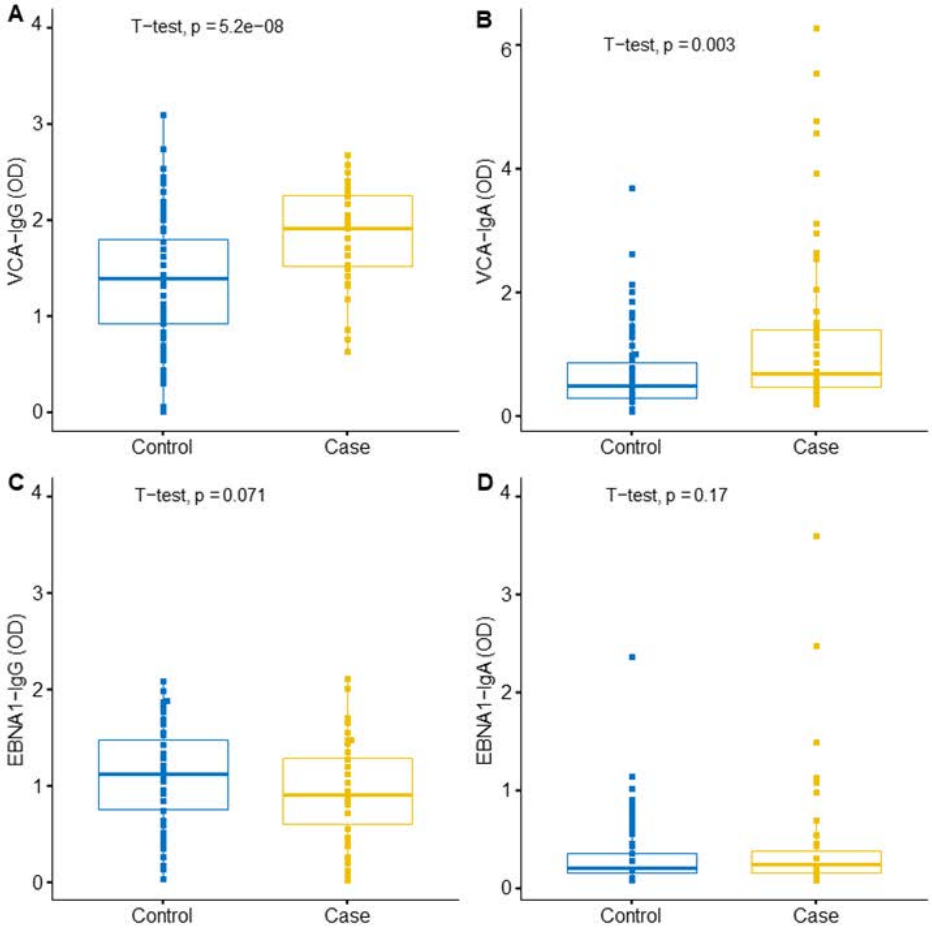
2.7 Supplementary information



Supplementary Figure 1 Spearman correlation between the average immunoglobulin G (IgG) antibody responses for the nine anti-EBV IgG antibodies presented in Table 2-2.



Supplementary Figure 2 Supervised analysis with sPLS-DA identified EBV antigens driving the discrimination between NK/T-cell lymphoma (NKTCL) cases and control. A) Sample plot depicting the first two sPLS-DA component with 95% confidence level ellipse plot; **B)** pyramid bar plot displays the top 10 EBV antigens selected on the first component of the sPLS-DA model. The length of the bar represents the importance of each antigen on the first component (importance from the bottom to top) in discriminating cases and controls. Colour indicates the class type (cases-controls) where the mean of the standardize signal intensity of the antigen is maximal. **C)** heatmaps of the 10 discriminant features identified by sPLS-DA with both rows and columns ordered using hierarchical (average linkage) clustering shows cluster C1 enriched in NKT cases and cluster C2 enriched in controls.



Supplementary Figure 3 Signal intensity for the four anti-EBV IgG antibodies tested by ELISA kits between NK/T-cell lymphoma (NKTCL) and controls, for **A)** VCA-IgG, **B)** VCA-IgA, **C)** EBNA1-IgG and **D)** EBNA1-IgA. P values from the t-test are listed.

Table S1 Differences in IgG and IgA antibodies between NK-T cell lymphoma (NKTCL) and controls in Hong Kong and Taiwan (0.0002<t test *p*-values<0.05) [a]

EBV Protein	Array sequence	Antibody type	NKTCL mean	NKTCL SD	Control mean	Control SD	Fold Change	t test *P*
EBNA3A	CAA24856.1-92243-92602	IgG	1.41	0.42	1.15	0.46	1.23	0.0003
THY.KINASE	YP_001129497.1-133399-131576	IgG	1.58	0.45	1.31	0.41	1.20	0.0003
BZLF1	YP_001129467.1-91697-91197	IgG	1.49	0.42	1.24	0.39	1.20	0.0004
EBNA3B	CAA24858.1-95353-95709	IgG	1.35	0.41	1.10	0.48	1.23	0.0005
BVRF1	YP_001129499.1-133954-135666	IgG	1.32	0.33	1.14	0.33	1.17	0.0006
BVRF2	YP_001129501.1-136465-138282	IgG	1.74	0.40	1.50	0.46	1.16	0.0007
BRLF1	YP_001129468.1-93725-91908	IgG	1.31	0.50	1.05	0.36	1.25	0.0010
EBNA3B	CAA24858.1-95788-98247	IgG	1.60	0.44	1.37	0.49	1.17	0.0019
EBNA3B	YP_001129464.1-83074-83430	IgG	1.43	0.33	1.25	0.39	1.14	0.0020
BLRF2	YP_001129461.1-76771-77259	IgA	1.27	0.24	1.15	0.25	1.11	0.0020
BDRF1	AFY97974.1-136284-137321	IgG	1.29	0.51	1.04	0.46	1.24	0.0030
LMP1	AFY97906.1-168167-168081	IgG	1.19	0.30	1.04	0.30	1.14	0.0034
BBLF1	AFY97956.1-108555-108328	IgG	1.72	0.41	1.52	0.46	1.13	0.0039
BcLF1	AFY97965.1-125044-120899-1	IgG	0.99	0.34	0.82	0.36	1.20	0.0048
BVRF2	YP_001129501.1-136465-138282	IgA	1.14	0.22	1.04	0.22	1.10	0.0050
BILF2	YP_001129503.1-139063-138317	IgG	1.50	0.31	1.35	0.41	1.12	0.0052
BWRF1	CAA24873.1-40189-41340	IgG	1.09	0.32	0.95	0.27	1.15	0.0057
BPLF1	CAA24839.1-71527-62078-2	IgG	1.93	0.40	1.73	0.53	1.11	0.0058
BARF1	YP_001129453.1-66746-67654	IgG	1.26	0.33	1.10	0.36	1.14	0.0062
BBLF1	YP_001129480.1-109516-109289	IgG	1.90	0.41	1.70	0.49	1.11	0.0063

BMRF1	AFY97929.1-67486-68700	IgA	1.28	0.28	1.16	0.28	1.11	0.0082
BDLF4	YP_001129488.1-117560-116883	IgG	1.66	0.32	1.51	0.44	1.10	0.0089
EBNA-LP	YP_001129440.1-20824-20955	IgG	1.26	0.35	1.11	0.32	1.13	0.0097
EBNA1	AFY97913.1-95532-97457	IgG	0.98	0.27	0.86	0.34	1.14	0.0098
EBNA3B	YP_001129464.1-83509-86532-1	IgG	1.48	0.41	1.31	0.43	1.13	0.0108
BBRF1	YP_001129476.1-102746-104587	IgG	1.33	0.39	1.17	0.35	1.14	0.0110
BZLF1	CAA24861.1-102338-102210	IgG	1.50	0.52	1.29	0.45	1.16	0.0113
VCA_p18	synthetic peptide	IgG	1.83	0.42	1.64	0.57	1.12	0.0130
BMRF1	YP_001129454.1-67745-68959	IgA	1.32	0.27	1.21	0.26	1.09	0.0135
BHRF1	YP_001129442.1-42204-42779	IgG	1.32	0.28	1.20	0.34	1.10	0.0137
BORF1	YP_001129451.1-63084-64178	IgG	1.13	0.29	1.02	0.26	1.11	0.0148
BZLF1	CAA24861.1-103155-102655	IgG	1.16	0.41	1.00	0.33	1.16	0.0148
LMP1	YP_401722.1-168670-168584	IgG	1.18	0.34	1.04	0.37	1.13	0.0149
BSLF2/BMLF1	YP_001129456.1-71967-70589	IgG	1.25	0.38	1.10	0.37	1.14	0.0151
BDLF3	YP_001129490.1-119605-118901	IgG	1.62	0.43	1.45	0.49	1.12	0.0158
BRRF2	AFY97943.1-93884-95497	IgG	1.91	0.48	1.72	0.51	1.11	0.0182
BFRF1	YP_001129446.1-46719-47729	IgG	1.35	0.42	1.20	0.34	1.13	0.0183
BALF2	YP_001129510.1-165796-162410-2	IgG	1.31	0.37	1.17	0.32	1.12	0.0192
BDLF3	AFY97964.1-118644-117940	IgG	1.63	0.45	1.45	0.52	1.12	0.0215
BLRF2	YP_001129461.1-76771-77259	IgG	2.04	0.35	1.89	0.52	1.08	0.0217
BRRF2	YP_001129470.1-94844-96457	IgG	2.00	0.43	1.83	0.50	1.09	0.0255
BZLF1	YP_001129467.1-90855-90724	IgG	1.59	0.49	1.41	0.44	1.12	0.0257
BcLF1	CAA24794.1-137466-133321-1	IgG	1.15	0.27	1.05	0.29	1.10	0.0270
BZLF1	YP_001129467.1-90855-90724	IgA	1.12	0.20	1.05	0.17	1.07	0.0283
BPLF1	YP_001129449.1-59370-49906-2	IgG	1.39	0.29	1.28	0.37	1.09	0.0302

LMP1	YP_001129515.1-170111-170025	IgG	1.01	0.20	0.93	0.26	1.08	0.0330
BALF2	YP_001129510.1-165796-162410-1	IgA	0.96	0.23	0.88	0.24	1.09	0.0333
EBNA3B	YP_001129464.1-83509-86532-2	IgG	1.45	0.44	1.30	0.50	1.12	0.0416
BALF5	YP_001129507.1-157772-154725-1	IgG	1.04	0.24	0.96	0.24	1.08	0.0420
BFRF3	CAA24838.1-61507-62037	IgG	1.96	0.35	1.83	0.52	1.07	0.0436
LMP1	YP_401722.1-168507-167702	IgG	0.94	0.32	0.83	0.38	1.13	0.0475
BcRF1	YP_001129494.1-126004-128256	IgG	0.98	0.21	1.05	0.27	0.93	0.0490

a. The table is ordered by t test P value (lowest to highest). Antibodies with $0.0002<P<0.05$ are listed.

Table S2 Differences in IgG antibodies against EBNA1 between NK-T cell lymphoma (NKTCL) and controls in Hong Kong and Taiwan.

EBV protein and array sequence	Antibody type	t test *P*	NKTCL mean (SD)	Control mean (SD)	NKTCL positivity	Control positivity
EBNA1 (synthetic peptide)	IgG	0.59	1.09 (0.50)	1.04 (0.48)	60.8%	60.4%
EBNA1 (AFY97842.1-95349-97142)	IgG	0.78	0.79 (0.17)	0.79 (0.33)	17.6%	10.4%
EBNA1 (CAA24816.1-107950-109875)	IgG	0.41	1.18 (0.28)	1.22 (0.37)	68.6%	76.0%
EBNA1 (AFY97913.1-95532-97457)	IgG	0.01	0.98 (0.27)	0.86 (0.34)	39.2%	26.6%

Chapter 3

A generalized proteome-wide Epstein-Barr Virus Antibody Signature predicts classical Hodgkin's Lymphoma in geographically and ethnically distinct populations

This chapter describes antibody signatures to delineate the EBV status of cHL cases to predict the EBV status of cHL tumours from a case-control study from an East Asian study population. These study findings demonstrated the generalizability of antibody markers previously reported in a European study population. Two IgG markers, together with patient demographics identified that accurately predicted the EBV status of cHL cases independent of geographic location and ethnic diversity of study populations. The work presented in this chapter comprises a publication pending submission.

3 A generalized proteome-wide Epstein-Barr Virus Antibody Signature predicts classical Hodgkin's Lymphoma in geographically and ethnically distinct populations

Yomani D. Sarathkumara[1*], Rena R. Xian[3*], Zhiwei Liu[2*], Kelly J. Yu[2], John K. C. Chan[4], Yok-Lam Kwong[5], Anna E. Coghill[6], Ruth M. Pfeiffer[2], Nathaniel Rothman[2†], Richard F. Ambinder[3†], Allan Hildesheim[2†], Carla Proietti[1†], Qing Lan[2 †], Denise L. Doolan[1†]

[1] Centre for Molecular Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia

[2] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA

[3] Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, USA

[4] Department of Pathology, Queen Elizabeth Hospital, Hong Kong, SAR, China

[5] Queen Mary Hospital, The University of Hong Kong, Hong Kong, SAR, China

[6] Cancer Epidemiology Program, Division of Population Sciences, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

***** These authors contributed equally as first authors

† These authors contributed equally as last authors

3.1 Abstract

The underlying contribution of EBV in the development of cHL and associated EBV-specific humoral immune responses is not well understood. In this study, we comprehensively characterized IgA and IgG antibody responses to 202 protein sequences representing the complete EBV proteome in cHL patients from East Asia. We identified 12 novel EBV specific IgG antibodies that were highly elevated in EBV-positive cHL cases from Hong Kong and Taiwan. We discovered 14 IgG antibodies that were significantly elevated in EBV-positive cHL compared with controls and 12 of these antibodies remained significantly associated with EBV-positive cHL after adjusting for demographics. Most of these antibodies predicted EBV-positive cHL in a European population, demonstrating their generalizability.

Our previous study in a European study population identified 18 antibodies (16 IgG and two IgA) that were elevated in EBV-positive cHL cases compared to disease-free controls and a signature of six IgG-markers that were elevated in EBV-positive compared to EBV-negative cHL cases; most of those findings were replicated in the East Asian study reported in this thesis. Specifically, all six antigens from the six-IgG signature were highly elevated among East Asians, and seven of 18 antibodies associated with EBV-positive cHL in Europeans retained high predictive power in East Asians.

Importantly, the combination of two IgG markers (BdRF1and BZLF1) together with patient demographics identified as highly discriminatory in the European study was able to effectively predict the EBV status of cHL cases from East Asia. Our data support the hypothesis that the aetiology of EBV-positive cHL is similar across populations and demonstrate possible use of EBV antibody biomarkers to predict EBV status of cHL tumours.

Key words: EBV proteome, Hodgkin lymphoma, antibody patterns, generalizability

3.2 Introduction

EBV is an oncogenic virus associated with epithelial carcinomas of the stomach and nasopharynx and lymphomas, including natural killer/T-cell lymphoma (NKTCL), Burkitt's, and a subset of classical Hodgkin lymphomas (1). Classical Hodgkin's Lymphoma (cHL) is a B-cell-derived malignancy. Histologically, cHL is characterized by large mononuclear Hodgkin cells and multinucleated Reed-Sternberg (HRS) cells (2, 3). Genomic evidence suggests that Epstein-Barr virus (EBV) status is aetiologically important for cHL pathogenesis (4-8), and elevated antibody levels for EBV antigens support the potential involvement of EBV in cHL pathogenesis (9). However, how EBV contributes to the pathogenesis of cHL remains to be elucidated.

Genomic, serological, and epidemiological studies of cHL suggest that EBV-associated (EBV-positive) and non-associated (EBV-negative) cases may represent two distinct aetiological entities (10-13). The role of EBV in cHL and its progression in cHL is not well understood (14, 15). Analysis of the antibody responses against all the proteins expressed by the virus in individuals with cHL should provide insights into the aetiological role of EBV in cHL pathogenesis. Additionally, identifying the subset of EBV proteins expressed at different stages of the EBV life cycle that are potential antibody targets associated with cHL could define an EBV-based antibody risk stratification signature for EBV-positive cHL. The current need to obtain tumour tissues remains a challenge to determine EBV status for cHL. Thus, it becomes of importance to identify sensitive and specific biomarkers of disease. Most serological studies to date have investigated antibody responses against only the major EBV-encoded antigen complexes, including the viral capsid antigen (VCA), early antigen (EA), EBV nuclear antigen 1 (EBNA-1), EBV nuclear antigen 2 (EBNA-2) and latent membrane protein 1 (LMP-1) (16, 17). However, the complete EBV genome translates approximately 100 open reading frames, most of which have not been investigated (18-26).

We applied our custom EBV proteome microarray to comprehensively evaluate both IgG and IgA antibody responses against the complete EBV proteome (27) represented by 202 peptide sequences for each of the known open reading frames from 86 EBV proteins EBV as well as predicted splice variants to screen sera from individuals with several EBV-associated cancers including nasopharyngeal carcinoma (NPC) (27), endemic Burkitt lymphoma (eBL) (28), and extranodal natural killer/T-cell lymphoma (NKTCL) (29).

We recently applied this EBV proteome microarray to evaluate the IgG and IgA immune responses against the complete proteome of EBV in individuals with cHL from a

European population (30) and demonstrated aberrant EBV-specific antibody profiles among EBV-positive cHL cases. We identified 16 IgG and two IgA antibodies that were significantly elevated in EBV-positive cHL cases compared to disease-free controls. In addition, we identified six IgG-markers that were highly elevated in EBV-positive compared to EBV-negative cHL cases. Amongst those, two IgG-markers comprised of BdRF1(VCAp40)-IgG and BZLF1(Zta)-IgG achieved high accuracy (area under the curve [AUC] = 0.75, 95% CI = 0.68, 0.83) in distinguishing EBV-positive from EBV-negative tumours in this European study population when modelled with patient demographics (*i.e.*, age group, sex, and study area).

We hypothesized that the aetiology of EBV-positive cHL is similar across populations. Therefore, we evaluated the antibody responses associated with EBV-positive cHL by comparing IgA and IgG responses between i) EBV-positive cHL cases and controls; ii) EBV-negative cHL cases and controls; and iii) EBV-positive cHL and EBV-negative cHL cases using samples collected from East Asian populations from Hong Kong and Taiwan. We then tested the generalizability of antibodies identified in the present study using our previous data generated from an European population (30). Similarly, we evaluated the generalizability of EBV antibodies previously identified in the European population in the East Asian study population. Finally, we tested whether the two IgG-markers identified as the best classifiers of EBV-positive *vs*. EBV-negative tumours in the European population could maintain their discriminatory power in geographically and ethnically distinct East-Asian populations.

3.3 Materials and Methods

3.3.1 Study population

Subjects were part of the AsiaLymph study, a multi-centre, hospital-based case-control study conducted in Hong Kong and Taiwan between 2012 and 2017 (29). Eligible cases were aged between 18 and 79 years at diagnosis and living in the geographic region served by the partnering hospital at the time of cancer diagnosis. Cases with a prior history of lymphoma were excluded. Blood and buccal cell collection were performed at the time of diagnosis and before cancer therapy. Controls were apparently healthy individuals with no symptoms of disease (disease-free/cHL-free controls) who were a subset of patients seen at the same partnering hospital for diseases/conditions that were not associated with risk factors under study, including injuries and selected diseases of the circulatory, digestive, genitourinary, and central nervous system. Patients with a history of any lymphoma were not eligible for controls.

Our study cohort included all histologically diagnosed cHLcases in the two regions (N=140). Additionally, we randomly selected 60 subjects who were frequency-matched to cHL cases on sex, age (+/- 5 years), date of enrolment (within 3 months), and region (Hong Kong/Taiwan) from all controls recruited in the two regions (N = 1496; 1119 from Hong Kong and 377 from Taiwan). EBV status of the cHL cases was determined by the standard immunohistochemical staining of tumour biopsies for EBV latent membrane antigen (LMP)-1 and/or in situ hybridization for EBV-encoded small RNAs (EBERs) or using EBV DNA viral load (31), as described detailed in the supplementary information (**Supporting Information Figure S1**, **Table S1**), leading to a total of 35 EBV-positive HL cases and 92 EBV-negative HL cases in the analysis.

The clinical study was approved by institutional review boards (IRB) at each of the participating sites and by the US National Institutes of Health and the US National Cancer Institute IRB (IRB:11CN206). Written informed consent was obtained from all participants. All laboratory testing was conducted under a protocol approved by the James Cook University Human Research Ethics Committee (H7696).

Our previous European study (30) was comprised of 139 EBV-positive cHL cases, 70 EBV-negative cHL cases and 141 disease-free controls selected from case-control studies conducted in in the UK (11, 32), Denmark and Sweden (33). EBV status of these cHL tumour samples were determined using immunohistochemical staining for LMP-1 and/or in situ hybridization for EBV-encoded small RNAs (EBERs) of the biopsies of tumour samples. Control samples were frequency-matched to EBV-positive cHL cases on sex, age (+/- 5 years) and study area. EBV-positive cHL and EBV-negative cHL cases were also matched on their clinical stage (30).

3.3.2 EBV custom proteome microarray

Our comprehensive custom EBV proteome microarray comprised 202 protein sequences, including 199 EBV protein sequences from five EBV strains (AG876, Akata, B95-8, Mutu, and Raji) representing nonredundant open reading frames and predicted splice variants from 86 EBV proteins (27-30). Also included were three synthetic EBV peptides (VCAp18, EBNA-1, and EA p47) representing the current gold-standard for detecting EBV-specific antibody responses and putative cancer biomarkers (27-30). Four "noDNA" (no translated protein) spots were included in the array to correct for person-specific background (*i.e.*, *E. coli* reactivity). Each microarray slide was printed with 16 arrays per slide.

For each study participant, plasma samples were tested blinded to case-control status by lab personnel for both IgA and IgG antibody responses as described in detail previously (27-30). Briefly, antibody responses were detected with biotin-conjugated goat anti-human IgG (1:1000 dilution) or IgA (1:500 dilution) secondary antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) and visualized with a streptavidin-conjugated SureLight® P3 (Columbia Biosciences, Columbia, MD, USA) antibody (1:200 dilution). After testing, air-dried probed slides were scanned on an Axon GenePix 4300B (Molecular Devices). Raw fluorescence intensities were corrected for spot-specific background using the Axon GenePix Pro 7 software, and corrected data were variant log-transformed using variance stabilizing normalization (VSN) transformation in Gmine (34). The array output was standardized to the person-specific background by dividing VSN values with the individual's cut-off (mean ±1.5 SD of the four "no DNA" spots), and the transformed data was referred to as the standardized signal intensity (SSI).

3.3.3 Statistical analysis

All statistical analyses were performed using R statistical software (https://www.r-project.org/, RStudio Inc., Boston, USA, Version 1.4.1103). For *p*-values, Benjamini and Hochberg false discovery rate (FDR) was applied to correct the statistical significance of multiple testing for 202 antibodies, and FDR<0.05 was considered significant when all 202 antibodies were considered. For other analyses presented (*i.e.*, generalizability of the antibody signatures described below), nominal *p*-values without FDR correction were used.

We calculated proportions for demographics variables (*i.e.*, sex, age, study area) between study groups (EBV-positive cHL *vs*. EBV-negative cHL and EBV-positive cHL *vs*. disease-free controls) at 95% confidence intervals to identify disparities between study groups.

Differences in the mean SSI for the IgG and IgA responses against the 202 EBV sequences in the array were compared between i) EBV-positive cHL cases and controls; ii) EBV-negative cHL cases and controls; and iii) EBV-positive cHL and EBV-negative cHL cases, using unpaired t-tests. In addition, linear regression analyses were carried out to test the association between anti-EBV antibody responses (*i.e.*, SSI) and case-control status or EBV status only for the HL cases by adjusting for sex, age, and study area.

The performance of the 18 EBV antigens identified as significantly associated with EBV-positive cHL in the European Caucasian population in our previous study (30) was investigated using the Area under the Receiver operative Curve (AUC), to test their ability,

along with the subject's demographics (*i.e.*, age group, sex and study area) to distinguish EBV-positive cHL from controls in the current study population. Similarly, the ability of the six IgG antibody markers previously identified as significantly elevated in EBV-positive cHL compared with EBV-negative cHL European cases (30) was tested together with the subject's demographics by the AUC to measure how well these IgG markers distinguish EBV-positive from EBV-negative cHL in the current Asian study population. Unpaired t-test and linear regression adjusted by age, sex, and area were also used to (i) validate the differences in antibody levels observed in European population (for the 18-antibody signature) in the east Asian population or (ii) to validate the differences in antibody levels observed in in East Asia population (for the 6-antibody signature) in the European population, and p value (not adjusted) <0.05 was considerate significant for these validation analysis.

In addition, the sensitivity and specificity of the two-marker combination, BdRF1(VCAp40)-IgG and BZLF1(Zta)-IgG, which best-classified tumour EBV status in our European study (30) were evaluated by AUC in the current Asian study population. The ability of this two-marker combination (as continuous levels) was evaluated with the subject's demographics (*i.e.*, age, sex, and study area) to classify the EBV status of cHL tumours (positive=1, negative=0).

Likewise, the antibodies found to be significantly elevated in EBV-positive cHL compared with controls or between EBV-positive cHL and EBV-negative cHL in the present study were tested in the European study population by unpaired t-test and linear regression adjusted by age, sex, and area. For this analysis we used IgG and IgA antibody data generated in our previous study (30), using the same custom EBV protein microarray platform which measured IgG and IgA antibody responses against 202 protein sequences on sera from139 EBV-positive cHL cases, 70 EBV-negative cHL cases and 141 population-based controls (30). The p values (not adjusted) <0.05 were considerate significant for testing the generalizability of antibodies.

3.4 Results

In the initial study design, disease free control samples (n=60) were matched by age (p >0.05) and sex (p = 0.294) to overall cHL cases (n=140) regardless of their EBV status. However, imbalances between groups emerged when we sub-grouped cHL cases according to their EBV status (EBV-positive cHL cases, n= 35 and EBV-negative cHL, n=92) were considered for analysis indicating a significant lower frequency of EBV-positive cases compared with EBV-negative cases in the overall study population. Age category 2 (0-39 years, $p = 5.254\times10^{-9}$), category 3 (54-100 years, $p = 3.667\times10^{-11}$) and sex (p = 0.002) were significantly different between EBV-positive cHL and EBV-negative cHL after stratification of cHL cases based on the EBV status. In addition, all age groups and sex were found to be disproportionate between EBV-positive cHL and control samples (**Table 3-1)**. The demographic characteristics of the 35 EBV-positive cHL cases, 92 EBV-negative cHL cases and 60 disease-free controls are presented in **Table 3-1**.

3.4.1 EBV-positive cHL *vs.* disease-free controls

Difference in the EBV antibody repertoire between EBV-positive cHL and disease-free controls in the East Asian study population

We identified 14 IgG antibodies (**Figure 3-1A, Table S2, Table 3-2**) but no IgA antibodies (**Table S3**), which were significantly elevated in EBV-positive cHL cases compared to disease-free controls, when comparing differences in the mean standardized signal intensity (SSI) for IgG and IgA antibodies against each of the 202 array sequences in the East Asian population, (FDR p <0.05, t-test).

Of these 14 IgG antibodies, 10 remained significantly associated with EBV-positive cHL cases when adjusting for age, sex, and study area by linear regression model (OR>1.3, FDR p <0.05, linear regression). Two additional IgG antibodies, EBNA3A and BDLF3, were found to be significantly associated with EBV-positive cHL by linear regression models that adjusted for age, sex, and study area (ORs > 1.5, FDR p <0.05, linear regression) (**Table 3-2**). All 16 IgG antibodies identified from this analysis showed high discriminative power with an Area under the curve (AUC) ranging from 0.80 to 0.85 (**Table 3-2**). The most pronounced SSI differences between EBV-positive cHL cases and controls were observed for IgG antibodies against sequences representing latent membrane protein 1 (LMP1) (FDR p = 0.037, t-test; AUC=0.85), Epstein–Barr nuclear antigen 3A (EBNA3A) (FDR p = 0.037, t-test; AUC=0.82),

EBV nuclear antigen leader protein (EBNA-LP) (FDR p = 0.037, t-test, AUC=0.81) and Thymidine Kinase (TK) (FDR p = 0.037, t-test; AUC=0.82). All those antigens showed an OR>1.4 (**Table 3-2**, **Supporting Information Figs. S2A–S2D)**.

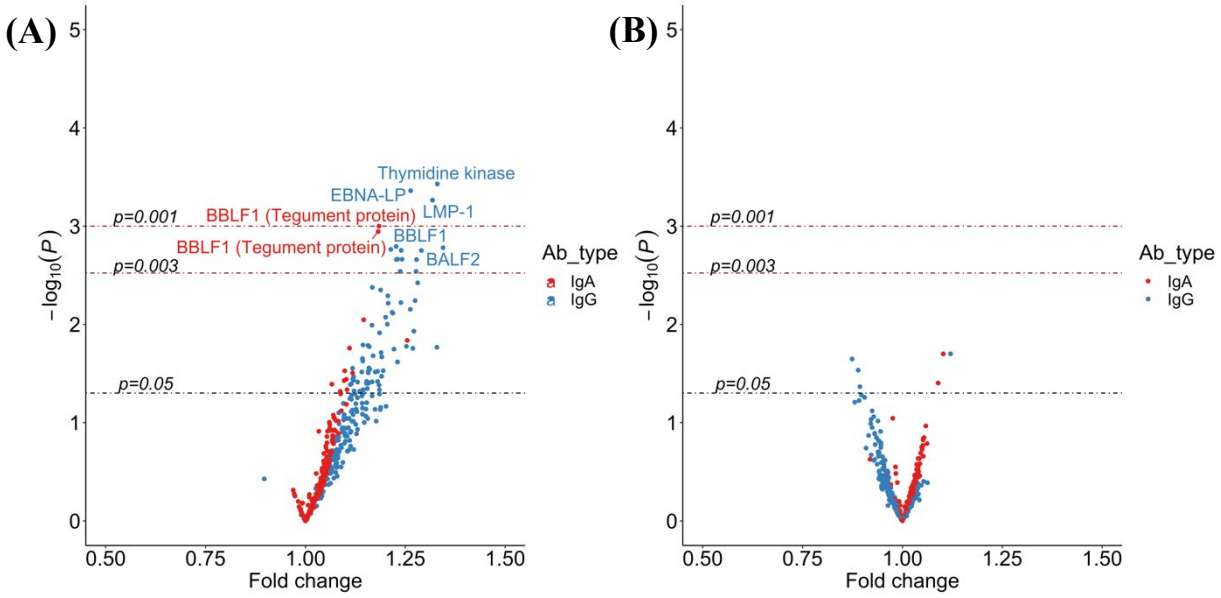


Figure 3-1 Case-control differences in the mean antibody responses for IgA and IgG. Case-control differences in the mean antibody response for EBV-positive cHL cases *vs.* controls.

(A) and between EBV-negative cHL cases vs. controls (B). The x-axis of the volcano plot displays the fold change (case vs. control ratio of standardized signal intensity) for 202 array sequences for each antibody type (red, IgA; blue, IgG). The y-axis illustrates the p-value corresponding to the t-test for a difference in standardized signal intensity (SSI) between groups. (A) 14 IgG antibodies were significantly elevated in EBV-positive cHL cases compared to controls (FDR<0.05). Seven antibodies with the significant p-values are highlighted. (B) No anti-EBV antibodies were significantly elevated in EBV-negative cHL cases compared to controls. The dashed lines represent the statistically significant p-value thresholds (FDR p <0.05).

Table 3-1 Population characteristics by case-control status.

Characteristic	EBV-positive cHL (n=35) n (%)	EBV-negative cHL (n=92) n (%)	*p-value* test of proportion between EBV-positive and EBV-negative cHL	Controls (n=60) n (%)	*p* value test of proportion between EBV-positive cHL and Controls
Study area					
Taiwan	2 (5.7)	7 (7.6)	1	4 (6.7)	1
Hong Kong	33 (94.3)	85 (92.4)		56 (93.3)	
Sex					
Female	8 (22.9)	51 (55.4)	0.002	27 (45.0)	0.053
Male	27 (77.1)	41 (44.6)		33 (55.0)	
Age at diagnosis (years)					
0-39	3 (8.6)	63 (68.5)	5.254×10-09	22 (36.7)	0.006
39-54	5 (14.3)	16 (17.4)	0.8779	22 (36.7)	0.036
54-100	27 (77.1)	13 (14.1)	3.667×10-11	16 (26.7)	5.26×10-06

Abbreviation: cHL, classical Hodgkin Lymphoma.

Table 3-2 EBV proteins (name and microarray sequence) for the IgG antibody responses significantly elevated in EBV-positive cHL cases compared with disease-free controls in the East Asian study population.

Protein name	Array sequence	EBV life cycle	t-test *p*	t-test adj *p* FDR	AUC	95% CI	OR	L95	U95	Lin.reg *p*-value	Lin.reg adj *p* FDR
THY.KINASE	YP_001129497.1-133399-131576	Early lytic	3.72×10-4	0.037	0.82	0.734-0.908	1.56	1.21	2.01	0.001	0.041
EBNA-LP	YP_001129440.1-20824-20955	Latent	4.36×10-4	0.037	0.81	0.727-0.899	1.35	1.14	1.61	0.001	0.041
LMP1	YP_401722.1-168507-167702	Latent	0.001	0.037	0.85	0.768-0.928	1.43	1.21	1.69	0.000	0.012
BBLF1	AFY97956.1-108555-108328	Early lytic	0.002	0.037	0.81	0.719-0.899	1.46	1.12	1.9	0.006	0.061
BALF2 (EA(D)_p138)	YP_001129510.1-165796-162410-1	Early lytic	0.002	0.037	0.83	0.743-0.918	1.49	1.17	1.89	0.001	0.042
LMP2A	YP_001129436.1-167587-167942	Latent	0.002	0.037	0.80	0.715-0.891	1.23	1.07	1.42	0.006	0.061
LF2	YP_001129504.1-151808-150519	Early lytic	0.002	0.037	0.81	0.717-0.893	1.32	1.11	1.57	0.002	0.043
EBNA3A	YP_001129463.1-80447-82888	Latent	0.002	0.037	0.82	0.734-0.908	1.74	1.28	2.38	0.001	0.041
BPFL1	YP_001129449.1-59370-49906-2	Late lytic	0.002	0.037	0.82	0.726-0.909	1.39	1.13	1.7	0.002	0.043
BDLF2	YP_001129491.1-120928-119666	Glycoprotein	0.002	0.037	0.81	0.721-0.899	1.43	1.14	1.8	0.003	0.046
BPFL1	YP_001129449.1-59370-49906-3	Late lytic	0.002	0.037	0.80	0.706-0.893	1.32	1.1	1.58	0.004	0.051
BLLF1 (gp350/220)	YP_001129462.1-79936-7727. r	Glycoprotein	0.002	0.037	0.82	0.737-0.909	1.45	1.17	1.81	0.001	0.042
LMP2B	AFY97910.1-1026-1196	Latent	0.003	0.042	0.81	0.718-0.903	1.30	1.09	1.54	0.004	0.051
EBNA3A	YP_401669.1-80382-82877	Latent	0.003	0.042	0.82	0.727-0.904	1.67	1.24	2.27	0.001	0.042
EBNA3A	AFY97915.1-80252-82747	Latent	0.004	0.051	0.81	0.721-0.901	1.63	1.2	2.2	0.002	0.043
BDLF3	AFY97964.1-118644-117940	Glycoprotein	0.008	0.065	0.81	0.724-0.901	1.57	1.18	2.09	0.003	0.046

FDR correction method: Benjamini and Hochberg. The table is ordered by significant t-test *p*-value (lowest to highest). ORs were adjusted for age (continuous), sex (M/F), and study area (Hong Kong and Taiwan) using a linear regression model. FDR *p*-value obtained for t-tests, Adj *p* FDR; is the FDR *p*-value obtained for the linear regression model adjusted by age, sex, and study area for all 202 antibodies. The antibodies found in our previous 18-antibody signature (30) are highlighted in bold and the others are novel antibodies identified from East Asia. The area under the Receiver operative Curve (AUC) was calculated for each antibody to test their ability along with the subject's demographics (*i.e.*, age group, sex, and study area) to distinguish EBV-positive cHL from controls in the current study.

Evaluation of the generalizability of the Asian EBV antibodies associated with EBV-positive cHL cases in the European population

The 16 antibodies (all IgG) identified as being elevated in the EBV-positive cHL in the Asian population were tested for their association with EBV-positive cHL (n=139) (compared to controls (n=141)) in the European population studied previously (30).

Among these 16 antibodies, we identified 11 antibodies that were significantly elevated in EBV-positive cHL compared with controls in the European population in the univariate analysis (p <0.05, t-test), and 10 of them remained significantly associated with EBV-positive cHL when adjusted for age, sex, and study area (nominal p <0.05, linear regression) (**Table S4**).

Evaluation of the generalizability of the European EBV antibodies associated with EBV-positive cHL cases in the Asian population

In our previous European study (30), 18 antibodies (16 IgG; 2 IgA) were significantly elevated in EBV-positive cHL cases *vs*. disease-free controls. We aimed to test the discrimination power of those 18 antibodies responses in the current geographically and ethnically distinct East-Asian populations from Hong Kong and Taiwan. Of the 18 European antibodies, 15 antibodies were consistently elevated in EBV-positive cHL cases compared with disease-free controls in the East-Asian populations (p < 0.05, t-test) (**Table 3-3**). After adjustment for age, sex, and residential area by linear regression model, 14 of the 15 antibodies were significantly associated with the EBV-positive cHL (p < 0.05, linear regression).

The most discriminative antibodies were LMP-1, TK, BDLF3 (glycoprotein 150), and BALF2[EA(D)_p138] with an AUC ≥ 0.80 and OR>1.3 (**Table 3-3, Supporting Information Figure S2C–S2F)**.

Table 3-3 Discrimination power (AUC and 95%CI) of the 18-antibodies identified in the European study in distinguishing EBV-positive classical Hodgkin lymphoma (cHL) from controls in the Asian population.

Protein name	Array sequence	IgG/ IgA	t-test *p*-value	EBV+ cHL mean	Control mean	AUC	95% CI	OR	L95	U95	Lin. reg *p*-value
Thymidine Kinase (Early lytic)	YP_001129497.1-133399-131576	IgG	3.72×10-4	1.91	1.44	**0.82**	0.73-0.91	1.56	1.21	2.01	8.10×10-4
LMP1 (Oncogene)	YP_401722.1-168507-167702	IgG	0.001	1.25	0.95	**0.85**	0.77-0.93	1.43	1.21	1.69	5.93×10-5
BBLF1 (Tegument protein)	YP_001129480.1-109516-109289	IgA	0.001	1.32	1.11	0.77	0.67-0.86	1.23	1.09	1.39	0.002
BBLF1 (Tegument protein)	AFY97956.1-108555-108328	IgA	0.001	1.30	1.10	0.76	0.66-0.86	1.22	1.08	1.38	0.002
BBLF1 (Tegument protein)	AFY97956.1-108555-108328	IgG	0.002	2.15	1.75	**0.81**	0.72-0.90	1.46	1.12	1.9	0.006
BALF2 (EA(D)_p138)	YP_001129510.1-165796-162410-1	IgG	0.002	1.48	1.10	**0.83**	0.74-0.92	1.49	1.17	1.89	0.001
BFLF2 (Late lytic)	YP_001129443.1-44763-43807	IgG	0.006	1.35	1.12	0.79	0.70-0.89	1.26	1.07	1.48	0.006
BDLF3 (glycoprotein 150)	AFY97964.1-118644-117940	IgG	0.008	2.07	1.70	**0.81**	0.72-0.90	1.57	1.18	2.09	0.003
BDLF3 (glycoprotein 150)	YP_001129490.1-119605-118901	IgG	0.010	2.06	1.71	**0.81**	0.72-0.90	1.52	1.16	2.00	0.004
BHRF1 (Bcl-2 homolog)	YP_001129442.1-42204-42779	IgG	0.016	1.59	1.37	0.76	0.67-0.86	1.26	1.05	1.50	0.014
BdRF1 (VCA_p40)	AFY97974.1-136284-137321_US	IgG	0.018	1.51	1.19	**0.80**	0.70-0.89	1.39	1.07	1.80	0.017
BBRF1 (Late lytic)	YP_001129476.1-102746-104587	IgG	0.024	1.69	1.38	0.79	0.69-0.88	1.33	1.01	1.75	0.047
BcLF1 (VCA_p160)	CAA24794.1-137466-133321-1	IgG	0.030	1.43	1.19	0.79	0.69-0.88	1.26	1.03	1.54	0.029
BBLF1 (Tegument protein)	YP_001129480.1-109516-109289	IgG	0.040	2.20	1.90	0.79	0.69-0.88	1.31	0.98	1.75	0.071
BBRF3 (glycoprotein M)	YP_001129479.1-107679-108896	IgG	0.049	1.66	1.45	0.79	0.69-0.88	1.27	1.02	1.58	0.034
BcLF1 (VCA_p160)	AFY97965.1-125044-120899-1	IgG	0.068	1.21	1.00	0.78	0.68-0.87	1.22	0.98	1.5	0.073
BcLF1 (VCA_p160)	YP_001129493.1-126005-121860-1	IgG	0.082	1.42	1.24	0.78	0.68-0.87	1.18	0.97	1.42	0.102
BARF1 (Oncogene)	YP_001129453.1-66746-67654	IgG	0.100	1.47	1.28	0.78	0.68-0.87	1.20	0.96	1.50	0.108

FDR correction method: Benjamini and Hochberg. The table is ordered by significant t-test *p*-value (lowest to highest) obtained for all 202 antibodies. ORs were adjusted for age, sex, and study area (Hong Kong and Taiwan) in the linear regression models for all 202 antibodies. The area under the Receiver operative Curve (AUC) was calculated for each antibody to test their ability along with the subject's demographics (*i.e.*, age group, sex, and study area) to distinguish EBV-positive cHL from controls in the current study. AUC values $\geq$ 0.80 highlighted in bold.

3.4.2 EBV-negative cHL cases *vs.* disease-free controls

Difference in the EBV antibody repertoire between EBV-negative cHL and controls in the East Asian study population

Similar to our study findings in the European population (30), no significant differences between EBV-negative cHL cases compared to disease-free controls could be identified for either IgG or IgA in our EBV proteome-wide analysis (**Figure 3-1B, Table S5, Table S6**).

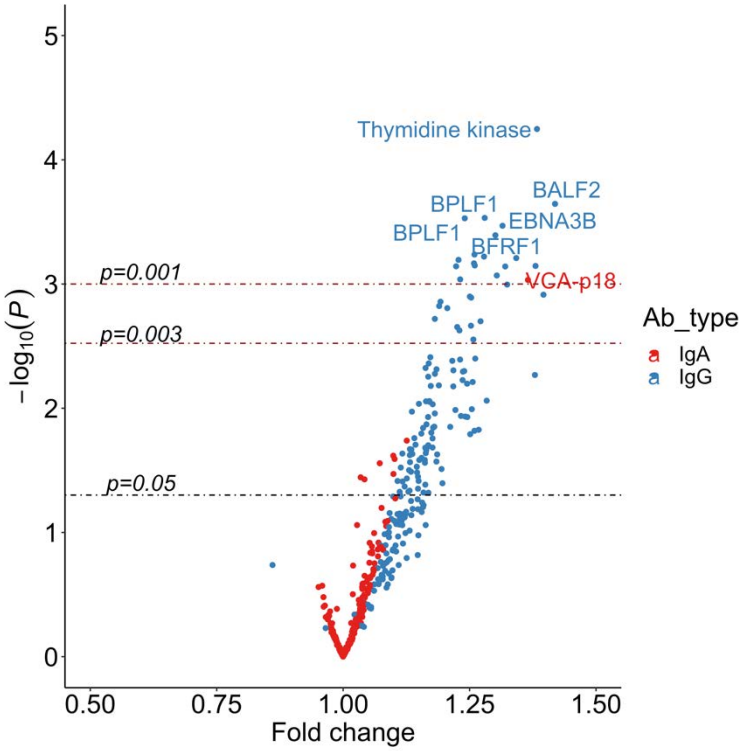


Figure 3-2 Case–case differences in standardized signal intensity (SSI) for EBV-positive cHL *vs.* EBV-negative cHL.

The x-axis of the volcano plot displays the fold change (case vs. control ratio of standardized signal intensity) for 202 protein sequences for each antibody type (red, IgA; blue, IgG). The y-axis illustrates the p-value corresponding to the t-test for a difference in SSI between EBV-positive vs. EBV-negative cHL. 74 IgG antibodies were significantly elevated in EBV-positive cHL compared to EBV-negative cHL cases (FDR<0.05). The dashed lines represent the statistically significant p-value thresholds. The antibodies with the significant p-values are highlighted.

3.4.3 EBV-positive cHL cases *vs.* EBV-negative cHL cases

Differences in the EBV antibody repertoire between EBV-positive cHL cases and EBV-negative cHL cases in the East Asian study population

The current study examined IgG and IgA antibody responses against 202 EBV protein sequences between EBV-positive cHL cases (n=35) and EBV-negative cHL cases (n=92) in East Asian study populations. We found 76 of the 202 IgG antibodies (**Figure 3-2, Table S7**), but no IgA antibodies (**Table S8**) that were significantly elevated in EBV-positive compared with EBV-negative cHL cases by t-tests (FDR p <0.05, t-test). Notably, we observed that only one IgG antibody against BALF2 remained borderline associated with EBV-positive cHL when adjusted for age, sex, and study area by linear regression models (FDR p=0.046, linear regression).

Evaluation of the generalizability of the Asian EBV antibodies in predicting EBV-positive cHL cases from EBV-negative cases in the European population

The 76 antibodies identified as significantly elevated in EBV-positive cHL cases compared with EBV-negative cHL cases in the Asian population were tested for their capacity to discriminate EBV-positive cHL from EBV-negative cHL cases in our previously studied European population (30).

Out of the 76 IgG antibodies that we identified as significantly elevated in the Asian EBV-positive cHL compared to EBV-negative cHL cases, we found 45 (p < 0.05, linear regression) were also highly associated among EBV-positive cHL in the European population after adjusted for demographics (age, sex, and study area) (**Table S9**).

Generalizability of the BALF2-IgG identified in the Asian population in predicting EBV-positive cHL cases from EBV-negative cases in the European population

The ability of BALF2-IgG alone in predicting the two groups, EBV-positive cHL from EBV-negative cHL cases in the East Asian cohort was high with an AUC of 0.73 (95% CI = 0.63, 0.83) compared to an AUC of 0.61 (95% CI = 0.53, 0.69) in the European population. When we included the subject's demographics in the models (age, sex, and study area), the discriminatory ability of BALF2-IgG improved to an AUC of 0.90 (95% CI = 0.84, 0.96) in classifying EBV status of cHL in East Asia compared to an AUC of 0.67 (95% CI = 0.59, 0.75) in the European population (**Supporting Information Figure S4)**. In the East Asian study population, demographics (age, sex, and study area) alone predicted the EBV status with an

AUC of 0.87 (95% CI = 0.80, 0.95) whilst it reported an AUC of 0.63 (95% CI = 0.55, 0.71) in the European population.

Generalizability of six antibodies and two IgG-markers in predicting EBV-positive cHL from EBV-negative cHL cases in the Asian population

In our previous European study (30), six IgG antibody markers were identified as significantly elevated in EBV-positive compared with EBV-negative cHL cases. Four of these IgG antibodies (BVRF2, BBRF1, BdRF1 and BARF1) were also consistently elevated in EBV-positive cHL cases compared with EBV-negative cHL subjects in the current geographically and ethnically distinct East-Asian population and remained significantly associated with EBV-positive cHL cases when adjusted for demographics ($p < 0.05$, linear regression) (**Table 3-4**).

Two of these six IgG markers, BdRF1(VCA p40) and BZLF1(Zta), were shown to effectively classify EBV-positive from EBV-negative cHL cases in the European population (30). We evaluated the ability of this same two-marker combination to classify the EBV status of the cHL patients in our East Asian cohort. Predictive performance for classifying EBV-positive *vs.* EBV-negative cHL was significantly improved using these two antibodies (continuous variable, AUC = 0.90, 95% CI = 0.83, 0.96, Model 1) in the model together with subject demographics compared with the subject's demographics (age, sex, and study area) alone (AUC=0.87, 95% CI = 0.80, 0.95, Model 2, **Supporting Information Figure S3)**. BdRF1-IgG marker together with patient's demographics predicted EBV status of cHL with an AUC of 0.90 (Model 5, 95% CI = 0.83, 0.96, Model 3) whilst BZLF1-IgG together with demographics had an AUC of 0.87 (Model 4). When we only included age in the model (Model 5), we obtained an AUC of 0.86 in predicting the EBV status of cHL patients, consistent with the higher incidence of cHL in the elderly (35-37). Conversely, the models which included only sex (Model 6), or only subject location (Model 7) have a lower ability to predict the EBV status of cHL patients, with an AUC of 0.66 (Model 6) and AUC of 0.51 (Model 7), respectively. (**Supporting Information Figure S3)**.

Summarised results of the antibodies that were significantly elevated and/or associated with EBV-positive cHL from European and East Asian studies are listed in the **Table 3-5**.

Table 3-4 Discrimination power (AUC and 95%CI) of the six-antibodies identified in the European study in distinguishing EBV-positive classical Hodgkin lymphoma (cHL) from EBV-negative cHL in the Asian population.

Protein name	Array sequence	t-test *p*-value	EBV+ cHL mean	EBV- cHL mean	AUC	95% CI	OR	L95	U95	Lin. reg *p*-value
BBRF1 (Late Lytic)	YP_001129476.1-102746-104587	0.001	1.69	1.23	0.90	0.83-0.96	1.48	1.15	1.91	0.003
BVRF2 (Viral Capsid)	YP_001129501.1-136465-138282.r	0.001	0.96	0.72	0.89	0.83-0.96	1.20	1.03	1.40	0.023
BdRF1 (VCA_p40)	AFY97974.1-136284-137321_US	0.001	1.51	1.08	0.90	0.83-0.96	1.52	1.16	1.99	0.003
BARF1 (Oncogene)	YP_001129453.1-66746-67654	0.012	1.47	1.19	0.89	0.82-0.96	1.27	1.02	1.58	0.036
BZLF (Zta)	CAA24861.1-102338-102210	0.087	1.42	1.22	0.87	0.80-0.95	1.12	0.88	1.44	0.366
BKRF4 (Late lytic)	YP_001129474.1-99676-100329	0.204	1.08	1.00	0.88	0.81-0.95	1.04	0.90	1.21	0.571

FDR correction method: Benjamini and Hochberg. Proteins are ordered by significant t-test *p*-values (lowest to highest) obtained for all 202 antibodies. ORs were adjusted for age, sex, and study area (Hong Kong and Taiwan) in the linear regression models for all 202 antibodies. The area under the Receiver operative Curve (AUC) was calculated for each antibody to test their ability along with the subject's demographics (*i.e.*, age group, sex, and study area) to distinguish EBV-positive cHL from controls in the current study. All six antibodies had AUC values $\geq$ 0.87.

Table 3-5 Table of results summary of significantly elevated/ associated antibodies with EBV-positive cHL cases in Europe and East Asia

	EBV-positive cHL vs Controls	EBV-positive cHL vs EBV-negative cHL
European study (Liu *et al.,* 2020)	• 18-antibodies (16 IgG and 2 IgA) highly elevated in EBV-positive cHL by t-tests and when adjusted for age, sex, and study location	• Six IgG antibodies highly elevated in EBV-positive cHL by t-tests and when adjusted for age, sex, and study location • Two IgG markers were able to best classify EBV status of cHL cases (AUC = 0.75, 95% CI = 0.68, 0.83)
Generalization of the antibodies in East Asian study population	• 15 out of 18-antibodies significantly elevated in EBV-positive cHL cHL by t-tests. 14 of the 15 antibodies were significantly associated with EBV-positive cHL when adjusted for age, sex, and study location	• Four of the six IgG were consistently elevated in EBV-positive cHL by t-tests and when adjusted for age, sex, and study location • Two IgG markers were able to classify EBV status of cHL cases with high accuracy (AUC = 0.90, 95% CI = 0.83, 0.96)
East Asian study	• 14 IgG antibodies highly elevated in EBV-positive cHL by t-tests • 10 out of 14 with additional two IgG antibodies were significantly associated with EBV-positive cHL when adjusted for age, sex, and study location	• 76 IgG antibodies highly elevated in EBV-positive cHL by t-tests • Only BALF2-IgG antibody was significantly associated with EBV-positive cHL when adjusted for age, sex, and study location and best classified EBV status of cHL cases

Generalization of the antibodies in European study population	• 11 IgG antibodies that were significantly elevated in EBV-positive cHL by t-tests and 10 of them remained significantly associated with EBV-positive cHL when adjusted for age, sex, and study area	• 47 out of 76 IgG antibodies highly elevated in EBV-positive cHL by t-tests • 45 out of 76 IgG antibodies were highly associated among EBV-positive cHL when adjusted for age, sex, and study area

3.5 Discussion

It is well established that a proportion of classical Hodgkin lymphoma (cHL) tumours are associated with EBV status (38). Different lines of evidence indicate that the two entities, EBV-positive cHL and EBV-negative cHL, represent two distinct diseases. This includes the demonstration of aberrant titres of EBV-specific antibodies and the expression of several viral antigens (*i.e.*, EBNA1, LMP-1, and LMP-2) in the malignant cells with plausible pathogenic functions (7-9, 23, 39). Irrespective of EBV's association with cHL, the biological mechanism underlying the contribution of the virus to disease pathogenesis is not yet understood. Here, we sought to further this understanding by a comprehensive evaluation of antibodies against entire EBV proteome.

Historically, serological studies on EBV have almost exclusively evaluated antibody responses against the three well-established EBV biomarkers; viral capsid antigen VCA, EBV nuclear antigen (EBNA)-1, and early antigen (EA) (40). Antibody responses against VCA-IgG, EBNA1-IgG, EA(D)-IgG (40) and IgA antibodies against VCA and EA among patients with EBV-positive cHL have been reported previously (18, 25). Accordingly, we previously undertook a comprehensive study of antibody responses to the complete EBV proteome in a European population, using a proteome array representing all EBV proteins and known splice variants (30). We reported 18 antibodies comprising of 16 IgG and two IgA markers that were significantly elevated in EBV-positive cHL cases compared to disease-free controls in the European study population (30). To the best of our knowledge, this is the only study evaluating EBV proteome-wide antibody responses in cHL.

In this study, we identified 14 IgG antibodies from the entire repertoire of 202 EBV proteome sequences spotted on our proteome array that were significantly elevated in EBV-positive cHL cases compared to controls in the Asian population. However, only 11 of the 14 and two additional IgG antibodies remained significantly associated with EBV-positive cHL cases in East Asia when adjusted for sex, age, and residential area. The most significant IgG antibody responses associated with EBV-positive cHL were against latent proteins sequences representing LMP-1, EBNA3A, and EBNA-LP, suggesting possible roles of these latent proteins in disease pathogenesis. Moreover, glycoproteins (BLLF1 (gp350/220), BDLF2, BDLF3) and early lytic proteins such as Thymidine kinase (TK), BALF2 and LF2 had strong associations with EBV-positive cHL. Interestingly, LMP-1, BDLF3, TK and BALF2 were among the 18 antibodies that were significantly elevated in EBV-positive cHL cases compared to disease-free controls in our European study (30).

Historically, the presence of LMP-1 has been determined to be a characteristic feature of Reed-Sternberg cells in tumour-specific Hodgkin lymphoma (41, 42). LMP-1 expression by immunohistochemistry is a surrogate marker of EBV-positive cHL (43, 44). LMP-1 mimics CD40 directing constitutive activation of NF-κB signalling, essential for EBV-induced transformation, enhancing B-cell survival (45, 46). Elevated responses against these early lytic proteins suggest that EBV-positive cHL patients have increased exposure to lytic viral replication and virion maturation (47, 48); although we cannot determine whether such lytic activity was present before cHL diagnosis in this retrospective study. However, antibody responses against EBNA-LP and EBNA-3A associated with EBV-positive cHL have not been previously reported. EBNA-LP is recognized as the first latent protein in resting B-cells after EBV infection (49). It is known to facilitate EBNA2-mediated transcriptional activation and is essential for EBV-mediated B-cell immortalization (50). EBNA-3A typically functions as a transcriptional regulator involved in B-cell transformation and immortalization (51). Recently, Liu *et al*. reported an association between significantly elevated IgG antibodies against EBNA3A and NKTCL in a similar study population from Hong Kong and Taiwan (30).

Glycoprotein 150, encoded by the BDLF3 gene has been recognized as a new viral immune evasion molecule (52) whilst BDLF2 is a glycoprotein that binds BMRF2 suggesting it is important for infection of epithelial cells (53, 54). BLLF1 (gp350/220) is an abundantly expressed glycoprotein in the viral envelope and plays a critical role in virus entry via endocytosis and infection of target cells. It has been shown to be involved in generating neutralizing antibodies *in vivo* (55). We observed more robust and broader overall IgG antibody responses compared to IgA in the current study. These results indicate systemic exposure to EBV infection of circulating B cells and therefore represent biologically relevant markers of lymphoid tumours, as supported by our previous findings on IgG predominance in cHL endemic BL and NKTCL (28-30).

We hypothesized that the aetiology of EBV-positive cHL is similar across populations. To test that, we assessed the generalizability of our current EBV antibodies that were significantly elevated in EBV-positive cHL in the European study population (30). We found 10 of the 16 IgG antibodies remained significantly associated with EBV-positive cHL cases compared to controls when adjusted for age, sex, and study area. Similarly, we found that the 14 out of the 18 antibodies (30) were significantly associated with EBV-positive cHL in our East Asian population when adjusted for subject's demographics with high AUCs ranged from 0.76 to 0.85.

In accordance with our study findings in the European population (30), we observed no differences in the anti-EBV antibody profile between EBV-negative cHL cases and disease-free controls in the current East Asian population. This supports our hypothesis that differences in the EBV antibody profile are specific to patients with EBV-positive cHL and are not universally observed as part of a systematically dysregulated immune response in all cHL cases. We further demonstrated that anti-EBV antibody profiles are exclusively specific to EBV-positive cHL cases and not observed in all cHL cases regardless of the EBV status.

Epidemiological evidence suggests that EBV-positive and EBV-negative cHL are distinct diseases (2, 12, 18, 23, 30). Individuals with a history of developing infectious mononucleosis during adulthood caused by primary EBV infection have a higher risk of developing EBV-positive cHL (14). However, risk stratifications for cHL are challenged by the current need to obtain tumour tissues to determine EBV status. Hence there is a need for a sensitive and specific biomarker of disease risk.

We only found IgG antibodies against BALF2 [EA(D)_p138] associated with EBV-positive cHL when adjusted for subject's demographics in the East Asian population although 76 IgG antibodies were elevated among EBV-positive cHL compared to EBV-negative cHL. However, when we tested these antibodies among the EBV-positive cHL and EBV-negative cHL using the European study population (30), 45 out of the 76 IgG antibodies remained significantly associated with EBV-positive cHL when adjusted for subject's demographics.

In our European study (30), we found six IgG markers that were significantly elevated in EBV-positive cHL cases compared to EBV-negative cHL cases. Among these, we identified a two-marker IgG combination, BdRF1(VCAp40)-IgG and BZLF1(Zta)-IgG, along with patient demographics, that best classified EBV-positive from EBV-negative cHL cases. Then, we replicated this finding in the current ethnically distinct study cohort. We achieved a higher AUC of 0.90 for the two-marker combination that classified EBV status in our cHL samples better than patient demographics alone. These results allowed us to identify a possible serology-based two IgG marker combination that could potentially classify cHL cases based on their EBV status in diverse study populations. However, we also reported that BdRF1-IgG with patient's demographics had an AUC of 0.90 in the current East Asian study population. Our results also indicated that age was a crucial individual covariate in predicting the EBV status of these cHL patients, consistent with previous study reports (35-37). The impaired immunity over latent infection is suggested to be responsible for the occurrence of EBV-positive cHL in older patients (56).

In addition, BALF2 [EA(D)_p138]-IgG was found in the current study to classifying EBV status of the cHL patients in the current study population with an AUC of 0.73. When we tested the discrimination power of this antibody in the European study population, the AUC decreased to an AUC=0.61. BALF2 gene has been suggested to play an important role in the EBV-lytic cycle induction (57).

Our study is limited by its retrospective nature and small numbers of samples analysed by proteome array. A prospective study design would be required to explore the anti-EBV antibody profiles before disease onset. Additional information on the history of infectious mononucleosis, education level, cigarette smoking, and other potential confounding factors for study subjects was also unavailable. Another technical limitation is that the proteome array is not designed to detect antibody responses against conformational epitopes, so information on post-translational processing associated with cHL is lacking. Still, our findings are not likely to introduce false associations due to this phenomenon. Our custom proteome microarray results from previous studies have been validated using multiplex serology (58) and ELISA specifically tested for synthetic VCAp18, EBNA1, and EAdp47 peptides (27, 29).

Our results supported the hypothesis that the aetiology of EBV-positive cHL is similar across populations in different geographic locations. The successful replication of our previous and current study findings implicates that the associations of EBV-positive cHL with specific anti-EBV antibodies are reproducible independent of geographic location and ethnic diversity. Antibody responses directed to EBV proteins involved with viral replication, maturation, EBV-induced transformation, and B-cell immortalization provide biologic plausibility for future research on EBV-positive cHL pathogenesis. Future studies should focus on improving our understanding of how antibody responses may reflect the role of EBV in cHL and enhance the risk stratification of patients with cHL.

3.6 References

1. Young LS, Rickinson AB. Epstein–Barr virus: 40 years on. Nature Reviews Cancer. 2004;4(10):757-68.

2. Glaser SL, Lin RJ, Stewart SL, Ambinder RF, Jarrett RF, Brousset P, et al. Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. International Journal of Cancer. 1997;70(4):375-82.

3. Pileri S, Ascani S, Leoncini L, Sabattini E, Zinzani P, Piccaluga P, et al. Hodgkin's lymphoma: the pathologist's viewpoint. Journal of Clinical Pathology. 2002;55(3):162-76.

4. Herbst H, Steinbrecher E, Niedobitek G, Young LS, Brooks L, Muller-Lantzsch N, et al. Distribution and phenotype of Epstein-Barr virus-harboring cells in Hodgkin's disease. Blood. 1992.

5. Hummel M, Anagnostopoulos I, Dallenbach F, Korbjuhn P, Dimmler C, Stein H. EBV infection patterns in Hodgkin's disease and normal lymphoid tissue: expression and cellular localization of EBV gene products. British Journal of Haematology. 1992;82(4):689-94.

6. Weiss LM, Chen YY, Liu XF, Shibata D. Epstein-Barr virus and Hodgkin's disease. A correlative in situ hybridization and polymerase chain reaction study. The American Journal of Pathology. 1991;139(6):1259-65.

7. Wu TC, Mann RB, Charache P, Hayward SD, Staal S, Lambe BC, et al. Detection of EBV gene expression in Reed-Sternberg cells of Hodgkin's disease. International Journal of Cancer. 1990;46(5):801-4.

8. Weiss LM, Movahed LA, Warnke RA, Sklar J. Detection of Epstein–Barr Viral Genomes in Reed–Sternberg Cells of Hodgkin's Disease. New England Journal of Medicine. 1989;320(8):502-6.

9. Levine PH, Ablashi DV, Berard CW, Carbone PP, Waggoner DE, Malan L. Elevated antibody titers to Epstein-Barr virus in Hodgkin's disease. Cancer. 1971;27(2):416-21.

10. Hjalgrim H, Engels E. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. Journal of Internal Medicine 2008;264(6):537-48.

11. Jarrett R, Krajewski A, Angus B, Freeland J, Taylor P, Taylor G, et al. The Scotland and Newcastle epidemiological study of Hodgkin’s disease: impact of histopathological review and EBV status on incidence estimates. Journal of Clinical Pathology. 2003;56(11):811-6.

12. Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y, Chabrier A, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein–Barr virus status–defined subgroups. Journal of the National Cancer Institute. 2012;104(3):240-53.
13. Evans A, Comstock G. Presence of elevated antibody titres to Epstein-Barr virus before Hodgkin's disease. The Lancet. 1981;317(8231):1183-6.
14. Ambinder RF. Epstein-barr virus and hodgkin lymphoma. ASH Education Program Book. 2007;2007(1):204-9.
15. Jha HC, Pei Y, Robertson ES. Epstein–Barr virus: diseases linked to infection and transformation. Frontiers in Microbiology. 2016;7:1602.
16. Farina A, Santarelli R, Gonnella R, Bei R, Muraro R, Cardinali G, et al. The BFRF1 gene of Epstein-Barr virus encodes a novel protein. Journal of Virology. 2000;74(7):3235-44.
17. Middeldorp JM. Epstein-Barr virus-specific humoral immune responses in health and disease, 2015. 289-323 p.
18. Chang ET, Zheng T, Lennette ET, Weir EG, Borowitz M, Mann R, et al. Heterogeneity of risk factors and antibody profiles in Epstein-Barr virus genome-positive and-negative Hodgkin lymphoma. The Journal of Infectious Diseases. 2004;189(12):2271-81.
19. Coghill AE, Hildesheim A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. American journal of epidemiology. 2014;180(7):687-95.
20. de Sanjosé S, Bosch R, Schouten T, Verkuijlen S, Nieters A, Foretova L, et al. Epstein-Barr virus infection and risk of lymphoma: Immunoblot analysis of antibody responses against EBV-related proteins in a large series of lymphoma subjects and matched controls. International Journal of Cancer. 2007;121(8):1806-12.
21. Evans AS, Gutensohn NM. A population-based case-control study of EBV and other viral antibodies among persons with Hodgkin's disease and their siblings. International Journal of Cancer. 1984;34(2):149-57.
22. Lehtinen T, Lumio J, Dillner J, Hakama M, Knekt P, Lehtinen M, et al. Increased risk of malignant lymphoma indicated by elevated Epstein-Barr virus antibodies—a prospective study. Cancer Causes & Control. 1993;4(3):187-93.
23. Levin LI, Chang ET, Ambinder RF, Lennette ET, Rubertone MV, Mann RB, et al. Atypical prediagnosis Epstein-Barr virus serology restricted to EBV-positive Hodgkin lymphoma. Blood, The Journal of the American Society of Hematology. 2012;120(18):3750-5.

24. Meij P, Vervoort MB, Bloemena E, Schouten TE, Schwartz C, Grufferman S, et al. Antibody responses to Epstein-Barr virus-encoded latent membrane protein-1 (LMP1) and expression of LMP1 in juvenile Hodgkin's disease. Journal of Medical Virology. 2002;68(3):370-7.
25. Mueller N, Evans A, Harris NL, Comstock GW, Jellum E, Magnus K, et al. Hodgkin's disease and Epstein-Barr virus. New England Journal of Medicine. 1989;320(11):689-95.
26. Mueller NE, Lennette ET, Dupnik K, Birmann BM. Antibody titers against EBNA1 and EBNA2 in relation to Hodgkin lymphoma and history of infectious mononucleosis. International Journal of Cancer. 2012;130(12):2886-91.
27. Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clinical Cancer Research. 2018;24(6):1305-14.
28. Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, et al. The association between the comprehensive Epstein–Barr virus serologic profile and endemic Burkitt lymphoma. Cancer Epidemiology and Prevention Biomarkers. 2020;29(1):57-62.
29. Liu Z, Sarathkumara YD, Chan JKC, Kwong Y-L, Lam TH, Ip DKM, et al. Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma. Scientific Reports. 2021;11(1):23664.
30. Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. International Journal of Cancer. 2020;147(3):608-18.
31. Shamay M, Kanakry JA, Low JSW, Horowitz NA, Journo G, Ahuja A, et al. CpG methylation in cell-free Epstein-Barr virus DNA in patients with EBV-Hodgkin lymphoma. Blood Advances. 2020;4(8):1624-7.
32. Johnson PC, McAulay KA, Montgomery D, Lake A, Shield L, Gallagher A, et al. Modeling HLA associations with EBV-positive and-negative H odgkin lymphoma suggests distinct mechanisms in disease pathogenesis. International journal of cancer. 2015;137(5):1066-75.
33. Smedby KE, Hjalgrim H, Melbye M, Torrång A, Rostgaard K, Munksgaard L, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. Journal of the National Cancer Institute. 2005;97(3):199-209.
34. Proietti C, Zakrzewski M, Watkins TS, Berger B, Hasan S, Ratnatunga CN, et al. Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) Web-Server. Scientific Reports. 2016;6(1):1-15.

35. Asano N, Yamamoto K, Tamaru J-I, Oyama T, Ishida F, Ohshima K, et al. Age-related Epstein-Barr virus (EBV)–associated B-cell lymphoproliferative disorders: comparison with EBV-positive classic Hodgkin lymphoma in elderly patients. Blood. 2009;113(12):2629-36.
36. Jarrett RF, Stark GL, White J, Angus B, Alexander FE, Krajewski AS, et al. Impact of tumor Epstein-Barr virus status on presenting features and outcome in age-defined subgroups of patients with classic Hodgkin lymphoma: a population-based study. Blood. 2005;106(7):2444-51.
37. Stark GL, Wood KM, Jack F, Angus B, Proctor SJ, Taylor PR, et al. Hodgkin's disease in the elderly: a population-based study. British Journal of Haematology. 2002;119(2):432-40.
38. Massini G, Siemer D, Hohaus S. EBV in Hodgkin lymphoma. Mediterranean Journal of Hematology and Infectious Diseases. 2009;1(2).
39. Küppers R. The biology of Hodgkin's lymphoma. Nature Reviews Cancer. 2009;9(1):15-27.
40. De Paschale M, Clerici P. Serological diagnosis of Epstein-Barr virus infection: Problems and solutions. World Journal of Virology. 2012;1(1):31-43.
41. Andersson J. Epstein-Barr virus and Hodgkin's lymphoma. Herpes. 2006;13(1):12-6.
42. Gulley ML, Glaser SL, Craig FE, Borowitz M, Mann RB, Shema SJ, et al. Guidelines for interpreting EBER in situ hybridization and LMP1 immunohistochemical tests for detecting Epstein-Barr virus in Hodgkin lymphoma. American Journal of Clinical Pathology. 2002;117(2):259-67.
43. Hashmi AA, Hussain ZF, Hashmi KA, Zafar MI, Edhi MM, Faridi N, et al. Latent membrane protein 1 (LMP1) expression in Hodgkin lymphoma and its correlation with clinical and histologic parameters. World Journal of Surgical Oncology. 2017;15(1):89-.
44. Santisteban-Espejo A, Perez-Requena J, Atienza-Cuevas L, Moran-Sanchez J, Fernandez-Valle MDC, Bernal-Florindo I, et al. Prognostic Role of the Expression of Latent-Membrane Protein 1 of Epstein-Barr Virus in Classical Hodgkin Lymphoma. Viruses. 2021;13(12).
45. Huen DS, Henderson SA, Croom-Carter D, Rowe M. The Epstein-Barr virus latent membrane protein-1 (LMP1) mediates activation of NF-kappa B and cell surface phenotype via two effector regions in its carboxy-terminal cytoplasmic domain. Oncogene. 1995;10(3):549-60.

46. Bargou RC, Emmerich F, Krappmann D, Bommert K, Mapara MY, Arnold W, et al. Constitutive nuclear factor-kappaB-RelA activation is required for proliferation and survival of Hodgkin's disease tumor cells. Journal of Clinical Investigation. 1997;100(12):2961-9.
47. Chiu Y-F, Sugden B, Chang P-J, Chen L-W, Lin Y-J, Lan Y-C, et al. Characterization and Intracellular Trafficking of Epstein-Barr Virus BBLF1, a Protein Involved in Virion Maturation. Journal of Virology. 2012;86(18):9647-55.
48. Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al. Human herpesviruses: biology, therapy, and immunoprophylaxis: Cambridge University Press; 2007.
49. Szymula A, Palermo RD, Bayoumy A, Groves IJ, Ba abdullah M, Holder B, et al. Epstein-Barr virus nuclear antigen EBNA-LP is essential for transforming naïve B cells, and facilitates recruitment of transcription factors to the viral genome. PLOS Pathogens. 2018;14(2):e1006890.
50. Wang C, Zhou H, Xue Y, Liang J, Narita Y, Gerdt C, et al. Epstein-Barr Virus Nuclear Antigen Leader Protein Coactivates EP300. Journal of Virology. 2018;92(9):e02155-17.
51. Cludts I, Farrell PJ. Multiple functions within the Epstein-Barr virus EBNA-3A protein. Journal of Virology. 1998;72(3):1862-9.
52. Gram AM, Oosenbrug T, Lindenbergh MF, Büll C, Comvalius A, Dickson KJ, et al. The Epstein-Barr Virus Glycoprotein gp150 Forms an Immune-Evasive Glycan Shield at the Surface of Infected Cells. PLoS Pathogens. 2016;12(4):e1005550.
53. Gore M, Hutt-Fletcher LM. The BDLF2 protein of Epstein–Barr virus is a type II glycosylated envelope protein whose processing is dependent on coexpression with the BMRF2 protein. Virology. 2009;383(1):162-7.
54. Tugizov SM, Berline JW, Palefsky JM. Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. Nature medicine. 2003;9(3):307-14.
55. Thorley-Lawson D, Poodry C. Identification and isolation of the main component (gp350-gp220) of Epstein-Barr virus responsible for generating neutralizing antibodies in vivo. Journal of Virology. 1982;43(2):730-6.
56. Grywalska E, Rolinski J, editors. Epstein-barr virus–associated lymphomas. Seminars in Oncology; 2015: Elsevier.
57. Decaussin G, Leclerc V, Ooka T. The lytic cycle of Epstein-Barr virus in the nonproducer Raji line can be rescued by the expression of a 135-kilodalton protein encoded by the BALF2 open reading frame. Journal of Virology. 1995;69(11):7309-14.

58. Simon J, Liu Z, Brenner N, Yu KJ, Hsu W-L, Wang C-P, et al. Validation of an Epstein-Barr Virus Antibody Risk Stratification Signature for Nasopharyngeal Carcinoma by Use of Multiplex Serology. Journal of Clinical Microbiology. 2020;58(5):e00077-20.

3.7 Supplementary Information:

Additional details EBV-based markers for distinguishing between EBV-positive and EBV-negative HL tumours

In the present study, the limited availability of tumour tissues from patients with classical Hodgkin lymphoma (HL) precluded us from determining tumour EBV status using immunohistochemical staining of tumour biopsies for in situ hybridization for EBV-encoded small RNAs (EBERs). A tool requiring only blood samples would obviate the need for tumour tissue to conduct work accounting for HL EBV status and could, therefore, prove very useful in future epidemiological research.

We aimed to determine the optimal EBV-based markers from plasma to distinguish between EBV-positive and EBV-negative HL tumours. We utilized two approaches – EBV DNA load by real-time polymerase chain reaction (PCR) and CpG methylation index (calculated as the absolute copy number of EBV DNA recovered from the high-salt eluate divided by the sum of the absolute copy number of EBV DNA in all fractions) (1).

We tested for a total of 134 HL cases with sufficient volume of plasma (>1mL) from the present study. Plasma samples were handled, and DNAs were extracted following the proposal published previously (1, 2). Briefly, EBV copy number was determined by real-time PCR with a primer pair and probe corresponding to the BamH-W region of the EBV genome (sense: CCCAACACTCCACCACACC, antisense: TCTTAGGAGCTGTCCGAGGG, and probe: 5′-[6-FAM]CACACACTACACACACCCACCCGTCTC [BHQ-1]-3′) were used (2). Extracted DNA was mixed with 950 ng carrier K562 DNA and added to 10 µL MBD-Bead slurry (MethylMiner DNA Enrichment Kit, Invitrogen, Carlsbad, CA), incubated together on a rotating mixer for 1 hour, and washed. Unbound, low-salt washes and a high-salt eluate (2000 mM NaCl) were collected. EBV was quantitated by PCR in the high-salt eluate and in the washes and flow through. Fractions were ethanol precipitated, resuspended in water, and PCR amplified with primer pair and probe corresponding to the BamH-W region of the EBV genome (amplicon length, 76 bp; repeat located between 14649 and 33137 on the reference sequence NC_007605.1). The EBV methylation index was calculated as the absolute copy number of EBV DNA recovered from the high-salt eluate divided by the sum of the absolute copy number of EBV DNA in all fractions.

Of 134 cases, 70 had known EBV status (24 EBV-positive and 46 EBV-negative HL) based on the gold standard (EBERs staining using tumour tissues). Based on the 70 cases, the area under the receiver operating characteristic (ROC) curve (AUC) of EBV DNA load was 0.89 (95% confidence interval [CI]=0.80-0.98), which was slightly higher than that of methylation index (AUC=0.85, 95%CI=0.76-0.95; *P* for difference= 0.28). Therefore, we concluded that EBV DNA load is an optimal approach to distinguishing EBV tumour status of HL.

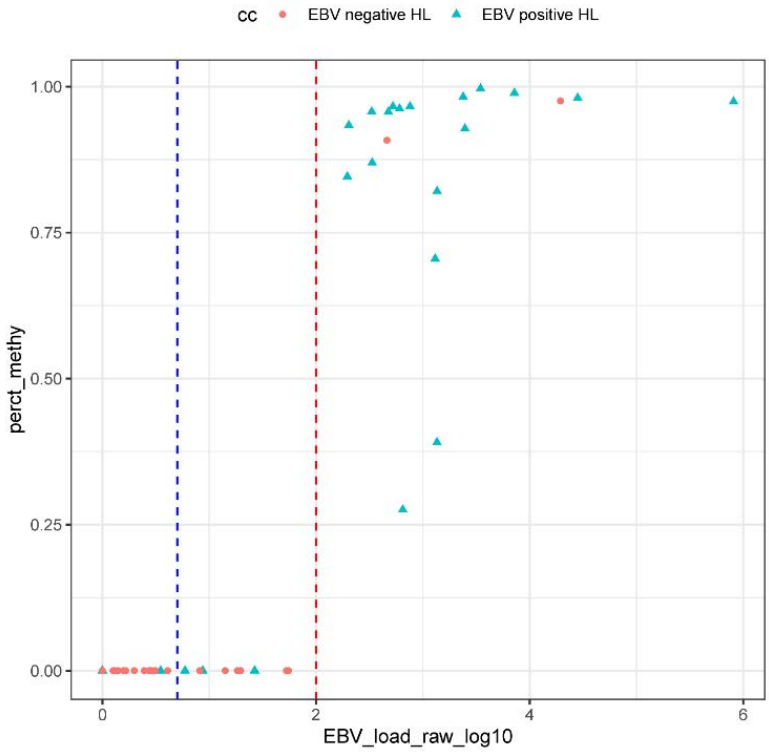


Figure S1. Distribution of EBV DNA load and methylation index among 24 EBV-positive HL patients and 46 EBV-negative HL patients

We then determined the optimal cut-offs for calling EBV-positive and EBV-negative tumours. We set two cut-offs based on the distributions shown in **Supplementary Figure S1**, that patients with an EBV DNA load of ≥ 100 copies/mL were defined as having EBV-positive HL tumours, and that patients with an EBV DNA load of <5 copies/mL were defined as having EBV-negative HL tumours. As shown in **Supplementary Table S1**, the positive predictive value (PPV) for calling an EBV-positive HL tumour at a cut-off of 100 copies/mL was 90%, and the PPV for calling an EBV-negative HL tumour at a cut-off of 5 copies/mL was 92.5%.

Table S1. Number of patients with EBV-positive and EBV-negative HL at different EBV DNA loads

Cut-off (copies/mL)	EBV-positive HL (n=24)	EBV-negative HL (n=46)
100+	18	2
5+ - <100	3	7
<5	3	37

Finally, we applied these two cut-offs to the remaining 64 samples with unknown EBV tumour status. We found that 11 HL cases would be grouped to EBV positive (*i.e.*, plasma EBV DNA load >100 copies/mL), and 44 would be grouped to EBV negative (*i.e.*, plasma EBV DNA load <5 copies/mL). These cases have been added to the cases with EBV status determined based on EBERs staining (*i.e.*, 24 EBV-positive and 48 EBV-negative HL), leading to a total of 35 EBV-positive HL cases and 92 EBV-negative HL cases in the final analysis.

References

1. Shamay M, Kanakry JA, Low JSW, Horowitz NA, Journo G, Ahuja A, et al. CpG methylation in cell-free Epstein-Barr virus DNA in patients with EBV-Hodgkin lymphoma. Blood Adv. 2020;4(8):1624-7.

2. Kanakry JA, Li H, Gellert LL, Lemas MV, Hsieh WS, Hong F, et al. Plasma Epstein-Barr virus DNA predicts outcome in advanced Hodgkin lymphoma: correlative analysis from a large North American cooperative group trial. Blood. 2013;121(18):3547-53.

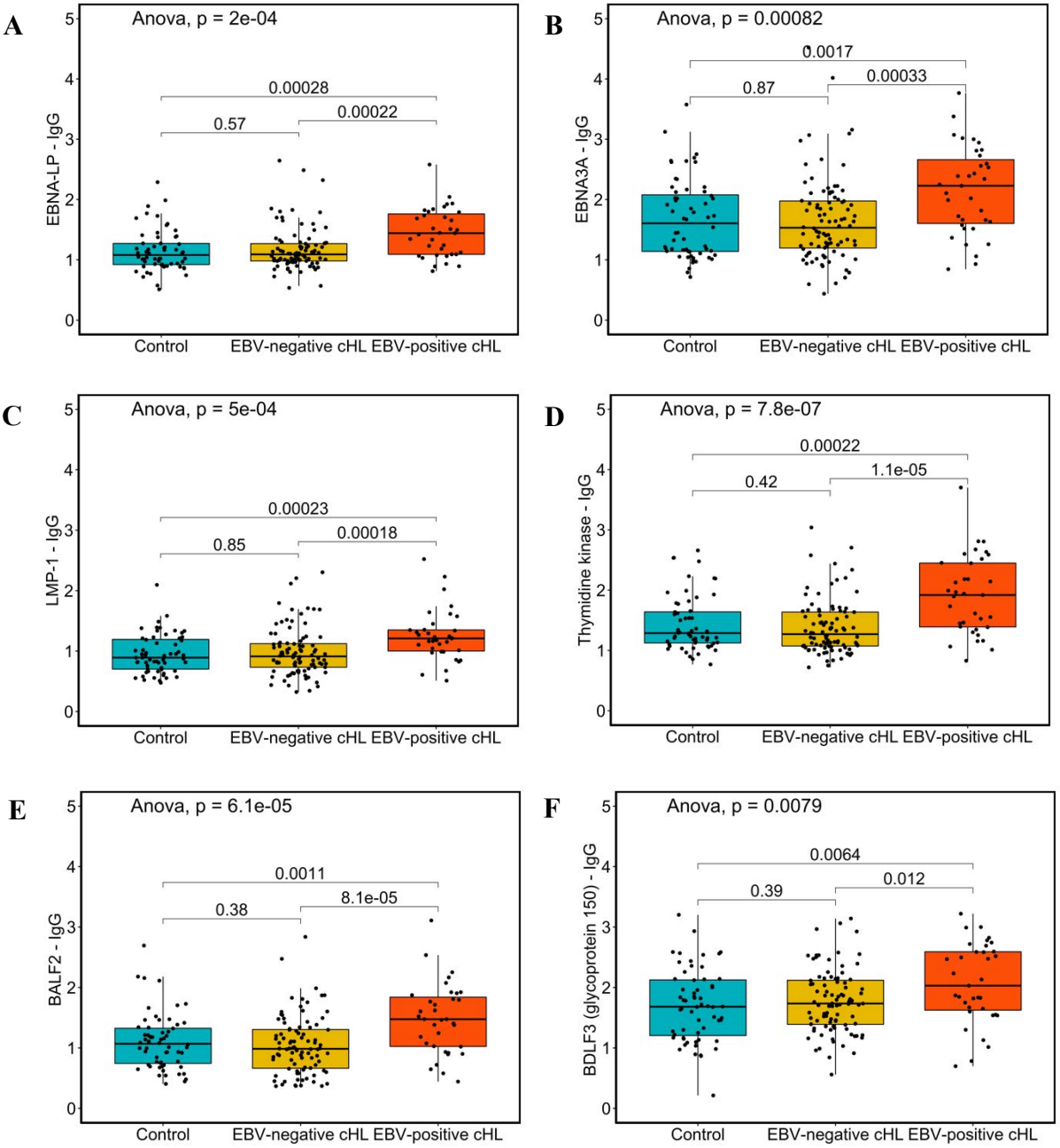
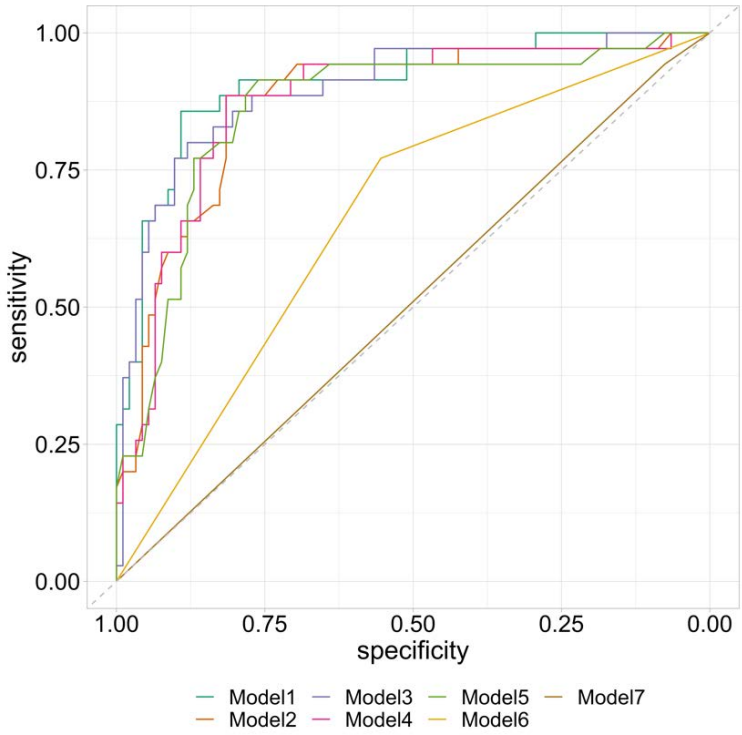


Figure S2. Average standardized signal intensity for the four anti-EBV IgG antibodies with the lowest *p* values for the comparison between EBV-positive cHL, EBV-negative cHL cases and controls. (A) EBNA-LP; (B) EBNA3A; (C) LMP-1; (D) Thymidine Kinase; (E) BALF2 and (F) Thymidine Kinase. Box plots show the average standardized signal intensity comparison between EBV-positive cHL, EBV-negative cHL cases and controls. *p*-values from the global ANOVA test and each two-way t-test (*i.e.*, EBV-positive cHL *vs*. EBV-negative cHL) are listed. Boxes show interquartile range (IQR).

A



B

	Model description	AUC	95% CI
Model 1	predicting EBV status of cHL with two antibodies (BZLF1-IgG and BdRF-IgG) + demographics (age, sex, and study area)	0.90	0.83-0.96
Model 2	predicting EBV status of cHL with demographics	0.87	0.80-0.95
Model 3	predicting EBV status of cHL with BdRF-IgG and demographics	0.90	0.83-0.96
Model 4	predicting EBV status of cHL with BZLF1-IgG and demographics	0.87	0.80-0.95
Model 5	predicting EBV status of cHL with subject's age alone	0.86	0.79-0.94
Model 6	predicting EBV status of cHL with subject's sex alone	0.66	0.58-0.75
Model 7	predicting EBV status of cHL with subject's study area alone	0.51	0.46-0.56

Figure S3. Receiver Operating Curves (ROC), for classifying cHL tumours as either EBV positive or negative (A) using two serological markers, BZLF1-IgG and BdRF-IgG, and subject's demographics. **(B)** Table describing the different models together with area under the ROC curve (AUC) and 95% confidence intervals (CI) for discriminating EBV status of cHL tumours.

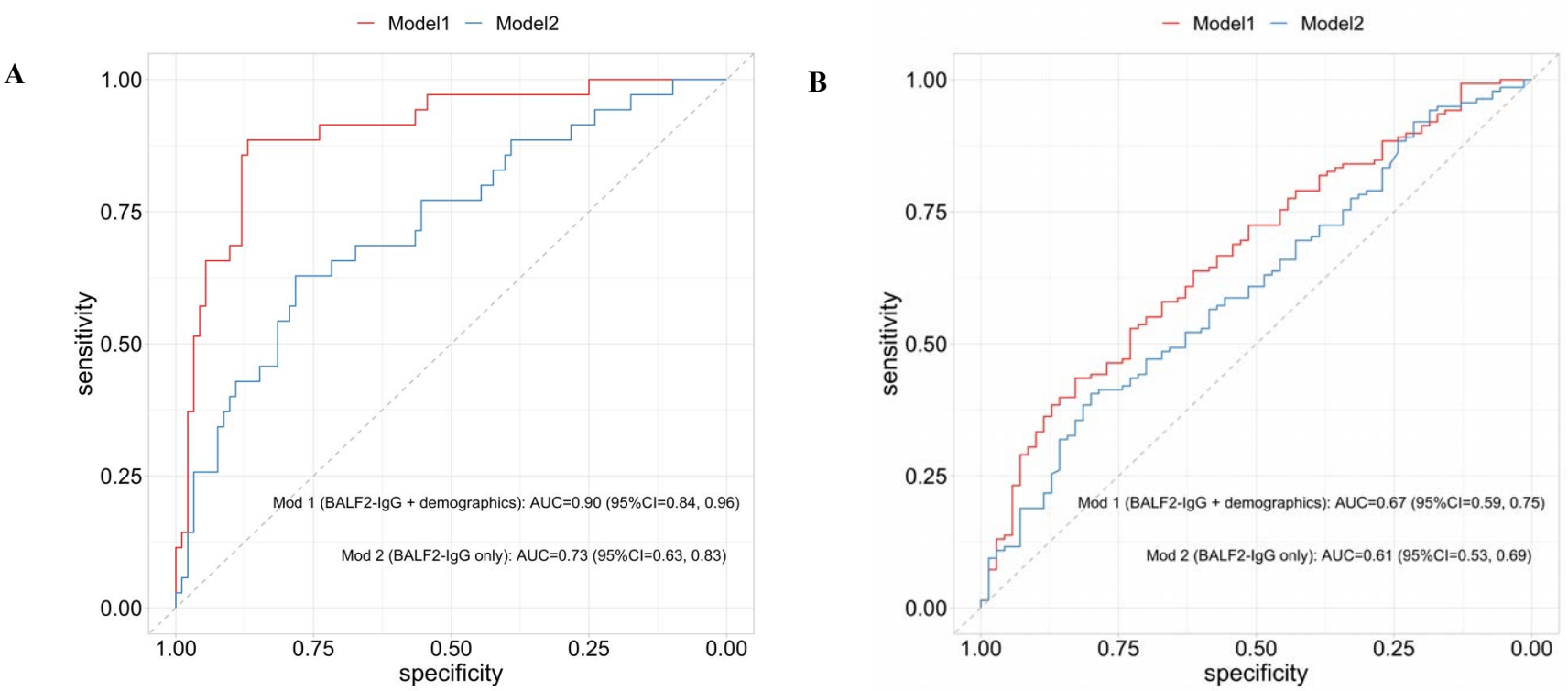


Figure S4. Receiver Operating Curve (ROC) for classifying cHL tumours as either EBV positive or negative using BALF2-IgG and subject demographics. (A). ROC curves for BALF2-IgG in the East Asia. Model 1: combination of BALF2-IgG and patient demographics (area under the curve [AUC] = 0.90, 95% CI = 0.84, 0.96). Model 2: BALF2-IgG alone had an AUC of 0.73 (95% CI = 0.63, 0.83). (B). ROC curves for BALF2-IgG antibody in the European study. Model 1: a combination of BALF2-IgG and patient demographics (AUC = 0.67, 95% CI = 0.59, 0.75). Model 2: BALF2-IgG alone had an AUC of 0.61 (95% CI = 0.53, 0.69).

Table S2 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgG antibody responses comparisons on mean difference between EBV-positive cHL cases and controls. Table is ordered by t-test p-value (lowest to highest) for the first 110 antibodies. The antibodies that were previously identified as significantly different from EBV-positive cHL cases compared to controls in our European study are highlighted in bold. *14 antibodies (t-test adj p FDR <0.05). ORs were adjusted for age, sex and study area (Hong Kong and Taiwan). T-test Adj p FDR: FDR p-value obtained for t-tests, Adj p FDR: is the FDR p-value obtained for the linear regression model.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test adj *p* FDR	EBV+ cHL mean	Control mean	OR	L95	U95	Lin. reg *p*-value	Lin. reg adj *p* FDR
THY.KINASE*	YP_001129497.1-133399-131576	Early lytic	IgG	3.72E-04	0.037	1.91	1.44	1.56	1.21	2.01	0.001	0.041
EBNA-LP*	YP_001129440.1-20824-20955	Latent	IgG	4.36E-04	0.037	1.45	1.14	1.35	1.14	1.61	0.001	0.041
LMP1*	YP_401722.1-168507-167702	Latent	IgG	0.001	0.037	1.25	0.95	1.43	1.21	1.69	5.93E-05	0.012
BBLF1*	AFY97956.1-108555-108328	Early lytic	IgG	0.002	0.037	2.15	1.75	1.46	1.12	1.9	0.006	0.061
BALF2 (EA(D)_p138)*	YP_001129510.1-165796-162410-1	Early lytic	IgG	0.002	0.037	1.48	1.10	1.49	1.17	1.89	0.001	0.042
LMP2A*	YP_001129436.1-167587-167942	Latent	IgG	0.002	0.037	1.26	1.04	1.23	1.07	1.42	0.006	0.061
LF2*	YP_001129504.1-151808-150519	Early lytic	IgG	0.002	0.037	1.48	1.19	1.32	1.11	1.57	0.002	0.043
EBNA3A*	YP_001129463.1-80447-82888	Latent	IgG	0.002	0.037	2.15	1.66	1.74	1.28	2.38	0.001	0.041
BPFL1*	YP_001129449.1-59370-49906-2	Late lytic	IgG	0.002	0.037	1.72	1.40	1.39	1.13	1.7	0.002	0.043
BDLF2*	YP_001129491.1-120928-119666	Glycoprotein	IgG	0.002	0.037	1.89	1.52	1.43	1.14	1.8	0.003	0.046
BPFL1*	YP_001129449.1-59370-49906-3	Late lytic	IgG	0.002	0.037	1.62	1.32	1.32	1.1	1.58	0.004	0.051
BLLF1 (gp350/220)*	YP_001129462.1-79936-77276 re	Glycoprotein	IgG	0.002	0.037	1.55	1.21	1.45	1.17	1.81	0.001	0.042

LMP2B*	AFY97910.1-1026-1196	Latent	IgG	0.003	0.042	1.39	1.12	1.3	1.09	1.54	0.004	0.051
EBNA3A*	YP_401669.1-80382-82877	Latent	IgG	0.003	0.042	2.09	1.64	1.67	1.24	2.27	0.001	0.042
EBNA3A	AFY97915.1-80252-82747	Latent	IgG	0.004	0.051	2.00	1.56	1.63	1.2	2.2	0.002	0.043
LMP2A	YP_001129436.1-871-951	Latent	IgG	0.004	0.053	1.40	1.20	1.19	1.04	1.36	0.013	0.093
EBNA3B	YP_001129464.1-83074-83430	Latent	IgG	0.004	0.053	1.62	1.36	1.3	1.07	1.58	0.010	0.078
BVRF2	YP_001129501.1-136465-138282	Late lytic	IgG	0.005	0.057	2.07	1.72	1.47	1.14	1.9	0.004	0.051
BFRF1A	AFY97921.1-46216-46623	Other/Unknown	IgG	0.006	0.059	0.70	0.55	1.17	1.04	1.31	0.010	0.078
EBNA3B	YP_001129464.1-83509-86532-2	Latent	IgG	0.006	0.059	1.73	1.40	1.4	1.1	1.8	0.009	0.076
BFLF2	YP_001129443.1-44763-43807	Late lytic	IgG	0.006	0.059	1.35	1.12	1.26	1.07	1.48	0.006	0.061
BVRF2	YP_001129501.1-136465-138282 re	Late lytic	IgG	0.007	0.064	0.96	0.76	1.23	1.05	1.43	0.010	0.078
EBNA3B	CAA24858.1-95788-98247	Latent	IgG	0.008	0.065	1.93	1.58	1.47	1.12	1.93	0.007	0.061
BDLF3	AFY97964.1-118644-117940	Latent	IgG	0.008	0.065	2.07	1.70	1.57	1.18	2.09	0.003	0.046
LMP1	YP_401722.1-168670-168584-m2	Latent	IgG	0.008	0.068	1.37	1.14	1.3	1.08	1.57	0.007	0.061
BDLF3	YP_001129490.1-119605-118901	Glycoprotein	IgG	0.010	0.076	2.06	1.71	1.52	1.16	2	0.004	0.051
BPFL1	CAA24839.1-71527-62078-2	Late lytic	IgG	0.010	0.076	2.32	1.99	1.44	1.09	1.88	0.011	0.079
EBNA3A	CAA24856.1-92243-92602	Latent	IgG	0.012	0.085	1.50	1.18	1.47	1.13	1.91	0.005	0.059
BILF2	YP_001129503.1-139063-138317	Glycoprotein	IgG	0.012	0.085	1.75	1.47	1.33	1.07	1.67	0.014	0.093
BRRF2	AFY97943.1-93884-95497	Late lytic	IgG	0.016	0.100	2.31	2.02	1.38	1.06	1.79	0.020	0.107
BHRF1	YP_001129442.1-42204-42779	Early lytic	IgG	0.016	0.100	1.59	1.37	1.26	1.05	1.5	0.014	0.093
BSRF1	YP_001129458.1-74770-75426	Other/Unknown	IgG	0.017	0.100	1.52	1.22	1.31	1.04	1.64	0.023	0.120
EBNA3C	YP_001129465.1-86654-87013	Latent	IgG	0.017	0.100	1.08	0.93	1.14	1	1.29	0.045	0.154
EAD	EA(D) 0.1	Early lytic	IgG	0.017	0.100	0.23	0.18	1.05	1	1.1	0.072	0.184
BdRF1 (VCA_p40)	AFY97974.1-136284-137321_US	Late lytic	IgG	0.018	0.100	1.51	1.19	1.39	1.07	1.8	0.017	0.102
BRLF1 (Rta)	YP_001129468.1-93725-91908	Immediate early lytic	IgG	0.018	0.100	1.48	1.21	1.26	1.01	1.56	0.040	0.154

CAPSID	YP_001129451.1-63084-64178	Late lytic	IgG	0.019	0.106	1.37	1.15	1.22	1.01	1.47	0.043	0.154
EBNA3C	CAA24859.1-98371-98730	Latent	IgG	0.021	0.111	1.27	1.09	1.16	0.97	1.38	0.110	0.222
EBNA3B	CAA24858.1-95353-95709	Latent	IgG	0.021	0.111	1.51	1.26	1.32	1.05	1.66	0.018	0.105
BDLF4	YP_001129488.1-117560-116883	Early lytic	IgG	0.022	0.113	1.74	1.52	1.22	1	1.48	0.054	0.165
LMP2A	YP_001129436.1-540-788	Latent	IgG	0.024	0.116	1.22	1.06	1.12	0.99	1.28	0.083	0.193
BBRF1	YP_001129476.1-102746-104587	Late lytic	IgG	0.024	0.116	1.69	1.38	1.33	1.01	1.75	0.047	0.154
FGAM-synthase	YP_001129438.1-1736-5692-2	Other/Unknown	IgG	0.028	0.123	1.49	1.29	1.21	1.01	1.45	0.041	0.154
BALF5 (DNA polymerase)	YP_001129507.1-157772-154725-2	Early lytic	IgG	0.028	0.123	1.27	1.13	1.15	1.02	1.3	0.024	0.122
FGAM-synthase	YP_001129438.1-1736-5692-1	Other/Unknown	IgG	0.029	0.123	1.24	1.07	1.17	1.01	1.37	0.041	0.154
BcLF1 (VCA_p160)	CAA24794.1-137466-133321-1	Late lytic	IgG	0.030	0.123	1.43	1.19	1.26	1.03	1.54	0.029	0.131
EBNA3B	YP_001129464.1-83509-86532-1_US	Latent	IgG	0.030	0.123	1.81	1.53	1.36	1.06	1.76	0.020	0.107
LF2	AFY97966.1-125043-127295	Early lytic	IgG	0.030	0.123	1.14	0.98	1.12	0.97	1.3	0.118	0.234
BZLF1 (Zta)	YP_001129467.1-91697-91197	Immediate early lytic	IgG	0.030	0.123	1.53	1.30	1.28	1.03	1.6	0.027	0.128
EBNA-LP	YP_001129440.1-29887-29952	Latent	IgG	0.030	0.123	1.23	1.08	1.14	0.99	1.32	0.068	0.183
A73	AFY97981.1-156513-156598	Other/Unknown	IgG	0.031	0.123	0.83	0.71	1.1	0.98	1.23	0.100	0.216
BZLF1 (Zta)	CAA24861.1-103155-102655	Immediate early lytic	IgG	0.034	0.129	1.29	1.08	1.26	1.03	1.53	0.026	0.124
BILF1	YP_001129506.1-154125-153187	Early lytic	IgG	0.034	0.129	1.33	1.19	1.12	0.98	1.29	0.110	0.222
BGLF5 (DNAse)	AFY97955.1-109922-108510	Early lytic	IgG	0.036	0.129	1.21	1.08	1.09	0.96	1.24	0.187	0.318
BRRF2	YP_001129470.1-94844-96457	Late lytic	IgG	0.037	0.129	2.38	2.13	1.31	1.01	1.71	0.044	0.154
DUTPASE	YP_001129459.1-76320-75484	Early lytic	IgG	0.038	0.129	1.66	1.44	1.24	1.01	1.53	0.044	0.154
BALF4	YP_001129508.1-160348-157775re	Glycoprotein	IgG	0.038	0.129	0.86	0.73	1.14	1	1.29	0.055	0.165

BFRF1	YP_001129446.1-46719-47729	Late lytic	IgG	0.038	0.129	1.72	1.47	1.29	1.01	1.66	0.046	0.154
BZLF2	CAA24860.1-102116-101445	Glycoprotein	IgG	0.039	0.129	2.26	2.00	1.36	1.04	1.77	0.026	0.124
VCA-p18	VCAp18 0.1	Late lytic	IgG	0.039	0.129	2.04	1.76	1.37	1.03	1.84	0.036	0.154
BBLF1	YP_001129480.1-109516-109289	Early lytic	IgG	0.040	0.129	2.20	1.90	1.31	0.98	1.75	0.071	0.183
BNLF2B	CAA24811.1-167303-166998	Late lytic	IgG	0.040	0.129	0.86	0.74	1.13	1	1.27	0.045	0.154
BLRF2 (VCA_p23)	YP_001129461.1-76771-77259	Early lytic	IgG	0.040	0.129	2.32	2.08	1.25	0.97	1.6	0.084	0.193
BGLF2	YP_001129486.1-115415-114405	Early lytic	IgG	0.041	0.129	1.43	1.29	1.14	0.99	1.31	0.081	0.192
UNCHARACTERIZED	CAA24880.1-59808-61583	Other/Unknown	IgG	0.041	0.129	0.82	0.69	1.16	1.03	1.31	0.019	0.107
LMP1	YP_001129515.1-169948-169188	Latent	IgG	0.045	0.138	1.13	0.98	1.2	1.04	1.39	0.015	0.099
BBLF2	CAA24824.1-119080-117515	Early lytic	IgG	0.046	0.138	1.40	1.23	1.18	0.98	1.41	0.078	0.189
EBNA2	YP_001129441.1-36201-37565 re	Latent	IgG	0.046	0.138	1.65	1.39	1.33	1.02	1.73	0.040	0.154
BZLF2	YP_001129466.1-90630-89959	Glycoprotein	IgG	0.048	0.138	2.29	2.04	1.33	1.03	1.73	0.033	0.148
BGLF1	CAA24832.1-128374-126851	Other/Unknown	IgG	0.048	0.138	1.44	1.26	1.19	1	1.41	0.057	0.165
BBRF3	YP_001129479.1-107679-108896	Glycoprotein	IgG	0.049	0.139	1.66	1.45	1.27	1.02	1.58	0.034	0.150
A73	AFY97981.1-159642-159726	Other/Unknown	IgG	0.050	0.139	1.06	0.92	1.15	1	1.32	0.054	0.165
BRRF1	YP_001129469.1-93724-94656	Immediate early lytic	IgG	0.051	0.139	1.40	1.24	1.18	1.01	1.38	0.043	0.154
BALF5 (DNA polymerase)	CAA24805.1-156746-153699-1	Early lytic	IgG	0.051	0.139	0.95	0.80	1.17	1	1.36	0.046	0.154
BPFL1	CAA24839.1-71527-62078-4_US	Late lytic	IgG	0.052	0.139	0.73	0.64	1.09	0.99	1.2	0.092	0.205
BFRF3 (VCA_p18)	AFY97924.1-49199-49729	Late lytic	IgG	0.052	0.139	2.29	2.05	1.28	0.99	1.65	0.059	0.166
BORF2	YP_001129452.1-64253-66733	Early lytic	IgG	0.053	0.139	1.03	0.88	1.11	0.95	1.3	0.176	0.306
CAPSID	YP_001129492.1-121844-120939	Late lytic	IgG	0.054	0.139	1.47	1.29	1.21	1	1.47	0.054	0.165
LMP1	YP_401722.1-168670-168584-m1	Latent	IgG	0.056	0.142	1.03	0.91	1.17	1.03	1.33	0.020	0.107
BFRF3 (VCA_p18)	YP_001129448.1-49335-49865	Late lytic	IgG	0.057	0.142	2.39	2.15	1.28	0.99	1.67	0.066	0.181
BFRF3 (VCA_p18)	CAA24838.1-61507-62037	Late lytic	IgG	0.057	0.142	2.27	2.04	1.27	0.98	1.65	0.071	0.183

BALF5 (DNA polymerase)	AFY97980.1-156149-153102-2	Early lytic	IgG	0.058	0.143	1.29	1.16	1.15	1	1.32	0.059	0.166
BALF5 (DNA polymerase)	YP_001129507.1-157772-154725-1	Early lytic	IgG	0.063	0.151	0.98	0.85	1.12	0.98	1.27	0.103	0.216
LMP1	AFY97987.1-168367-167562	Latent	IgG	0.063	0.151	1.07	0.93	1.19	1.04	1.37	0.016	0.099
BVLF1	YP_001129500.1-136454-135636	Early lytic	IgG	0.064	0.151	1.38	1.26	1.12	0.98	1.28	0.088	0.198
BZIP/BZLF1 (Zta)	CAA24861.1-102530-102423	Immediate early lytic	IgG	0.064	0.151	0.92	0.82	1.11	0.99	1.25	0.083	0.193
BcLF1 (VCA_p160)	AFY97965.1-125044-120899-1	Late lytic	IgG	0.068	0.159	1.21	1.00	1.22	0.98	1.5	0.073	0.184
EBNA2	CAA24806.1-159322-156749_US	Latent	IgG	0.070	0.160	0.74	0.63	1.11	0.97	1.27	0.133	0.258
BZLF1 (Zta)	YP_001129467.1-90855-90724	Immediate early lytic	IgG	0.071	0.160	1.63	1.42	1.23	0.99	1.53	0.066	0.181
EBNA3C	CAA24859.1-98805-101423 re	Latent	IgG	0.073	0.160	0.95	0.82	1.15	1	1.34	0.056	0.165
BcLF1 (VCA_p160)	YP_001129493.1-126005-121860-2	Late lytic	IgG	0.073	0.160	1.17	1.04	1.12	0.95	1.31	0.180	0.311
BOLF1	CAA24841.1-75239-71520-1_US	Late lytic	IgG	0.073	0.160	0.93	0.79	1.19	1.01	1.39	0.037	0.154
EBNA3A	YP_001129463.1-80026-80361	Latent	IgG	0.074	0.160	1.48	1.25	1.27	0.98	1.65	0.075	0.187
BLLF1 (gp350/220)	YP_001129460.1-76393-76701	Glycoprotein	IgG	0.074	0.160	1.05	0.94	1.06	0.94	1.2	0.328	0.444
BMRF1 (EA(D)_p47/54)	YP_001129454.1-67745-68959	Early lytic	IgG	0.077	0.162	1.93	1.71	1.24	0.95	1.61	0.111	0.222
BBLF2/3	AFY97868.1-104653-104048_US	Early lytic	IgG	0.077	0.162	1.05	0.94	1.07	0.95	1.21	0.272	0.387
BZLF1 (Zta)	YP_001129467.1-91045-90941	Immediate early lytic	IgG	0.079	0.165	1.08	0.99	1.07	0.97	1.18	0.159	0.297
BcLF1 (VCA_p160)	YP_001129493.1-126005-121860-1	Late lytic	IgG	0.082	0.167	1.42	1.24	1.18	0.97	1.42	0.102	0.216
LMP1	YP_001129515.1-170457-170190	Latent	IgG	0.083	0.167	1.04	0.95	1.07	0.96	1.19	0.252	0.367

BPFL1	CAA24839.1-71527-62078-3_US	Late lytic	IgG	0.083	0.167	0.90	0.80	1.1	0.98	1.22	0.109	0.222
BDLF3	AFY97990.1-85953-86312	Glycoprotein	IgG	0.086	0.171	0.98	0.85	1.18	1.01	1.39	0.045	0.154
BXRF1	YP_001129498.1-133398-134144	Early lytic	IgG	0.088	0.174	1.31	1.19	1.08	0.95	1.24	0.238	0.361
BOLF1	AFY97840.1-62772-59044-2_US	Late lytic	IgG	0.091	0.175	1.09	0.98	1.03	0.91	1.16	0.639	0.690
EBNA3B	AFY97829.1-82733-83089	Latent	IgG	0.091	0.175	1.23	1.12	1.06	0.92	1.21	0.441	0.534
EBNA1	AFY97913.1-95532-97457_US	Latent	IgG	0.091	0.175	1.12	0.97	1.19	0.99	1.42	0.069	0.183
BARF1	YP_001129512.1-166530-167195	Other/Unknown	IgG	0.094	0.179	1.29	1.18	1.13	0.99	1.3	0.076	0.187
BKRF4	AFY97946.1-98716-99369	Late lytic	IgG	0.096	0.180	0.98	0.83	1.13	0.95	1.35	0.175	0.306
EBNA1	AFY97842.1-95349-97142_US	Latent	IgG	0.097	0.180	0.85	0.76	1.07	0.96	1.2	0.241	0.361
BALF2 (EA(D)_p138)	YP_001129510.1-165796-162410-2	Early lytic	IgG	0.097	0.180	1.43	1.28	1.11	0.93	1.33	0.239	0.361
BARF1	YP_001129453.1-66746-67654	Other/Unknown	IgG	0.100	0.183	1.47	1.28	1.2	0.96	1.5	0.108	0.222

Table S3 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgA antibody responses comparisons on mean difference between EBV-positive cHL cases and controls. Table is ordered by t-test p-value (lowest to highest) for the top 10 antibodies. The antibodies that were previously identified as significantly different from EBV-positive cHL cases compared to controls in our European study are highlighted in bold. ORs were adjusted for age, sex, and study area (Hong Kong and Taiwan). T-test Adj p FDR; FDR p-value obtained for t-tests, Adj p FDR; is the FDR p-value obtained for the linear regression model.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test adj *p* FDR	EBV+ cHL mean	Control mean	OR	L95	U95	Lin. reg *p*-value	Lin. reg adj *p* FDR
BBLF1	YP_001129480.1-109516-109289	Early lytic	IgA	1.0E-03	1.1E-01	1.32	1.11	1.23	1.09	1.39	1.5E-03	2.5E-01
BBLF1	AFY97956.1-108555-108328	Early lytic	IgA	1.1E-03	1.1E-01	1.30	1.10	1.22	1.08	1.38	2.5E-03	2.5E-01
BRRF2	AFY97943.1-93884-95497	Late lytic	IgA	9.0E-03	6.0E-01	1.45	1.27	1.18	1.02	1.36	3.0E-02	8.6E-01
VCA-p18	VCAp18 0.1	Late lytic	IgA	1.5E-02	7.0E-01	1.15	0.92	1.25	1.03	1.52	2.5E-02	8.6E-01
THY.KINASE	YP_001129497.1-133399-131576	Early lytic	IgA	1.7E-02	7.0E-01	0.95	0.85	1.06	0.98	1.15	1.3E-01	9.9E-01
EAD	EA(D) 0.1	Early lytic	IgA	3.0E-02	7.9E-01	0.62	0.56	1.07	1.01	1.12	2.4E-02	8.6E-01
EBNA3A	YP_001129463.1-80447-82888	Latent	IgA	3.1E-02	7.9E-01	1.34	1.20	1.17	1.02	1.33	2.8E-02	8.6E-01
BRRF2	YP_001129470.1-94844-96457	Late lytic	IgA	3.6E-02	7.9E-01	1.56	1.42	1.13	0.99	1.3	8.1E-02	9.9E-01
BPFL1	YP_001129449.1-59370-49906-3	Late lytic	IgA	3.7E-02	7.9E-01	0.87	0.79	1.04	0.96	1.12	3.4E-01	9.9E-01
EBNA2	YP_001129441.1-36201-37565 re	Latent	IgA	4.1E-02	7.9E-01	1.01	0.95	1.07	1.01	1.14	2.9E-02	8.6E-01

Table S4 Comparison of current antibody signature from East Asia in the European study population between EBV-positive cHL and Controls. Table is ordered by t-test p-value (lowest to highest). The antibodies that were previously identified as significantly different from EBV-positive cHL cases compared to controls in our European study are highlighted in bold. ORs were adjusted for age, sex and study area (Hong Kong and Taiwan). The area under the Receiver operative Curve (AUC) was calculated for the 11 antibodies ($p < 0.05$, t-test) in European study to test their ability along with the subject's demographics (i.e., age, sex, and study area) to classify EBV-positive cHL from controls.

Protein name	**Array sequence**	Asian study					European study						
		t-test *p*-value	**OR**	**L95**	**U95**	**Lin. reg *p*-value**	**t-test *p*-value**	**AUC**	**95% CI**	**OR**	**L95**	**U95**	**Lin. reg *p*-value**
THY.KINASE	YP_001129497.1-133399-131576	3.72×10-4	1.56	1.21	2.01	0.001	2.23×10-4	0.608	0.542-0.674	1.19	1.08	1.3	3.49×10-4
EBNA-LP	YP_001129440.1-20824-20955	4.36×10-4	1.35	1.14	1.61	0.001	0.205	-	-	1.03	0.98	1.08	0.207
LMP1	YP_401722.1-168507-167702	0.001	1.43	1.21	1.69	5.93×10-5	0.002	0.603	0.537-0.670	1.08	1.03	1.14	0.003
BBLF1	AFY97956.1-108555-108328	0.002	1.46	1.12	1.9	0.006	0.335	-	-	1.01	0.98	1.05	0.381
BALF2 (EA(D) p138)	YP_001129510.1-165796-162410-1	0.002	1.49	1.17	1.89	0.001	0.001	0.605	0.539-0.671	1.15	1.06	1.26	0.002
LMP2A	YP_001129436.1-167587-167942	0.002	1.23	1.07	1.42	0.006	0.001	0.602	0.535-0.668	1.11	1.04	1.19	0.001
LF2	YP_001129504.1-151808-150519	0.002	1.32	1.11	1.57	0.002	0.002	0.601	0.535-0.667	1.08	1.03	1.14	0.003

EBNA3A	YP_001129463.1-80447-82888	0.002	1.74	1.28	2.38	0.001	1.89×10-4	0.571	0.504-0.639	1.16	1.07	1.25	0.000
BPFL1	YP_001129449.1-59370-49906-2	0.002	1.39	1.13	1.7	0.002	0.001	0.547	0.479-0.615	1.07	1.02	1.11	0.002
BDLF2	YP_001129491.1-120928-119666	0.002	1.43	1.14	1.8	0.003	0.044	0.569	0.501-0.636	1.13	1	1.27	0.055
BPFL1	YP_001129449.1-59370-49906-3	0.002	1.32	1.1	1.58	0.004	0.246	-	-	1.02	0.99	1.05	0.301
BLLF1 (gp350/220)	YP_001129462.1-79936-7727. r	0.002	1.45	1.17	1.81	0.001	1.17×10-4	0.629	0.564-0.694	1.14	1.06	1.21	1.60×10-4
LMP2B	AFY97910.1-1026-1196	0.003	1.3	1.09	1.54	0.004	1.46×10-4	0.669	0.606-0.733	1.14	1.07	1.22	1.63×10-4
EBNA3A	YP_401669.1-80382-82877	0.003	1.67	1.24	2.27	0.001	0.441	-	-	0.96	0.87	1.06	0.387
EBNA3A	AFY97915.1-80252-82747	0.004	1.63	1.2	2.2	0.002	0.005	0.526	0.459-0.594	1.05	1.01	1.09	0.009
BDLF3	AFY97964.1-118644-117940	0.008	1.57	1.18	2.09	0.003	0.058	-	-	1.03	1	1.06	0.071

Table S5 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgG antibody responses comparisons on mean difference between EBV-negative cHL cases and controls. Table is ordered by t-test p-value (lowest to highest) for the first 10 antibodies. T-test Adj p FDR; FDR p-value obtained for t-tests.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test adj *p* FDR	EBV-cHL mean	Control mean
BARF1	YP_001129512.1-166530-167195	Other/Unknown	IgG	0.374	0.937	1.137	1.180
BARF1	YP_001129453.1-66746-67654	Other/Unknown	IgG	0.112	0.937	1.186	1.284
BBRF3	YP_001129479.1-107679-108896	Glycoprotein	IgG	0.649	0.937	1.488	1.455
BcLF1 (VCA_p160)	YP_001129493.1-126005-121860-2	Late lytic	IgG	0.801	0.951	1.054	1.040
FGAM-synthase	YP_001129438.1-1736-5692-1	Other/Unknown	IgG	0.958	0.972	1.069	1.071
BFRF1A	YP_001129445.1-46352-46759	Other/Unknown	IgG	0.631	0.937	0.918	0.900
BSRF1	YP_001129458.1-74770-75426	Other/Unknown	IgG	0.590	0.937	1.187	1.217
CAPSID	YP_001129492.1-121844-120939	Late lytic	IgG	0.199	0.937	1.213	1.292
BMRF1 (EA(D)_p47/54)	YP_001129454.1-67745-68959	Early lytic	IgG	0.453	0.937	1.642	1.708
BVRF1	YP_001129499.1-133954-135666	Late lytic	IgG	0.201	0.937	1.219	1.295

Table S6 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgA antibody responses comparisons on mean difference between EBV-negative cHL cases and controls. Table is ordered by t-test p-value (lowest to highest) for the first 10 antibodies. T-test Adj p FDR; FDR p-value obtained for t-tests.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test adj *p* FDR	EBV-cHL mean	Control mean
BGLF5 (ALK.EXONUCLEASE)	YP_001129480.1-109516-109289	Early lytic	IgA	0.020	0.975	1.225	1.111
BGLF5 (ALK.EXONUCLEASE)	AFY97956.1-108555-108328	Early lytic	IgA	0.039	0.975	1.195	1.097
BKRF4	YP_001129474.1-99676-100329	Late lytic	IgA	0.090	0.975	0.928	0.951
BBLF4	CAA24821.1-114259-111830_US	Early lytic	IgA	0.108	0.975	1.116	1.054
BZLF1 (Zta)	CAA24861.1-102338-102210	Immediate early lytic	IgA	0.142	0.975	1.087	1.032
BDLF2	YP_001129491.1-120928-119666	Glycoprotein	IgA	0.145	0.975	1.221	1.159
EBNA-LP	YP_401636.1-35590-35694	Latent	IgA	0.149	0.975	1.068	1.015
BRRF2	AFY97943.1-93884-95497	Late lytic	IgA	0.163	0.975	1.347	1.268
BRRF2	YP_001129470.1-94844-96457	Late lytic	IgA	0.171	0.975	1.490	1.418
BALF2 (EA(D)_p138)	YP_001129510.1-165796-162410-2	Early lytic	IgA	0.177	0.975	0.839	0.800

Table S7 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgG antibody responses comparisons on mean difference between EBV-positive and EBV-negative cHL. Table is ordered by t-test p-value (lowest to highest) for the significant antibodies. The six IgG antibodies that were previously identified as significantly different from EBV-positive cHL cases compared to EBV-negative cHL in our European study are highlighted in bold. 76 IgG antibodies were significantly elevated among EBV-positive cHL cases compared to EBV-negative cHL (FDR p <0.05, t-test). ORs were adjusted for age, sex and study area (Hong Kong and Taiwan). T-test Adj p FDR; FDR p-value obtained for t-tests, Adj p FDR; is the FDR p-value obtained for the linear regression model.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test adj *p* FDR	EBV+ cHL mean	EBV- cHL mean	OR	L95	U95	Lin. reg *p*-value	Lin. reg adj *p* FDR
THY.KINASE	YP_001129497.1-133399-131576	Early lytic	IgG	5.65E-05	9.73E-03	1.91	1.38	1.44	1.12	1.85	4.95E-03	1.26E-01
BALF2 (EA(D)_p138)	YP_001129510.1-165796-162410-1	Early lytic	IgG	2.26E-04	9.73E-03	1.48	1.04	1.63	1.27	2.1	2.26E-04	4.57E-02
BPFL1	YP_001129449.1-59370-49906-2	Late lytic	IgG	2.94E-04	9.73E-03	1.72	1.34	1.33	1.08	1.64	8.60E-03	1.58E-01
BPFL1	CAA24839.1-71527-62078-2	Late lytic	IgG	2.96E-04	9.73E-03	2.32	1.87	1.52	1.15	2	3.74E-03	1.26E-01
EBNA3B	YP_001129464.1-83509-86532-2	Latent	IgG	3.40E-04	9.73E-03	1.73	1.32	1.4	1.1	1.78	7.87E-03	1.58E-01
BFRF1	YP_001129446.1-46719-47729	Late lytic	IgG	4.05E-04	9.73E-03	1.72	1.32	1.41	1.13	1.76	3.02E-03	1.26E-01
BVRF2	YP_001129501.1-136465-138282	Late lytic	IgG	5.79E-04	9.73E-03	2.07	1.64	1.48	1.13	1.95	5.62E-03	1.26E-01

BDLF2	YP_001129491.1-120928-119666	Glycoprotein	IgG	6.01E-04	9.73E-03	1.89	1.48	1.33	1.04	1.72	2.71E-02	1.78E-01
BVRF2	YP_001129501.1-136465-138282 re	Late lytic	IgG	6.18E-04	9.73E-03	0.96	0.72	1.2	1.03	1.4	2.31E-02	1.78E-01
BDLF4	YP_001129488.1-117560-116883	Early lytic	IgG	6.39E-04	9.73E-03	1.74	1.41	1.25	1.01	1.56	4.57E-02	2.25E-01
BPFL1	YP_001129449.1-59370-49906-3	Late lytic	IgG	6.81E-04	9.73E-03	1.62	1.29	1.28	1.05	1.55	1.71E-02	1.78E-01
LF2	YP_001129504.1-151808-150519	Early lytic	IgG	7.09E-04	9.73E-03	1.48	1.17	1.31	1.1	1.56	3.39E-03	1.26E-01
BBRF1	YP_001129476.1-102746-104587	Late lytic	IgG	7.15E-04	9.73E-03	1.69	1.23	1.48	1.15	1.91	3.28E-03	1.26E-01
LMP2A	YP_001129436.1-167587-167942	Latent	IgG	7.20E-04	9.73E-03	1.26	1.03	1.19	1.02	1.38	2.59E-02	1.78E-01
EBNA3A	YP_401669.1-80382-82877	Latent	IgG	7.22E-04	9.73E-03	2.09	1.58	1.49	1.06	2.1	2.36E-02	1.78E-01
EBNA3A	YP_001129463.1-80447-82888	Latent	IgG	8.53E-04	1.08E-02	2.15	1.65	1.48	1.05	2.1	2.91E-02	1.84E-01
EBNA-LP	YP_001129440.1-20824-20955	Latent	IgG	9.18E-04	1.09E-02	1.45	1.17	1.24	1.03	1.48	2.55E-02	1.78E-01
EBNA3A	AFY97915.1-80252-82747	Latent	IgG	1.01E-03	1.13E-02	2.00	1.51	1.46	1.05	2.04	2.55E-02	1.78E-01
BdRF1 (VCA_p40)	AFY97974.1-136284-137321_US	Late lytic	IgG	1.22E-03	1.24E-02	1.51	1.08	1.52	1.16	1.99	2.74E-03	1.26E-01
BFLF2	YP_001129443.1-44763-43807	Late lytic	IgG	1.26E-03	1.24E-02	1.35	1.08	1.23	1.04	1.46	1.69E-02	1.78E-01

LMP2B	AFY97910.1-1026-1196	Latent	IgG	1.29E-03	1.24E-02	1.39	1.11	1.21	1.01	1.44	3.88E-02	2.13E-01
BLRF2 (VCA_p23)	YP_001129461.1-76771-77259	Early lytic	IgG	1.39E-03	1.28E-02	2.32	1.95	1.27	0.95	1.7	1.07E-01	3.27E-01
BILF1	YP_001129506.1-154125-153187	Early lytic	IgG	1.50E-03	1.32E-02	1.33	1.12	1.18	1.01	1.37	4.16E-02	2.13E-01
EBNA3B	YP_001129464.1-83074-83430	Latent	IgG	1.56E-03	1.32E-02	1.62	1.34	1.13	0.9	1.41	2.86E-01	4.54E-01
LMP2A	YP_001129436.1-871-951	Latent	IgG	1.91E-03	1.54E-02	1.40	1.18	1.17	1.01	1.36	4.12E-02	2.13E-01
LMP1	YP_401722.1-168507-167702	Latent	IgG	1.99E-03	1.55E-02	1.25	0.98	1.34	1.09	1.64	5.34E-03	1.26E-01
BBLF4	YP_001129475.1-102801-100372 re	Early lytic	IgG	2.17E-03	1.60E-02	1.14	0.91	1.18	1	1.39	5.06E-02	2.39E-01
LMP1	YP_401722.1-168670-168584-m2	Latent	IgG	2.22E-03	1.60E-02	1.37	1.12	1.19	0.97	1.45	9.22E-02	3.03E-01
BBLF2	CAA24824.1-119080-117515	Early lytic	IgG	2.36E-03	1.65E-02	1.40	1.14	1.14	0.96	1.36	1.30E-01	3.31E-01
EBNA3B	CAA24858.1-95353-95709	Latent	IgG	2.80E-03	1.88E-02	1.51	1.20	1.12	0.86	1.46	3.89E-01	5.19E-01
BGLF5 (DNAse)	AFY97955.1-109922-108510	Early lytic	IgG	3.89E-03	2.47E-02	1.21	1.03	1.14	0.98	1.33	8.67E-02	3.01E-01
BFRF1A	AFY97921.1-46216-46623	Other/Unknown	IgG	3.98E-03	2.47E-02	0.70	0.56	1.17	1.04	1.32	1.08E-02	1.78E-01
CAPSID	YP_001129451.1-63084-64178	Late lytic	IgG	4.06E-03	2.47E-02	1.37	1.11	1.26	1.05	1.51	1.50E-02	1.78E-01

CAPSID	YP_001129492.1-121844-120939	Late lytic	IgG	4.15E-03	2.47E-02	1.47	1.21	1.21	1	1.47	5.23E-02	2.39E-01
BRRF2	AFY97943.1-93884-95497	Late lytic	IgG	4.36E-03	2.52E-02	2.31	1.98	1.28	0.94	1.73	1.17E-01	3.29E-01
BGLF2	YP_001129486.1-115415-114405	Early lytic	IgG	4.74E-03	2.56E-02	1.43	1.23	1.18	1.01	1.39	4.22E-02	2.13E-01
BPFL1	YP_001129449.1-59370-49906-4 re	Late lytic	IgG	4.86E-03	2.56E-02	1.19	1.00	1.14	0.97	1.33	1.22E-01	3.29E-01
EBNA3B	CAA24858.1-95788-98247	Latent	IgG	4.95E-03	2.56E-02	1.93	1.58	1.17	0.88	1.56	2.87E-01	4.54E-01
EBNA3B	YP_001129464.1-83509-86532-1_US	Latent	IgG	5.02E-03	2.56E-02	1.81	1.46	1.25	0.96	1.64	1.06E-01	3.27E-01
BFRF2	YP_001129447.1-47636-49411 re	Late lytic	IgG	5.07E-03	2.56E-02	0.77	0.63	1.14	1.01	1.28	3.31E-02	2.00E-01
BBLF1	AFY97956.1-108555-108328	Early lytic	IgG	5.28E-03	2.60E-02	2.15	1.82	1.23	0.93	1.62	1.55E-01	3.52E-01
EAD	EA(D) 0.1	Early lytic	IgG	5.40E-03	2.60E-02	0.23	0.17	1.05	1	1.1	6.77E-02	2.68E-01
BXRF1	YP_001129498.1-133398-134144	Early lytic	IgG	5.57E-03	2.62E-02	1.31	1.12	1.15	0.99	1.34	7.21E-02	2.75E-01
BLLF1 (gp350/220)	YP_001129462.1-79936-77276 redesigned	Glycoprotein	IgG	5.94E-03	2.73E-02	1.55	1.25	1.34	1.04	1.71	2.44E-02	1.78E-01
EBNA2	YP_001129441.1-36201-37565 re	Latent	IgG	6.15E-03	2.75E-02	1.65	1.32	1.3	0.99	1.7	6.05E-02	2.60E-01
BZLF1 (Zta)	CAA24861.1-103155-102655	Immediate early lytic	IgG	6.35E-03	2.75E-02	1.29	1.03	1.26	1.04	1.53	2.06E-02	1.78E-01

BILF2	YP_001129503.1-139063-138317	Glycoprotein	IgG	6.55E-03	2.75E-02	1.75	1.47	1.28	1.03	1.59	2.54E-02	1.78E-01
BZLF2	CAA24860.1-102116-101445	Glycoprotein	IgG	6.61E-03	2.75E-02	2.26	1.93	1.27	0.95	1.69	1.09E-01	3.28E-01
BZLF1 (Zta)	YP_001129467.1-91697-91197	Immediate early lytic	IgG	6.67E-03	2.75E-02	1.53	1.25	1.29	1.03	1.62	2.73E-02	1.78E-01
BSRF1	YP_001129458.1-74770-75426	Other/Unknown	IgG	8.69E-03	3.38E-02	1.52	1.19	1.34	1.06	1.69	1.56E-02	1.78E-01
BHRF1	YP_001129442.1-42204-42779	Early lytic	IgG	8.76E-03	3.38E-02	1.59	1.36	1.2	0.99	1.47	6.69E-02	2.68E-01
BZLF2	YP_001129466.1-90630-89959	Glycoprotein	IgG	8.80E-03	3.38E-02	2.29	1.97	1.25	0.93	1.67	1.35E-01	3.31E-01
BGLF5 (DNAse)	YP_001129481.1-110883-109471	Early lytic	IgG	8.87E-03	3.38E-02	1.19	1.02	1.15	1	1.33	5.33E-02	2.39E-01
BRRF2	YP_001129470.1-94844-96457	Late lytic	IgG	9.21E-03	3.41E-02	2.38	2.07	1.26	0.93	1.69	1.35E-01	3.31E-01
BBLF2/3	AFY97868.1-104653-104048_US	Early lytic	IgG	9.29E-03	3.41E-02	1.05	0.89	1.14	0.99	1.32	6.92E-02	2.69E-01
EBNA1	AFY97913.1-95532-97457_US	Latent	IgG	1.02E-02	3.66E-02	1.12	0.89	1.18	0.97	1.44	1.02E-01	3.23E-01
BALF4	YP_001129508.1-160348-157775 re	Glycoprotein	IgG	1.03E-02	3.66E-02	0.86	0.70	1.14	1	1.3	5.44E-02	2.39E-01
BVLF1	YP_001129500.1-136454-135636	Early lytic	IgG	1.07E-02	3.71E-02	1.38	1.21	1.12	0.97	1.31	1.33E-01	3.31E-01
FGAM-synthase	YP_001129438.1-1736-5692-2	Other/Unknown	IgG	1.10E-02	3.77E-02	1.49	1.26	1.18	0.98	1.41	7.71E-02	2.83E-01

BRLF1 (Rta)	YP_001129468.1-93725-91908	Immediate early lytic	IgG	1.15E-02	3.77E-02	1.48	1.20	1.25	1	1.56	5.16E-02	2.39E-01
BARF1	YP_001129453.1-66746-67654	Other/Unknown	IgG	1.17E-02	3.77E-02	1.47	1.19	1.27	1.02	1.58	3.58E-02	2.07E-01
BVRF1	YP_001129499.1-133954-135666	Late lytic	IgG	1.18E-02	3.77E-02	1.42	1.22	1.14	0.95	1.37	1.49E-01	3.48E-01
BOLF1	CAA24841.1-75239-71520-1_US	Late lytic	IgG	1.18E-02	3.77E-02	0.93	0.74	1.21	1.04	1.41	1.41E-02	1.78E-01
BALF3	AFY97901.1-160450-158393_US	Early lytic	IgG	1.35E-02	4.22E-02	1.06	0.91	1.14	0.99	1.3	6.31E-02	2.60E-01
A73	AFY97981.1-159642-159726	Other/Unknown	IgG	1.41E-02	4.22E-02	1.06	0.90	1.11	0.97	1.27	1.29E-01	3.31E-01
EBNA3C	CAA24859.1-98805-101423 re	Latent	IgG	1.41E-02	4.22E-02	0.95	0.78	1.1	0.94	1.29	2.20E-01	4.05E-01
VCA-p18	VCAp18 0.1	Late lytic	IgG	1.43E-02	4.22E-02	2.04	1.73	1.26	0.94	1.69	1.21E-01	3.29E-01
LF2	AFY97966.1-125043-127295	Early lytic	IgG	1.43E-02	4.22E-02	1.14	0.97	1.12	0.95	1.32	1.83E-01	3.70E-01
BRRF1	YP_001129469.1-93724-94656	Immediate early lytic	IgG	1.44E-02	4.22E-02	1.40	1.21	1.16	0.98	1.37	8.19E-02	2.95E-01
BKRF4	AFY97946.1-98716-99369	Late lytic	IgG	1.49E-02	4.30E-02	0.98	0.77	1.17	0.99	1.38	6.27E-02	2.60E-01
EBNA3A	YP_001129463.1-80026-80361	Latent	IgG	1.52E-02	4.32E-02	1.48	1.18	1.37	1.06	1.76	1.70E-02	1.78E-01
BMRF1 (EA(D)_p47/54)	YP_001129454.1-67745-68959	Early lytic	IgG	1.57E-02	4.41E-02	1.93	1.64	1.16	0.89	1.52	2.66E-01	4.44E-01

BTRF1	CAA24796.1-139642-140916	Other/Unknown	IgG	1.60E-02	4.42E-02	1.42	1.23	1.09	0.92	1.31	3.27E-01	4.86E-01
EBNA2	CAA24806.1-159322-156749_US	Latent	IgG	1.62E-02	4.42E-02	0.74	0.59	1.09	0.95	1.25	2.29E-01	4.05E-01

Table S8 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgA antibody responses comparisons on mean difference between EBV-positive and EBV-negative cHL. Table is ordered by t-test p-value (lowest to highest) for the first 10 antibodes. ORs were adjusted for age, sex, and study area (Hong Kong and Taiwan). T-test Adj p FDR; FDR p-value obtained for t-tests, Adj p FDR; is the FDR p-value obtained for the linear regression model.

Protein name	Array sequence	EBV life cycle	IgG/ IgA	t-test *p*	t-test Adj *p* FDR	EBV+ cHL mean	EBV- cHL mean	OR	L95	U95	Lin. reg *p*-value	Lin. reg adj *p* FDR
VCA-p18	VCAp18 0.1	Late lytic	IgA	9.27E-04	1.87E-01	1.15	0.84	1.23	1.01	1.49	4.19E-02	9.95E-01
EBNA3A	YP_001129463.1-80447-82888	Latent	IgA	1.82E-02	9.43E-01	1.34	1.19	1.09	0.94	1.25	2.48E-01	9.95E-01
EAD	EA(D) 0.1	Early lytic	IgA	2.41E-02	9.43E-01	0.62	0.56	1.03	0.97	1.1	3.54E-01	9.95E-01
BMRF1 (EA(D)_p47/54)	YP_001129454.1-67745-68959	Early lytic	IgA	2.56E-02	9.43E-01	1.42	1.29	1.04	0.91	1.19	5.80E-01	9.95E-01
EBNA2	YP_001129441.1-36201-37565 re	Latent	IgA	2.77E-02	9.43E-01	1.01	0.94	1.04	0.97	1.12	2.72E-01	9.95E-01
BLRF2 (VCA_p23)	YP_001129461.1-76771-77259	Early lytic	IgA	3.38E-02	9.43E-01	1.44	1.31	1.06	0.92	1.23	4.34E-01	9.95E-01
BKRF4	YP_001129474.1-99676-100329	Late lytic	IgA	3.60E-02	9.43E-01	0.96	0.93	1.02	0.98	1.06	2.98E-01	9.95E-01
LMP2A	YP_001129436.1-1026-1196	Latent	IgA	3.74E-02	9.43E-01	1.10	1.05	1.05	0.99	1.1	8.35E-02	9.95E-01
BPFL1	CAA24839.1-71527-62078-2	Late lytic	IgA	5.32E-02	9.77E-01	1.47	1.34	1.11	0.93	1.32	2.45E-01	9.95E-01
BOLF1	CAA24841.1-75239-71520-1_US	Late lytic	IgA	6.34E-02	9.77E-01	0.94	0.88	1.06	0.97	1.16	2.01E-01	9.95E-01

Table S9 Comparison of current antibody signature from East Asia in the European study population between EBV-positive cHL and EBV-negative cHL. Table is ordered by t-test p-value (lowest to highest). The antibodies that were previously identified as significantly different from EBV-positive cHL cases compared to controls in our European study are highlighted in bold. ORs were adjusted for age, sex, and study area (Hong Kong and Taiwan). Antibodies highlighted in bold are having p-value <0.05 by t-tests and linear regression models in the European study.

	Asian Study					European Study				
Protein name	**t-test *p***	**OR**	**L95**	**U95**	**Lin. reg *p*-value**	**t-test *p***	**OR**	**L95**	**U95**	**Lin. reg *p*-value**
THY.KINASE	5.65E-05	1.44	1.12	1.85	0.005	**1.25E-05**	1.27	1.13	1.43	**1.11E-04**
BRLF1 (Rta)	0.012	1.25	1	1.56	0.052	**5.59E-05**	1.33	1.17	1.52	**3.88E-05**
LMP2A	0.001	1.19	1.02	1.38	0.026	**1.01E-04**	1.14	1.05	1.24	**0.002**
BZLF2	0.007	1.27	0.95	1.69	0.109	**2.82E-04**	1.2	1.08	1.33	**0.001**
BRRF2	0.009	1.26	0.93	1.69	0.135	**3.75E-04**	1.28	1.14	1.44	**6.24E-05**
EBNA3A	0.001	1.48	1.05	2.1	0.029	**0.001**	1.19	1.09	1.31	**2.03E-04**
EBNA3C	0.014	1.1	0.94	1.29	0.220	**0.001**	1.15	1.05	1.27	**0.004**
LMP1	0.002	1.19	0.97	1.45	0.092	**0.001**	1.08	1.03	1.12	**0.002**
BLLF1 (gp350/220)	0.006	1.34	1.04	1.71	0.024	**0.001**	1.15	1.05	1.25	**0.002**
BARF1	0.012	1.27	1.02	1.58	0.036	**0.002**	1.21	1.07	1.36	**0.002**
BVRF2	0.001	1.48	1.13	1.95	0.006	**0.002**	1.18	1.05	1.32	**0.005**
BFRF1	4.05E-04	1.41	1.13	1.76	0.003	**0.003**	1.06	1.02	1.11	**0.003**
BBRF1	0.001	1.48	1.15	1.91	0.003	**0.003**	1.06	1.02	1.09	**0.003**
BFRF1A	0.004	1.17	1.04	1.32	0.011	**0.003**	1.21	1.06	1.37	**0.004**
BVRF2	0.001	1.2	1.03	1.4	0.023	**0.004**	1.06	1.02	1.1	**0.007**
LMP2B	0.001	1.21	1.01	1.44	0.039	**0.004**	1.13	1.04	1.24	**0.005**

FGAM-synthase	0.011	1.18	0.98	1.41	0.077	**0.004**	1.22	1.09	1.36	**0.001**
BPFL1	2.94E-04	1.33	1.08	1.64	0.009	**0.004**	1.08	1.03	1.13	**0.003**
BALF2 (EA(D)_p138)	2.26E-04	1.63	1.27	2.1	2.26E-04	**0.005**	1.15	1.03	1.28	**0.013**
BZLF2	0.009	1.25	0.93	1.67	0.135	**0.009**	1.1	1.03	1.18	**0.008**
LF2	0.001	1.31	1.1	1.56	0.003	**0.009**	1.08	1.01	1.15	**0.022**
CAPSID	0.004	1.26	1.05	1.51	0.015	**0.009**	1.05	1.01	1.08	**0.010**
EBNA2	0.016	1.09	0.95	1.25	0.229	**0.010**	1.09	1.02	1.16	**0.010**
EBNA3A	0.001	1.46	1.05	2.04	0.025	**0.011**	1.06	1.01	1.11	**0.030**
BVLF1	0.011	1.12	0.97	1.31	0.133	**0.011**	1.11	1.02	1.21	**0.016**
VCA-p18	0.014	1.26	0.94	1.69	0.121	**0.012**	1.09	1.02	1.18	**0.017**
EBNA3B	0.005	1.25	0.96	1.64	0.106	**0.013**	1.06	1.02	1.11	**0.009**
BFLF2	0.001	1.23	1.04	1.46	0.017	**0.014**	1.06	1.01	1.1	**0.013**
BGLF5 (DNAse)	0.009	1.15	1	1.33	0.053	**0.017**	1.21	1.05	1.39	**0.009**
BRRF2	0.004	1.28	0.94	1.73	0.117	**0.018**	1.03	1	1.07	**0.030**
BdRF1 (VCA_p40)	0.001	1.52	1.16	1.99	0.003	**0.020**	1.14	1.02	1.27	**0.018**
LF2	0.014	1.12	0.95	1.32	0.183	**0.021**	1.03	1	1.07	**0.035**
BILF2	0.007	1.28	1.03	1.59	0.025	**0.021**	1.04	1.01	1.08	**0.023**
BALF4	0.010	1.14	1	1.3	0.054	**0.026**	1.2	1.03	1.4	**0.018**
BSRF1	0.009	1.34	1.06	1.69	0.016	**0.028**	1.06	1.01	1.11	**0.032**
BFRF2	0.005	1.14	1.01	1.28	0.033	**0.029**	1.05	1.01	1.1	**0.028**
LMP1	0.002	1.34	1.09	1.64	0.005	**0.029**	1.07	1	1.14	0.053
BVRF1	0.012	1.14	0.95	1.37	0.149	**0.030**	1.09	1.01	1.17	**0.026**
BDLF2	0.001	1.33	1.04	1.72	0.027	**0.031**	1.19	1.02	1.39	**0.026**
BILF1	0.002	1.18	1.01	1.37	0.042	**0.032**	1.04	1	1.08	**0.047**
BXRF1	0.006	1.15	0.99	1.34	0.072	**0.032**	1.04	1	1.09	0.065
EBNA3B	0.002	1.13	0.9	1.41	0.286	**0.033**	1.06	1	1.12	**0.042**

BHRF1	0.009	1.2	0.99	1.47	0.067	**0.040**	1.07	1.01	1.14	**0.025**
EAD	0.005	1.05	1	1.1	0.068	**0.043**	1.05	1	1.11	**0.045**
BKRF4	0.015	1.17	0.99	1.38	0.063	**0.043**	1.06	1	1.11	**0.035**
BZLF1 (Zta)	0.007	1.29	1.03	1.62	0.027	**0.047**	1.13	1.01	1.27	**0.040**
BZLF1 (Zta)	0.006	1.26	1.04	1.53	0.021	**0.048**	1.04	1	1.08	**0.037**
BPFL1	0.005	1.14	0.97	1.33	0.122	0.050	1.04	1	1.08	0.062
A73	0.014	1.11	0.97	1.27	0.129	0.061	1.05	0.99	1.11	0.111
BGLF2	0.005	1.18	1.01	1.39	0.042	0.068	1.04	0.99	1.08	0.097
BOLF1	0.012	1.21	1.04	1.41	0.014	0.086	1.06	0.99	1.14	0.095
BBLF1	0.005	1.23	0.93	1.62	0.155	0.087	1.04	1	1.08	**0.037**
LMP2A	0.002	1.17	1.01	1.36	0.041	0.095	1.07	1	1.15	0.068
BBLF4	0.002	1.18	1	1.39	0.051	0.104	1.09	0.99	1.2	0.068
BBLF2/3	0.009	1.14	0.99	1.32	0.069	0.106	1.05	0.98	1.13	0.171
EBNA-LP	0.017	1.16	1.02	1.32	0.027	0.124	1.06	1	1.13	0.072
EBNA2	0.006	1.3	0.99	1.7	0.061	0.129	1.06	0.98	1.15	0.160
EBNA3B	0.005	1.17	0.88	1.56	0.287	0.137	1.05	0.99	1.12	0.088
BGLF5 (DNAse)	0.004	1.14	0.98	1.33	0.087	0.139	1.03	0.99	1.06	0.122
BLRF2 (VCA_p23)	0.001	1.27	0.95	1.7	0.107	0.142	1.09	0.98	1.2	0.112
BTRF1	0.016	1.09	0.92	1.31	0.327	0.147	1.03	0.99	1.07	0.157
EBNA3B	0.003	1.12	0.86	1.46	0.389	0.151	1.05	0.98	1.12	0.202
EBNA1	0.010	1.18	0.97	1.44	0.102	0.160	1.04	0.98	1.1	0.165
CAPSID	0.004	1.21	1	1.47	0.052	0.174	1.04	0.99	1.09	0.105
BBLF2	0.002	1.14	0.96	1.36	0.130	0.192	1.03	0.99	1.07	0.115
EBNA-LP	0.001	1.24	1.03	1.48	0.025	0.250	1.05	0.98	1.11	0.162
BPFL1	0.017	1.08	0.96	1.21	0.220	0.259	1.04	0.99	1.09	0.154
BPFL1	0.001	1.28	1.05	1.55	0.017	0.280	1.02	0.98	1.06	0.375

BPFL1	2.96E-04	1.52	1.15	2	0.004	0.323	1.03	0.96	1.1	0.460
BRRF1	0.014	1.16	0.98	1.37	0.082	0.379	1.04	0.96	1.13	0.355
EBNA3A	0.015	1.37	1.06	1.76	0.017	0.420	1.02	0.97	1.08	0.402
EBNA3B	3.40E-04	1.4	1.1	1.78	0.008	0.439	1.05	0.93	1.18	0.451
EBNA3A	0.001	1.49	1.06	2.1	0.024	0.591	0.98	0.86	1.11	0.733
BALF3	0.013	1.14	0.99	1.3	0.063	0.594	1.06	0.91	1.23	0.475
BMRF1 (EA(D)_p47/54)	0.016	1.16	0.89	1.52	0.266	0.635	0.98	0.89	1.09	0.763
BDLF4	0.001	1.25	1.01	1.56	0.046	0.867	1.02	0.95	1.09	0.581

Chapter 4

Characterization of the proteome-wide Epstein-Barr virus antibody responses after T-cell immunotherapy in patients with EBV-associated lymphomas

This chapter details the evaluation of systemic IgG and IgA antibody responses against the entire EBV proteome in treatment responders and non-responders to Epstein-Barr virus-specific T-cell (EBVST) immunotherapy in Phase I clinical protocols. This is the first report to our knowledge that comprehensively evaluated humoral immune responses in EBV-positive lymphoma patients treated with EBV-specific T-cell therapy. The work presented in this chapter comprises a publication pending submission.

4 Characterization of the proteome-wide Epstein-Barr virus antibody responses after T-cell immunotherapy in patients with EBV-associated lymphomas

Yomani D. Sarathkumara[1], Nathan W. Van Bibber[2], Zhiwei Liu[3], Sweta K. Sinha[2], Cliona M Rooney[4†], Anna E. Coghill[2†], Carla Proietti[1†], Denise L. Doolan[1 †]

[1] Centre for Molecular Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia

[2] Cancer Epidemiology Program, Division of Population Sciences, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

[3] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA

[4] Baylor College of Medicine, Houston, TX, USA

[*] These authors contributed equally as first authors

[†] These authors contributed equally as last authors

4.1 Abstract

Epstein–Barr virus (EBV) is associated with a diverse range of lymphomas and several viral antigens expressed by EBV-associated lymphomas have been targeted for T-cell directed immunotherapy. Recent studies have demonstrated promising safety and clinical efficacy using EBV-specific T cells (EBVSTs) to treat EBV-associated malignancies, including the most immunogenic EBV lymphoma in post-transplant recipients. However, there are significant numbers of individuals who do not respond or show poor clinical improvement (non-responders). To identify additional EBV antigens that may be effective targets for T cell immunotherapy, we comprehensively characterized IgA and IgG antibody responses to the complete EBV proteome in EBV-associated lymphoma patients treated with EBVSTs infusions in Phase I clinical trials. Differences in antibody profiles between responders and non-responders were demonstrated at pre, 2-weeks. 4-weeks, and 3-months. Unexpectedly, overall mean antibody responses were decreased in responders and elevated in non-responders. Six anti-EBV antibodies, three IgA (BGLF3, BALF2, BBLF2/3) and three IgG (BGLF2, LF1, BGLF3), that were significantly decreased from pre to 3-months post-treatment in responders but significantly increased among non-responders were identified. Our data suggest the potential of these antibodies as novel targets for EBVST immunotherapy.

Key words: EBV-specific T-cells, Antibodies, Immunotherapy, Responders

4.2 Introduction

Epstein-Barr virus (EBV) is associated with a diverse range of malignancies including Burkitt's lymphoma, Hodgkin lymphoma, extranodal natural killer/T-cell lymphoma and post-transplant lymphoproliferative disease (PTLD) (1). EBV is also associated with cancers arising from epithelial cells (*i.e.*, nasopharyngeal carcinoma (NPC) and gastric carcinoma) but is rarely found in smooth muscle cell tumours (*i.e.*, leiomyosarcoma) (2, 3).

EBV-associated lymphomas express viral antigens which represent potential targets for T-cell immunotherapy as oncolytic agents, since virus-specific T-cells (VSTs) can provide immunostimulatory effects and confer anti-tumour activity (4). In recent years, antigen-specific T-cell immunotherapy has proved to be a promising therapeutic approach, predominantly for hematologic malignancies (5). EBV-specific T-cells (EBVSTs) infusion has been shown to be a safe and effective treatment for patients with EBV-associated PTLD (6, 7). More recently, EBVSTs directed to viral antigens generated from transplant donors have been effective in treating EBV-associated PTLD in both hematopoietic stem cell transplant (HSCT) and solid organ transplant (SOT) recipients (8-10). Clinical trials using EBV-CTLs for PTLD after HSCT has shown efficacy in 85% (11/13 patients) achieved complete remissions that were sustained without recurrence. Another study using donor derived EBV-CTLs or third party EBV-CTLs showed an overall response rate of 71%, with 10 patients achieved complete remissions out of 14 (10).

Importantly, epitope spreading beyond the initially targeted EBV antigens has been observed in patients achieving clinical responses following cytotoxic T lymphocytes (CTLs) treatment, and this breadth appeared critical for maintaining antitumor response (11).

EBV-associated malignancies demonstrate distinct gene expression profiles (12). Expression patterns of a limited number of EBV genes characterize EBV latency types and the type of tumour (13). Of the various types, type III latency has been the most successfully treated by adoptive T-cell therapy as compared to other latency types (14). Type II latency tumours express EBV nuclear antigen (EBNA1) and latent membrane proteins (LMP 1 and 2) which are less immunogenic, whilst type I express EBNA1 which is poorly immunogenic and more challenging to treat with EBVSTs (15). However, T-cell immunotherapy directed against the LMP1 and/or LMP2 antigens was shown to be a durable, safe, and effective approach without significant toxicity in Hodgkin lymphoma and extranodal NK/T-cell lymphoma, which typically expresses type II latency (16-18).

Despite recent advances in the rapidly growing field of T-cell immunotherapy, subsets of patients are still not responding or partially responding to treatment and are relapsing, challenges achieving remission in the long term (19). The need for the biomarkers that could distinguish individuals who did respond from those who did not respond to immunotherapy would be valuable.

Immuno-oncology research has to date focused only on T-cell responses. The role of B-cells and humoral responses in cancer immunotherapies has not been studied well. EBV-associated malignancies are associated with aberrant antibody responses to EBV proteins and antigen complexes (20, 21). IgG and IgA antibodies against EBV capsid antigen (VCA), EBNA1, and early antigen (EA) have been recognized as major antigen targets and commonly used in serology–based screening for EBV-associated malignancies (22).

Improved understanding of the repertoire of EBV antigens targeted by immunotherapy would broaden the potential of immunotherapy for the treatment EBV-associated tumours. As this approach becomes more broadly accessible in the future, an improved understanding of which EBV antigens to be targeted and controlled to reach optimum outcomes of immunotherapy is highly essential to accelerate the development of targeted therapies for EBV-positive lymphomas.

The systematic characterization of EBV antigens as novel immunotherapy targets for the treatment of EBV lymphomas could be achieved by whole-proteome profiling. We have previously developed a comprehensive microarray comprising the complete EBV proteome and applied this to characterize humoral responses in EBV-associated cancers (23-26).

Herein, we used our custom protein microarray representing the complete EBV proteome to measure IgG and IgA antibody responses targeting 199 different EBV protein sequences in 56 patients on autologous or third-party EBV-specific T-cell immunotherapy clinical protocols conducted by the Baylor College of Medicine, USA. This study represents the first evaluation of this multiplex tool in the context of treatment outcomes following T-cell immunotherapy. Our objective was to determine the differences in antibody responses to EBV proteins between responders and non-responders for EBVSTs immunotherapy. Then, we evaluated how these immune markers changed between pre- and post-treatment timepoints in these individuals. These results have the potential to elucidate which EBV proteins are 'seen' by the immune system and correlate with effective patient outcomes, a crucial piece of information to inform effective therapeutic treatment.

4.3 Materials and Methods

4.3.1 Subjects and samples

Archived plasma samples were selected from patients diagnosed with EBV-positive lymphomas receiving autologous, or allogeneic (third-party) peptide stimulated infusion of EBV-specific T-cells (EBVSTs) collected from a multi-centre Phase I T-cell immunotherapy clinical study (GRALE, PREVALE and MABEL) conducted by Baylor College of Medicine at the Baylor College of Medicine, Baylor St. Luke's Medical Center, Harris County Hospital District Ben Taub, Mayo Clinic - Minnesota, Texas Children's Hospital, Texas Children's Hospital General Clinical Research Center, or the Methodist Hospital (**Table 4-1**). The clinical studies were approved by the Institutional Review Boards of the Baylor College of Medicine, USA (BCM IRB), the National Cancer Institute (NCI), representative institutes, and affiliated hospitals. Ethical approval was granted for the recruitment of patients and the generation of their autologous EBVSTs for immunotherapy (ethics number H-15280); or for the recruitment of healthy donors for allogeneic (third-party) EBVST for immunotherapy (ethics number H-151520). Written informed consent was obtained from all participants. All laboratory testing was conducted under a protocol approved by James Cook University Human Research Ethics Committee (H7696).

The 56 patients diagnosed with EBV-positive lymphoma treated with EBVSTs used in our study were categorized as positive responders (n=36) or non-responders (n=20) based on treatment outcomes at follow-up assessment (4). The samples were collected at pre-treatment and post-treatment timepoints (2-weeks, 4-weeks, and 3-months). Paired plasma samples were not available for analysis from all individuals.

Table 4-1 EBV-specific T-cells (EBVSTs) clinical trials targeting EBV-associated malignancies (modified from Sharma et al., 2020 (4))

Sponsor Center	Clinical trial ID Phase Name	EBVST characteristics Donor source/ Manufacturing/ Target specificities, if described
Baylor College of Medicine Houston, TX, USA	NCT01555892 GRALE Phase I	Autologous Peptide stimulated LMP 1 and 2, EBNA1, BARF1

Baylor College of Medicine Houston, TX, USA	NCT02973113 PREVALE Phase I	Autologous Peptide stimulated LMP, EBNA1, BARF1
Baylor College of Medicine Houston, TX, USA	NCT02287311 MABEL Phase I	Allogeneic (Third-party) Peptide stimulated LMP1 and 2, EBNA1, BARF1

4.3.2 EBV custom protein microarray

Plasma samples were probed using a custom EBV protein microarray targeting IgA and IgG antibodies against 202 EBV protein sequences, as previously described (23-27). Briefly, our comprehensive EBV protein microarray contains 199 EBV protein sequences generated from five different EBV strains (AG876, Akata, B95-8, Mutu, and Raji) representing nonredundant open reading frames and predicted splice variants from 86 EBV proteins. Also included on the array were peptide sequences representing three synthetic EBV peptides (VCAp18, EBNA-1, and EA p47) considered putative cancer biomarkers. Four "noDNA" (no translated protein) spots were included in the array to correct for person-specific background.

Plasma samples from each study participant (**Table 4-2**) were tested blinded on this EBV protein microarray as described previously (23-26). Briefly, antibody responses were detected with biotin-conjugated goat anti-human IgG (1:1000 dilution) or IgA (1:500 dilution) secondary antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) and visualized with streptavidin-conjugated SureLight® P3 (Columbia Biosciences, Columbia, MD, USA) (1:200 dilution) antibody. After probing, air-dried probed slides were scanned on an Axon GenePix 4300B (Molecular Devices). Raw fluorescence intensities were corrected for spot-specific background using the Axon GenePix Pro 7 software, and data were variant log-transformed using variance stabilizing normalization (VSN) transformation in Gmine (28). The array output was then standardized, referred to as the standardized signal intensity (SSI), to the person-specific background using the individual's cut-off (mean ±1.5 SD of the four "no DNA" spots). Positivity was defined as a standardized signal intensity > 1.0, and output was further categorized into positive (1) and negative (0) responses.

A cut-off of coefficient of variation [CV] < 30%) was selected based on nineteen duplicate samples for quality control and assessment of array reproducibility as applied in our previous studies (23-25). We excluded array spots with CVs ≤ 30% from the analysis, leaving 74 IgG and 202 IgA markers.

4.3.3 Statistical analysis

All statistical testing was performed using R Studio (https://www.r-project.org/, RStudio Inc., Boston, MA, USA, Version 2022.07.1). For all the *p*-values, $p < 0.05$ was considered significant.

IgG and IgA antibodies with CVs > 30% were excluded from our analyses, consistent with our previous reports (23-25). Differences in the mean standardized signal intensity (SSI) for IgG and IgA antibodies between the responders (n=36) and non-responders (n=20) were tested using unpaired-t-tests at each timepoint (pre, 2-weeks, 4-weeks, and 3-months). The odds ratios (ORs) and 95% CIs for the association between each anti-EBV antibody variable (*i.e.*, categorized into binary responses (positive=1, negative=0)) and treatment response status at follow-up (positive response=1, no response=0) were calculated using logistic regression models adjusted for sex, age at enrolment, and diagnosis (T-cell lymphomas, Hodgkin's lymphoma, B-cell lymphomas, and other). Following analyses were performed using standardized signal intensity (SSI) data.

Random Forest models were applied to identify the antibody markers that were able to most effectively discriminate responders from non-responders for immunotherapy treatment at each timepoint. Random forest is a supervised machine learning classifier to improve the predictive accuracy of a dataset (29). The most critical antibodies were selected using the R statistical package randomForest (Version 4.7-1.1) (30). Within each timepoint, the top 12 antibodies with the highest values for the two randomForest prediction metrics, MeanDecreaseAccuracy (MDA) and MeanDecreaseGini (MDG) were determined. Missing values were imputed using the function rfImput. The overlapping antibody markers by the two metrics (MDA and MDG) at each timepoint were then considered the most critical markers in discriminating groups.

The total change in antibody (Ab) response was evaluated between the pre and 3-months ($Ab_{3mo}-Ab_{pre}$) timepoints among responders and non-responders for treatment. Total change in antibody response was defined as the difference in antibody response between the 3-months timepoint and baseline (pre timepoint). We included only patients (n = 41 (for IgG), n = 42 (for IgA)) for whom we had complete matching paired data for the pre and 3-months timepoints and excluded patients missing either pre or 3-months timepoint. The odds ratios (ORs) and 95% CIs for the association between each anti-EBV antibody variable (*i.e.*, $Ab_{3mo}-Ab_{pre,}$ SSI as continuous levels) and treatment response status as the outcome variable

were calculated using logistic regression models adjusted for sex, age at enrolment, and diagnosis (T-cell lymphomas, Hodgkin's lymphoma, B-cell lymphomas, and other).

4.4 Results

All patients included in the EBVST immunotherapy study were diagnosed with EBV-associated tumours (n=56) and classified as either responders (n=36) or non-responders (n=20) at treatment follow-up. The distribution of baseline demographic characteristics among the responders (n=36) *vs*. non-responders (n=20) is shown in **Table 4-2**. We evaluated antibody at responses across different time points, for pre-treatment or 2-weeks, 4-weeks, or 3-months post-treatment.

Table 4-2 Baseline characteristics of all individuals and by responders versus non-responders status for EBVSTs treatment

Characteristics	All subjects (n=56*) n (%)	Responders (n=36) n (%)	Non-responders (n=20) n (%)
Sex			
Female	17 (30.36)	10 (27.78)	7 (35.0)
Male	39 (69.64)	26 (72.22)	13 (65.0)
Age groups (years)			
0-30	24 (42.86)	17 (47.22)	7 (35.0)
31-60	21 (37.5)	12 (33.33)	9 (45.0)
61-100	11(19.64)	7 (19.44)	4 (20.0)
Ethnic groups			
Hispanic	12 (21.43)	9 (25.0)	3 (15.0)
Non-Hispanic	44 (78.57)	27 (75.0)	17 (85.0)
Diagnosis			
T-cell lymphomas	13 (23.21)	8 (22.22)	6 (30.0)
Hodgkin's lymphoma	19 (33.93)	14 (38.89)	5 (25.0)
B-cell lymphomas	13 (23.21)	7 (19.44)	5 (25.0)
Other	11 (19.64)	7 (19.44)	4 (20.0)
Clinical trial protocol			
GRALE	38 (67.86)	28 (77.78)	10 (50.0)
MABEL	15 (26.78)	6 (16.67)	9 (45.0)
Other	3 (3.57)	2 (5.55)	1 (5.0)

4.4.1 Differential anti-EBV antibody profiles between responders *vs.* non-responders for EBVSTs immunotherapy

We evaluated the IgG and IgA antibody responses at pre-, 2-weeks, 4-weeks, and 3-month timepoints between responders and non-responders for EBVST immunotherapy (**Figure 1**). IgG and IgA antibodies that were significantly different between responders and non-responders per each timepoint by either t-tests, logistic regression or Random Forest analyses are reported in **Table S1**.

4.4.1.1 Antibody profile at pre-EBVSTs immunotherapy

At pre-treatment, IgG antibodies against EBV nuclear antigen leader protein (EBNA-LP) and IgA antibodies for lytic gene BGLF3.5 were significantly elevated among the non-responders as compared to responders (p <0.05, t-test) (**Figure 4-1**). IgG antibodies against Epstein–Barr nuclear antigen (EBNA3B) remained borderline associated with the non-responders' group when adjusted for age, sex, and diagnosis in the logistic regression model (p =0.04). Random Forest analysis identified five additional IgA and three IgG antibodies as important discriminative markers between groups (**Table S1**). Both t-tests and random forest metrices recognized EBNA-LP-IgG as an important marker at pre-treatment. However, none of these antibodies overlapped by all metrics (*i.e.*, t-tests, logistic regression output, MDA and MDG).

4.4.1.2 Antibody profile at 2-weeks post-EBVSTs immunotherapy

IgG antibodies against LMP1, BGLF3.5 and two variants of BMRF1 and IgA antibodies against BGRF1/BDRF1 and EBNA3A were highly elevated in non-responders (p <0.05, t-test) at 2-weeks after immunotherapy treatment (**Figure 4-1**). BGLF3.5-IgG, LMP1-IgG and BGRF1/BDRF1-IgA remained significantly associated with the non-responder's group when adjusted for age, sex, and diagnosis (p <0.05, logistic regression) (**Table S1)**. In addition, IgG antibodies for two variants of latent membrane protein 2A (LMP2A) and BNLF2A were significantly associated with the no response group when adjusted for demographics (p <0.05, logistic regression). BGRF1/BDRF1-IgA and LMP1-IgG were overlapped in t-tests, logistic regression output, MDA and MDG of Random Forest analysis (**Table S1)**.

Interestingly, LMP1, LMP2A, BNLF2A, BGLF2 markers were observed to be significantly elevated at 3-month timepoint among non-responders (p <0.05, t-test).

4.4.1.3 Antibody profile at 4-weeks post-EBVSTs immunotherapy

Notably, IgA antibodies against BPLF1, the late lytic gene, were significantly associated with the non-responders' group (p = 0.02, t-test; p = 0.02, logistic regression) and overlapped by MDA and MDG in Random Forest analysis (**Figure 4-1, Table S1)**. However, none of the important markers at 2-weeks timepoint were elevated at 4-weeks post-treatment. Subsequently, antibody responses seemed low at 4-weeks post-EBVST infusion in all individuals (**Figure 4-1**).

4.4.1.4 Antibody profile at 3-months post-EBVSTs immunotherapy

More pronounced IgG responses, but not IgA, were apparent at the 3-month timepoint in patients who did not respond to treatment at follow-up (**Figure 4-1)**. 36 IgG antibodies were identified as significantly elevated among non-responders for treatment (p <0.05, t-test). Eight of these 36 IgG antibodies and two additional IgG antibodies were significantly associated with the no-response group by adjusted logistic regression models p (<0.05, logistic regression) (**Table S1)**.

Of all IgG antibodies, four variants of LMP2A oncoprotein were significantly associated with non-responders when adjusted for demographics in logistic regression models. LMP1, BKRF2 and two variants of BGRF1/BDRF1, were among the other IgG antibodies highly elevated among non-responders at the 3-month timepoint (p <0.05, t-test; <0.05, logistic regression). One of the BGRF1/BDRF1-IgG overlapped in all three metrics (*i.e.*, logistic regression output, MDA and MDG) (**Table S1)**. Interestingly, our results demonstrated very high overall IgG response rates in patients treated with EBVST immunotherapy who had no complete responses at follow-up.

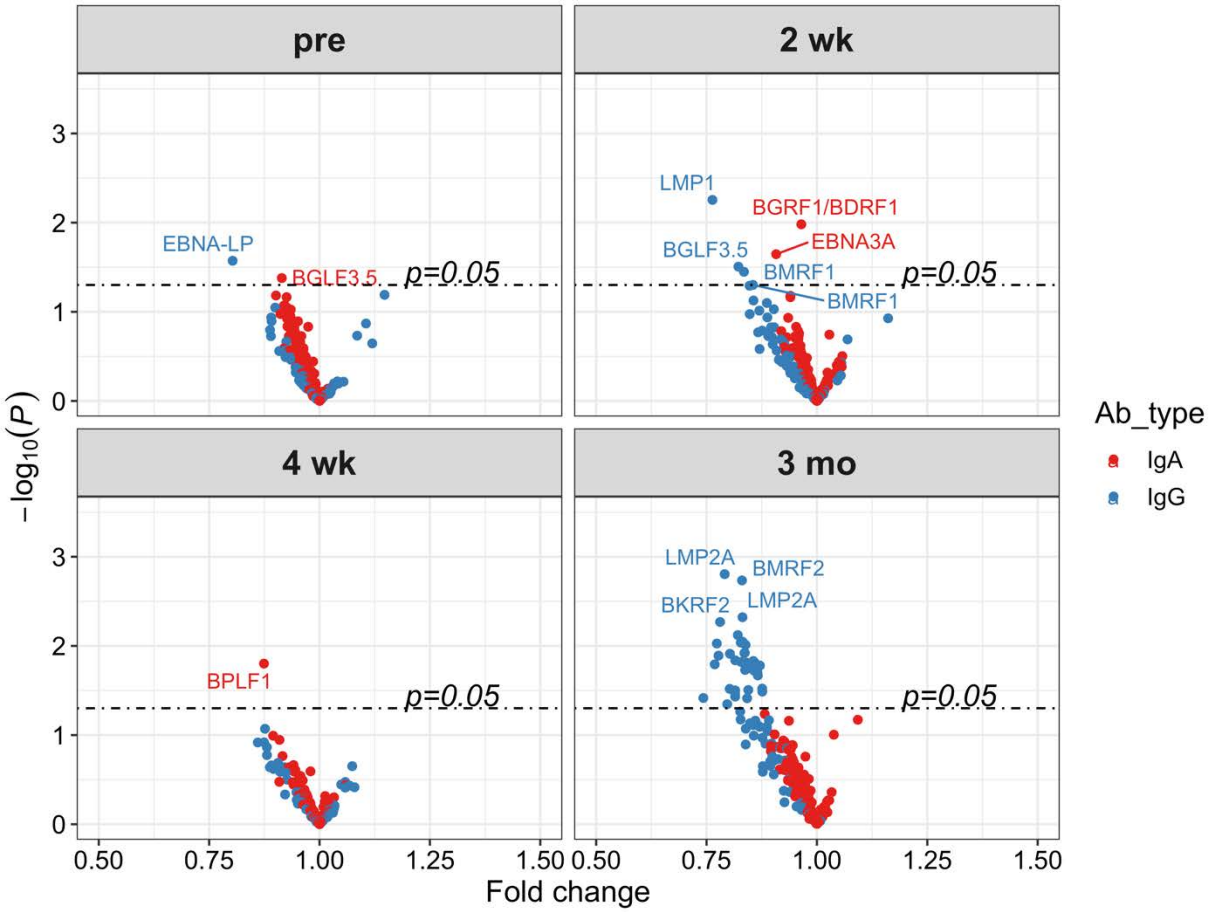


Figure 4-1 Differential IgA and IgG responses between treatment responders and non-responders.

Differences in the mean antibody response for treatment responders vs. non-responders at pre, 2-weeks (2 wk), 4-weeks (4 wk), and 3-months (3 mo). The x-axis of the volcano plot displays the fold change (responders vs. non-responders ratio of standardized signal intensity) for all antibodies with CV < 30%. (red, IgA; blue, IgG). The y-axis illustrates the p-value corresponding to the t-test for a difference in standardized signal intensity (SSI) between treatment responders and non-responders. The dashed lines represent the statistically significant p-value threshold. The antibodies with the smallest p-values are highlighted.

4.4.2 Change in Antibody responses between responders and non-responders at pre and 3-months post-EBVSTs immunotherapy

Total change in antibody response was evaluated between the pre and 3-months ($Ab_{3mo}-Ab_{pre}$) timepoints among responders and non-responders for EBVSTs immunotherapy. We excluded 15 for IgG and 14 for IgA individuals missing either pre or 3-months timepoints, leaving n = 41 (for IgG) and n = 42 (for IgA) matching paired data for pre-treatment and 3 months post-treatment to analyze the total change in antibody response.

Six anti-EBV antibodies (3-IgA; BGLF3, BALF2, BBLF2/3 and 3-IgG; BGLF2, LF1, BGLF3) were significantly associated with treatment response at the end of follow-up (p <0.05, logistic regression) when adjusted for age, sex, and diagnosis (**Table 4-3**). Ribbon plots in **Figure 4-2** show the changes in average responses per each antibody marker.

Importantly, the no response group demonstrated a significant increase in mean antibody responses at 3-months compared to pre-treatment for all six antibody markers, indicating the elicitation of robust antibody responses for EBVSTs infusions in individuals who did not clinically respond to treatment (non-responders). Notably, in contrast, the responders to treatment showed a significant decrease in antibody responses at 3-months compared to pre-treatment, consistent with a role for these antibodies in destroying tumour cells.

Both IgG (p=0.046, OR=0.075, 95% CI = 0.004, 0.713) and IgA (p=0.036, OR=0.011, 95% CI = 0, 0.564) responses for BGLF3 were highly significantly associated with non-responders for mean change in antibody responses. BGLF3-IgG was significantly elevated at 3-month timepoint in no-response group in univariate analysis (p <0.05, t-test), and was identified as an important marker by random forest metrics. BGLF2-IgG was highly elevated in non-responders at 3-months (p <0.05, t-test) whilst it was identified as an important marker at 2-weeks post-immunotherapy by MDA and MDG.

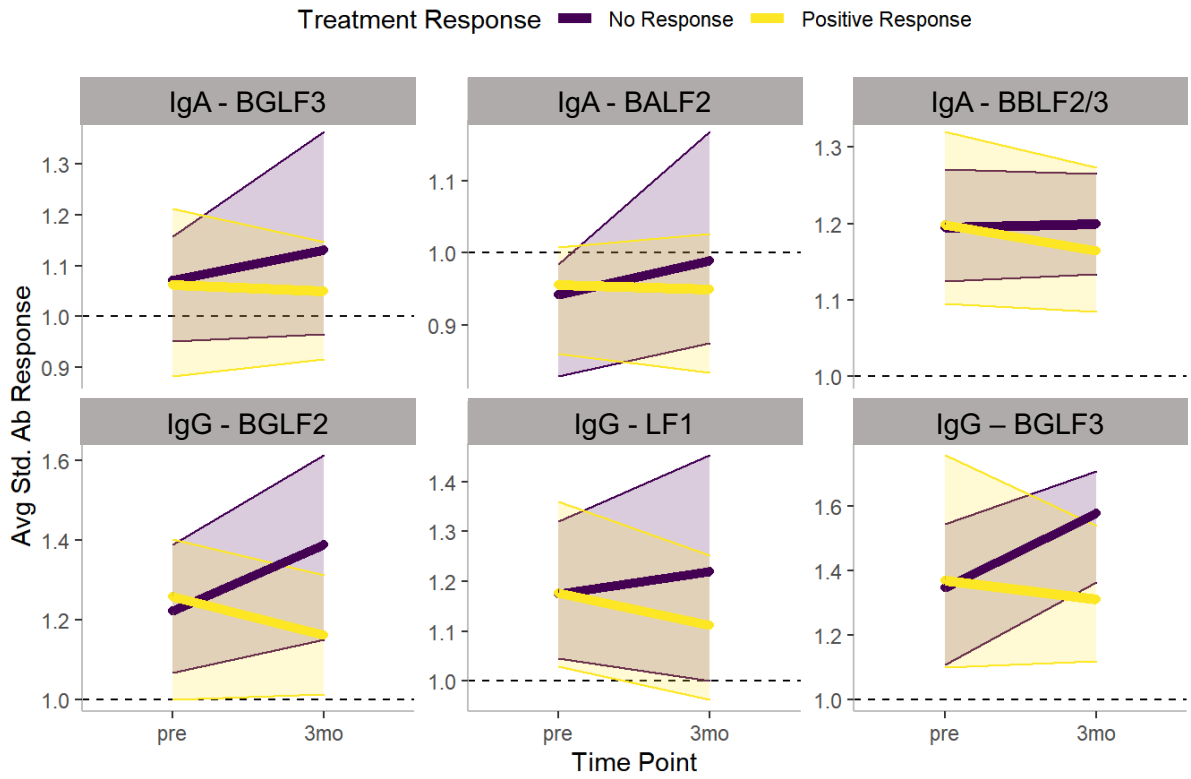


Figure 4-2 Markers significant for total change in antibody response between the pre and 3-months timepoints.

The ribbon plots show the change in the mean of antibody responses per each significant antibody marker between the pre and 3-months timepoints ($p < 0.05$, logistic regression). Purple indicates no response group, whilst yellow indicates a positive response group. The solid lines indicate the mean standardized antibody response per timepoint, and the ribbons include the range between show the 75th and 25th percentile range of standardized antibody response.

Table 4-3 Table of significant odds ratios for the total change in antibody response between the pre and 3-months ($Ab_{3mo}-Ab_{pre}$) timepoints

Array sequence	Protein name	Life cycle	IgG/ IgA	Log. reg *p*-value	OR (95% CI)
YP_001129484.1-113481-112483	BGLF3	Late lytic	IgA	0.036	0.011 (0-0.564)
YP_001129510.1-165796-162410-2	BALF2	Early lytic	IgA	0.036	0 (0-0.154)
AFY97950.1-104968-104363	BBLF2/3	Early lytic	IgA	0.043	0 (0-0.324)
YP_001129486.1-115415-114405	BGLF2	Early lytic	IgG	0.042	0.06 (0.002-0.681)
YP_001129505.1-153178-151769	LF1	Other/Unknown	IgG	0.045	0.036 (0.001-0.7)
YP_001129484.1-113481-112483	BGLF3	Late lytic	IgG	0.046	0.075 (0.004-0.713)

4.5 Discussion

The adaptive immune system plays a critical role in cancer, and it is now well established that antigen-specific CTLs are important in anti-cancer immune response either by destroying tumour cells or controlling tumour growth. This provides the foundation for developing immunotherapies for cancer treatment (31). Indeed, there has been a dramatic expansion in antigen-specific T-cell based therapeutic approaches for the treatment of hematologic malignancies over the past three decades (5). EBVSTs infusion have been specifically used to treat EBV-associated lymphomas for more than 20 years (14, 15).

EBV-positive tumours express a number of viral latency-associated antigens which can be targeted for T-cell immunotherapy. Adoptive transfer of EBVSTs has been demonstrated as a promising treatment option for immunogenic type III latency-derived PTLD, commonly occurring in transplant immunocompromised recipients with weakened immune systems (10). However, most research to date on EBVSTs have targeted only a limited number of viral antigens expressed in latency, namely LMP1, LMP2 and EBNA1 (32). Thus, broadening the spectrum of viral antigens by proteome-wide profiling can be a promising approach for the identification of novel target antigens for immunotherapy, for EBVSTs for EBV-associated lymphomas. To address this, we utilized a custom protein microarray consisting predicted sequences from 86 proteins and splice variants representing the complete EBV proteome identified across the five EBV strains to measure both IgG and IgA antibody responses. In addition, to identifying a broad spectrum of viral antigens as target antigens for immunotherapy, this approach would provide insights into differences in humoral antibody profiles between responders and non-responders to EBVSTs immunotherapy. This is the first report to our knowledge, comprehensively evaluating patterns of anti-EBV antibody responses in EBV-positive lymphoma patients undergoing EBV-specific T-cell therapy.

Profound differences in the anti-EBV antibody profiles between responders and non-responders to EBVST immunotherapy were demonstrated, with the mean antibody responses elevated in non-responders and decreased in responders. Notably, we identified 3-IgA (BGLF3, BALF2, BBLF2/3) and 3-IgG (BGLF2, LF1, BGLF3) antibodies that were significantly decreased from pre to 3 months in individuals who responded to treatment, suggesting that these antibodies play an important role in targeting EBV-positive tumour cells.

Conversely, we identified different sets of antibodies that were significantly increased in patients who did not respond at pre-treatment and timepoints (2-weeks, 4-weeks, and 3-months). At baseline pre-EBVSTs immunotherapy, EBNA3B-IgG was significantly associated

with non-responders for treatment when adjusted for sex, age, and diagnosis in logistic regression models. EBNA3B mutations suggest promoting EBV-driven B cell lymphomagenesis and immune evasion (33, 34). Late lytic (BGLF3.5-IgG, BNLF2A-IgG, BGRF1/BDRF1-IgA) and latent (LMP2A-IgG, LMP-1-IgG) antibodies were significantly associated with no response group for treatment when adjusted for demographics in logistic regression model at 2-weeks timepoint. Moreover, IgA antibodies against capsid scaffold protein BGRF1/BDRF1 and IgG antibodies for LMP-1 latent protein demonstrated significant associations with non-responders for treatment as per adjusted logistic regression output, MeanDecreaseAccuracy (MDA) and MeanDecreaseGini (MDG) measures for the Random Forest constructed with the most differentially expressed antibodies at 2-weeks post-EBVST immunotherapy. At 4-weeks post-therapy, BPLF1-IgA was identified as a marker associated with non-responders by all three matrices (logistic regression output, MDA and MDG). BPLF1 is expressed at the late phase of lytic EBV infection and has been recognized as contributing to immune evasion (35).

The most pronounced elevations in antibody positivity at 3-months timepoint for responders compared with non-responder for EBVST immunotherapy were observed for IgG antibodies. Although the univariate analysis of antibody profiles between responder *vs*. non-responders for treatment identified 36 IgG antibodies (p <0.05, t-test) as highly elevated, only ten IgG antibodies associated with individuals who had no response for treatment by adjusted logistic regression models. Our results suggested a pivotal role of LMP2A-IgG in non-responders to the immunotherapy treatment after 3 months. LMP2A has been described as a facilitator of B-cell survival, promoting virus persistence, supports B-cell activation and transformation (36-39). In a clinical study on metastatic melanoma, patients treated with immune checkpoint inhibitors were tested for anti-EBV EBNA-1-IgG serum antibodies between responders and non-responders at pre-treatment compared to post-treatment. Total IgG anti-EBNA1 antibody levels were found to be similar in responders and non-responders (40). Peripheral autoantibodies against tumour-associated proteins have been explored as markers for cancer and to predict clinical outcomes (41-43).

Our overall findings reveal that patients with positive clinical improvement who responded to EBVSTs immunotherapy were characterized by an overall low level of antibody responses whereas patients who did not respond to treatment had elevated antibodies post-treatment. Our data are consistent with previous reports showing persistently high EBV viremia levels and decreased frequency of EBV-specific CTLs in PTLD patients who poorly responded to immunotherapy, suggesting impaired T-cell recognition of tumour targets (44, 45). Recent

studies have demonstrated that the tumour microenvironment and its immune cells play an essential role in regulating the responses contributing efficacy of immunotherapy (46-48). The tumour microenvironment consists of various cell types including cancer cells, immune cells, and stromal cells (48). The tumour microenvironment plays a dual role in cancer. It can suppress tumour progression by killing cancer cells or inhibiting their outgrowth through the action of anti-tumour immune cells, such as T-cells and NK cells (47). In some cases of EBV-driven malignancies, the tumour microenvironment is modulated for viral benefit leading to tumour progression and resistance to immunotherapy by EBV latent genes interfering with the innate and adaptive immunity (13). Therefore, resistance to immunotherapy in EBV-associated malignancies has been suggested due to the changes in the tumour microenvironment to suppress or anergize EBV-specific T-cell activity (14, 32, 49).

We hypothesized that the elevated anti-EBV antibodies inversely associated with the clinical improvement of patients could be explained by the alterations in the hostile tumour microenvironment. However, little is known about the role of B-cells and antibodies in the tumour microenvironment of EBV-associated cancers (50). The role of B-cells and B-cell mediated antibody responses (representing the humoral arm of the adaptive immune system) in facilitating anti-tumour activity and its contribution to immunotherapies are relatively understudied in the field of immune-oncology (51). In tumour immunology, B-cells have known functions such as antigen presentation and producing tumour-specific antibodies (52, 53). It is known that B-cells that produce antigen-specific antibodies can exert anti-tumour responses by antibody-dependent cell cytotoxicity (ADCC) and activation of the complement cascade (54). Some studies suggest that B-cells generate inhibitory factors that impede or suppress the immune system's ability to kill cancerous cells (55-57). B-cells hindering anti-tumour responses have been demonstrated in mouse models of cancer (58-60). However, more knowledge of the full range of B-cell and antibody functions in the tumour microenvironment is needed to deliver a complete picture (61).

One limitation of our study is the small sample size, which was dictated by the phase 1 clinical study design. Thus, it is important to follow-up on our initial findings suggesting that non-responders to immunotherapy had elevated antibody responses, in larger patient cohorts from clinical immunotherapy trials. Additionally, because of the limited sample sizes, our analyses were not able to consider differences between immunotherapy protocols, autologous and allogenic EBVSTs, tumour types, or EBV exposure in donors derived EBVSTs (third-party) in clinical outcomes.

The current study is the first of its kind to describe the antibody responses against complete EBV proteome in EBV-associated lymphomas treated with EBVST immunotherapy. We demonstrated decreased antibody responses in responders for EBVSTs immunotherapy and identified six antibodies that could be potential targets of EBVSTs. Importantly, our results showed that non-responders had elevated antibody levels post-treatment, suggesting that antibody responses play an important role in hindering anti-tumour responses EBVSTs activity.

Despite the promising results from recent clinical trials of immunotherapies, still there are numbers of non-responders for treatments which necessitate identifying additional targets for cancer immunotherapies to achieve successful treatment outcomes. The diverse immune escape strategies related to EBV infection substantially impact disease progression and response to immunotherapies in EBV-associated malignancies. Thus, understanding the immune evasion mechanisms of EBV and identifying particular genes targeted by host immunity may provide valuable insights for reversing immune suppression to treat EBV-associated lymphomas. Our data suggest the potential of this sort of work for broadening target viral antigens filling an essential niche in targeted therapy for EBV-associated lymphomas. Future studies are required to identify predictive biomarkers to distinguish responders and non-responders to EBVST immunotherapy, which could facilitate novel targeted immunotherapy applications and precision medicine.

4.6 References

1. Hudnall SD, Ge Y, Wei L, Yang N-P, Wang H-Q, Chen T. Distribution and phenotype of Epstein–Barr virus-infected cells in human pharyngeal tonsils. Modern Pathology. 2005;18(4):519-27.
2. Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop AD. The Immunology of Epstein-Barr Virus–Induced Disease. Annual Review of Immunology. 2015;33(1):787-821.
3. Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. Frontiers in Oncology. 2019;9:713.
4. Sharma S, Leung WK, Heslop HE. Virus-specific T cells for malignancies - then, now and where to? Current Stem Cell Reports. 2020;6(2):17-29.
5. Huang J, Huang X, Huang J. CAR-T cell therapy for hematological malignancies: Limitations and optimization strategies. Frontiers in Immunology. 2022;13.
6. Rooney CM, Ng CYC, Loftin S, Smith CA, Li C, Krance RA, et al. Use of gene-modified virus-specific T lymphocytes to control Epstein-Barr-virus-related lymphoproliferation. The Lancet. 1995;345(8941):9-13.
7. Rooney CM, Smith CA, Ng CYC, Loftin SK, Sixbey JW, Gan Y, et al. Infusion of Cytotoxic T Cells for the Prevention and Treatment of Epstein-Barr Virus–Induced Lymphoma in Allogeneic Transplant Recipients. Blood. 1998;92(5):1549-55.
8. Heslop HE, Slobod KS, Pule MA, Hale GA, Rousseau A, Smith CA, et al. Long-term outcome of EBV-specific T-cell infusions to prevent or treat EBV-related lymphoproliferative disease in transplant recipients. Blood. 2010;115(5):925-35.
9. Doubrovina E, Oflaz-Sozmen B, Prockop SE, Kernan NA, Abramson S, Teruya-Feldstein J, et al. Adoptive immunotherapy with unselected or EBV-specific T cells for biopsy-proven EBV+ lymphomas after allogeneic hematopoietic cell transplantation. Blood, The Journal of the American Society of Hematology. 2012;119(11):2644-56.
10. Bollard CM, Rooney CM, Heslop HE. T-cell therapy in the treatment of post-transplant lymphoproliferative disease. Nature reviews Clinical Oncology. 2012;9(9):510-9.
11. Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. Journal of Clinical Oncology. 2014;32(8):798-808.

12. Middeldorp JM, Brink AATP, van den Brule AJC, Meijer CJLM. Pathogenic roles for Epstein–Barr virus (EBV) gene products in EBV-associated proliferative disorders. Critical Reviews in Oncology/Hematology. 2003;45(1):1-36.
13. Bauer M, Jasinski-Bergner S, Mandelboim O, Wickenhauser C, Seliger B. Epstein-Barr Virus-Associated Malignancies and Immune Escape: The Role of the Tumor Microenvironment and Tumor Cell Evasion Strategies. Cancers (Basel). 2021;13(20).
14. Gottschalk S, Rooney CM. Adoptive T-Cell Immunotherapy. Current Topics in Microbiology and Immunology. 2015;391:427-54.
15. Heslop HE, Sharma S, Rooney CM. Adoptive T-Cell Therapy for Epstein-Barr Virus-Related Lymphomas. Journal of Clinical Oncology. 2021;39(5):514-24.
16. Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. Journal of Clinical Oncology. 2014;32(8):798-808.
17. Cho SG, Kim N, Sohn HJ, Lee SK, Oh ST, Lee HJ, et al. Long-term Outcome of Extranodal NK/T Cell Lymphoma Patients Treated With Postremission Therapy Using EBV LMP1 and LMP2a-specific CTLs. Molecular Therapy. 2015;23(8):1401-9.
18. Gottschalk S, Edwards OL, Sili U, Huls MH, Goltsova T, Davis AR, et al. Generating CTLs against the subdominant Epstein-Barr virus LMP1 antigen for the adoptive immunotherapy of EBV-associated malignancies. Blood. 2003;101(5):1905-12.
19. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nature Reviews Gastroenterology & Hepatology. 2019;16(6):361-75.
20. Henle G, Henle W. Epstein-Barr virus-specific IgA serum antibodies as an outstanding feature of nasopharyngeal carcinoma. International Journal of Cancer. 1976;17(1):1-7.
21. Fachiroh J, Schouten T, Hariwiyanto B, Paramita DK, Harijadi A, Haryana SM, et al. Molecular Diversity of Epstein-Barr Virus IgG and IgA Antibody Responses in Nasopharyngeal Carcinoma: A Comparison of Indonesian, Chinese, and European Subjects. The Journal of Infectious Diseases. 2004;190(1):53-62.
22. Coghill AE, Hildesheim A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. American Journal of Epidemiology. 2014;180(7):687-95.

23. Liu Z, Sarathkumara YD, Chan JKC, Kwong Y-L, Lam TH, Ip DKM, et al. Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma. Scientific Reports. 2021;11(1):23664.
24. Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. International Journal of Cancer. 2020;147(3):608-18.
25. Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, et al. The association between the comprehensive Epstein–Barr virus serologic profile and endemic Burkitt lymphoma. Cancer Epidemiology and Prevention Biomarkers. 2020;29(1):57-62.
26. Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clinical Cancer Research. 2018;24(6):1305-14.
27. Liu Z, Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, et al. Patterns of interindividual variability in the antibody repertoire targeting proteins across the Epstein-Barr virus proteome. The Journal of Infectious Diseases. 2018;217(12):1923-31.
28. Proietti C, Zakrzewski M, Watkins TS, Berger B, Hasan S, Ratnatunga CN, et al. Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) Web-Server. Scientific Reports. 2016;6(1):1-15.
29. Biau G, Scornet E. A random forest guided tour. TEST. 2016;25(2):197-227.
30. Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002;2(3):18-22.
31. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. British Journal of Cancer. 2021;124(2):359-67.
32. Anna M, Riccardo T, Riccardo D, Debora M, Elena M, Patrizia C, et al. The interplay between Epstein-Barr virus and the immune system: a rationale for adoptive cell therapy of EBV-related disorders. Haematologica. 2010;95(10):1769-77.
33. White RE, Rämer PC, Naresh KN, Meixlsperger S, Pinaud L, Rooney C, et al. EBNA3B-deficient EBV promotes B cell lymphomagenesis in humanized mice and is found in human tumors. Journal of Clinical Investigation. 2012;122(4):1487-502.
34. Bhattacharjee S, Ghosh Roy S, Bose P, Saha A. Role of EBNA-3 Family Proteins in EBV Associated B-cell Lymphomagenesis. Frontiers in Microbiology. 2016;7.
35. van Gent M, Braem SG, de Jong A, Delagic N, Peeters JG, Boer IG, et al. Epstein-Barr virus large tegument protein BPLF1 contributes to innate immune evasion through interference with toll-like receptor signaling. PLoS Pathogens. 2014;10(2):e1003960.

36. Brielmeier M, Mautner J, Laux G, Hammerschmidt W. The latent membrane protein 2 gene of Epstein—Barr virus is important for efficient B cell immortalization. Journal of General Virology. 1996;77(11):2807-18.
37. Beaufils P, Choquet D, Mamoun RZ, Malissen B. The (YXXL/I) 2 signalling motif found in the cytoplasmic segments of the bovine leukaemia virus envelope protein and Epstein-Barr virus latent membrane protein 2A can elicit early and late lymphocyte activation events. The EMBO journal. 1993;12(13):5105-12.
38. Mancao C, Hammerschmidt W. Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. Blood. 2007;110(10):3715-21.
39. Rancan C, Schirrmann L, Hüls C, Zeidler R, Moosmann A. Latent Membrane Protein LMP2A Impairs Recognition of EBV-Infected Cells by CD8+ T Cells. PLOS Pathogens. 2015;11(6):e1004906.
40. Fässler M, Diem S, Mangana J, Hasan Ali O, Berner F, Bomze D, et al. Antibodies as biomarker candidates for response and survival to checkpoint inhibitors in melanoma patients. Journal for ImmunoTherapy of Cancer. 2019;7(1):50.
41. Wang D-R, Wu X-L, Sun Y-L. Therapeutic targets and biomarkers of tumor immunotherapy: response versus non-response. Signal Transduction and Targeted Therapy. 2022;7(1):331.
42. Zhou J, Zhao J, Jia Q, Chu Q, Zhou F, Chu X, et al. Peripheral Blood Autoantibodies Against to Tumor-Associated Antigen Predict Clinical Outcome to Immune Checkpoint Inhibitor-Based Treatment in Advanced Non-Small Cell Lung Cancer. Frontiers in Oncology. 2021;11.
43. Lu H, Goodell V, Disis ML. Humoral Immunity Directed against Tumor-Associated Antigens As Potential Biomarkers for the Early Diagnosis of Cancer. Journal of Proteome Research. 2008;7(4):1388-94.
44. Van Esser JWJ, Niesters HGM, Thijsen SFT, Meijer E, Osterhaus ADME, Wolthers KC, et al. Molecular quantification of viral load in plasma allows for fast and accurate prediction of response to therapy of Epstein–Barr virus-associated lymphoproliferative disease after allogeneic stem cell transplantation. British Journal of Haematology. 2001;113(3):814-21.
45. Doubrovina E, Oflaz-Sozmen B, Prockop SE, Kernan NA, Abramson S, Teruya-Feldstein J, et al. Adoptive immunotherapy with unselected or EBV-specific T cells for biopsy-proven EBV+ lymphomas after allogeneic hematopoietic cell transplantation. Blood. 2012;119(11):2644-56.

46. Zhang J, Shi Z, Xu X, Yu Z, Mi J. The influence of microenvironment on tumor immunotherapy. FEBS Journal. 2019;286(21):4160-75.
47. LV B, Wang Y, Ma D, Cheng W, Liu J, Yong T, et al. Immunotherapy: Reshape the Tumor Immune Microenvironment. Frontiers in Immunology. 2022;13.
48. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature. 2013;501(7467):346-54.
49. Sinha D, Srihari S, Beckett K, Le Texier L, Solomon M, Panikkar A, et al. 'Off-the-shelf' allogeneic antigen-specific adoptive T-cell therapy for the treatment of multiple EBV-associated malignancies. Journal for ImmunoTherapy of Cancer. 2021;9(2):e001608.
50. Tan GW, Visser L, Tan LP, van den Berg A, Diepstra A. The Microenvironment in Epstein-Barr Virus-Associated Malignancies. Pathogens. 2018;7(2).
51. Kim SS, Sumner WA, Miyauchi S, Cohen EEW, Califano JA, Sharabi AB. Role of B Cells in Responses to Checkpoint Blockade Immunotherapy and Overall Survival of Cancer Patients. Clinical Cancer Research. 2021;27(22):6075-82.
52. DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B, et al. Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. Clinical Immunology. 2018;187:37-45.
53. Bruno TC, Ebner PJ, Moore BL, Squalls OG, Waugh KA, Eruslanov EB, et al. Antigen-Presenting Intratumoral B Cells Affect CD4+ TIL Phenotypes in Non–Small Cell Lung Cancer PatientsTIL-Bs Present Antigen to CD4 TILs in NSCLC. Cancer immunology research. 2017;5(10):898-907.
54. Kinker GS, Vitiello GAF, Ferreira WAS, Chaves AS, Cordeiro de Lima VC, Medina TDS. B Cell Orchestration of Anti-tumor Immune Responses: A Matter of Cell Localization and Communication. Frontiers in Cell and Developmental Biology. 2021;9:678127.
55. Shalapour S, Font-Burgada J, Di Caro G, Zhong Z, Sanchez-Lopez E, Dhar D, et al. Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. Nature. 2015;521(7550):94-8.
56. Kessel A, Haj T, Peri R, Snir A, Melamed D, Sabo E, et al. Human CD19+CD25high B regulatory cells suppress proliferation of CD4+ T cells and enhance Foxp3 and CTLA-4 expression in T-regulatory cells. Autoimmunity Reviews. 2012;11(9):670-7.
57. Khan AR, Hams E, Floudas A, Sparwasser T, Weaver CT, Fallon PG. PD-L1hi B cells are critical regulators of humoral immunity. Nature Communications. 2015;6(1):1-16.

58. Affara NI, Ruffell B, Medler TR, Gunderson AJ, Johansson M, Bornstein S, et al. B cells regulate macrophage phenotype and response to chemotherapy in squamous carcinomas. Cancer cell. 2014;25(6):809-21.
59. Shalapour S, Lin X-J, Bastian IN, Brain J, Burt AD, Aksenov AA, et al. Inflammation-induced IgA+ cells dismantle anti-liver cancer immunity. Nature. 2017;551(7680):340-5.
60. Ammirante M, Luo J-L, Grivennikov S, Nedospasov S, Karin M. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. Nature. 2010;464(7286):302-5.
61. Bruno TC. New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. Nature. 2020;577(7791):474-6.

4.7 Supplementary Information

Table S1 EBV proteins on microarray (name, microarray sequences and EBV life cycle) for IgG and IgA antibody responses comparisons between responders and non-responders or EBVST immunotherapy treatment for each timepoint. ORs were adjusted for age, sex, and diagnosis. Two randomForest prediction metrics, Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) are reported for the overlapping antibody markers.

Array sequence	Protein name	Life cycle	*P*-value	Responders Mean	Non-responders Mean	Time point	IgG/ IgA	Log. reg *p*-value	OR (95% CI)	MDA	MDG
Pre-EBVST immunotherapy											
YP_001129440.1-20824-20955	EBNA-LP	Latent	0.03	1.04	1.29	pre	IgG	-	-	2.336	0.213
YP_001129483.1-112496-112035	BGLF3.5	Late lytic	0.04	1.15	1.26	pre	IgA	-	-	-	-
AFY97829.1-82733-83089	EBNA3B	Latent	0.33	1.03	1.10	pre	IgG	0.04	0.2 (0.04-0.83)	-	-
YP_001129454.1-67745-68959	BMRF1	Early lytic	0.11	1.45	1.59	pre	IgA	-	-	2.461	0.493
AFY97929.1-67486-68700	BMRF1	Early lytic	0.07	1.42	1.57	pre	IgA	-	-	3.439	0.319
CAA24877.1-48504-49967	EBNA2	Latent	0.08	1.16	1.26	pre	IgA	-	-	2.584	0.278
YP_001129515.1-169948-169188	LMP1	Latent	0.07	1.20	1.29	pre	IgA	-	-	3.154	0.187
CAA24856.1-92243-92602	EBNA3A	Latent	0.09	1.23	1.32	pre	IgA	-	-	2.931	0.221
AFY97978.1-151556-150147	LF1	Other/Unknown	0.13	1.05	1.17	pre	IgG	-	-	3.181	0.196
AFY97832.1-35377-35409	EBNA-LP	Latent	0.06	0.73	0.64	pre	IgG	-	-	2.702	0.266
2-weeks post-EBVST immunotherapy											
YP_001129485.1-117754-118890	BGRF1/BDRF1	Late lytic	0.01	0.97	1.01	2 wk	IgA	0.01	0.18 (0.04-0.63)	3.899	0.368

YP_001129463.1-80447-82888	EBNA3A	Latent	0.02	1.23	1.36	2 wk	IgA	-	-	2.868	0.275
YP_001129436.1-1574-1680	LMP2A	Latent	0.15	1.06	1.18	2 wk	IgG	0.02	0.17 (0.04-0.66)	-	-
YP_001129436.1-871-951	LMP2A	Latent	0.20	1.19	1.30	2 wk	IgG	0.04	0.12 (0.01-0.74)	-	-
YP_001129483.1-112496-112035	BGLF3.5	Late lytic	0.03	0.76	0.92	2 wk	IgG	0.04	0.15 (0.02-0.8)	-	-
YP_001129454.1-67745-68959	BMRF1	Early lytic	0.05	1.97	2.30	2 wk	IgG	-	-	-	-
AFY97929.1-67486-68700	BMRF1	Early lytic	0.04	1.83	2.19	2 wk	IgG	-	-	3.178	0.226
AFY97906.1-168167-168081	LMP1	Latent	0.01	0.99	1.30	2 wk	IgG	0.01	0.12 (0.02-0.5)	3.396	0.372
AFY97988.1-166888-166706	BNLF2A	Late lytic	0.24	1.20	1.29	2 wk	IgG	0.02	0.07 (0-0.49)	-	-
CAA24807.1-161678-159312	BALF3	Late lytic	0.07	1.12	1.19	2 wk	IgA	-	-	4.96	0.351
YP_001129509.1-162392-160335	BALF3	Late lytic	0.23	1.30	1.36	2 wk	IgA	-	-	2.789	0.219
YP_001129449.1-59370-49906-3	BPLF1	Late lytic	0.07	1.32	1.54	2 wk	IgG	-	-	3.671	0.282
YP_001129486.1-115415-114405	BGLF2	Early lytic	0.08	1.19	1.34	2 wk	IgG	-	-	3.532	0.271
4-weeks post-EBVST immunotherapy											
YP_001129449.1-59370-49906-3	BPLF1	Late lytic	0.02	0.86	0.99	4 wk	IgA	0.02	0.11 (0.01-0.62)	3.007	0.389
YP_001129476.1-102746-104587	BBRF1	Late lytic	0.23	1.05	1.12	4 wk	IgA	0.04	0.22 (0.04-0.84)	-	-
YP_001129507.1-157772-154725-2	BALF5	Early lytic	0.32	1.11	1.15	4 wk	IgA	-	-	3.762	0.216
YP_001129442.1-42204-42779	BHRF1	Early lytic	0.65	1.24	1.28	4 wk	IgA	-	-	2.669	0.195
YP_001129504.1-151808-150519	LF2	Early lytic	0.17	1.29	1.46	4 wk	IgG	-	-	2.222	0.314
YP_001129510.1-165796-162410-1	BALF2	Early lytic	0.95	1.20	1.21	4 wk	IgA	-	-	2.715	0.205
3-months post-EBVST immunotherapy											
YP_001129436.1-1026-1196	LMP2A	Latent	0.02	1.05	1.24	3 mo	IgG	0.04	0.13 (0.01-0.74)	-	-
YP_001129485.1-117754-118890	BGRF1/BDRF1	Late lytic	0.01	0.87	1.11	3 mo	IgG	0.02	0.17 (0.03-0.73)	2.773	0.216
YP_001129470.1-94844-96457	BRRF2	Late lytic	0.02	2.15	2.54	3 mo	IgG	-	-	-	-
YP_001129440.1-20824-20955	EBNA-LP	Latent	0.04	0.97	1.31	3 mo	IgG	-	-	3.159	0.253
YP_001129500.1-136454-135636	BVLF1	Late lytic	0.01	1.24	1.45	3 mo	IgG	-	-	-	-
YP_001129455.1-68964-70037	BMRF2	Glycoprotein	0.00	1.14	1.38	3 mo	IgG	-	-	-	-

YP_001129438.1-1736-5692-2	FGAM	Other/Unknown	0.03	1.35	1.65	3 mo	IgG	-	-	-	-
YP_001129496.1-131574-129454	BXLF2	Glycoprotein	0.02	1.35	1.54	3 mo	IgG	-	-	-	-
YP_001129486.1-115415-114405	BGLF2	Early lytic	0.02	1.16	1.39	3 mo	IgG	-	-	-	-
YP_001129448.1-49335-49865	BFRF3	Late lytic	0.03	2.37	2.70	3 mo	IgG	-	-	-	-
YP_001129484.1-113481-112483	BGLF3	Late lytic	0.01	1.31	1.58	3 mo	IgG	-	-	3.468	0.222
YP_001129436.1-360-458	LMP2A	Latent	0.00	1.09	1.31	3 mo	IgG	0.04	0.07 (0-0.6)	-	-
YP_001129436.1-540-788	LMP2A	Latent	0.01	0.94	1.15	3 mo	IgG	0.04	0.19 (0.03-0.88)	-	-
YP_001129439.1-9659-10171	BcRF1	Late lytic	0.02	1.12	1.31	3 mo	IgG	-	-	-	-
YP_001129498.1-133398-134144	BXRF1	Late lytic	0.02	1.12	1.33	3 mo	IgG	-	-	-	-
YP_001129489.1-117772-117539	BDLF3.5	Glycoprotein	0.03	0.91	1.11	3 mo	IgG	-	-	-	-
YP_001129436.1-871-951	LMP2A	Latent	0.00	1.10	1.39	3 mo	IgG	0.02	0.06 (0-0.48)	-	-
YP_001129461.1-76771-77259	BLRF2	Late lytic	0.01	2.18	2.61	3 mo	IgG	-	-	-	-
YP_001129488.1-117560-116883	BDLF4	Early lytic	0.03	1.48	1.75	3 mo	IgG	-	-	-	-
YP_001129506.1-154125-153187	BILF1	Glycoprotein	0.01	1.14	1.37	3 mo	IgG	-	-	-	-
YP_001129504.1-151808-150519	LF2	Early lytic	0.02	1.14	1.38	3 mo	IgG	-	-	-	-
YP_001129466.1-90630-89959	BZLF2	Glycoprotein	0.01	2.02	2.47	3 mo	IgG	-	-	-	-
YP_001129440.1-35441-35473	EBNA-LP	Latent	0.02	0.83	1.09	3 mo	IgG	-	-	3.991	0.225
YP_001129472.1-98500-98913	BKRF2	Glycoprotein	0.01	0.92	1.18	3 mo	IgG	0.05	0.19 (0.03-0.91)	-	-
YP_001129512.1-166530-167195	BARF1	Early lytic	0.02	1.13	1.31	3 mo	IgG	-	-	-	-
YP_001129479.1-107679-108896	BBRF3	Glycoprotein	0.01	1.53	1.97	3 mo	IgG	-	-	-	-
YP_001129454.1-67745-68959	BMRF1	Early lytic	0.04	1.98	2.36	3 mo	IgG	-	-	-	-
CAA24827.1-122341-120929	BGLF5	Early lytic	0.01	0.91	1.09	3 mo	IgG	-	-	-	-
AFY97906.1-168167-168081	LMP1	Latent	0.04	1.09	1.34	3 mo	IgG	0.04	0.16 (0.02-0.79)	-	-
CAA24829.1-124938-125915	BGRF1/BDRF1	Late lytic	0.03	0.92	1.15	3 mo	IgG	0	0.1 (0.02-0.44)	-	-
CAA24839.1-71527-62078-3	BPLF1	Late lytic	0.05	0.76	0.96	3 mo	IgG	-	-	-	-
AFY97988.1-166888-166706	BNLF2A	Late lytic	0.02	1.17	1.36	3 mo	IgG	-	-	-	-

CAA24860.1-102116-101445	BZLF2	Glycoprotein	0.01	1.95	2.43	3 mo	IgG	-	-	-	-
CAA24838.1-61507-62037	BFRF3	Late lytic	0.02	2.25	2.59	3 mo	IgG	-	-	-	-
AFY97924.1-49199-49729	BFRF3	Late lytic	0.03	2.27	2.59	3 mo	IgG	-	-	-	-
YP_401715.1-160908-158851	BALF3	Late lytic	0.01	1.23	1.46	3 mo	IgG	-	-	-	-
YP_001129474.1-99676-100329	BKRF4	Late lytic	0.13	0.91	1.08	3 mo	IgG	0.05	0.25 (0.06-0.94)	-	-
AFY97980.1-156149-153102-2	BALF5	Early lytic	0.12	1.17	1.30	3 mo	IgG	0.05	0.1 (0.01-0.7)	-	-
AFY97929.1-67486-68700	BMRF1	Early lytic	0.06	1.37	1.55	3 mo	IgA	-	-	3.335	0.28
YP_001129498.1-133398-134144	BXRF1	Late lytic	0.44	0.94	0.99	3 mo	IgA	-	-	5.765	0.316
YP_001129507.1-157772-154725-2	BALF5	Early lytic	0.07	1.12	1.19	3 mo	IgA	-	-	2.532	0.419

Chapter 5

Molecular biomarkers of latent and active tuberculosis in Papua New Guinea

This chapter describes the first study on whole-blood host transcriptomic changes associated with different tuberculosis (TB) infection states in Papua New Guinea (PNG). It reports the discovery of gene expression signatures which discriminate active TB from latent TB and from disease-free healthy individuals in Balimo, a remote community in the Western Province of PNG. The work presented in this chapter comprises a publication pending submission.

5 Molecular biomarkers of latent and active tuberculosis in Papua New Guinea

Yomani D. Sarathkumara[1], Jessica Scott[1,2], Daniel Pelowa[3], Robert Dowi[3], William Pomat[4], Catherine M. Rush[1,2], Jeffrey Warner[1,2], Carla Proietti[1*], Denise L. Doolan[1*]

[1] Centre for Molecular Therapeutics, Australian Institute of Tropical Health of Medicine, James Cook University, Cairns, QLD, Australia

[2] College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, QLD, Australia

[3] Balimo District Hospital, Balimo, Western Province, Western Province, Papua New Guinea

[4] Papua New Guinea Institute of Medical Research, Goroka, Eastern Highlands Province, Papua New Guinea

*** Corresponding Author:** Email: denise.doolan@jcu.edu.au (DLD); Carla.proietii@jcu.edu.au (CP)

[†] These authors contributed equally as last authors

Competing Interest Statement: The authors declare no competing interest.

5.1 Abstract

Tuberculosis (TB) is a major global public health threat causing millions of deaths and clinical cases worldwide. Papua New Guinea (PNG) is recognized as a country with one of the highest tuberculosis burdens in Oceania, with limited diagnostic facilities contributing to this epidemic. TB is mainly diagnosed by symptom-based clinical examinations. Therefore, asymptomatic individuals carrying latent TB infection often remain undiagnosed and confound estimates of the actual TB burden. This study aimed to identify whole-blood transcriptomic signatures associated with different TB infection states, using samples collected from Balimo, a remote region of the Western Province of PNG with a very high TB burden. Classification of TB exposure and disease categories was defined by clinical measures and by the interferon-gamma release assay (IGRA) to assess exposure to *Mycobacterium tuberculosis*. We studied active pulmonary TB (PTB) patients (culture and PCR+); clinically diagnosed TB patients undergoing anti-TB treatments (IGRA+); past clinical TB patients who had completed anti-TB treatments; individuals with latent TB (IGRA+), and healthy controls (IGRA-). We performed whole-blood RNA-sequencing (RNA-seq) and evaluated differentially expressed genes (DEGs) between study groups and assessed their biological and functional processes.

We identified 20 and 29 significant DEGs in active PTB compared to healthy controls and latent TB, respectively. Of these, six genes were common to both comparisons; one of these genes, Triggering Receptor Expressed on Myeloid cells-like 4 (*TREML4*), was downregulated, and the other six genes; *SEPT4, AC098613.1, CCDC144NL, CCRL2*, and *APOL4* were upregulated. We identified 14 genes as the signature of active PTB. Of these, *C1QA*, *C1QB*, *C1QC*and *ANKRD22* upregulation have been previously identified in active TB. Importantly, 16 novel genes were identified as the signature of latent TB compared to controls. Additionally, three distinct genes (*CDC42BPA*, *TTTY14*, *LINC02573*) were identified as differentially expressed in active PTB compared to latent TB. Gene ontology analysis showed DEGs found in active PTB were enriched in immunological responses, including host T-cell activation. Disease ontology revealed that these DEGs were associated with primary bacterial infections, including active and latent TB. Our results identified novel transcriptomic signatures that were specific to latent and active PTB and genes that discriminated active from latent TB. This foundational study provides important insights into host transcriptomic signatures for stratifying patients according to their TB status and improves our understanding of immune changes underlying TB infection from an ethnically diverse population in PNG.

Key words: Active PTB, Latent TB, Blood transcriptomics, RNA-seq

5.2 Introduction

Tuberculosis (TB) caused by the bacterium *Mycobacterium tuberculosis* (MTB) remains a major cause of long-term morbidity and mortality globally. It is the second leading infectious killer after SARS-CoV2 (1), responsible for an estimated 10.6 million cases and 1.6 million deaths worldwide in 2021 (2). TB is also the leading cause of deaths among individuals infected with HIV (1, 2). The incidence rate of TB has been increasing predominantly in many resource-limited countries, with a 3.6% increase between 2020 and 2021, which has been further exacerbated by the COVID-19 pandemic (3).

Papua New Guinea (PNG) is recognized as one of the highest TB burden countries in the Western Pacific region for the number of TB cases recorded (2). There were an estimated 42,000 TB cases, with 5,100 deaths in HIV-negative individuals and 370 deaths in HIV-positive individuals in 2021, as reported by the World Health Organization (WHO) (4). HIV-positive cases account for about 5-10% of TB cases in PNG (5). PNG is among the top 30 high burden countries accounting for TB and multi-drug/rifampicin-resistant TB (MDR/RR-TB), a strain that accounts for 86-90% of the annual estimated TB incidences worldwide (6). Furthermore, the TB burden is much higher in remote PNG communities than the national incident rates, with an estimated incidence of 1,290 per 100,000 people in the Gulf Province and about 550 per 100,000 in the Western Province (5, 7).

Pulmonary TB (PTB) is the most common clinical manifestation of TB, and the lungs are the predominant site of infection. A majority of the reported TB cases are PTB; however, the disease can be disseminated to other organs and sites of the body, which is then referred to as extrapulmonary TB (EP-TB) (8, 9). A recent nationwide epidemiological study in PNG reported that 42.4% of all TB cases notified in 2016 were EP-TB, and the proportion of PTB was 27.3% (10). A hospital-based paediatric study undertaken in the Madang Province reported that 384 individuals (52.3%) had EP-TB from 734 patients diagnosed with presumed TB (11). In another retrospective cohort study conducted in West Sepik Province of PNG between 2014-2016, 43.1% and 56.9% were clinically diagnosed as extrapulmonary and pulmonary TB cases, respectively. However, of all TB cases, only 26.5% were bacteriologically confirmed (12).

Most individuals infected with MTB remain asymptomatic but carry the pathogen with a risk of subsequent progression to clinical disease, a condition termed latent tuberculosis (LTB). One-third of the global population has LTB, and at least 10% of those will progress to an active case during their lifetime, with a high impact on morbidity and mortality (13).

Smear microscopy is the fundamental laboratory method for PTB diagnosis. It is often the only diagnostic test available in resource-limited settings to identify the presence of MTB from sputum samples. This technique is fast and inexpensive but has low sensitivity (14-16). Imaging techniques such as X-rays to identify PTB associated lung abnormalities and to evaluate extrapulmonary TB sites are preferable complementary clinical diagnostic tools recommended by the WHO (17). However, chest X-rays or other radiographic findings and imaging techniques are not usually available in resource-limited settings. Historically, bacterial culture is the gold standard method of MTB detection (14). The disadvantages of bacteriological confirmation by culture are that it is time-consuming, expensive, and technically complex (15). Conventional culture in solid medium requires a longer time (up to eight weeks), whilst liquid culture often uses mycobacteria growth indicator tubes (MGIT), with results taking about two weeks. Liquid culture systems are, however, more prone to contamination and have an increased chance of detecting a variety of other mycobacterial species, including non-tuberculous mycobacteria (NTM), and thus require differentiation and species identification (18). More recently, WHO-approved rapid diagnostic molecular methods such as the Xpert® MTB/RIF or MTB/RIF Ultra are increasingly being used to detect active TB (14).

Both the tuberculin skin test (TST) and the interferon gamma-release assays (IGRA) tests are used to test whether a person has latent TB infection by reporting the presence of specific T-cell responses (19). Unfortunately, neither TST nor IGRA test can accurately differentiate latent from active TB disease (20, 21). Hence, the current challenge in TB prevalent countries is identifying asymptomatic, latently infected individuals. Underdiagnosis or overdiagnosis of EP-TB and latent TB burden are common in resource-limited countries like PNG because the diagnosis is predominantly achieved by symptom-based screening and clinical examinations (22). Even if laboratory facilities are available, diagnosis of EP-TB still may be challenged by difficulties in obtaining clinical specimens from extrapulmonary sites however, detecting the presence of infection in asymptomatic people is more challenging (23, 24).

In some rural areas of PNG, the tuberculosis incidence is higher than the annual national incidence rate, and tuberculosis infections are highly misdiagnosed. The Balimo region, in the Middle Fly District of the Western Province, is known to have a high TB burden in the country (25). This area is typical of a rural, remote, tropical community with a population of about 40,000 people (26). TB incidence rate in this area is at 700-900/100,000 population. In Balimo, approximately 75% of all cases are EP-TB disease, with about 98% of the EP-TB cases on

treatment with no laboratory evidence of TB (X-rays are not available, not able to obtain a clinical sample frequently), whilst about 40% of PTB are on treatment with no laboratory confirmation (AFB negative microscopy or when no smear available for testing). The emergence of drug-resistant (DR-TB) has also been reported from Balimo (27). In Balimo, there are no epidemiological reports describing the latent TB burden.

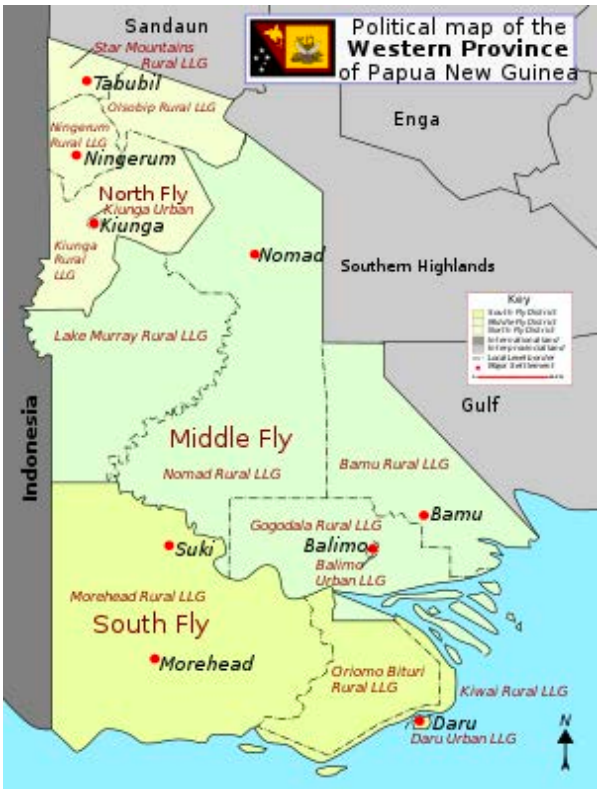


Figure 5-1 Western Province of Papua New Guinea (PNG)

The map showing Balimo region in the south-east of the Middle Fly District. (Image source: Wikimedia Commons).

The primary health facility in this area is Balimo District Hospital (BDH). It provides clinical services by a full-time health care worker and nursing staff, and diagnostic facilities and functions are only limited to smear microscopy available for TB diagnosis (25-27). Lack of roads, restricted vehicles, and primary mode of transport by boats and associated travel costs lead to challenges in accessing TB diagnosis and treatment in this isolated rural community (28). In the BDH hospital setting, TB diagnoses depend on presumptive diagnosis based on clinical investigations and presented symptoms, therefore the latent TB burden is unknown. BDH lacks the resources and laboratory capacity to facilitate culture and molecular techniques. Hence, the absence of diagnostic tools to identify actual TB cases, including latent TB,

contributes to the TB epidemic burden in the rural Balimo region and many other parts of the country (25, 26).

Recent studies on host transcriptomic mRNA signatures reveal transcriptomics as a promising platform for discovering biomarkers for TB diagnosis and progression (29). Several studies have revealed distinguishing gene expression of patients with active or latent TB patients compared to non-infected healthy controls (30-33), highlighting the potential to use transcriptomic signatures as biomarkers of different infection states. In particular, whole-blood transcriptomics analysis using RNA sequencing (RNA-seq) has recently been used to identify diagnostic and mechanistic immune signatures of both latent and active infection (34-38). The identification of biomarkers using such high-throughput techniques would more accurately discriminate the different infection status among the entire spectrum of MTB infections. Hence, it would be very beneficial for the development of specific diagnostic tools to identify individuals at risk, particularly in a high TB burden country where culture and other diagnostics tests are unavailable in remote regions.

The study of molecular immune signatures of TB through host transcriptomics can address several gaps in knowledge related to high TB burden in rural PNG. Firstly, identifying specific molecular signatures that are associated with TB can help in developing more accurate diagnostic tools and treatment strategies tailored to the unique context of PNG. Secondly, the genetic diversity of MTB in PNG is poorly understood. Therefore, host transcriptomics can provide insight into the interaction between the host and the bacterium and may help to identify genetic markers associated with MTB that in PNG. Overall, there is a lack of understanding regarding the immune responses to MTB in PNG. Host transcriptomics can provide a more comprehensive understanding of the immune response to MTB, including the identification of immune pathways that are associated or inhibited in response to TB infection. This information can help to develop strategies for improving the immune response to TB in PNG.

In the current study, we profiled the host transcriptome of individuals with a spectrum of MTB infectious states in Balimo. We compared the host transcriptional differences between groups to identify status-specific biomarker signatures. In particular, the study aimed at identifying a unique transcriptional profile of active PTB infection and latent infection in PNG.

5.3 Materials and Methods

5.3.1 Ethics statement

The study was conducted with the permission and support of the Middle Fly District Health Services and the Church Health Services. The study protocols were approved by the Medical Research Advisory Committee (MRAC) in PNG (MRAC No. 17.02) and the Human Research Ethics Committee, James Cook University (JCU) (H7696). Written informed consent was obtained from all study participants.

5.3.2 Subjects and samples selection

A total of 178 study participants were recruited over a two-week study period in January 2020 from the Balimo region in the Middle Fly District of the Western Province of PNG. Relevant information on participant medical history and demographic data were obtained using an interview-based questionnaire.

Blood samples were collected from all study participants, and individuals suspected of TB also provided a spontaneous sputum sample for TB culture. Whole-blood was collected by venepuncture into four QuantiFERON-TB Gold Plus (QFT®-Plus) blood collection tubes (1mL volume) (QIAGEN, Hilden, Germany) and a PAXgene® blood collection tube (2.5mL) (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland). QFT®-Plus blood collection tubes were incubated at 37°C at the Balimo District Hospital (BDH), transported at room temperature, and assayed at the Australian Institute of Tropical Health and Medicine, Townsville (AITHM) facility, Townsville. Whole-blood collected into PAXgene® collection tubes were transported at room temperature to Townsville and stored at -80°C until use.

Study subjects were classified into five groups based on the available clinical and laboratory diagnosis as determined by microbiological methods, PCR and IGRA. The five participant groups (**Figure 5-1**) were:

Group 1; Healthy - no history of evidence of MTB exposure (IGRA negative, asymptomatic with no symptoms of TB including fever, cough, or sputum production no clinical diagnosis), n=21

Group 2; Latent TB - evidence of MTB exposure (IGRA positive, asymptomatic, no clinical diagnosis), n=21

Group 3; Past clinical TB patients - clinically diagnosed TB and completed anti-TB treatment (IGRA positive, clinically diagnosed, treated), n=21

Group 4. Clinically active TB treatments - clinically diagnosed TB patients currently on treatment (IGRA positive, clinically diagnosed, ongoing treatment), n=16

Group 5. Active pulmonary TB (PTB) patients with clinical symptoms, laboratory confirmed, not yet treated (IGRA positive or negative, clinically diagnosed, *M. tuberculosis* culture positive and PCR positive, untreated), n=8

Blood RNA samples collected from all patients in these groups were randomized into different batches using randomizr (version 0.20.0) package in R prior to RNA extraction for RNA-seq analysis to ensure random distribution of samples amongst the study groups.

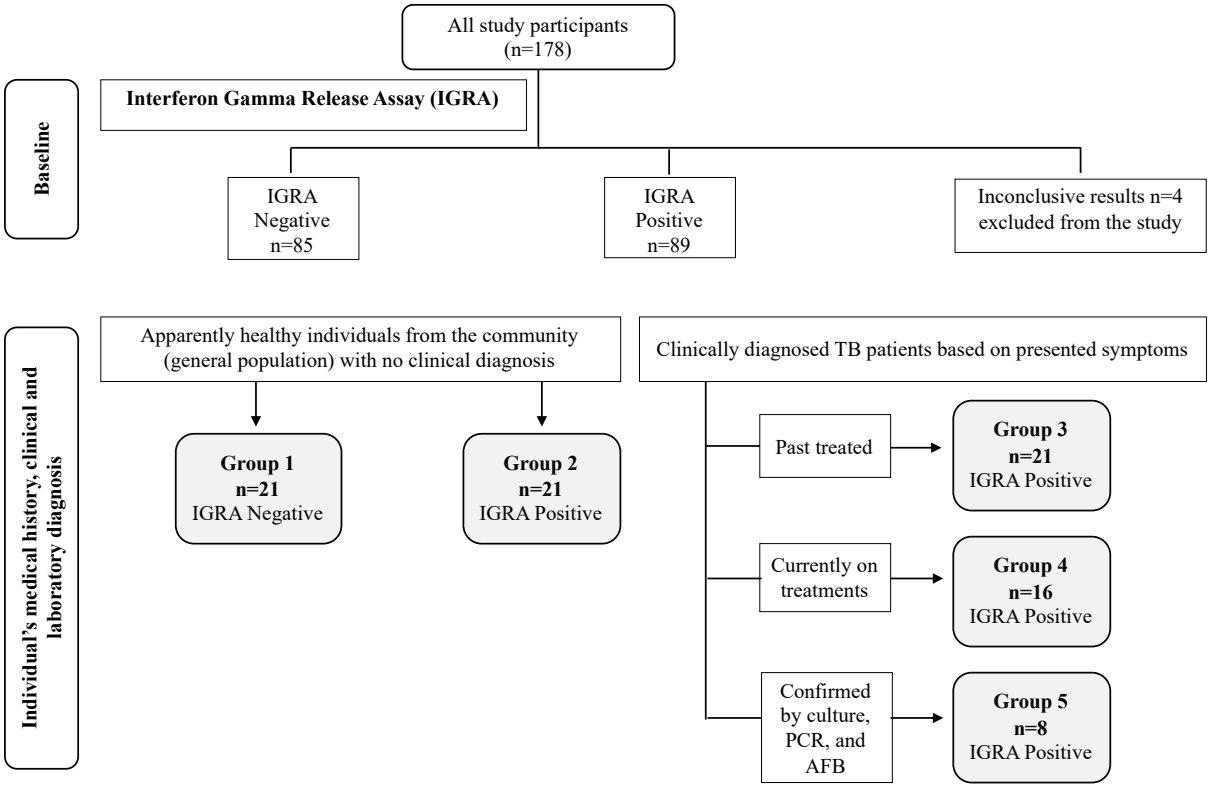


Figure 5-2. Overview of study design and subject selection.

All study participants were tested by Interferon Gamma Release Assay (IGRA). Groups were categorized according to the available lab and clinical diagnosis information. RNA-seq analysis was performed for blood RNA samples from individuals in Groups 1-5. PTB= pulmonary TB.

5.3.3 Interferon Gamma Release Assay (IGRA) - QFT®-Plus assay

QFT®-Plus blood collection tubes included a Nil tube (unstimulated negative control), TB1 and TB2 tubes (containing peptides from the MTB–complex–associated ESAT-6 and CFP-10 antigens to assess $CD4^+$ T cell responses, or both $CD4^+$ and $CD8^+$ T cell response, respectively) and a Mitogen tube (positive control) (39).

QFT®-Plus blood collection tubes following incubation, were centrifuged for 15 min at 3000×*g,* and the plasma was transferred into 1.5 mL microcentrifuge tubes, were then transported from BDH to Townsville and stored at -80°C. Plasma samples were thawed on ice, and interferon-gamma (IFN-γ) was assayed by ELISA (QFT®-Plus; Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions, using 50 µL volume from each Nil, TB1 Antigen, TB2 Antigen, and Mitogen sample. The optical density (OD) of each well was measured using the SpectroNano Star spectrophotometer plate reader (BMG Labtech, Ortenberg, Germany). OD values were analyzed using the QFT®-Plus Analysis Software (Version 2.71.2), which assess quality control of the assay, generates a standard curve, and provides a test result for each subject. Test results from the QFT®-Plus assay were considered positive when the IFN-γ concentration of the TB antigen tube (TB1 or TB2) minus Nil is ≥0.35 IU/mL and ≥25% of the Nil value. A negative result was defined as antigen–nil < 0.35 IU/mL or <25% of nil when mitogen ≥ 0.5 IU/mL. Results were considered indeterminate if nil > 8 IU/mL, or antigen–nil ≥ 0.35 IU/mL and < 25% of nil when the nil was ≤8.0 IU/mL and the mitogen response was <0.5 IU/mL (40).

5.3.4 Microbiological methods and confirmation of MTB by PCR

Sputum samples were inoculated onto Mycobacterial Growth Indicator Tube™ (MGIT) (Becton Dickinson, MD, USA) and Löwenstein-Jensen (LJ) solid culture media. MGITs were incubated at 37ºC for up to 7 weeks, whilst solid cultures were incubated for 8-10 weeks.

DNA was extracted from positive cultures using a commercially available High Pure PCR Template Preparation Kit (Roche Molecular Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. As per published protocols (41), two TaqMan real-time polymerase chain reaction (qPCR) assays were used to confirm *Mycobacterium* species (*IS6110* assay) or *M. tuberculosis* complex (MTBC) (*senX3-regX3* assay) on all DNA extracts. Samples that were reactive in both the *IS6110* and *senX3-regX3* assays were considered MTB-positive (27).

5.3.5 RNA Extraction

Total RNA from whole-blood samples collected in PAXgene® tubes was extracted using PAXgene® Blood RNA Kit (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland) according to the manufacturer's instructions which included DNase I treatment. Briefly, frozen blood specimens in PAXgene® tubes were incubated for a minimum of 2 hrs at room temperature (RT) before processing to allow complete lysis of blood cells. PAXgene® blood collection tubes were inverted several times and then centrifuged at 4800×*g* for 10 min at RT. The pellets were resuspended in 4 mL of RNase-free water by vortexing until the pellets were no longer visible. After washing with RNase-free water, the pellet was dissolved in 350 µL resuspension buffer and incubated with 300 µL binding buffer and 40 µL proteinase K for 10 min at 55°C in a shaker-incubator. The lysate was transferred into a PAXgene® shredder spin column and centrifuged at 18000×*g* for 3 min. The flow-through fraction was mixed with 350 µL ethanol and transferred to a PAXgene® RNA spin column. After washing the column with washing buffer 1, samples were incubated with 10 µL of DNase I in 70 µL of DNA digestion buffer (RDD) for 15 min. PAXgene® RNA spin columns were washed with washing buffer 2, and RNA was eluted with 40 µL of RNase-free water. Extracted total RNA was used for RNA quality control (QC) assessments and stored in a couple of aliquots at -80°C until further processing. The concentration of extracted total RNA was measured using NanoPhotometer® N60 (Implen, München, Germany) and RNA purity was assessed by the A260/A280 ratio.

5.3.6 RNA-seq library preparation, sequencing, and data pre-processing

The library preparation, sample QC and sequencing on total RNA samples were performed at the Australian Genomics Research Facility (AGRF, Melbourne, Victoria, Australia). RNA libraries were synthesized using the TruSeq Stranded Total RNA with Ribo-Zero Plus Library Prep Kit (Illumina Inc., CA, USA), followed by paired-end sequencing on a NovaSeq S4, 300 cycles to generate read lengths of 150bp.

Data were generated with the Illumina bcl2fastq version 2.20.0.422 pipeline. The FASTQ reads were assessed for a quality check using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). The data were also screened for any Illumina adapter/overrepresented sequences and cross-species contamination. The cleaned trimmed sequence reads were then aligned to the human reference genome, GRCh38 using

STAR version 2.7.0e (42) with default parameters generating BAM files required for downstream analyses.

5.3.7 Differential gene expression analyses

Differential gene expression (DGE) analysis was employed to define the changes in gene expression profiles between the different clinical groups, as follows:

- Active PTB patients without treatment versus healthy controls (IGRA-) (Group 5 versus Group 1)
- Active PTB patients without treatment versus latent TB (IGRA+) (Group 5 vs Group 2)
- Latent TB (IGRA+) patients vs healthy controls (IGRA-) (Group 2 versus Group 1)
- Clinically active TB patients under treatment vs healthy controls (IGRA-) (Group 4 vs Group 1)
- Past clinical TB vs healthy controls (IGRA-) (Group 3 vs Group 1)
- Clinically active TB patients under treatment vs healthy controls (IGRA-) (Group 4 vs Group 1)
- Active PTB patients without treatment versus latent TB (IGRA+) (Group 5 vs Group 2)
- Clinically active TB patients under treatment versus latent TB (IGRA+) (Group 4 vs Group 2)
- Past clinical TB patients versus latent TB (IGRA+) (Group 3 vs Group 2)
- Active PTB patients without treatment versus past clinical TB patients (Group 5 vs Group 3)
- Past TB versus clinically active TB patients under treatment (Group 3 vs Group 4)
- Active PTB patients without treatment versus clinically active TB patients under treatment (Group 5 vs Group 4)

ConsensusDE (version 1.12.0) (43) in R (version 4.1.2)/Bioconductor package was used to generate a summarised table containing read counts from all RNA-seq experiments from the BAM files generated using STAR using the buildSummarized function in consensusDE, with a sample table provided that described the experimental design.

The summarised experiment object generated from consensusDE was used for subsequent differential expression (DE) analysis using the R/Bioconductor DESeq2 package (version 1.34.0) (44). Briefly, we created a DESeqDataSet object using the matrix of counts

(summarised experiment) and metadata using a design formula indicating the design of the experiment by testing for the effect of different groups, controlling for the effect of the subject's sex. The pre-filtering method was applied to the data sets to filter low read counts and genes with less than 10 reads were removed from the analysis.

Differential gene expression analysis was performed using the DESeq function, generating all possible pairwise comparisons as listed above. Significantly differentially expressed genes (DEGs) were defined as those with an adjusted *p*-value <0.05 and log 2-fold change >1 between each pairwise comparison.

5.3.8 Gene ontology and functional enrichment analyses for DEGs

Gene Ontology (GO) enrichment analysis was performed on the list of DEGs to demonstrate biological and functional processes associated with our defined clinical groups. The clusterProfiler package (version 4.2.2) was used to inform the function of DEGs and their biological process (45). Disease Ontology (DO) annotates and integrates DEGs associated with human diseases or conditions (46). DO enrichment analysis was therefore performed to identify the DEGs related to various diseases using the R/Bioconductor package DOSE (disease ontology semantic and enrichment analysis) (version 3.20.1) (47). In addition, the Metascape (https://metascape.org/gp/) online tool (48) was used for enrichment analysis of the DEGs, and Metascape's DisGeNET (https://www.disgenet.org/) (49) database used for disease ontology.

5.4 Results

5.4.1 Clinical and demographic characteristics of the study population

The study design is summarized in **Figure 5-1**. Of the 178 study participants from Balimo, Western Province, PNG, 85 individuals (47.75%) had a positive IGRA test. For RNA-seq analysis, we selected 87 individuals classified into five clinical groups based on available clinical and laboratory data (**Figure 5-1**). We randomly selected individuals from the community matched on sex and age (+/- 5 years) as healthy controls (Group 1, n=21, IGRA-) and latent TB (Group 1, n=21, IGRA+). Clinical data informed categorization of individuals who were previously clinically diagnosed TB patients who had completed anti-TB treatment (Group 3, n=21, IGRA+) or clinically diagnosed TB patients currently undergoing TB treatment (Group 4, n=16, IGRA+). Finally, there were eight individuals (Group 5, IGRA+/-) who were positive by sputum MTB culture and confirmed by PCR to have an active PTB infection. The demographic, clinical and laboratory features of these 87 study subjects in the 5 different groups are reported in **Table 5-1**.

5.4.2 Differential gene expression profiles in blood RNA between clinical groups

Active PTB patients without treatment versus healthy controls (IGRA-)

First, we evaluated differences in gene expression profiles between active PTB patients (Group 5) *vs*. IGRA- healthy controls (Group 1, IGRA-) to identify a gene signature for active PTB. We identified 20 significantly differentially expressed genes (DEGs); of those 16 genes were upregulated and 4 genes were downregulated in active PTB (Group 5) (**Figure 5-2A**, listed in **Supplement Table 1**). We identified 14 significantly DEGs as a signature of active PTB compared to healthy controls (**Figure 5-3, Table 5-2**). Among these were upregulated complement genes *C1QA*, *C1QB* and *C1QC* associated with active TB, suggesting a role for complement during active TB infection (50) and *ANKRD22* (ankyrin repeat domain 22), upregulation is suggested the enhanced inflammatory response in TB (51, 52).

Active PTB patients without treatment versus latent TB (IGRA+)

A total of 29 genes were identified as significantly differentially expressed between active PTB (Group 5) and latent TB (Group 2, IGRA+); of these 7 DEGs were upregulated and 22 were downregulated (**Figure 5-2B**, **Supplement Table 2**).

Collectively, there were 20 and 29 genes significantly differentially expressed in the active PTB cases compared to healthy control and latent TB groups, respectively. As per the Venn diagram, there were six overlapping genes between active PTB vs healthy controls and active PTB vs latent TB (**Figure 5-3)**, five of these genes; namely, *SEPT4, AC098613.1, CCDC144NL, CCRL2*, and *APOL4* were upregulated. Interestingly, *TREML4* (Triggering Receptor Expressed on Myeloid cells-like 4), known to modulate inflammatory responses, was downregulated in both comparisons (active PTB *vs*. healthy controls or latent TB) (**Figure 5-3, Table 5-2**) (53). Additionally, three distinct genes (*CDC42BPA*, *TTTY14*, *LINC02573*) were found were differentially expressed (upregulated or downregulated) in active TB compared to latent TB.

Latent TB (IGRA+) patients vs healthy controls (IGRA-)

We identified 36 significantly DEGs (29 upregulated and 7 downregulated) in latent TB (Group 2) when comparing latent TB with healthy controls (Group 1) (**Figure 5-2C**, listed in **Supplement Table 1**). We found 20 overlapped significant DEGs between active PTB vs latent TB and latent TB vs healthy controls (**Figure 5-3, Table 5-2)**. However, none of these significant genes were previously identified in the context of TB diagnosis or progression. There were 16 novel significantly DEGs in latent TB compared to healthy controls (Group 1) (**Figure 5-3, Table 5-2**).

Clinically active TB patients under treatment vs healthy controls (IGRA-)

A total of 60 significant DEGs were identified when comparing clinically diagnosed TB patients on anti-TB treatments (Group 4, IGRA+) with healthy controls (Group 1, IGRA-) and 56 of these genes were downregulated in clinical TB patients (**Figure 5-2D**, listed in **Supplement Table 4**).

Past clinical TB vs healthy controls (IGRA-)

There were no significant DEGs when comparing past clinically diagnosed TB (Group 3, IGRA+) with healthy controls (Group 1, IGRA-) (**Figure S1D**).

Other pairwise comparisons that were performed showing DEGs results inconsequential from a diagnostic perspective, are shown in the volcano plots in **Figure S1**.

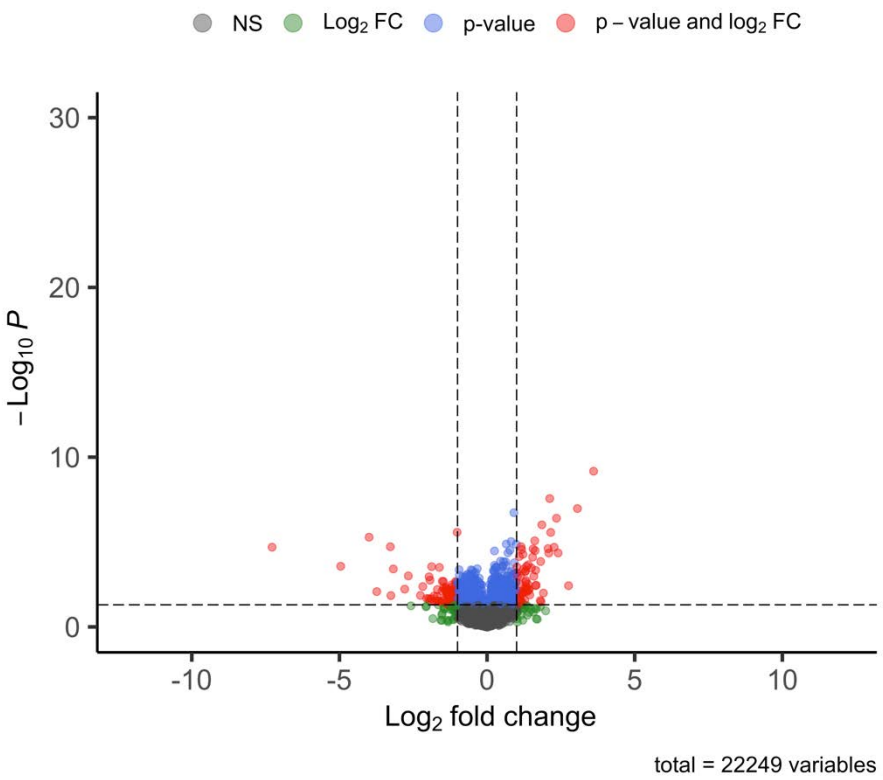
Table 5-1 Demographic and clinical data of the study cohort

	G1. Healthy Controls (n=21)	G2. Latent TB (n=21)	G3. Past clinical TB patients (n=21)	G4. Current clinical TB patients (n=16)	G5. Active PTB patients (n=8)
Sex, n (%)					
Female	15 (17.42)	14 (66.67)	12 (57.14)	7 (43.75)	7 (87.50)
Male	6 (28.57)	7 (33.33)	9 (42.85)	9 (56.25)	1 (12.50)
Age mean (range)	49 (16-72)	48 (18-68)	45 (25-58)	41 (17-63)	48 (18-63)
IGRA, n (%)					
Positive	-	21 (100)	21 (100)	16 (100)	6 (75)
Negative	21 (100)	-	-	-	2 (25)
Culture, n (%)					
Positive	-	-	-	-	8 (100%)
Negative	-	-	-	-	-

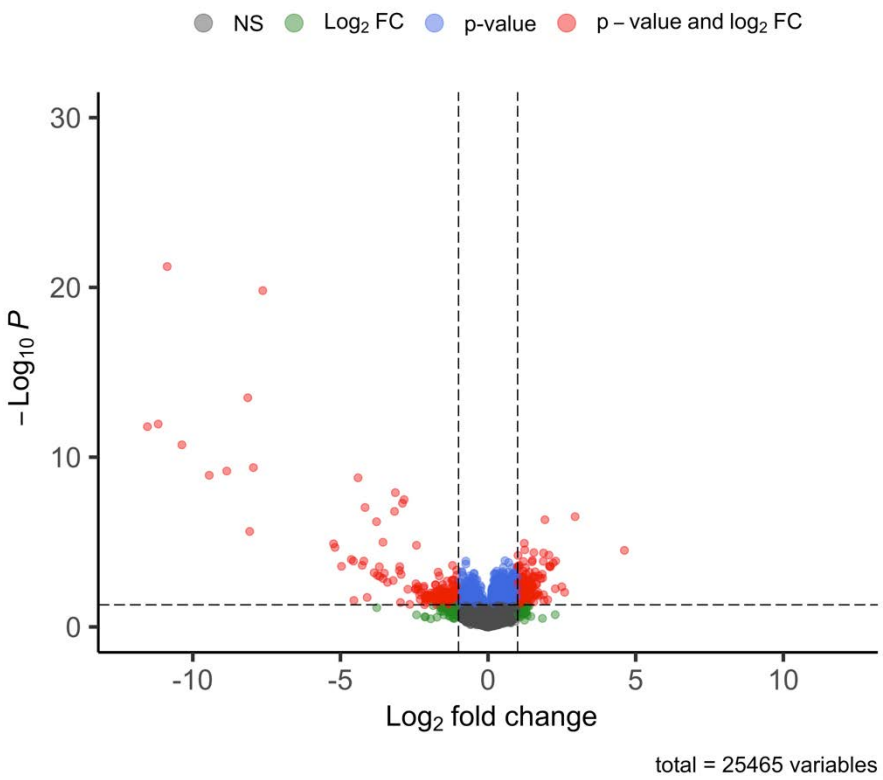
TB, Tuberculosis; IGRA, Interferon Gamma Release Assay

B

A **Active PTB vs. Healthy controls**



Active PTB vs. Latent TB



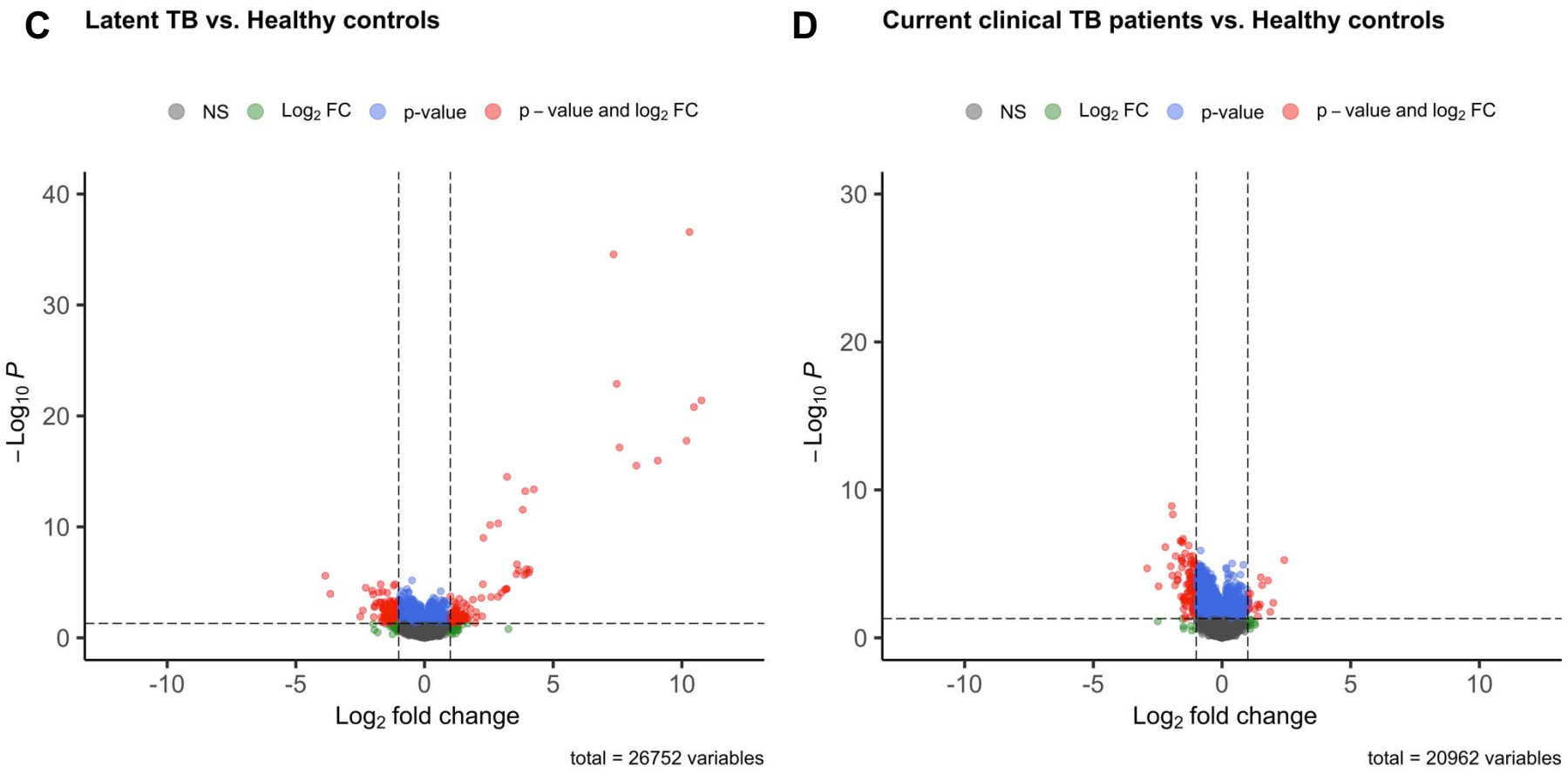


Figure 5-3 Volcano plots showing differentially expressed genes between groups.

The distribution of gene expression fold changes in; (A) active PTB relative to healthy controls (IGRA-), (B) active PTB relative to latent TB (IGRA+), (C) Latent TB (IGRA+) relative to healthy controls (IGRA-), and (D) clinically diagnosed TB patients (IGRA+) relative to Control IGRA-. Significantly expressed genes with log 2-fold change >1 and p-value ≤ 0.05 are shown in red. The vertical lines correspond to 1.0-fold up and down, and the horizontal line represents a p-value of 0.05. NS= non-significant, FC= fold change.

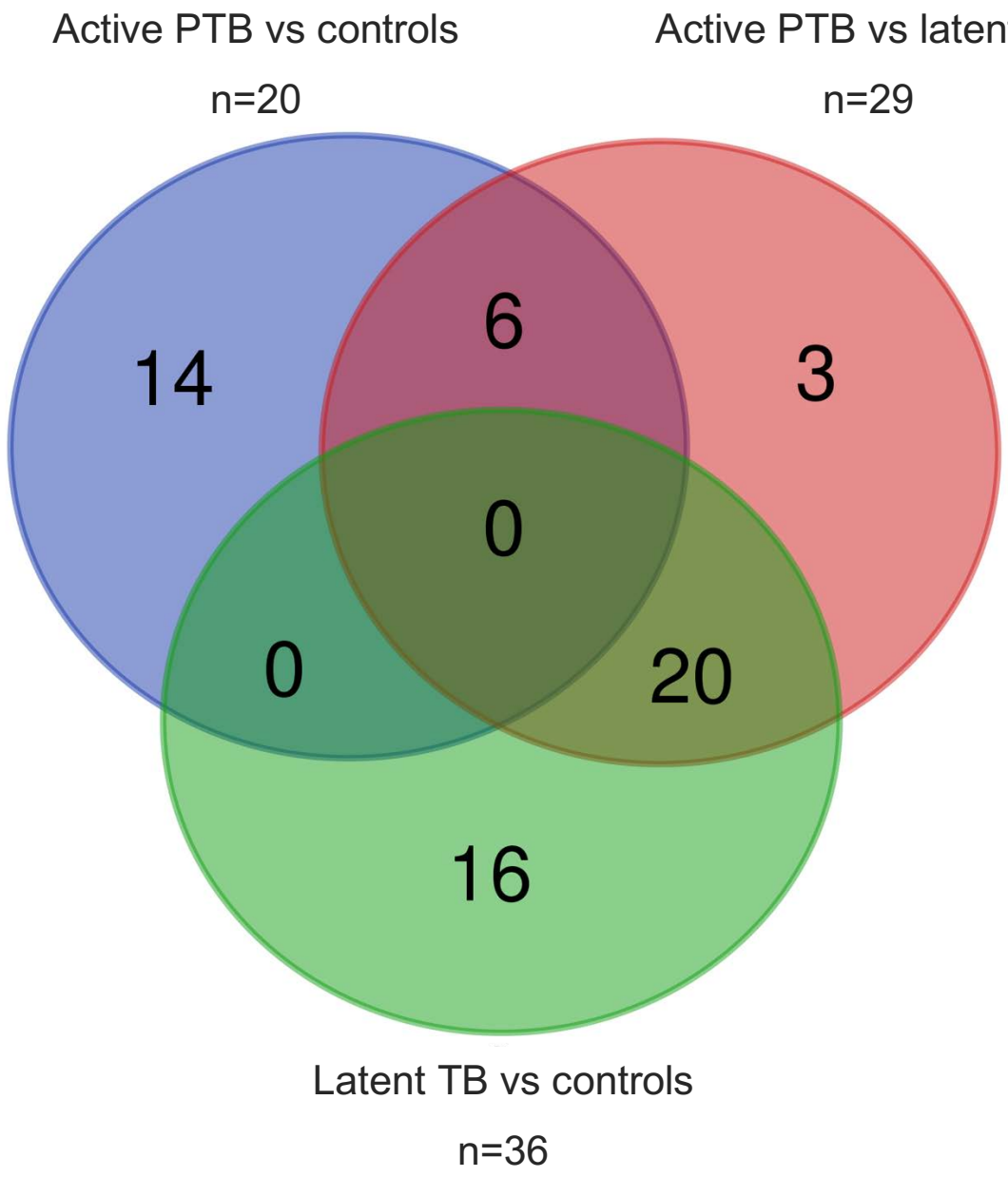


Figure 5-4 Venn diagram of significantly differentially expressed genes identified in the three pair-wise comparisons.

Significantly DEGs between active PTB vs. healthy controls IGRA- (n=20), active PTB vs. latent TB IGRA+ (n=20), and Latent TB vs. healthy controls IGRA- (n=36) and overlap between these groups are demonstrated.

Table 5-2 Summary of significantly differentially expressed genes identified in the three pair-wise comparisons [a]

Name of the comparison	Total no. of genes	Gene symbol
Active PTB vs healthy controls and Active PTB vs latent TB	6	APOL4, CCDC144NL, TREML4, AC098613.1, CCRL2, SEPT4
Active PTB vs latent TB and Latent TB vs healthy controls	20	LINC00278, DDX3Y, KDM5D, TXLNGY, BCORP1, FAM224B, AC010889.1, USP9Y, RPS4Y1, TTTY10, AL121872.1, TTTY15, ANOS2P, EIF1AY, ZFY, AL139042.1, PRKY, NLGN4Y, FAM224A, UTY
Active PTB vs healthy controls	14	AL512770.1, C1QA, AC004053.1, AL137005.1, RMI2, SLC30A8, LINC02470, PDCD1LG2, C1QB, ANKRD22, AP000662.1, ALMS1P1, CDCP1, C1QC
Active PTB vs latent TB	3	CDC42BPA, TTTY14, LINC02573
Latent TB vs healthy controls	16	KDM5DP1, PRY, IGKV3D-7, TRPC6, AC010086.1, GYG2P1, PRY2, PPP1R17, PRYP4, STAC, AC007431.1, GGT5, PRYP3, AC010737.1, TMSB4Y, ZFY-AS1

a. Significantly DEGs between active PTB *vs*. healthy controls IGRA- (n=20), active PTB *vs*. latent TB IGRA+ (n=20), and Latent TB *vs*. healthy controls IGRA- (n=36) and overlap between these groups from Venn diagram are listed in this table.

5.4.3 Biological relevance of the differentially expressed genes

GO analysis and DO analysis assessed the biological relevance of the genes that were significantly differentially expressed between the different clinical groups.

Active PTB patients without treatment versus healthy controls (IGRA-)

GO analysis by clusterProfiler revealed that the DEGs between active PTB (Group 5) compared with healthy controls (Group 1) were mainly enriched in antigen processing and presentation of peptide antigen, regulation of defence response to virus by host and T-cell activation via T-cell receptor contact with antigen bound to MHC molecule on the antigen-

presenting cell (**Figure 5-4A**). More importantly, Metascape's enrichment analysis, based on the information in the DisGeNET database for gene-disease/disorder, showed that the top 100 DEGs for immunological functions were associated with active and latent tuberculosis (**Figure 5-4B**).

Active PTB patients without treatment versus latent TB (IGRA+)

DEGs identified in the comparison between active PTB patients (Group 5) with the IGRA+ controls (Group 2) were mainly enriched in immunological responses to virus by the host and demonstrated gamma-delta T-cell activation (**Figure 5-4C**). DO of the DEGs associated diseases suggested that these DEGs were closely linked with primary bacterial infectious diseases, including tuberculosis (**Figure 5-4D**).

Latent TB (IGRA+) patients vs healthy controls (IGRA-)

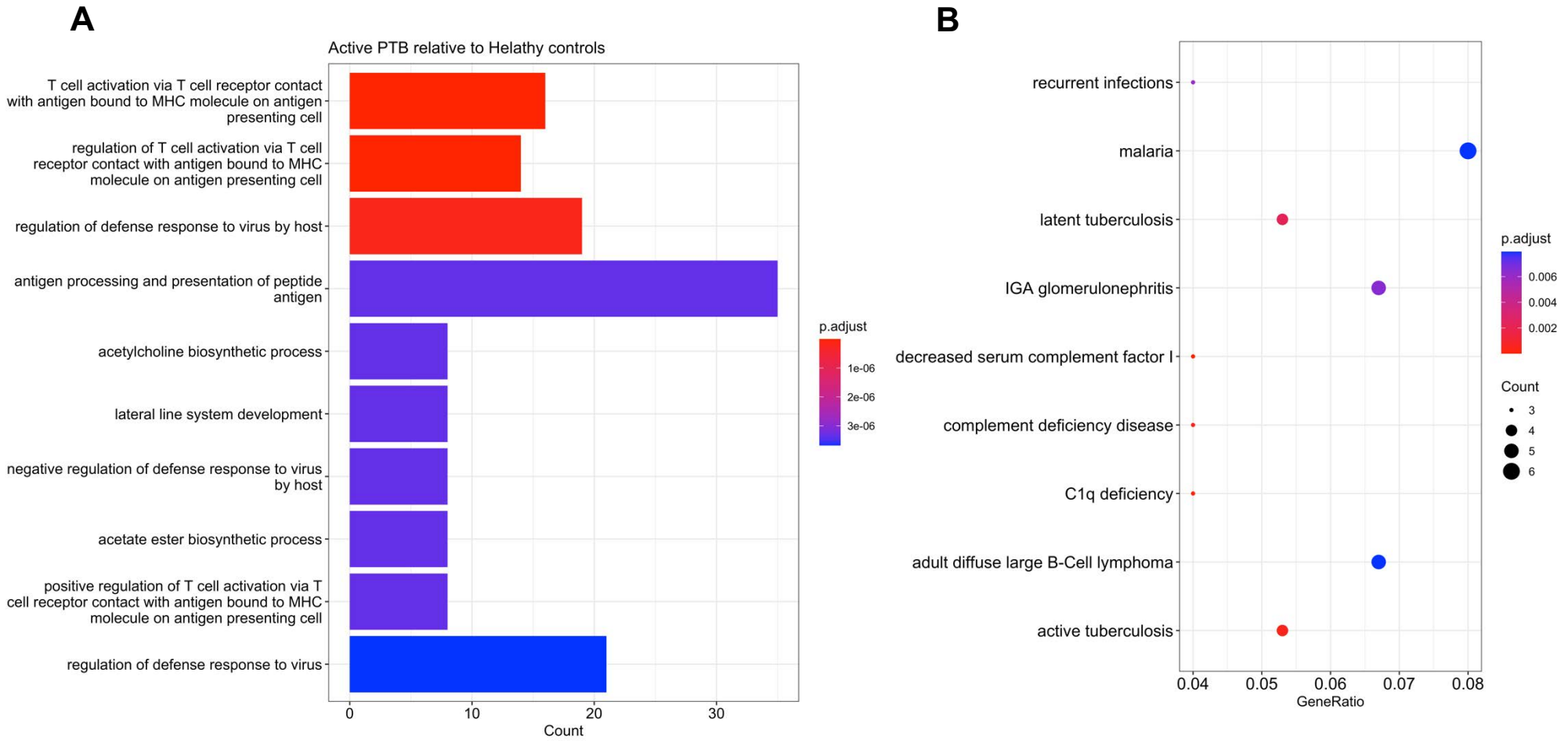
DEGs obtained from the comparison between latent TB IGRA+ (Group 2) and healthy controls IGRA- (Group 1), were not associated with any immunological pathway or any related diseases (**Figure S3**).

Clinically active TB patients under treatment vs healthy controls (IGRA-)

DEGs identified in the comparison between IGRA+ clinically diagnosed TB patients (Group 4) with IGRA- (Group 1) controls were enriched in functional metabolic activities (**Figure S2A**). However, the DO of the DEGs suggested a link between the genes with multiple diseases, including viral, bacterial and immunodeficiencies. (**Figure S2B**).

5.4.4 Similarities of biomarker genes of TB from other study populations

As shown in **Table 5-3**, we herein identified genes that were previously reported in gene expression analysis studies on TB diagnosis and progression. *SEPT4, ANKRD22*, *APOL4*, *PDCD1LG2*, *CCRL2*, *C1QA*, *C1QB* and *C1QC* were highly expressed in our active PTB group from PNG that have been identified in previous studies on TB gene expression analysis from various study populations (**Table 5-3)**. *SEPT4, CCRL2* and *APOL4* were found in the six overlapping genes in active PTB *vs*. latent TB or healthy controls (**Table 5-2**), although there are no significant signatures specific to latent TB in our study that have been previously associated with latent TB.


A
Active PTB relative to Helathy controls
T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
regulation of T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
regulation of defense response to virus by host
antigen processing and presentation of peptide antigen
acetylcholine biosynthetic process
lateral line system development
negative regulation of defense response to virus by host
acetate ester biosynthetic process
positive regulation of T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
regulation of defense response to virus
0
10
20
30
Count
p.adjust
1e-06
2e-06
3e-06
B
recurrent infections
malaria
latent tuberculosis
IGA glomerulonephritis
decreased serum complement factor I
complement deficiency disease
C1q deficiency
adult diffuse large B-Cell lymphoma
active tuberculosis
0.04
0.05
0.06
0.07
0.08
GeneRatio
p.adjust
0.006
0.004
0.002
Count
3
4
5
6

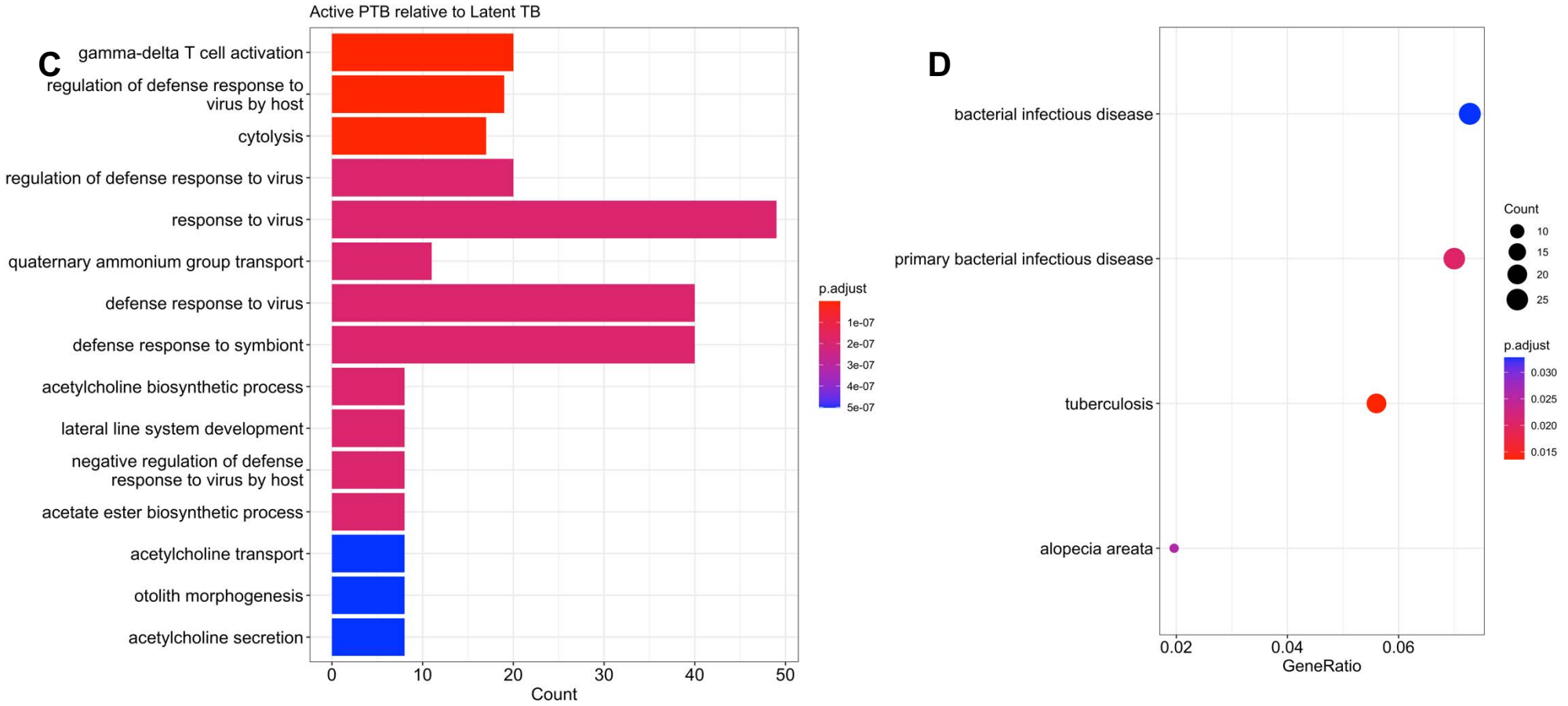


Figure 5-5 Functional enrichment analysis; gene and disease ontology of DEGs.

(A) Bar graph of the top 10 GO terms for the DEGs based biological process (BP) ordered by statistical significance, **(B)** Dot plot showing DO processes for the top DEGs of the relevant diseases between active PTB (Group 5) compared to the healthy controls (Group 1, IGRA-). **(C)** Bar graph of the top 15 GO terms for the DEGs based on BP ordered by statistical significance, **(D)** Dot plot showing DO processes for the top DEGs related to active PTB (Group 5) compared to the latent TB (Group 2, IGRA+). Colours indicate the Benjamini–Hochberg corrected *p*-value, and dot sizes indicate the number of genes contributing to the enrichment of the term.

Table 5-3 Previously identified genes of interest for TB diagnosis and progression

Gene Symbol	Name	Function	Target for	Study
C1QA	Complement C1q A chain	Complement system, host pathogen interaction, innate immune system	TB diagnosis	Petrilli et al. 2020 (54), Cai et al. 2014 (50),
C1QB	Complement C1q B chain	Complement system, host pathogen interaction, innate immune system	TB diagnosis and TB progression	Petrilli et al. 2020 (54), Cai et al. 2014 (50), Gliddon et al. (55), Blankley et al. 2016 (56), Bloom et al. 2013 (57)
C1QC	Complement C1q C chain	Complement system, host pathogen interaction, innate immune system	TB diagnosis	Cai et al. 2014 (50), Suliman et al. 2018 (58), Lubbers et al. 2018 (59)
SEPT4	Septin 4	Diverse cellular functions	TB progression	Suliman et al. 2018 (58), Zak et al. 2016 (35)
ANKRD22	Ankyrin repeat domain containing protein 22	Protein-protein interaction domain	TB diagnosis and TB progression	Bloom et al. 2013 (57), Suliman et al. 2018 (58), Zak et al. 2016 (35), de Araujo et al. 2016, (60), Natarajan et al. 2022 (51),

APOL4	Apolipoprotein L4	Play a role in lipid exchange and transport throughout the body	TB diagnosis	de Araujo et al. 2016, (60)
PDCD1LG2	Programmed cell death 1 ligand 2	Adaptive immune system, cell adhesion and lymphocyte activation	TB diagnosis	Petrilli et al. 2020 (54), Liang et al. 2021(61)
CCRL2	C-C Motif Chemokine Receptor Like 2	Chemokine signalling	TB progression	Petrilli et al. 2020 (54)

5.5 Discussion

Diagnostic uncertainty and clinical presentation mimicking TB is a significant feature contributing to the global TB epidemic. The identification of diagnostic signatures for TB that are additionally capable of distinguishing between the different states of infection (*i.e.*, active versus latent) would be of enormous value to improve public health.

A number of reports have identified putative host blood transcriptomic signatures for TB using various next-generation RNA-sequencing approaches to inform diagnosis and understanding of TB infection (30, 33, 62-64). To our knowledge, no host transcriptomic studies have been conducted in Papua New Guinea (PNG) despite that the country is an epicentre of TB. Hence, the current foundational study was undertaken to improve our understanding of TB and particularly in PNG by identifying changes in the host whole-blood transcriptome from presumptive TB patients, laboratory confirmed PTB, latent TB and healthy controls residing in the rural Balimo region of the Western province of PNG. In this region, there is a lack of basic bacteriology services and proper laboratory diagnosis required for differential diagnosis of tuberculosis (26).

In the current study, we profiled the transcriptomes of active PTB patients, clinically diagnosed TB patients on anti-TB treatments (IGRA+) compared to healthy (IGRA-) and latent TB (IGRA+) individuals from the community. We stratified study participants into clinical groups based on their clinical diagnosis, medical history, laboratory diagnosis by culture and PCR confirmation and according to their IGRA test results. We identified differentially expressed genes between the TB patient groups and healthy control groups. Importantly, we identified a host blood transcriptomic signature that differentiated individuals with active PTB from those with latent disease.

We identified 20 and 29 genes significantly differentially expressed in active PTB patients compared to healthy controls and latent TB, respectively. We found six overlapping genes, five of the genes were upregulated whilst *TREML4* was downregulated between both comparisons (active PTB *vs*. healthy controls or latent TB). We found 14 unique significant genes as the signature of active TB patients compared to healthy controls in this study cohort. Among these 14 genes signature, complement genes (*C1QA*, *C1QB* and *C1QC*) and *ANKRD22* upregulation have been identified previously in active TB disease and progression (**Table 3**).

Previous studies have shown complement component C1q as a biomarker to detect active TB infection (50, 59). It has also been indicated as a marker for differential diagnosis as C1q levels are highly elevated among TB patients compared to patients with sarcoidosis, leprosy,

and pneumonia (59). Subsequently, three distinct genes (*CDC42BPA*, *TTTY14*, *LINC02573*) were found were differentially expressed (upregulated or downregulated) in active TB compared to latent TB.

In latent TB compared to healthy controls, 36 significantly DEGs were identified. However, 20 of these genes were overlapped between active PTB *vs.* latent TB and latent TB *vs.* healthy controls comparisons. None of these significant genes were previously reported, thus we identified 16 novel genes (10 upregulated, 6 downregulated) as a signature of latent TB compared to healthy controls in PNG. Importantly, we found no significant DEGs when comparing past clinically diagnosed TB (Group 3, IGRA+) with healthy controls suggesting that these patients may have completely cleared infections after anti-TB treatment regimens.

Analysis of the DEG also uncovered specific immunological functions and pathways associated with each group that further improved our understanding of TB. Gene ontology and enrichment analyses of transcriptional profiles from active PTB patients differed from both latent TB and control samples and revealed that these genes were major contributors in immune responses such as antigen processing and presentation of peptide antigen, T-cell activation via T-cell receptor contact with antigen bound to MHC molecule on antigen-presenting cell, gamma-delta T-cell activation and regulation of defence response to virus by host. Moreover, active PTB patients appeared to elicit immune responses that corroborated ongoing infection.

On the other hand, between clinically diagnosed TB patients and IGRA- healthy controls, most DEGs appeared to mediate chemical reactions and pathways resulting in metabolic processes and inflammation. In contrast, IGRA+ latent TB and IGRA- healthy disease-free control individuals demonstrated pathways involving metabolic processes.

To investigate the gene-disease associations, we applied DO analysis to extract disease relevance from DEGs. DO analyses identified signatures of primary bacterial infectious diseases, including TB, in the active PTB group compared to latent TB and IGRA- healthy control groups. Thus, these results from DO confirm our bacterial culture and PCR confirmation of active PTB patients. We found no immune related functional genes other than metabolic pathways in latent TB compared to healthy controls.

Interestingly, DO analysis of clinically diagnosed TB patients revealed patterns also attributed to a number of other infectious diseases, including viral, bacterial, fungal, and autoimmune diseases. This may be due to enrichment of immune functions upregulated or downregulated shared across multiple diseases. Alternatively, due to the widespread of tropical diseases prevalent in Balimo region, patients mimicking TB-like symptoms at the time of clinical examinations, may have been misdiagnosed. PNG has been recognized for having the

highest prevalence for a larger number of tropical infectious diseases (65). Our findings further highlight that disease diagnostic uncertainty is a major challenge for estimating the actual TB burden in the Balimo region and facing the risk of an emerging drug-resistant TB epidemic.

Finally, and of high relevance to the field and laboratory, our data suggests that IGRA assays are acceptable and imperfect screening tests to diagnose latent TB (19, 20). Since that assay is based on an immune response to MTB antigens, a positive IGRA cannot accurately differentiate subclinical or active TB, and individuals with IGRA+ results do not progress to active TB disease (66-68).

Although our results demonstrate statistical differences between the different study groups, the results should be interpreted with caution since there are a number of limitations to our study that may have influenced outcomes. A major limitation was that most TB patients were clinically diagnosed based on their clinical presentation only and no laboratory confirmation was available in their health records. We also have no confirmation that the past TB patients who had received anti-TB treatment had cleared the infection completely. Our study was also limited by the relatively small sample sizes in the defined groups, emphasizing the importance of follow-up studies in larger cohorts with definitive TB diagnosis. It highlights the importance of employing basic bacteriology facilities at BDH and improving differential diagnosis of tropical infections and other immune related diseases in this region We also lacked information on BCG vaccination status, HIV status, autoimmune diseases, exposure to other NTM species and tropical infections in these individuals. In fact, the impact of these factors is considered to influence TB-associated host immune responses (34).

Other factors that are particularly relevant in a remote setting like Balimo include suboptimal logistics associated with sample collection, storage, and transportation for the laboratory facilities in Australia.

In summary, herein we comprehensively analyzed transcription profiles associated with different TB infection states in Balimo, a rural tropical region of PNG. Our study revealed heterogeneous blood transcriptomic profiles between individuals with active PTB or latent TB and healthy controls from the community. We identified 14 significant genes as the signature of active TB and 16 novel genes as the signature of latent TB compared to healthy controls. There were six overlapping genes between both comparisons (active PTB vs. healthy controls or latent TB). Three distinct genes were also identified in active TB compared to latent TB. However, additional work to identify the optimal minimal gene signature and subsequent validation in distinct cohorts is required.

Future work could build on the results reported herein, including modular and pathway analysis using Ingenuity Pathway Analysis of transcriptional data to identify the biological function of the significant modules and the relationship between modules and different clinical groups as the foundation for biomarker-discovery strategy. Additionally, our study results could be validated in publicly available applicable datasets on TB such as, Zak et al 2016 (35), Suliman et al. 2018 (58) and Petrilli et al. 2020 (54). In addition, our latent TB individuals could also be stratified based on the expression of previously published gene signatures of progressors to active TB infection.

In summary, this foundational study provides important insights into host transcriptomic signatures for stratifying patients according to their TB status and furthers our understanding of immune changes underlying TB infection from an ethnically diverse population in PNG. Importantly, our results identify putative transcriptional signatures of active TB and of latent TB in PNG which may have the ability to generalize between ethnically and geographically diverse population (*i.e*., between PNG and Africa).

5.6 References

1. Global tuberculosis report 2021 2021 [cited 2022 October 31]. Available from: https://www.who.int/publications/i/item/9789240037021.
2. Global tuberculosis report 2022 2022 [cited 2022 October 31]. Available from: https://www.who.int/publications/i/item/9789240061729.
3. Stop TB Partnership sounds the alarm on the TB funding crisis as TB deaths increase for a second year in a row and TB incidence rises for the first time in 20 years 2022 [cited 2022 25 Nov]. Available from: https://www.stoptb.org/news/stop-tb-partnership-sounds-alarm-tb-funding-crisis-tb-deaths-increase-second-year-row-and-tb#:~:text=The%20TB%20incidence%20rate%20(new,of%20the%20past%202%20decades.
4. Tuberculosis profile: Papua New Guinea Geneva2021 [cited 2022 October 31]. Available from: https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22PG%22.
5. McBryde E. Evaluation of risks of tuberculosis in Western Province Papua New Guinea. 2012.
6. WHO releases new global lists of high-burden countries for TB, HIV-associated TB and drug-resistant TB 2021 [cited 2022 October 31]. Available from: https://www.who.int/news/item/17-06-2021-who-releases-new-global-lists-of-high-burden-countries-for-tb-hiv-associated-tb-and-drug-resistant-tb.
7. Cross GB, Coles K, Nikpour M, Moore OA, Denholm J, McBryde ES, et al. TB incidence and characteristics in the remote gulf province of Papua New Guinea: a prospective study. BMC Infectious Diseases. 2014;14(1):1-10.
8. Lee JY. Diagnosis and treatment of extrapulmonary tuberculosis. Tuberculosis and Respiratory Diseases. 2015;78(2):47-55.
9. Transmission and Pathogenesis of Tuberculosis. Core curriculum on tuberculosis. 6 th ed. Atlanta: Centers for Disease Control and Prevention National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of Tuberculosis Elimination; 2013.
10. Aia P, Wangchuk L, Morishita F, Kisomb J, Yasi R, Kal M, et al. Epidemiology of tuberculosis in Papua New Guinea: analysis of case notification and treatment-outcome data, 2008–2016. Western Pacific Surveillance and Response Journal. 2018;9(2):9.

11. Watch V, Aipit J, Kote-Yarong T, Rero A, Bolnga JW, Lufele E, et al. The burden of presumed tuberculosis in hospitalized children in a resource-limited setting in Papua New Guinea: a prospective observational study. International health. 2017;9(6):374-8.
12. Kelebi T, Takarinda K, Commons R, Sissai B, Yowei J, Gale M. Gaps in tuberculosis care in West Sepik Province of Papua New Guinea. Public Health Action. 2019;9(Suppl 1):S68.
13. Getahun H, Matteelli A, Abubakar I, Aziz MA, Baddeley A, Barreira D, et al. Management of latent *Mycobacterium tuberculosis* infection: WHO guidelines for low tuberculosis burden countries. European Respiratory Journal. 2015;46(6):1563-76.
14. Organization WH. Early detection of tuberculosis: an overview of approaches, guidelines and tools. 2011.
15. Hopewell PC, Pai M, Maher D, Uplekar M, Raviglione MC. International Standards for Tuberculosis Care. The Lancet Infectious Diseases. 2006;6(11):710-25.
16. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. The Lancet Infectious Diseases. 2006;6(10):664-74.
17. Organization WH. Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. World Health Organization; 2016. Report No.: 9241511508.
18. Chihota VN, Grant AD, Fielding K, Ndibongo B, van Zyl A, Muirhead D, et al. Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. The International Journal of Tuberculosis and Lung Disease. 2010;14(8):1024-31.
19. Pai M, Riley LW, Colford JM. Interferon-γ assays in the immunodiagnosis of tuberculosis: a systematic review. The Lancet Infectious Diseases. 2004;4(12):761-76.
20. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma Interferon Release Assays for Detection of *Mycobacterium tuberculosis* Infection. Clinical Microbiology Reviews. 2014;27(1):3-20.
21. Mazurek GH, Weis SE, Moonan PK, Daley CL, Bernardo J, Lardizabal AA, et al. Prospective Comparison of the Tuberculin Skin Test and 2 Whole-Blood Interferon-γ Release Assays in Persons with Suspected Tuberculosis. Clinical Infectious Diseases. 2007;45(7):837-45.
22. Karki B, Kittel G, Bolokon Jr I, Duke T. Active community-based case finding for tuberculosis with limited resources: estimating prevalence in a remote area of Papua New Guinea. Asia Pacific Journal of Public Health. 2017;29(1):17-27.

23. Purohit M, Mustafa T. Laboratory diagnosis of extra-pulmonary tuberculosis (EPTB) in resource-constrained setting: state of the art, challenges and the need. Journal of Clinical and Diagnostic Research. 2015;9(4):EE01.
24. Esmail H, Barry CE, 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014;369(1645):20130437.
25. Diefenbach-Elstob T, Graves P, Dowi R, Gula B, Plummer D, McBryde E, et al. The epidemiology of tuberculosis in the rural Balimo region of Papua New Guinea. Tropical Medicine & International Health. 2018;23(9):1022-32.
26. Warner J, Rush C. Tropical fever in remote tropics: tuberculosis or melioidosis, it depends on the lab. Microbiology Australia. 2021;42(4):173-8.
27. Diefenbach-Elstob T, Guernier V, Burgess G, Pelowa D, Dowi R, Gula B, et al. Molecular evidence of drug-resistant tuberculosis in the Balimo region of Papua New Guinea. Tropical Medicine and Infectious Disease. 2019;4(1):33.
28. Diefenbach-Elstob T, Plummer D, Dowi R, Wamagi S, Gula B, Siwaeya K, et al. The social determinants of tuberculosis treatment adherence in a remote region of Papua New Guinea. BMC Public Health. 2017;17(1):70.
29. Singhania A, Wilkinson RJ, Rodrigue M, Haldar P, O'Garra A. The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. Nature Immunology. 2018;19(11):1159-68.
30. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973.
31. Constantoulakis P, Filiou E, Rovina N, Chras G, Hamhougia A, Karabela S, et al. In vivo expression of innate immunity markers in patients with *Mycobacterium tuberculosis* infection. BMC infectious diseases. 2010;10(1):1-10.
32. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. Journal of molecular medicine. 2007;85(6):613-21.
33. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Medicine. 2013;10(10):e1001538.

34. Burel JG, Babor M, Pomaznoy M, Lindestam Arlehamn CS, Khan N, Sette A, et al. Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. Frontiers in Immunology. 2019;10(221).
35. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
36. Kwan PKW, Periaswamy B, De Sessions PF, Lin W, Molton JS, Naftalin CM, et al. A blood RNA transcript signature for TB exposure in household contacts. BMC Infectious Diseases. 2020;20(1):403.
37. Ault RC, Headley CA, Hare AE, Carruthers BJ, Mejias A, Turner J. Blood RNA signatures predict recent tuberculosis exposure in mice, macaques and humans. Scientific Reports. 2020;10(1):16873.
38. Estévez O, Anibarro L, Garet E, Pallares Á, Barcia L, Calviño L, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. Frontiers in Immunology. 2020;11(1470).
39. QuantiFERON®-TB Gold Plus (QFT®-Plus) Blood Collection Hilden, Germany Qiagen; 2019 [Available from: https://www.qiagen.com/de/resources/resourcedetail?id=22bdba7c-4b2b-44cc-9a1c-715d264f87a0&lang=en.
40. QuantiFERON®-TB Gold Plus (QFT®-Plus) ELISA Package Insert Hilden, Germany: Qiagen; 2019 [cited 2022 Jul 25]. Available from: https://www.qiagen.com/de/resources/resourcedetail?id=8ad75fb6-2cae-4c0f-998c-f7809afe0640&lang=en.
41. Broccolo F, Scarpellini P, Locatelli G, Zingale A, Brambilla AM, Cichero P, et al. Rapid diagnosis of mycobacterial infections and quantitation of *Mycobacterium tuberculosis* load by two real-time calibrated PCR assays. Journal of Clinical Microbiology. 2003;41(10):4565-72.
42. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
43. Waardenberg AJ, Field MA. consensusDE: an R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. PeerJ. 2019;7:e8206.
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

45. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141.
46. Kurbatova N, Swiers R. Disease ontologies for knowledge graphs. BMC Bioinformatics. 2021;22(1):377.
47. Yu G, Wang L-G, Yan G-R, He Q-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2014;31(4):608-9.
48. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications. 2019;10(1):1523.
49. Piñero J, Saüch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. Computational and Structural Biotechnology Journal. 2021;19:2960-7.
50. Cai Y, Yang Q, Tang Y, Zhang M, Liu H, Zhang G, et al. Increased complement C1q level marks active disease in human tuberculosis. PLoS One. 2014;9(3):e92340.
51. Natarajan S, Ranganathan M, Hanna LE, Tripathy S. Transcriptional Profiling and Deriving a Seven-Gene Signature That Discriminates Active and Latent Tuberculosis: An Integrative Bioinformatics Approach. Genes 2022;13(4).
52. Johnson WE, Odom A, Cintron C, Muthaiah M, Knudsen S, Joseph N, et al. Comparing tuberculosis gene signatures in malnourished individuals using the TBSignatureProfiler. BMC Infectious Diseases. 2021;21(1):106.
53. Gonzalez-Cotto M, Guo L, Karwan M, Sen SK, Barb J, Collado CJ, et al. TREML4 Promotes Inflammatory Programs in Human and Murine Macrophages and Alters Atherosclerosis Lesion Composition in the Apolipoprotein E Deficient Mouse. Frontiers in Immunology. 2020;11.
54. Petrilli JD, Araújo LE, da Silva LS, Laus AC, Müller I, Reis RM, et al. Whole blood mRNA expression-based targets to discriminate active tuberculosis from latent infection and other pulmonary diseases. Scientific Reports. 2020;10(1):22072.
55. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Frontiers in Immunology. 2021;12:637164.
56. Blankley S, Graham CM, Turner J, Berry MPR, Bloom CI, Xu Z, et al. The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. PLoS One. 2016;11(10):e0162220.

57. Bloom CI, Graham CM, Berry MPR, Rozakeas F, Redford PS, Wang Y, et al. Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers. PLoS One. 2013;8(8):e70630.
58. Suliman S, Thompson EG, Sutherland J, Weiner J, 3rd, Ota MOC, Shankar S, et al. Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. American Journal of Respiratory and Critical Care Medicine. 2018;197(9):1198-208.
59. Lubbers R, Sutherland JS, Goletti D, de Paus RA, van Moorsel CHM, Veltkamp M, et al. Complement Component C1q as Serum Biomarker to Detect Active Tuberculosis. Frontiers in Immunology. 2018;9:2427.
60. de Araujo LS, Vaas LAI, Ribeiro-Alves M, Geffers R, Mello FCQ, de Almeida AS, et al. Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. Frontiers in Microbiology. 2016;7.
61. Liang S-K, Chien L-H, Chang G-C, Tsai Y-H, Su W-C, Chen Y-M, et al. Programmed Death Ligand 2 Gene Polymorphisms Are Associated With Lung Adenocarcinoma Risk in Female Never-Smokers. Frontiers in Oncology. 2021;11.
62. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
63. Lee S-W, Wu LS-H, Huang G-M, Huang K-Y, Lee T-Y, Weng JT-Y. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. BMC bioinformatics. 2016;17 Suppl 1(Suppl 1):3-.
64. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, et al. Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification. bioRxiv. 2019:583674.
65. Kline K, McCarthy JS, Pearson M, Loukas A, Hotez PJ. Neglected Tropical Diseases of Oceania: Review of Their Prevalence, Distribution, and Opportunities for Control. PLOS Neglected Tropical Diseases. 2013;7(1):e1755.
66. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. The Lancet Infectious Diseases. 2012;12(1):45-55.
67. Organization WH. Guidelines on the management of latent tuberculosis infection: World Health Organization; 2015.
68. Pai M, Behr M. Latent *Mycobacterium tuberculosis* Infection and Interferon-Gamma Release Assays. Microbiology Spectrum. 2016;4(5):4.5.24.

5.7 Supplementary Information

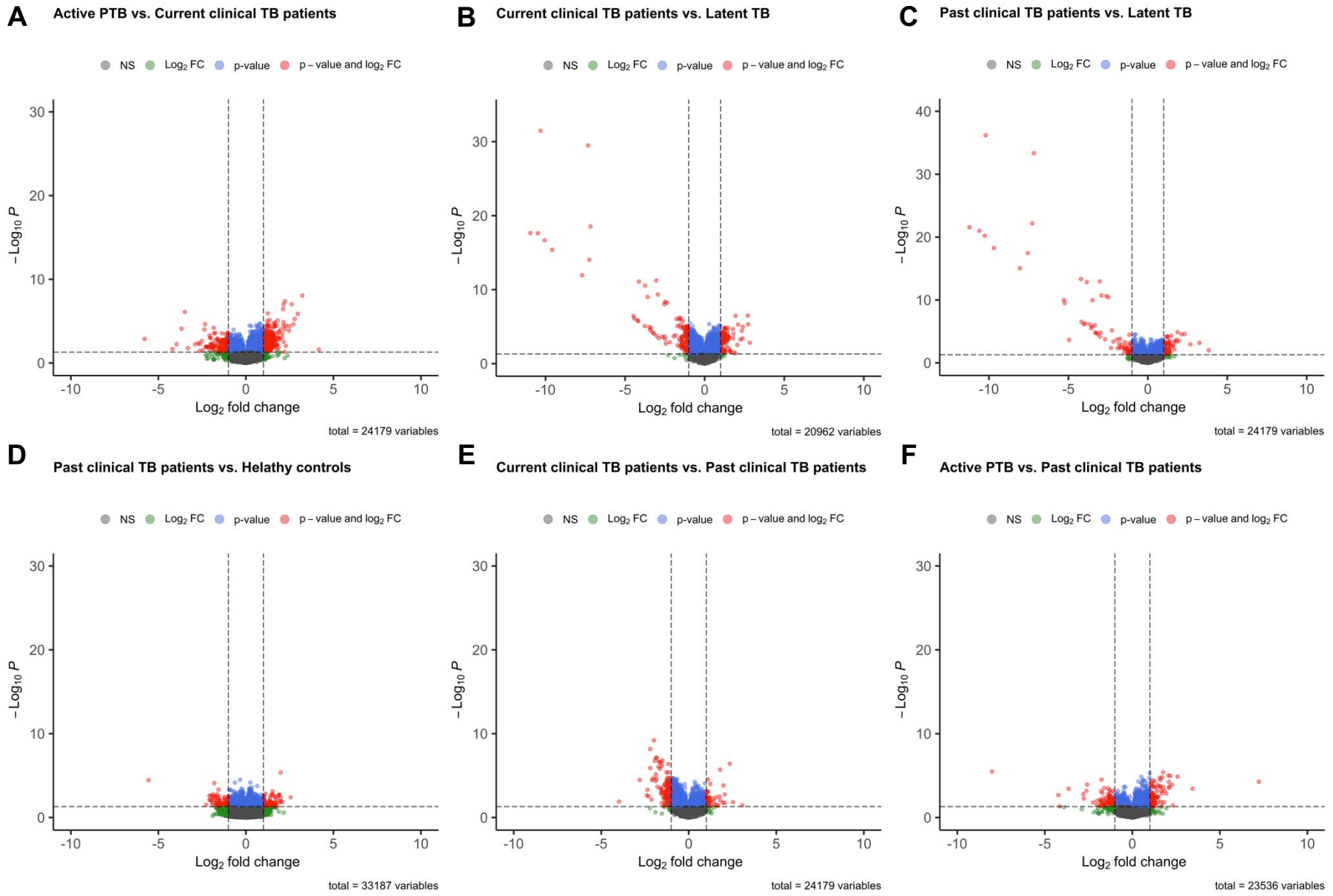


Figure S1 Volcano plots showing differentially expressed genes between groups. The distribution of gene expression fold changes in (A) active PTB relative to latent TB (IGRA+) (B) Current clinical TB patients (IGRA+) relative to latent TB (IGRA+) (C) Past clinical TB patients (IGRA+) relative to latent TB (IGRA+) (D) Past clinical TB patients (IGRA+) relative to healthy controls (IGRA-) (E) Current clinical TB patients (IGRA+) relative to past clinical TB patients (IGRA+) and (F) Active PTB relative to past clinical TB patients (IGRA+). Differentially expressed genes with log 2-fold change >1 and *p*-value ≤ 0.05 are shown in red. The vertical lines correspond to 1.0-fold up and down and the horizontal line represents a *p*-value of 0.05. NS= non-significant, FC= fold change.

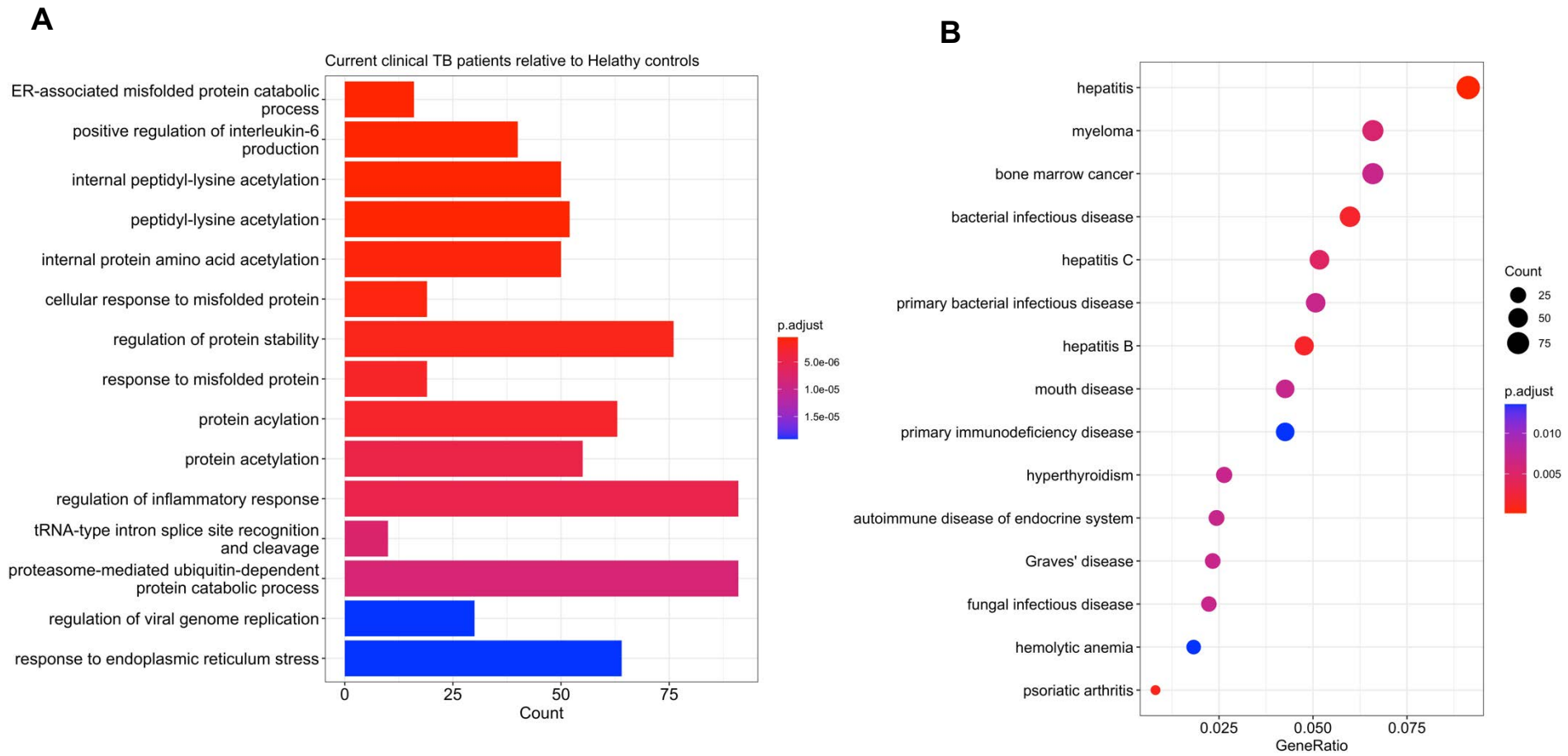


Figure S2 Functional enrichment analysis; gene and disease ontology of DEGs in current clinical TB patients (Group 4, IGRA+) compared to healthy controls (Group 1, IGRA-). (A) Bar graph of the top 15 GO terms for the DEGs based biological process (BP) ordered by statistical significance, **(B)** Dot plot showing top DO processes for the DEGs of the relevant diseases. Colours indicate the Benjamini–Hochberg corrected *p*-value and dot sizes indicate the number of genes contributing to the enrichment of the term.

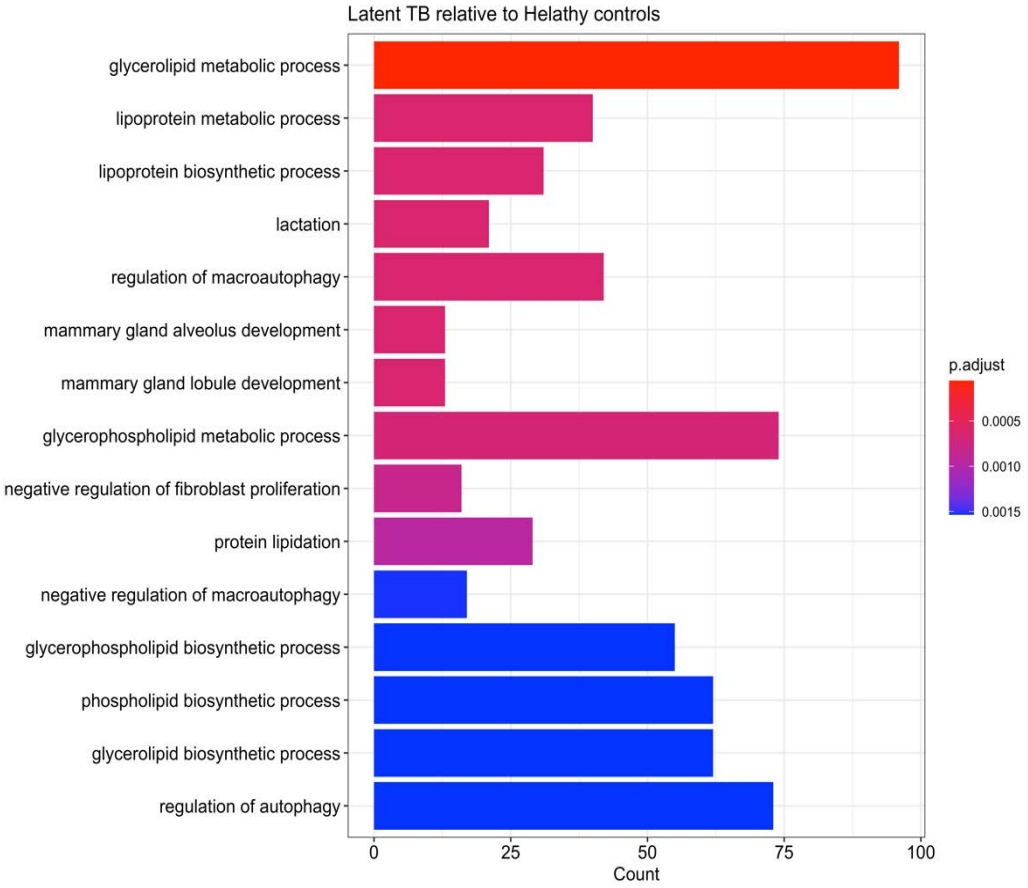


Figure S3 Functional enrichment analysis and gene ontology of DEGs in latent TB (Group 2, IGRA+) compared healthy controls (Group 1, IGRA-). Bar graph of the top 15 GO terms for the DEGs based biological process (BP) ordered by statistical significance Colours indicate the Benjamini–Hochberg corrected *p*-value and dot sizes indicate the number of genes contributing to the enrichment of the term.

Table S1 Significantly DEGs in active PTB relative to healthy controls (IGRA-). Significantly expressed genes with log 2-fold change >1 and adj *p*-value ≤ 0.05 are shown in the table.

gene_id	gene symbol	baseMean	log 2 FC	lfcSE	stat	pvalue	padj
ENSG00000159189	C1QC	15.895	3.608	0.585	6.173	6.72E-10	1.49E-05
ENSG00000205212	CCDC144NL	54.459	3.060	0.576	5.315	1.07E-07	0.001
ENSG00000235538	AL137005.1	7.033	2.406	0.589	4.084	4.42E-05	0.039
ENSG00000173369	C1QB	50.392	2.351	0.464	5.070	3.97E-07	0.002
ENSG00000254602	AP000662.1	15.394	2.272	0.533	4.262	2.03E-05	0.027
ENSG00000108387	SEPT4	85.675	2.155	0.459	4.691	2.71E-06	0.008
ENSG00000223552	AC098613.1	190.921	2.123	0.382	5.556	2.75E-08	0.000
ENSG00000228302	AL512770.1	11.545	2.092	0.513	4.080	4.51E-05	0.039
ENSG00000152766	ANKRD22	524.240	2.065	0.489	4.221	2.43E-05	0.030
ENSG00000100336	APOL4	70.233	1.855	0.379	4.897	9.71E-07	0.004
ENSG00000197646	PDCD1LG2	348.050	1.626	0.391	4.158	3.21E-05	0.032
ENSG00000163016	ALMS1P1	92.524	1.616	0.363	4.454	8.43E-06	0.019
ENSG00000173372	C1QA	67.052	1.559	0.370	4.209	2.56E-05	0.030
ENSG00000175643	RMI2	654.314	1.217	0.301	4.041	5.32E-05	0.044
ENSG00000121797	CCRL2	119.173	1.170	0.279	4.192	2.76E-05	0.031
ENSG00000163814	CDCP1	72.176	1.153	0.269	4.278	1.89E-05	0.027
ENSG00000248242	AC004053.1	300.131	-1.014	0.216	-4.696	2.65E-06	0.008
ENSG00000164756	SLC30A8	32.345	-3.275	0.766	-4.277	1.89E-05	0.027
ENSG00000225231	LINC02470	125.866	-3.995	0.877	-4.557	5.20E-06	0.013
ENSG00000188056	TREML4	123.877	-7.280	1.708	-4.263	2.01E-05	0.027

Table S2 Significantly DEGs in active PTB relative to latent TB (IGRA+). Significantly expressed genes with log 2-fold change >1 and adj *p*-value ≤ 0.05 are shown in the table.

gene_id	gene symbol	baseMean	log 2 FC	lfcSE	stat	pvalue	padj
ENSG00000234426	AL139042.1	1504.191	4.618	1.109	4.165	3.11E-05	0.029
ENSG00000205212	CCDC144NL	54.459	2.948	0.577	5.111	3.21E-07	4.81E-04
ENSG00000223552	AC098613.1	190.921	1.926	0.383	5.031	4.87E-07	0.001
ENSG00000108387	SEPT4	85.675	1.879	0.460	4.083	4.45E-05	0.039
ENSG00000100336	APOL4	70.233	1.552	0.379	4.094	4.23E-05	0.039
ENSG00000143776	CDC42BPA	406.311	1.244	0.297	4.183	2.88E-05	0.028
ENSG00000121797	CCRL2	119.173	1.225	0.280	4.380	1.18E-05	0.014
ENSG00000176728	TTTY14	1355.749	-2.425	0.561	-4.320	1.56E-05	0.017
ENSG00000229236	TTTY10	382.623	-2.843	0.514	-5.531	3.18E-08	6.23E-05
ENSG00000241859	ANOS2P	680.162	-2.899	0.532	-5.444	5.20E-08	9.47E-05
ENSG00000099725	PRKY	4322.019	-3.141	0.551	-5.695	1.23E-08	2.61E-05
ENSG00000165246	NLGN4Y	225.202	-3.168	0.604	-5.243	1.58E-07	2.52E-04
ENSG00000233236	LINC02573	7.321	-3.560	0.807	-4.410	1.04E-05	0.013
ENSG00000067646	ZFY	571.339	-3.777	0.758	-4.980	6.37E-07	0.001
ENSG00000215580	BCORP1	191.421	-4.166	0.780	-5.341	9.23E-08	1.57E-04
ENSG00000129824	RPS4Y1	999.169	-4.404	0.730	-6.029	1.65E-09	3.81E-06
ENSG00000230663	FAM224B	28.594	-5.185	1.218	-4.257	2.07E-05	0.021
ENSG00000233522	FAM224A	28.358	-5.232	1.199	-4.364	1.27E-05	0.014
ENSG00000183878	UTY	10161.518	-7.630	0.821	-9.290	1.55E-20	1.97E-16
ENSG00000131002	TXLNGY	1113.792	-7.948	1.272	-6.250	4.11E-10	1.31E-06
ENSG00000188056	TREML4	123.877	-8.073	1.711	-4.717	2.39E-06	0.003
ENSG00000114374	USP9Y	2206.045	-8.137	1.072	-7.591	3.18E-14	2.70E-10
ENSG00000260197	AC010889.1	51.098	-8.850	1.433	-6.176	6.58E-10	1.86E-06
ENSG00000012817	KDM5D	1252.994	-9.443	1.552	-6.084	1.17E-09	2.98E-06
ENSG00000198692	EIF1AY	360.043	-10.369	1.544	-6.715	1.89E-11	6.86E-08
ENSG00000231535	LINC00278	10778.548	-10.868	1.128	-9.631	5.91E-22	1.51E-17
ENSG00000233864	TTTY15	438.353	-11.173	1.570	-7.114	1.13E-12	5.74E-09
ENSG00000067048	DDX3Y	468.274	-11.535	1.633	-7.066	1.60E-12	6.78E-09
ENSG00000225012	AL121872.1	444.025	-12.252	1.705	-7.185	6.71E-13	4.27E-09

Table S3Significantly DEGs in Latent TB (IGRA+) relative to healthy controls (IGRA). Significantly expressed genes with log 2-fold change >1 and adj *p*-value ≤ 0.05 are shown in the table.

gene_id	gene symbol	baseMean	log 2 FC	lfcSE	stat	pvalue	padj
ENSG00000225012	AL121872.1	444.025	12.313	1.216	10.125	4.27E-24	3.81E-20
ENSG00000233864	TTTY15	438.353	10.763	1.113	9.671	3.99E-22	2.14E-18
ENSG00000012817	KDM5D	1252.994	10.469	1.099	9.529	1.58E-21	7.06E-18
ENSG00000231535	LINC00278	10778.548	10.299	0.807	12.762	2.66E-37	7.12E-33
ENSG00000067048	DDX3Y	468.274	10.180	1.160	8.773	1.74E-18	6.66E-15
ENSG00000198692	EIF1AY	360.043	9.066	1.093	8.296	1.07E-16	3.19E-13
ENSG00000260197	AC010889.1	51.098	8.233	1.007	8.173	3.02E-16	8.08E-13
ENSG00000131002	TXLNGY	1113.792	7.577	0.879	8.616	6.93E-18	2.32E-14
ENSG00000114374	USP9Y	2206.045	7.469	0.746	10.017	1.28E-23	8.58E-20
ENSG00000183878	UTY	10161.518	7.345	0.592	12.396	2.74E-35	3.67E-31
ENSG00000215580	BCORP1	191.421	4.247	0.562	7.555	4.18E-14	9.32E-11
ENSG00000233522	FAM224A	28.358	4.083	0.825	4.951	7.39E-07	0.001
ENSG00000230663	FAM224B	28.594	4.056	0.840	4.828	1.38E-06	0.002
ENSG00000226918	AC010086.1	13.586	3.961	0.796	4.978	6.41E-07	0.001
ENSG00000169789	PRY	21.657	3.947	0.823	4.798	1.61E-06	0.002
ENSG00000129824	RPS4Y1	999.169	3.917	0.522	7.506	6.12E-14	1.26E-10
ENSG00000169807	PRY2	20.145	3.883	0.822	4.726	2.29E-06	0.002
ENSG00000067646	ZFY	571.339	3.817	0.546	6.984	2.87E-12	5.49E-09
ENSG00000233070	ZFY-AS1	16.879	3.662	0.744	4.920	8.65E-07	0.001
ENSG00000206159	GYG2P1	22.040	3.592	0.695	5.167	2.38E-07	3.54E-04
ENSG00000154620	TMSB4Y	13.518	3.572	0.749	4.771	1.84E-06	0.002
ENSG00000099725	PRKY	4322.019	3.207	0.407	7.886	3.13E-15	7.60E-12
ENSG00000169763	PRYP3	9.835	3.192	0.775	4.118	3.82E-05	0.030
ENSG00000277438	KDM5DP1	9.030	3.182	0.774	4.110	3.96E-05	0.030
ENSG00000172283	PRYP4	9.346	3.125	0.767	4.075	4.60E-05	0.034
ENSG00000165246	NLGN4Y	225.202	2.865	0.436	6.575	4.85E-11	8.66E-08
ENSG00000241859	ANOS2P	680.162	2.548	0.390	6.526	6.75E-11	1.13E-07
ENSG00000229236	TTTY10	382.623	2.289	0.374	6.112	9.84E-10	1.55E-06
ENSG00000229308	AC010737.1	11.593	2.268	0.524	4.331	1.49E-05	0.013
ENSG00000106341	PPP1R17	20.342	-1.140	0.263	-4.331	1.48E-05	0.013
ENSG00000137672	TRPC6	494.148	-1.200	0.281	-4.268	1.97E-05	0.017
ENSG00000228325	IGKV3D-7	4.972	-1.626	0.408	-3.982	6.82E-05	0.046

ENSG00000099998	GGT5	134.948	-1.702	0.393	-4.332	1.48E-05	0.013
ENSG00000263499	AC007431.1	10.805	-2.034	0.505	-4.025	5.70E-05	0.041
ENSG00000144681	STAC	951.613	-2.284	0.548	-4.167	3.09E-05	0.026
ENSG00000234426	AL139042.1	1504.191	-3.853	0.819	-4.706	2.53E-06	0.003

Table S4 Significantly DEGs in clinically diagnosed TB patients (IGRA+) relative to healthy controls (IGRA-). Significantly expressed genes with log 2-fold change >1 and adj p-value ≤ 0.05 are shown in the table.

gene_id	gene symbol	baseMean	log 2 FC	lfcSE	stat	pvalue	padj
ENSG00000151388	ADAMTS12	83.326	2.416	0.532	4.543	5.56E-06	0.007
ENSG00000169397	RNASE3	38.501	1.787	0.467	3.824	1.31E-04	0.025
ENSG00000204385	SLC44A4	7.842	1.557	0.428	3.639	2.74E-04	0.035
ENSG00000169385	RNASE2	370.774	1.506	0.382	3.940	8.16E-05	0.022
ENSG00000128512	DOCK4	17847.983	-1.009	0.270	-3.742	1.83E-04	0.029
ENSG00000215244	AL137145.2	2292.316	-1.015	0.286	-3.548	3.87E-04	0.042
ENSG00000184557	SOCS3	1361.645	-1.022	0.273	-3.739	1.85E-04	0.029
ENSG00000248455	LINC02217	15.386	-1.038	0.300	-3.463	5.33E-04	0.049
ENSG00000123610	TNFAIP6	1273.588	-1.062	0.294	-3.614	3.01E-04	0.037
ENSG00000280832	GSEC	1342.475	-1.064	0.237	-4.485	7.28E-06	0.008
ENSG00000128918	ALDH1A2	4418.987	-1.073	0.275	-3.900	9.61E-05	0.022
ENSG00000211653	IGLV1-40	175.180	-1.078	0.291	-3.707	2.09E-04	0.030
ENSG00000180019	AC079741.1	8.996	-1.083	0.276	-3.923	8.76E-05	0.022
ENSG00000233431	AL359815.1	169.310	-1.099	0.256	-4.291	1.78E-05	0.011
ENSG00000163421	PROK2	8026.166	-1.101	0.251	-4.392	1.12E-05	0.009
ENSG00000120306	CYSTM1	5092.401	-1.114	0.252	-4.418	9.95E-06	0.009
ENSG00000187554	TLR5	3657.751	-1.125	0.254	-4.426	9.60E-06	0.009
ENSG00000261172	AC133919.2	19.010	-1.130	0.287	-3.931	8.44E-05	0.022
ENSG00000151726	ACSL1	48270.253	-1.131	0.242	-4.681	2.86E-06	0.005
ENSG00000162949	CAPN13	56.218	-1.165	0.335	-3.478	5.06E-04	0.048
ENSG00000198216	CACNA1E	2723.748	-1.166	0.301	-3.874	1.07E-04	0.022
ENSG00000124107	SLPI	156.244	-1.173	0.297	-3.952	7.75E-05	0.022
ENSG00000266642	AC024267.6	9.221	-1.185	0.313	-3.789	1.51E-04	0.027
ENSG00000148926	ADM	677.455	-1.188	0.262	-4.539	5.65E-06	0.007
ENSG00000158352	SHROOM4	131.225	-1.204	0.341	-3.533	4.11E-04	0.043
ENSG00000137757	CASP5	1881.888	-1.210	0.274	-4.418	9.94E-06	0.009
ENSG00000264400	RN7SL491P	14.835	-1.215	0.305	-3.979	6.93E-05	0.021

ENSG00000112053	SLC26A8	1220.262	-1.224	0.295	-4.147	3.36E-05	0.014
ENSG00000212743	AL137145.1	160.747	-1.232	0.313	-3.938	8.23E-05	0.022
ENSG00000268658	LINC00664	25.399	-1.239	0.351	-3.532	4.13E-04	0.043
ENSG00000259182	AC019254.1	80.256	-1.247	0.268	-4.658	3.19E-06	0.005
ENSG00000124102	PI3	368.294	-1.253	0.321	-3.909	9.27E-05	0.022
ENSG00000283597	FAM169B	571.202	-1.272	0.308	-4.132	3.60E-05	0.015
ENSG00000259293	LIPC-AS1	16.465	-1.279	0.349	-3.663	2.49E-04	0.034
ENSG00000251230	MIR3945HG	389.164	-1.295	0.259	-4.999	5.76E-07	0.002
ENSG00000132470	ITGB4	101.687	-1.306	0.354	-3.691	2.24E-04	0.032
ENSG00000280587	LINC01348	10.340	-1.307	0.318	-4.111	3.95E-05	0.016
ENSG00000183019	MCEMP1	50.995	-1.378	0.376	-3.662	2.50E-04	0.034
ENSG00000251139	AC084871.2	20.472	-1.428	0.323	-4.421	9.81E-06	0.009
ENSG00000268170	AC073342.2	156.491	-1.433	0.301	-4.761	1.93E-06	0.004
ENSG00000162551	ALPL	9182.638	-1.509	0.290	-5.196	2.03E-07	0.001
ENSG00000138772	ANXA3	3178.619	-1.540	0.302	-5.092	3.53E-07	0.001
ENSG00000229314	ORM1	64.275	-1.555	0.362	-4.296	1.74E-05	0.011
ENSG00000158089	GALNT14	2347.258	-1.559	0.345	-4.518	6.24E-06	0.007
ENSG00000215196	BASP1-AS1	1124.725	-1.564	0.345	-4.530	5.89E-06	0.007
ENSG00000165181	C9orf84	9.016	-1.568	0.340	-4.610	4.03E-06	0.006
ENSG00000183762	KREMEN1	3695.712	-1.580	0.309	-5.112	3.19E-07	0.001
ENSG00000254695	AC087379.2	108.973	-1.616	0.315	-5.137	2.79E-07	0.001
ENSG00000145555	MYO10	698.686	-1.709	0.445	-3.840	1.23E-04	0.024
ENSG00000254420	AP003086.2	284.242	-1.711	0.423	-4.041	5.32E-05	0.018
ENSG00000269711	AC008763.3	14.643	-1.743	0.455	-3.829	1.29E-04	0.025
ENSG00000285984		73.820	-1.801	0.386	-4.670	3.02E-06	0.005
ENSG00000095203	EPB41L4B	12.193	-1.823	0.502	-3.628	2.86E-04	0.036
ENSG00000254789	AC073172.1	3572.340	-1.912	0.326	-5.862	4.57E-09	4.79E-05
ENSG00000274173	AL035661.1	37.903	-1.931	0.482	-4.007	6.15E-05	0.020
ENSG00000254946	AC073172.2	24.905	-1.954	0.322	-6.077	1.22E-09	2.56E-05
ENSG00000239265	CLRN1-AS1	54.144	-1.993	0.460	-4.334	1.47E-05	0.010
ENSG00000249173	LINC01093	28.328	-2.202	0.445	-4.953	7.31E-07	0.002
ENSG00000204933	CD177P1	56.103	-2.464	0.686	-3.592	3.28E-04	0.038
ENSG00000204936	CD177	283.216	-2.909	0.683	-4.261	2.03E-05	0.011

Chapter 6

The effect of tropical temperatures on the quality of RNA extracted from stabilized whole blood samples

This chapter titled work 'The effect of tropical temperatures on the quality of RNA extracted from stabilized whole blood samples' has been published in *International Journal of Molecular Sciences* on 13th September 2022 (DOI: https://doi.org/10.3390/ijms231810609).

6 The effect of tropical temperatures on the quality of RNA extracted from stabilized whole blood samples

Yomani D. Sarathkumara[1], Daniel J. Browne[1], Ashton M. Kelly[1], David J. Pattinson[1], Catherine M. Rush[1,2], Jeffrey Warner[1,2], Carla Proietti[1], Denise L. Doolan[1*]

[1] Centre for Molecular Therapeutics, Australian Institute of Tropical Health of Medicine, James Cook University, Cairns, QLD, Australia

[2] College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, QLD, Australia

*** Corresponding Author:** Email: denise.doolan@jcu.edu.au (DLD)

Competing Interest Statement: The authors declare no competing interest.

6.1 Abstract

Whole-blood derived transcriptional profiling is widely used in biomarker discovery, immunological research and therapeutic development. Traditional molecular and high-throughput transcriptomic platforms, including molecular assays with quantitative-PCR (qPCR) and RNA-sequencing (RNA-seq), are dependent upon high-quality, high-quantity and intact RNA. However, collecting high-quality RNA from field studies can be challenging in remote tropical locations due to resource restrictions and logistics of post-collection processing. The current study tested the relative performance of two most widely used whole-blood RNA collection systems, PAXgene® and Tempus™, in optimal laboratory conditions as well as suboptimal conditions in tropical field sites, including the effects of extended storage times and high storage temperatures. We found that Tempus™ tubes maintained a slightly higher RNA quantity and integrity relative to PAXgene® tubes at suboptimal tropical conditions. Both PAXgene® and Tempus™ tubes gave similar RNA purity (A260/A280). Additionally, we found that Tempus™ tubes preferentially maintained the stability of mRNA transcripts for two reference genes tested, Succinate dehydrogenase complex, subunit A (SDHA) and TATA-box-binding protein (TBP) even when RNA quality decreased with storage length and temperature. However, both tube types preserved rRNA transcript 18S ribosomal RNA (18S) equally. Our results suggest that Tempus™ blood RNA collection tubes are preferable to PAXgene® for whole-blood collection in suboptimal tropical conditions for RNA-based studies in resource-limited settings.

Key words: PAXgene®, Tempus™, Blood RNA extraction, RT-qPCR, tropical, climate

6.2 Introduction

Gene expression profiles from whole-blood derived RNA have proven useful as a molecular signature reflecting physiological and pathological changes in the body in the body (1). Immune signatures for various diseases and metabolic states, including biomarkers of disease, disease prognosis or therapeutic efficacy, have been identified by blood transcriptional profiling (2-4). Since blood transcript profiles are reproducible, cost-effective, and easy to implement, they can be rapidly translated into clinical practice (5). High-quality, intact RNA is imperative for both traditional molecular diagnostics and high-throughput transcriptome sequencing techniques, such as quantitative PCR (qPCR) (6) and RNA-sequencing (RNA-seq) (7). High-quality RNA can be readily obtained from fresh blood when processed immediately following collection. However, RNA quality can be adversely impacted by processing delays and complex logistical issues. Remote tropical locations are particularly challenging due to locality and resource restrictions (8). Pre-extraction factors that influence gene expression or lead to RNA degradation include processing delays and storage conditions (*i.e.*, duration and temperature) (9-12) since blood RNA is highly susceptible to enzymatic degradation and oxidative damage (13, 14).

Commercially available blood collection systems with additives that stabilize RNA have been developed to address RNA stability, resulting in significantly enhanced quantity and quality of RNA extracted from whole-blood (15, 16). The two most widely used blood RNA stabilizing systems are PAXgene® and Tempus™ (17, 18). These systems contain proprietary solutions that lyse cells, inactivate RNases, and minimize changes in gene expression (19). Those systems are designed to facilitate long-term whole-blood storage at low temperatures (−80°C), removing the necessity of isolating RNA immediately post-collection and allowing batched processing (8, 20). However, each system has different stabilization efficiencies, thus impacting the resultant transcriptional profiles (21). Several studies have investigated the relative performance of PAXgene® and Tempus™ in specific suboptimal conditions (17, 20). There is, however, an unmet need to identify the optimal RNA stabilization system for use in suboptimal tropical conditions, including high temperatures and extended times before storage at −80°C.

In this study, we compared the effects of extended storage times and high temperatures simulating suboptimal tropical conditions before freezing of whole-blood RNA stabilized in either the PAXgene® or Tempus™ systems. Under several suboptimal tropical conditions, we found that Tempus™ blood RNA collection tubes resulted in higher RNA yields compared to

PAXgene®. Additionally, our results showed that Tempus™ tubes maintained a higher mRNA transcript stability for two reference genes tested, *Succinate dehydrogenase complex, subunit A* (*SDHA*) and *TATA-box-binding protein* (*TBP*). However, both tube types preserved rRNA transcript *18S ribosomal RNA* (*18S*) equally. Overall, our study suggests that Tempus™ blood RNA collection tubes are preferable to PAXgene® when collecting whole-blood RNA in suboptimal tropical conditions.

6.3 Materials and Methods

6.3.1 Sample Collection

Whole-blood was collected from healthy adult volunteer donors into Tempus™ (3 mL) Blood RNA tubes (Applied Biosystems, Foster City, CA, USA) or PAXgene® (2.5 mL) Blood RNA tubes (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland) according to manufacturer's instructions. Briefly, whole-blood was collected directly into each tube by standard venepuncture and immediately shaken vigorously for 10s to ensure that the stabilising reagent makes uniform contact with the sample as per the manufacturer's instructions.

6.3.2 Ethics approval and consent to participate

Protocols for obtaining volunteer blood samples were reviewed and approved by the James Cook University Human Research Ethics Committee (H7886). All participants provided written informed consent. All experiments were performed in compliance with the Declaration of Helsinki.

6.3.3 Experimental Design

Experiment A compared RNA yields, purity, and integrity from MagMAX™ extractions using spin columns or magnetic beads (**Figure 6-1**, top panel). All blood samples collected for Experiment A were kept for two hours at 25°C room temperature (RT) after collection and freshly extracted (*i.e.*, no storage at –80°C). Experiment B evaluated RNA yields, purity and integrity in whole-blood samples stored at different temperatures (25, 30, 35 or 40 °C) and storage times (0, 5, 7 or 10 days; **Figure 6-1**). All samples collected for Experiment B were frozen at –20°C overnight then transferred to –80°C until RNA extraction with magnetic beads-

based MagMAX™. All blood samples stored in PAXgene® tubes were thawed for 2h at room temperature whereas Tempus™ tubes were thawed for 30 min on ice prior RNA isolation.

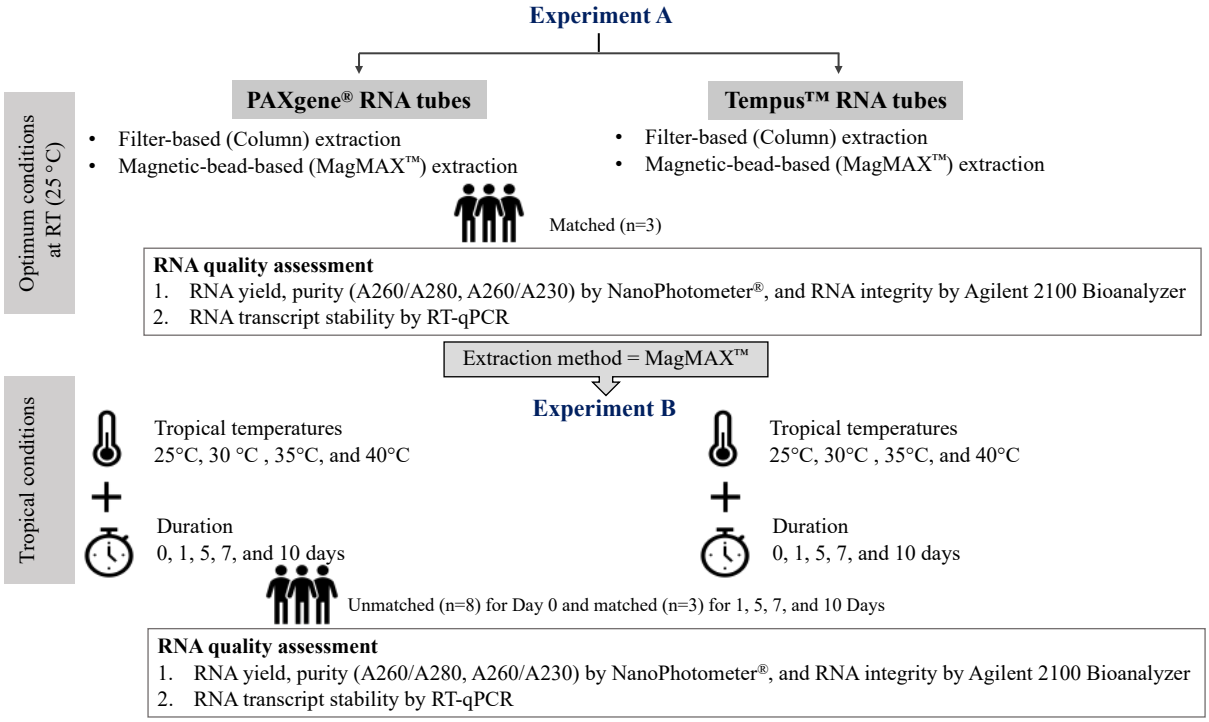


Figure 6-1 Experimental design.

Experiment A. Comparison of total RNA yields, purity, and integrity using spin column-based and magnetic bead-based kits for the two types of blood stabilization systems (PAXgene® Blood RNA Tubes and Tempus™ Blood RNA Tubes). All RNA samples were extracted fresh post collection. **Experiment B**. Systematic testing of different temperatures (25, 30, 35, or 40 °C) and storage times (0, 1, 5, 7, or 10 days) were immediately frozen at −80 °C for later extraction with matched samples (n = 3). Samples immediately frozen at −80 °C for later extraction is referred to "D1/Control". Unmatched samples (n = 8) collected in PAXgene® or Tempus™ RNA stabilizing systems and processed post collection immediately at optimal laboratory conditions (D0/Fresh).

6.3.4 RNA extraction

Column-based RNA purification

Total RNA from whole-blood collected in PAXgene® tubes was extracted according to manufacturer's instructions using PAXgene® Blood RNA Kit (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland), which included DNase I treatment. Total RNA was eluted in 40 µL elution buffer. According to the manufacturer's instructions, total RNA from blood collected in Tempus™ tubes was extracted using the Tempus™ Spin RNA Isolation Kit (Applied Biosystems, CA, USA) RNA was eluted in 90 µL of elution solution. DNase treatment was an optional step in the Tempus™ column extraction system, and therefore not included as genomic DNA contamination using this procedure is minimal (less than 0.05% by weight) according to the manufacturer's specifications.

Magnetic bead-based RNA purification (MagMAX™)

Total RNA from whole-blood was extracted from PAXgene® Blood RNA Tubes using MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit (Life Technologies, CA, USA) according to the manufacturer's protocol, including a TURBO™ DNase and protease step. RNA was isolated using MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit, compatible with Tempus™ Blood RNA tubes (Life Technologies, CA, USA) following the manufacturer's protocol with TURBO™ DNase treatment. All extracted RNA samples were stored at -80˚C. The technical characteristics of each extraction method are summarised in **Table S1**.

6.3.5 RNA yield, purity and integrity

RNA concentration (ng/µL), A260/A280 and A260/230 ratios to indicate RNA purity were measured by spectroscopic quantification using NanoPhotometer® N60 (Implen, München, Germany). RNA integrity was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the Eukaryote Total RNA Nano assay, complementing RNA 6000 NanoChip kit (Agilent Technologies, CA, USA), following the manufacturer's instructions. The RNA integrity number (RIN) was calculated by the Agilent 2100 Expert software (Version B.02.10.SI764, Agilent). The RIN ranges from 1 to 10; a RIN of fully intact RNA is 10, and a RIN of completely degraded RNA is 1.

6.3.6 Reverse transcription quantitative PCR (RT-qPCR)

Reverse Transcription

All reverse transcription (RT) reactions were conducted using the SuperScript IV™ First-Strand Synthesis System™ (ThermoFisher Scientific, Waltham, MA, USA). All samples were primed with 37.5 ng of random hexamers and 10mM dNTPs at 65˚C for 5 min, then 4˚C for 1 min. Reverse transcription was then performed using the SuperScript IV™ reverse-transcriptase (SSIV) for 10 min at 23˚C, 10 min at 50˚C, and 10 min at 85˚C. SSIV concentration was assessed at 20 U (20 units/µL RNA) reactions compared with 5 U reactions (**Figure S2**) as previously described (22). All subsequent RNA samples were reverse transcribed at 30 ng/µL (**Figure S3**) using 5 U reactions in 15 µL total volume reactions for test conditions. All cDNA samples were stored at 4˚C.

Quantitative PCR (qPCR)

qPCR was run with 5 µL total reaction volume using SsoAdvanced SYBR® SuperMix (Bio-Rad, Hercules, CA, USA), which facilitates excellent reaction efficiencies (6). All reactions contained 0.5 µM of desalt-grade primers (Sigma-Aldrich) with 0.75 ng/uL sample cDNA. Each sample was run in technical triplicate replicate, followed by a melt curve analysis to ensure primer specificity. Primers used for the RT-qPCR assays were sourced from Primer Bank™ (23) (**Table S2**). Primer efficiencies were calculated as per MIQE guidelines (24), as previously published from cDNA standards (6). Reaction efficiency was calculated from $\log_2$ dilutions of pooled cDNA from 1×10^6 unstimulated PBMCs. The PCR cycling program included an enzyme activation step at 95˚C for 2 min and then 40 cycles of annealing and extension at 95˚C for 15 s and 60˚C for 30 s, respectively. The cycle threshold (Ct) value was set to 0.3 ΔRN, and a pooled cDNA positive control was included across all plates to ensure reproducibility. qPCR was performed using the QuantStudio 5 real-time PCR system running QuantStudio Design and Analysis Software (v1.5.1, Applied Biosystems).

6.3.7 Statistical analysis

Statistical analyses were performed using R statistical software (https://www.r-project.org/, Version 1.4.1103). Unpaired t-tests were used to compare the data from spin column-based and magnetic bead-based MagMAX™ RNA extractions to determine any significant difference between RNA isolation systems for total RNA yield, A260/A280 ratios

and RNA integrity. Statistical significance was defined using *p*-values <0.05. Normal distribution of data and normality of residuals were evaluated using the Shapiro-Wilks test.

The average Ct values for each replicates/triplicate and targets with Ct-values > 35 or undefined were considered beyond the limit of detection (LOD) and removed from the analysis (17). Paired t-tests determined the differences of RNA yields normalized to input whole-blood volume, A260/A280 ratios, RIN values and Ct values for short and medium lengths of all housekeeping genes between the two RNA stabilization systems (PAXgene® and Tempus™) for RNA extracted on Fresh and Control conditions. Whole-blood samples processed post-collection immediately at optimal laboratory conditions without -80°C storage are referred to as 'D0/Fresh'. Samples frozen immediately at -80°C for later extraction are referred to as 'D1/Control'.

Multiple linear regression models were fit to investigate the overall relationship of independent variables at experimental conditions (*i.e.,* tube type, temperature, and storage time) on normalized RNA yields, A260/A280 ratios, RIN values or Ct values. To further explore the significant interaction between temperatures and tube type, separate analyses were performed at each temperature. To compare the changes of Ct values on Control and test conditions on PAXgene® and Tempus™ tube types, we performed a two-way analysis of variance with multiple comparisons.

6.4 Results

6.4.1 Magnetic-Bead and Spin-Column-based RNA purification systems extracted equivalent whole-blood RNA quantity, quality and purity.

The quantity, quality and purity of RNA extracted were evaluated using two standard RNA isolation systems (18) spin-column-based (PAXgene® QIAGEN/BD; Tempus™ Applied Biosystems) or magnetic-bead-based (MagMAX™ Life Technologies; compatible with either PAXgene® or Tempus™ collection tube) systems. RNA concentration (ng/µL), purity (A260/A280 and A260/A230) were evaluated by spectroscopic quantification using NanoPhotometer® N60 (Implen, München, Germany). RNA integrity number (RIN) was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA yield (ng) was normalized to whole-blood volume collected in each blood RNA collection system (*i.e.,* 2.5 mL in PAXgene®, *vs.* 3 mL in Tempus™). All RNA samples were extracted fresh post-collection for the comparison of RNA isolation protocols.

We found the RNA isolation protocol did not affect the amount of total RNA extracted, the RNA quality, or the purity of RNA (Normalized total RNA: $p = 0.875$, RIN: $p = 0.124$, A260/A280: $p = 0.101$, A260/A230: $p = 0.318$, MagMAX™ *vs.* columns, unpaired t-test) **(Figure S1**). Additionally, all samples had RIN of >7.0 and an A260/A280 ratio between 1.98–2.15, suggesting recovery of high-quality and high-purity RNA from both RNA isolation systems. A260/230 ratios were used as a secondary measurement detecting presence of residual phenol/ethanol, salts and carbohydrates that can affect RNA quality. We observed differences between column *vs*. MagMAX™ extraction methods, although differences were not significant. Comparatively, better values were obtained with Tempus™ column extractions. We found RNA yields obtained from the Tempus™ tubes were significantly lower than those obtained with PAXgene® tubes (p=0.008, unpaired t-test), suggesting that the PAXgene® tubes give higher RNA yields than Tempus™ when extracted in optimal laboratory conditions (**Figure S1**). Given the high quality of the extracted RNA and the advantages of extracting in a 96-well-plate format (6), magnetic bead-based (MagMAX™) RNA purification system compatible with PAXgene® and Tempus™ systems were used for the rest of the extractions.

6.4.2 A higher quantity of RNA was obtained using Tempus™ Blood RNA Tubes in suboptimal tropical conditions.

Next, we compared the quantity, quality, and purity of RNA extracted from whole-blood samples stored in suboptimal tropical conditions collected in PAXgene® and Tempus™ tubes. To simulate suboptimal tropical conditions, matched whole-blood samples were collected in either PAXgene® or Tempus™ tubes and stored at different temperatures (25, 30, 35 or 40˚C) for different lengths of time (0, 1, 5, 7 or 10 days) before storage at -80˚C for later extraction. These samples were compared to matched samples immediately frozen at -80˚C for later extraction (D1/Control) or unmatched samples collected and processed post-collection immediately in optimal laboratory conditions (D0/Fresh). The tube type had no effect on RNA quantity (*i.e.,* normalized RNA concentration) when extracted in Fresh ($p = 0.065$, paired t-test) or Control ($p = 0.274$, paired t-test) conditions (**Figure 6-2A**). However, we found that tube type significantly affected normalized RNA concentration ($p < 0.0001$; **Table 6-1, Model 1**) in samples subjected to suboptimal tropical conditions.

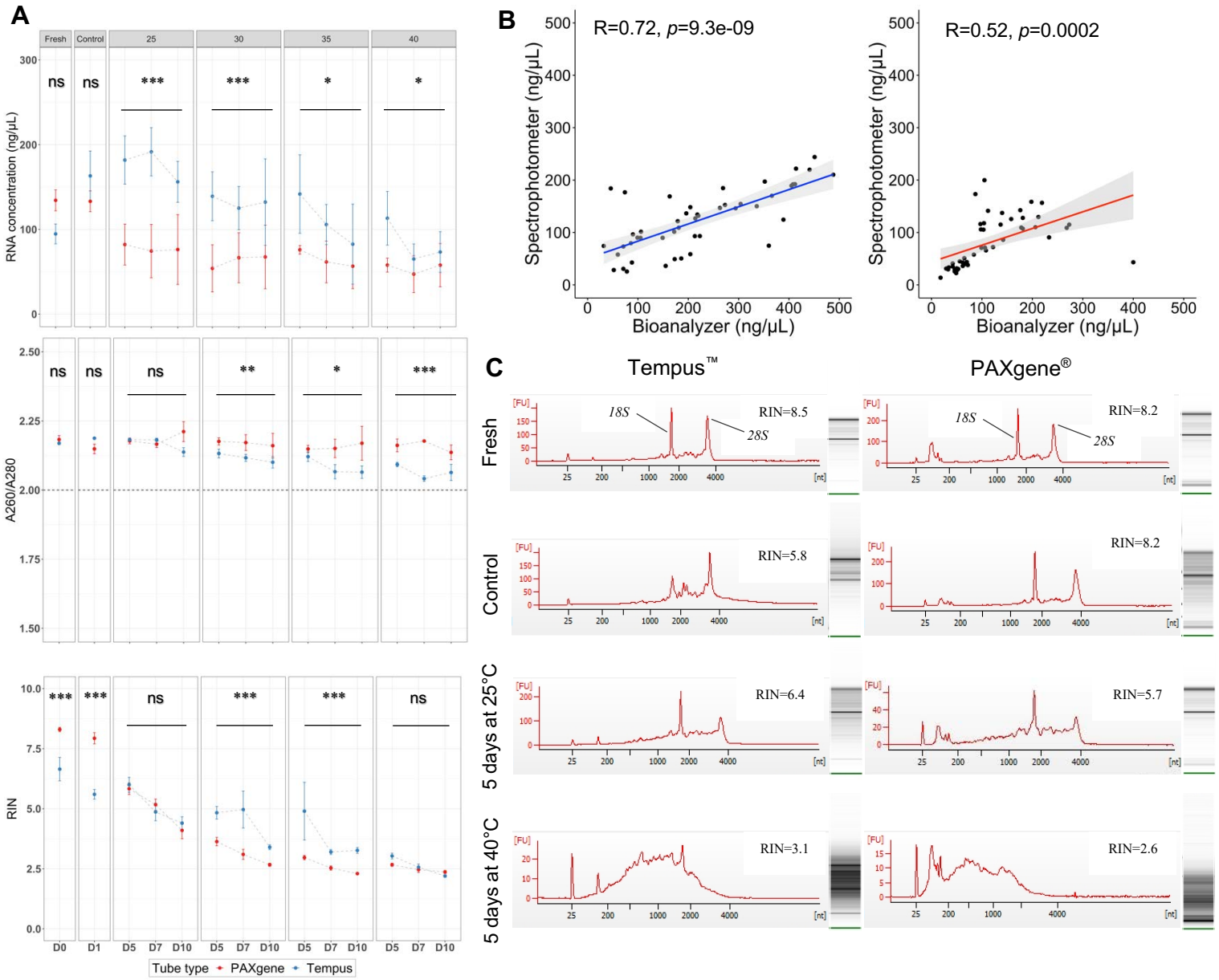

A
Fresh
Control
25
30
35
40
RNA concentration (ng/µL)
A260/A280
RIN
Tube type PAXgene Tempus
B
R=0.72, p=9.3e-09
R=0.52, p=0.0002
Spectrophotometer (ng/µL)
Bioanalyzer (ng/µL)
C
Tempus™
PAXgene®
Fresh
Control
5 days at 25°C
5 days at 40°C
18S
28S
RIN=8.5
RIN=8.2
RIN=5.8
RIN=8.2
RIN=6.4
RIN=5.7
RIN=3.1
RIN=2.6

Figure 6-2 Comparison RNA concentration, purity, and integrity between PAXgene® and Tempus™ blood collection systems under suboptimal tropical storage conditions.
(A) Mean RNA yield normalized to whole-blood volume (ng/µL), A260/A280 ratios determined spectrophotometrically, and RNA integrity number (RIN) across the different conditions [Fresh (n=8), Day1/Control; immediately frozen at -80°C for later extraction and various temperatures (25, 30, 35, 40°C) (matched subjects n=3)]. **(B)** Scatter plots revealed correlations between normalized RNA concentration (ng/uL) by Spectrophotometer or Bioanalyzer for PAXgene® (left panel) and Tempus™ (right panel). Pearson's correlation assessed correlations between variables. **(C)** Electropherograms showed two distinct peaks (*28S* and *18S*), gel images showed two bands comprising the *28S* and *18S* from high-quality RNA, and smears indicated RNA degradation. The uncropped non-quantitative gel per n is shown; *18S* and *28S* peaks of sample #3 (Fresh) and #2 (Control, 5 days at 25°C and 5 days at 40°C). Note the different scales in Figure 2C. The dashed line indicates OD260/A280 = 2.0 for high-quality RNA. Effect of tube type in the linear regression models for each temperature point are indicated; $^{***}p< 0.001$, $^{**}p< 0.01$, $^{*}p< 0.05$, ns – non-significant. RIN= RNA integrity number.

To investigate these data further, we applied three multiple linear regression models to evaluate the effects of explanatory variables (*i.e.*, tube type, storage times, temperature, and biological subject) on normalized RNA concentrations, A260/A280 ratios and RIN values (**Table 6-1**). Model 1 explained 76% of the variation in extracted RNA quantity ($R^2 = 0.764$, $p<2.2\times10^{-16}$). Tempus™ tubes had a significant effect on the RNA yield (p=1.15e-06), and temperature variation in tube type significantly impacted RNA yield (p=0.0003; **Table 6-1, Model 1**). Additionally, we applied separate linear models to each temperature condition to explore the effect of the temperature on the tube type (**Table S3**). We found that the concentration of RNA extracted from whole-blood collected in Tempus™ tubes was significantly greater than for PAXgene® tubes at all evaluated temperatures [p=7.16e-06 (25˚C); p=0.0004 (30˚C); p=0.033 (35˚C); p=0.023 (40˚C), **Table S3**]. As a secondary measurement of RNA concentration, we obtained the RNA concentration readings from Agilent Bioanalyzer. Similarly, Tempus™ tubes gave higher RNA yields when measured with Agilent Bioanalyzer (p=0.001; **Table S4**). Interestingly, RNA concentration measurements by spectrophotometer and bioanalyzer were more strongly correlated in Tempus™ than in

PAXgene® tubes [p=0.0002, R^2=0.270 (PAXgene®); p=9.3×10^{-9}, R^2=0.518 (Tempus™)]; **Figure 6-2B**), suggesting that tube-specific contents influence concentration measurements.

Table 6-1 Multiple linear regression models. Model 1 on normalized RNA concentration (ng/µL), Model 2 on A260/A280 ratios and Model 3 on RIN (expressed as $\log_2$ RIN)

Model 1 = RNA concentration (ng/µL)				
Explanatory variable	Estimate	Std. Error	t value	p value
(Intercept)	66.230	78.853	0.84	0.404
Tube type: Tempus	254.894	47.339	5.384	1.15e-06***
Days	5.049	9.997	0.505	0.615
Temperature	-0.106	2.358	-0.045	0.964
Tube type Tempus: Days	-5.859	3.342	-1.753	0.084
Tube type Tempus: Temperature	-4.650	1.228	-3.786	0.0003*
Days: Temperature	-0.170	0.299	-0.57	0.571
Adjusted R-squared: 0.764 (p value: <2.2e-16)				
Model 2 = A260/A280				
Explanatory variable	Estimate	Std. Error	t value	p value
(Intercept)	2.181	0.102	21.466	<2e-16***
Tube type: Tempus	0.161	0.061	2.638	0.011*
Days	0.008	0.013	0.606	0.547
Temperature	0.000	0.003	-0.085	0.933
Tube type Tempus: Days	-0.008	0.004	-1.951	0.055
Tube type Tempus: Temperature	-0.005	0.002	-3.075	0.003**
Days: Temperature	0.000	0.000	-0.567	0.572
Adjusted R-squared:0.565 (p value: 1.458e-10)				
Model 3 = $\log_2$(RIN)				
Explanatory variable	Estimate	Std. Error	t value	p value
(Intercept)	4.905	0.572	8.581	3.45e-12***
Tube type: Tempus	0.338	0.343	0.985	0.328
Days	-0.148	0.072	-2.047	0.045*
Temperature	-0.082	0.017	-4.777	1.10e-05***
Tube type Tempus: Days	-0.020	0.024	-0.836	0.407
Tube type Tempus: Temperature	0.002	0.009	0.236	0.814

Days: Temperature	0.002	0.002	1.057	0.295
Adjusted R-squared: 0.797 (p value: <2.2e-16)				

Independent multiple linear regression models: Model 1 on normalized RNA concentration, Model 2 on A260/A280 ratios and Model 2 on $\log_2$(RIN) values as dependent variable and tube types, days, temperatures and subjects as the independent variables. $^{***}p< 0.001$, $^{**}p< 0.01$, $^{*}p< 0.05$

When considering RNA purity, all extracted RNA samples had A260/A280 ratios >2, regardless of tube type, indicating a high purity under all test conditions (**Figure 6-2A**, middle panel). There was no difference in A260/A280 ratios between tube type for RNA extracted from Fresh and Control RNA samples ([p=0.480, (Fresh); p=0.111, (Control), paired t-test]. However, at tropical storage and temperature conditions, the tube type (p=0.011) and the storage temperature on tube type (p=0.003) had significant effects on A260/A280 ratios as per the multiple linear regression model (**Table 6-1**, **Model 2**). We used A260/230 ratio as a secondary measurement of RNA purity although with low quantities as ng/uL RNA A260/230 ratios are highly variable A260/230 ratios (**Figure S6**). A260/230 ratios are irrelevant for RNA quality and only describe dissolved solvent concentration relative to RNA. Linear models further indicated that the higher temperatures decreased purity as evidenced by A260/A280 ratio [30˚C (p=0.007), 35˚C (p=0.011) and 40˚C (p=4.89e-05), **Table S3**]. Taken together, these data demonstrated that higher RNA yields are extracted from Tempus™ blood tubes compared to PAXgene® tubes in suboptimal tropical conditions. These data also showed that RNA yields significantly decreased with increasing temperature in both PAXgene® and Tempus™ tubes.

6.4.3 A higher quality of RNA was obtained using PAXgene® tubes in optimal laboratory conditions.

We determined if high RNA quality was preserved using PAXgene® or Tempus™ tubes in suboptimal tropical conditions and compared the RNA quality with samples extracted in optimal laboratory conditions. RIN values declined over time and temperature, irrespective of the whole-blood collection tube (**Figure 6-2A)**. PAXgene® had significantly higher RIN values in 'Fresh' (p=0.013, paired t-test) and 'Control' conditions (p=0.001, paired t-test) compared to Tempus™. The electropherograms showed comparable results for different conditions applied on PAXgene® and Tempus™ systems (**Figure 6-2C)**. Ribosomal RNA bands were

clearly visible in 'Fresh' extractions and 'Control' samples. Most RNA eluates stored at room temperature (25°C) for 5-7 days obtained RIN values around 5–6 with visible *18S* and *28S* bands. In contrast, RNA stored at 40°C did not show distinct rRNA banding. These data demonstrated that higher quality RNA was obtained with PAXgene® tubes (compared to Tempus™) when RNA was extracted post-collection immediately or when samples were maintained at the optimal storage conditions as recommended by the manufacturers.

6.4.4 A higher quality of RNA was obtained using PAXgene® tubes in optimal laboratory conditions.

We determined if high RNA quality was preserved using PAXgene® or Tempus™ tubes in suboptimal tropical conditions and compared the RNA quality with samples extracted in optimal laboratory conditions. RIN values declined over time and temperature, irrespective of the whole-blood collection tube (**Figure 6-2A)**. PAXgene® had significantly higher RIN values in 'Fresh' (p=0.013, paired t-test) and 'Control' conditions (p=0.001, paired t-test) compared to Tempus™. The electropherograms showed comparable results for different conditions applied on PAXgene® and Tempus™ systems (**Figure 6-2C)**. Ribosomal RNA bands were clearly visible in 'Fresh' extractions and 'Control' samples. Most RNA eluates stored at room temperature (25°C) for 5-7 days obtained RIN values around 5–6 with visible *18S* and *28S* bands. In contrast, RNA stored at 40°C did not show distinct rRNA banding. These data demonstrated that higher quality RNA was obtained with PAXgene® tubes (compared to Tempus™) when RNA was extracted post-collection immediately or when samples were maintained at the optimal storage conditions as recommended by the manufacturers.

6.4.5 The highest quality RNA was obtained using Tempus™ Blood RNA tubes in suboptimal tropical conditions.

A multiple linear regression model was built to explore the effects of suboptimal tropical conditions on RNA integrity to determine the effect of tube type, storage temperature, and storage time on RNA integrity (measured by RIN values) (**Table 6-1, Model 3**). Model 3 explained approximately 79.7% (R^2 = 0.797) of the variation in RIN values. RIN values decreased significantly over time (p=0.045), decreasing by 0.862 per day (Model 3, Days Estimate =-0.148; $e^{-0.148}$=0.862). RIN values also decreased significantly with the increasing of temperature (p <0.0001), decreasing by 0.921 (Model 3, Temperature Estimate= -0.082; e^{-}

$^{0.082}$=0.921) for each degree Celsius (˚C) increase in temperature. These data demonstrated that RNA integrity significantly decreased with the length of storage time and temperature.

The impact of storage time on RNA integrity (RIN values) was also analyzed at each temperature for each tube type (**Table S3**). RIN values were significantly higher in Tempus™ than PAXgene® tubes at 30˚C (p=0.0001) and 35˚C (p=0.001). Although, no significant differences in RIN at 25˚C (p=0.787) and 40˚C (p=0.399) for tube types (**Figure 6-2A**). In agreement with previously published studies (9, 16), these results demonstrated that RNA integrity is temperature-sensitive, and both tube types produced low-quality RNA at increased storage times and temperatures. Nevertheless, our data suggest that Tempus™ tubes may provide better RNA integrity (higher RIN values) under certain suboptimal tropical conditions compared to PAXgene® tubes.

6.4.6 Tempus™ tubes maintain mRNA integrity across suboptimal tropical conditions.

In order to validate our RNA quality measurements, we quantified mRNA and rRNA extracted from PAXgene® and Tempus™ tubes using RT-qPCR. We tested the relative mRNA abundance of two human reference genes, *Succinate dehydrogenase complex, subunit* A (*SDHA*) and *TATA-box-binding protein* (*TBP*), and one rRNA transcript *18S ribosomal RNA* (*18S*). The RNA concentration of all samples was normalized at pre-cDNA synthesis (*i.e.,* at 30 ng/µL). Hence, an increasing cycle threshold (Ct) value indicated a decreasing relative transcript quality rather than abundance (25). We tested RT-qPCR primer sets designed to amplify different-sized fragments of the same target gene (*i.e.,* amplicons between 100-300 bp) and differences in the relative RNA quality (*i.e*., increased Ct values) would be expected to be intensified when assaying genes with primers amplifying larger amplicons.

There was no significant difference in the mean Ct values between the tube types for Control samples from smaller amplicons (*p=ns,* paired t-test*:* D1/Control, 100-200 bp**)** or with larger amplicons (*p=ns,* paired t-test*:* D1/Control, 200-300 bp, **Figure 6-3)**. In addition, we found no statistically significant difference between matched 'Fresh' and 'Control' samples between PAXgene® and Tempus™ tubes (**Figure S4**). However, the Ct value varied significantly at tropical storage conditions across all three tested genes across short (100-200 bp) and medium amplicons (200-300 bp) (**Figure 3**). Tempus™ tubes maintained significantly higher transcript stability, as indicated by lower Ct values obtained for three tested genes compared to PAXgene® tubes at suboptimal tropical conditions (**Table S5**).

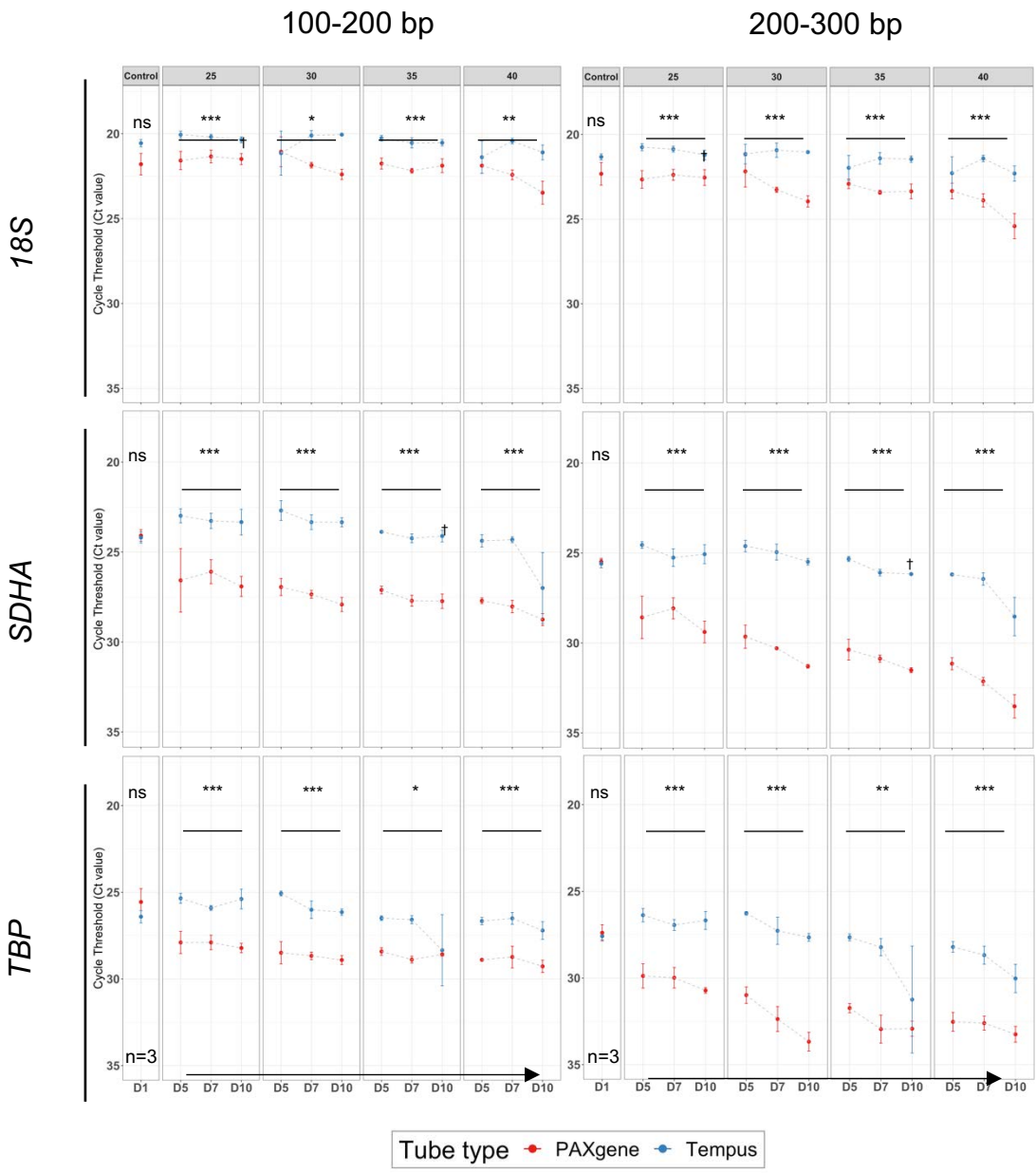


Figure 6-3 Cycle thresholds (Ct) of housekeeping genes assessing RNA from PAXgene® and Tempus™ stored at different temperatures and times.

The mean cycle threshold (Ct) values for *18S*, *SDHA* and *TBP* with100-200 base pair (bp) (short-amplicon) or 200-300 bp (medium-amplicon) lengths across different conditions [Control and at multiple storage temperatures and days (matched n=3)]. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns – non-significant. † = data from only two observations were potentially available due to LOD. Blue: Tempus™ Blood RNA tubes, Red: PAXgene® Blood RNA tubes.

Multiple comparison testing found that the *18S* rRNA Ct values were not statistically significantly influenced by incubation temperature or duration when rRNA was collected in either PAXgene® or Tempus™ tubes. In contrast, mRNA (*SDHA* and *TBP*) collected in PAXgene® tubes were significantly impacted by storage time and temperature compared to Tempus™ tubes (**Table S6**).

We showed that the RNA degraded samples, as indicated by the decreasing level of RIN had higher Ct values (**Figure 6-4A**). The Ct shifted towards higher cycle numbers for *SDHA* with larger amplicons than short and medium length amplicons, which was much more evident in PAXgene® tubes than in Tempus™ (**Figure 6-4A**). These results indicated that relative overall stability in terms of mRNA expression levels were maintained in Tempus™ compared to PAXgene® tubes. A similar relationship between RIN and Ct values was observed for *TBP* (**Figure S5**). However, **Figure S5** clearly shows that both tube types had Ct < 30 for all product lengths for *18S,* suggesting both PAXgene® and Tempus™ tubes preserved rRNA at suboptimal tropical conditions. As indicative of decreasing relative transcript quality, increasing Ct values were validated by correlating change in Ct values with RIN. We considered the change in Ct values (ΔCt) as the difference between samples collected under suboptimal tropical conditions and the mean of the 'Control' samples. Strong statistically significant correlations were found between ΔCt and RIN for all tested genes **(Figure 6-4B)**. These negative correlations indicated that with the decreasing RIN values, the ΔCt of 200-300bp amplicons increased, thus validating the use of RT-qPCR to assess the quality of the RNA. Taken together, these data demonstrate that Tempus™ collection tubes better maintain mRNA stability in suboptimal tropical conditions even if RIN is significantly decreasing.

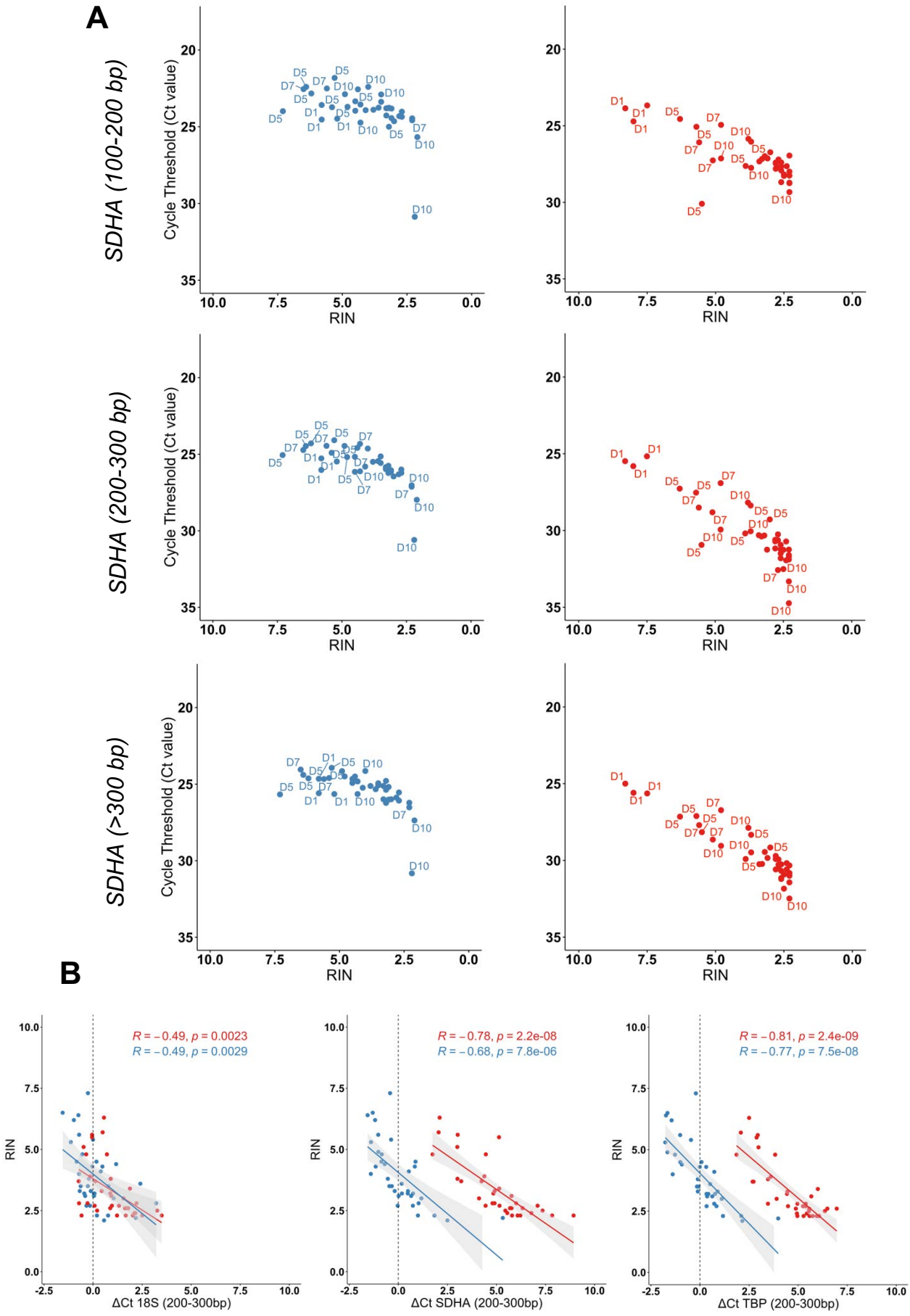

A
SDHA (100-200 bp)
SDHA (200-300 bp)
SDHA (>300 bp)
Cycle Threshold (Ct value)
RIN
B
R = −0.49, p = 0.0023
R = −0.49, p = 0.0029
R = −0.78, p = 2.2e-08
R = −0.68, p = 7.8e-06
R = −0.81, p = 2.4e-09
R = −0.77, p = 7.5e-08
ΔCt 18S (200-300bp)
ΔCt SDHA (200-300bp)
ΔCt TBP (200-300bp)

Figure 6-4 Cycle thresholds (Ct) in dependence on amplicon length and RNA integrity (RIN).

(A) Scatter plots show the Ct values in dependence on amplicon length and RNA integrity (RIN) for different lengths of SDHA amplicons differences for each tube type. **(B)** Spearman correlation of the ΔCt values with RIN value (RNA quality) for medium length amplicons (200-300 bp) of 18S, (left), SDHA (middle) and TBP (right). Blue: Tempus™ Blood RNA tubes, Red: PAXgene® Blood RNA tubes.

To test if the presence of PCR inhibitors, which are often co-extracted from whole-blood (*i.e.*, haemoglobin, lactoferrin, anticoagulants, etc.) (26, 27), could have contributed to these results, RT-qPCR was performed on a $\log_2$ serial dilution of undiluted extraction eluent. We considered that a trendline gradient of Ct values relative to the dilution greater than -3.3 (*i.e.,* E' < 100%) is indicative of the presence of PCR inhibitors (26). There was no apparent effect of inhibitors in both PAXgene® or Tempus™ tubes when the samples were diluted below 60ng/uL (**Figure S3**). These data demonstrated that our findings were unlikely to be a consequence of inhibitors present in the RT-qPCR reaction.

In summary, our data showed that Tempus™ tubes maintained a higher RNA quantity and integrity comparatively to PAXgene® tubes when RNA is stored in suboptimal tropical conditions. Furthermore, Tempus™ tubes maintained stability of mRNA in conditions where RNA samples were heavily degraded as indicated by RIN. Taken together, this study establishes that the Tempus™ blood RNA collection system resulted in a better quality of RNA and enhanced stability of mRNA when whole-blood samples are stored under suboptimal tropical conditions.

6.5 Discussion

Gene expression profiling with molecular techniques such as RT-qPCR and next-generation sequencing requires high-quality integral RNA. It is well established that the pre-analytical variables in blood sample collection and processing have profound effects on RNA quality that may consequently introduce substantial technical bias for molecular analysis (11, 28). Pre-analytical handling of blood samples and storage can be more challenging in tropical remote field study settings where freezing at -80°C immediately post-collection may not be an option. Here, we evaluated PAXgene® and Tempus™ blood RNA stabilization tubes for preserving RNA quantity, purity, quality, and gene transcript stability at suboptimal tropical conditions.

According to the respective manufacturers, PAXgene® blood RNA tubes effectively stabilize RNA for up to three days at room temperature, five days at 2–8°C and up to 11 years at –20°C or –70°C; whilst Tempus™ blood RNA tubes stabilize RNA for up to five days at room temperature, at least a week at 4 °C or –80 °C for long-term storage (29, 30). Duale *et al.* showed that the RNA yield, quality and integrity were stable up to six years of storage at −80°C in Tempus™ blood RNA tubes (8).

This study determined the impact of warm tropical temperatures (25, 30, 35 and 40˚C) and prolonged storage times (0, 5, 7 and 10 days) on total RNA yield, purity, quality and transcript stability of the two most widely used commercially available blood RNA stabilizing systems, PAXgene® and Tempus™. These conditions were selected to simulate condition in field sites in tropical or subtropical regions for a 10-day field trip, representing a challenging situation for preserving RNA. The performances of commercially available kits with columns (spin-column-based) and MagMAX™ (magnetic beads based) protocols were used to extract RNA from blood collected in PAXgene® and Tempus™ tubes. The total RNA yield, RNA integrity (RIN) and purity were used as performance measures.

In Experiment A, both Columns *vs.* MagMAX™ extraction methods had no significant differences in normalized total RNA yields, RIN and A260/280 ratios in PAXgene® and Tempus™ tubes. However, total RNA obtained from Tempus™ tubes were significantly lower than PAXgene® tubes. We observed comparable average OD 260/280 ratios and RIN >7 for column and MagMAX™ extraction methods for blood collected in PAXgene® and Tempus™ tubes. One limitation of Experiment A was that only two biological subjects were evaluated for Tempus™ column *vs.* MagMAX™ extractions due to the unavailability of Tempus™ blood collection tubes of the same batch. An A260/A280 ratio between 1.8 to 2.2 indicates highly

purified RNA with minimum DNA contamination (31). These data demonstrated that a similar quantity, quality and purity of RNA could be obtained using either spin-column or magnetic-bead RNA purifications when purifying RNA from whole-blood collected in either PAXgene® or Tempus™ systems. Higher RIN value indicates better RNA integrity, and RIN values above seven are considered ideal for high-throughput downstream applications (32, 33). However, RNA samples with a RIN of five have been used in gene expression studies (14, 34). MagMAX™ RNA extractions were used as the method of RNA purification for the rest of the study.

Both PAXgene® and Tempus™ tubes gave similar RNA yields and purity (A260/A280) when the samples were extracted Fresh or post-freezing, whilst RIN values were significantly higher in blood samples extracted from PAXgene® than Tempus™. The reason for this difference is not yet known. Overall RIN values were much lower in control samples compared to freshly extracted RNA. A previous study investigating the impact of storage duration (24, 32 and 40hr) and storage temperatures (24˚C, 4˚C and -80˚C) of whole-blood collected in heparin tubes on the qualities of DNA and RNA showed that RNA integrity declined dramatically when the samples were frozen (9). Freezing blood samples will lead to irreversible cellular damages causing osmotic and ice injuries of red blood cells due to water crystallization (35). Activated intracellular enzymes such as RNases can be released upon thawing resulting in RNA degradation (36).

The linear regression model on normalized RNA concentrations indicated that the Tempus™ tubes result in a higher RNA yield than PAXgene® while we demonstrated that the tube type and temperature significantly affect RNA yields. However, similar RNA purities were obtained from both tube types. A similar study by Duale et al., comparing PAXgene® *vs*. Tempus™ tubes stored for 0, 2, 5, and 7 days at RT (~22˚C) and then stored at -80°C until extraction, showed that RNA yields collected in the Tempus™ tubes were consistently higher than PAXgene® tubes. However, RNA quality (average 260/280 ratios and RIN values) was similar in both systems (17). Consistent with our study, other studies have reported that higher RNA yields were obtained with Tempus™ tubes compared to PAXgene® tubes (37-40). However, we observed low A260/230 ratios (less than 2.0) for both tube types in our study. This ratio is decreased in the presence of residual phenol, salts and carbohydrates that can affect the accuracy of downstream application and used as a secondary measurement for RNA purity (41). Historically, low A260/A230 ratios are reported attributed to the high salt content of the elution buffers contained in PAXgene® extraction kits (16,17,34,35) and as well as in Tempus™ extraction kits (16, 25, 42).

Our results indicated a gradual decrease of RNA quality in terms of RIN values over storage duration and increased temperatures in both tube types. However, RNA extracted from Tempus™ tubes had improved RINs compared to PAXgene® at suboptimal tropical conditions. RNA integrity was influenced mainly by increased storage temperatures at higher temperatures. It has been well documented that RNA molecules are sensitive to physical degradation due to high temperatures (43). However, the effects of higher temperatures on Tempus™ and PAXgene® blood RNA stabilizing systems have not been previously studied. Our data suggested that good quality RNA (average RIN >5) can be obtained in both tube types when samples are kept at 25˚C room temperature for up to 5 days of storage duration. Overall, we demonstrated that satisfactory amounts of good quality RNA can be achieved using blood RNA stabilizing systems in warm tropical temperatures (25-30˚C) and at storage times up to a week. When the storage temperatures are above 30°C, the RNA quality drops significantly and may not be adequate in downstream applications. RT-qPCR data further demonstrated that PAXgene® tubes do not preserve mRNA with the same efficiency as Tempus™ tubes, but both tubes equally preserved rRNA from degradation in suboptimal tropical conditions.

In most low-resource settings, microscopy and serological assays such as ELISAs remain the standard methods for diagnosis of tropical infections especially in low-income and middle-income countries, despite limited sensitivity and specificity. More sensitive molecular methods have potential to inform disease, diagnosis, and treatment, and to facilitate field-based intervention and biobanking studies (*i.e.*, large-scale field trials). This report provides important information to facilitate such studies. In particular, our data show that the Tempus™ blood RNA collection system resulted in higher quality RNA and maintained more consistent stability of mRNA when whole-blood samples were stored under suboptimal tropical conditions, as compared with the PAXgene® system.

We conclude that collection of whole-blood samples in Tempus™ tubes is the preferred system of choice for gene expression and molecular studies in rural and remote resource-limited settings where electricity and storage facilities are compromised. Our findings are especially relevant to facilitate diagnosis, treatment and interventions against disease that are prevalent in tropical countries, including neglected tropical diseases caused by helminth, bacterial, protozoan, viral, and fungal infections.

6.6 References

1. Mohr S, Liew C-C. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends in molecular medicine. 2007;13(10):422-32.
2. Zaas AK, Chen M, Varkey J, Veldman T, Hero III AO, Lucas J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host & Microbe. 2009;6(3):207-17.
3. Holcomb ZE, Tsalik EL, Woods CW, McClain MT. Host-based peripheral blood gene expression analysis for diagnosis of infectious diseases. Journal of Clinical Microbiology. 2017;55(2):360-8.
4. Gliddon HD, Herberg JA, Levin M, Kaforou M. Genome-wide host RNA signatures of infectious diseases: discovery and clinical translation. Immunology. 2018;153(2):171-8.
5. Gautam A, Donohue D, Hoke A, Miller SA, Srinivasan S, Sowe B, et al. Investigating gene expression profiles of whole blood and peripheral blood mononuclear cells using multiple collection and processing methods. PLoS One. 2019;14(12):e0225137.
6. Browne DJ, Brady JL, Waardenberg AJ, Loiseau C, Doolan DL. An analytically and diagnostically sensitive RNA extraction and RT-qPCR protocol for peripheral blood mononuclear cells. Frontiers in immunology. 2020;11:402.
7. Kukurba KR, Montgomery SB. RNA sequencing and analysis. Cold Spring Harbor Protocols. 2015;2015(11):pdb. top084970.
8. Duale N, Lipkin WI, Briese T, Aarem J, Rønningen KS, Aas KK, et al. Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank–evaluation of RNA quality and stability. BMC Research nNotes. 2014;7(1):1-10.
9. Huang L-H, Lin P-H, Tsai K-W, Wang L-J, Huang Y-H, Kuo H-C, et al. The effects of storage temperature and duration of blood samples on DNA and RNA qualities. PloS One. 2017;12(9):e0184692.
10. Malentacchi F, Pizzamiglio S, Wyrich R, Verderio P, Ciniselli C, Pazzagli M, et al. Effects of transport and storage conditions on gene expression in blood samples. Biopreservation and Biobanking. 2016;14(2):122-8.
11. Malentacchi F, Pazzagli M, Simi L, Orlando C, Wyrich R, Günther K, et al. SPIDIA-RNA: second external quality assessment for the pre-analytical phase of blood samples used for RNA based analyses. PloS One. 2014;9(11):e112293.
12. Qin J, Williams TL, Fernando MR. A novel blood collection device stabilizes cell-free RNA in blood during sample shipping and storage. BMC Research Notes. 2013;6(1):1-8.

13. Donohue DE, Gautam A, Miller S-A, Srinivasan S, Abu-Amara D, Campbell R, et al. Gene expression profiling of whole blood: a comparative assessment of RNA-stabilizing collection methods. PloS One. 2019;14(10):e0223065.
14. Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, et al. Impact of RNA degradation on gene expression profiling. BMC medical genomics. 2010;3(1):1-14.
15. Langelaan M, Dylus J, Bock E, Jongen B, Mertens A, Raijmakers M. Improved pre-analytical process for RNA isolation from whole blood samples. Ned Tijdschr Klin Chem Labgeneesk. 2014;39(3):164-5.
16. Feddersen S, Bastholt L, Pedersen SM. Stabilization of circulating thyroglobulin mRNA transcripts in patients treated for differentiated thyroid carcinoma. Annals of Clinical Biochemistry. 2017;54(5):558-66.
17. Duale N, Brunborg G, Rønningen KS, Briese T, Aarem J, Aas KK, et al. Human blood RNA stabilization in samples collected and transported for a large biobank. BMC Research Notes. 2012;5(1):1-9.
18. Aarem J, Brunborg G, Aas KK, Harbak K, Taipale MM, Magnus P, et al. Comparison of blood RNA isolation methods from samples stabilized in Tempus tubes and stored at a large human biobank. BMC Research Notes. 2016;9(1):1-14.
19. Yip L, Fuhlbrigge R, Atkinson MA, Fathman CG. Impact of blood collection and processing on peripheral blood gene expression profiling in type 1 diabetes. BMC genomics. 2017;18(1):1-16.
20. Weber DG, Casjens S, Rozynek P, Lehnert M, Zilch-Schöneweis S, Bryk O, et al. Assessment of mRNA and microRNA stabilization in peripheral human blood for multicenter studies and biobanks. Biomarker insights. 2010;5:BMI. S5522.
21. Menke A, Rex-Haffner M, Klengel T, Binder EB, Mehta D. Peripheral blood gene expression: it all boils down to the RNA collection tubes. BMC Research Notes. 2012;5(1):1-8.
22. Browne DJ, Kelly AM, Brady JL, Doolan DL. A high-throughput screening RT-qPCR assay for quantifying surrogate markers of immunity from PBMCs. Frontiers in immunology. 2022;13:962220-.
23. Spandidos A, Wang X, Wang H, Seed B. PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. Nucleic acids research. 2010;38(suppl_1):D792-D9.

24. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments. Oxford University Press; 2009.
25. Eikmans M, Rekers NV, Anholts JD, Heidt S, Claas FH. Blood cell mRNAs and microRNAs: optimized protocols for extraction and preservation. Blood, The Journal of the American Society of Hematology. 2013;121(11):e81-e9.
26. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors–occurrence, properties and removal. Journal of applied microbiology. 2012;113(5):1014-26.
27. Al-Soud WA, Rådström P. Purification and characterization of PCR-inhibitory components in blood cells. Journal of clinical microbiology. 2001;39(2):485-93.
28. Zhang H, Korenková V, Sjöback R, Švec D, Björkman J, Kruhøffer M, et al. Biomarkers for monitoring pre-analytical quality variation of mRNA in blood samples. PloS One. 2014;9(11):e111644.
29. PAXgene Blood RNA System. Technical Note 2018 [cited 2021 Mar 24]. Available from: https://www.preanalytix.com/storage/download/_ProductResources_/TechnicalNotes/PROM-7266-002_BD-7969_TN_Blood_RNA_System_RNA_stability_over_11_years_storage_1118_WW_WEB.pdf.
30. Applied Biosystems™ Tempus™ Blood RNA Tube [cited 2021 Mar 24]. Available from: https://www.thermofisher.com/order/catalog/product/4342792#/4342792.
31. Mlcochova H, Hezova R, Stanik M, Slaby O, editors. Urine microRNAs as potential noninvasive biomarkers in urologic cancers. Urologic Oncology: Seminars and Original Investigations; 2014: Elsevier.
32. Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Koues O, et al. Multi-perspective quality control of Illumina RNA sequencing data analysis. Briefings in functional genomics. 2017;16(4):194-204.
33. Puchta M, Boczkowska M, Groszyk J. Low RIN Value for RNA-Seq Library Construction from Long-Term Stored Seeds: A Case Study of Barley Seeds. Genes. 2020;11(10):1190.
34. Shen Y, Li R, Tian F, Chen Z, Lu N, Bai Y, et al. Impact of RNA integrity and blood sample storage conditions on the gene expression analysis. OncoTargets and therapy. 2018;11:3573.

35. Yang J, Sui X, Wen C, Pan C, Zhu Y, Zhang J, et al. A hemocompatible cryoprotectant inspired by freezing-tolerant plants. Colloids and Surfaces B: Biointerfaces. 2019;176:106-14.
36. Koh EJ, Yu SY, Kim SH, Kim SJ, Lee E-I, Hwang SY. Understanding confounding effects of blood handling strategies on RNA quality and transcriptomic alteration using RNA sequencing. BioChip Journal. 2021;15(2):187-94.
37. Asare AL, Kolchinsky SA, Gao Z, Wang R, Raddassi K, Bourcier K, et al. Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation. BMC genomics. 2008;9(1):1-10.
38. Häntzsch M, Tolios A, Beutner F, Nagel D, Thiery J, Teupser D, et al. Comparison of whole blood RNA preservation tubes and novel generation RNA extraction kits for analysis of mRNA and MiRNA profiles. PLoS One. 2014;9(12):e113298.
39. Matheson LA, Duong TT, Rosenberg AM, Yeung RS. Assessment of sample collection and storage methods for multicenter immunologic research in children. Journal of immunological methods. 2008;339(1):82-9.
40. Nikula T, Mykkänen J, Simell O, Lahesmaa R. Genome-wide comparison of two RNA-stabilizing reagents for transcriptional profiling of peripheral blood. Translational Research. 2013;161(3):181-8.
41. Roy D, Tomo S, Modi A, Purohit P, Sharma P. Optimising total RNA quality and quantity by phenol-chloroform extraction method from human visceral adipose tissue: A standardisation study. MethodsX. 2020;7:101113.
42. Carrillo-Ávila J, de la Puente R, Catalina P, Rejón J, Espín-Vallejo L, Valdivieso V, et al. Evaluation of RNA purification methods by using different blood stabilization tubes: identification of key features for epidemiological studies. BMC Research Notes. 2020;13(1):1-7.
43. Sidova M, Tomankova S, Abaffy P, Kubista M, Sindelka R. Effects of post-mortem and physical degradation on RNA integrity and quality. Biomolecular detection and quantification. 2015;5:3-9.

6.7 Supplementary information

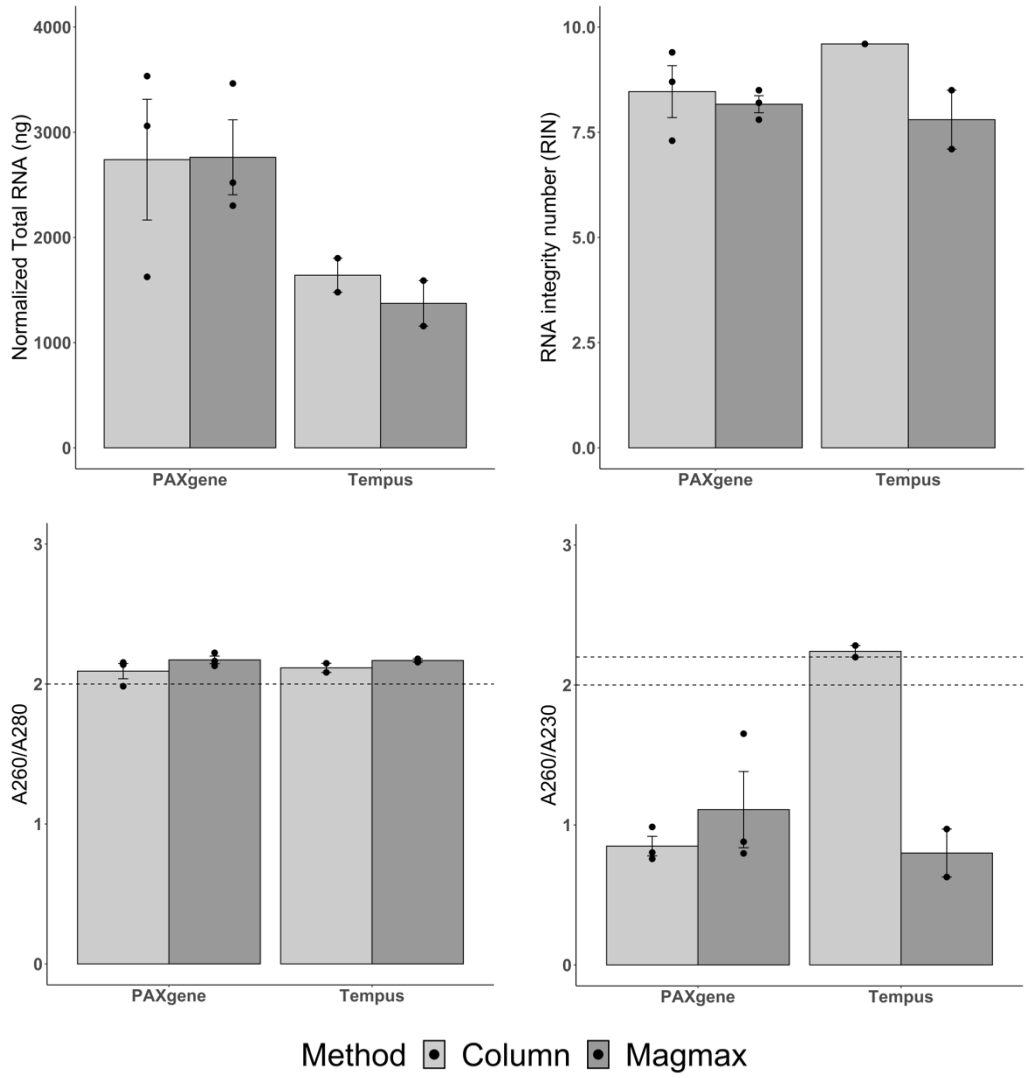


Figure S1: Normalized total RNA yield, integrity, and purity comparisons between RNA isolation protocols (silica-membrane column-based vs. magnetic beads) for PAXgene® and Tempus™ Blood RNA tubes. Bar graphs showing total RNA yields normalized to input whole blood volume, RNA Integrity Number (RIN), OD A260/280 and A260/230 ratios for silica-membrane column-based and magnetic beads based (MagMAX™) RNA purification protocols compatible with PAXgene® and Tempus™ blood collection tubes (matched n=3). Each bar represents the mean value, and the error bar indicates ± SE. Data from only two subjects were available on the Tempus™ collection system due to the unavailability of Tempus™ blood collection tubes from the same batch. The dashed lines at A260/A280 ratio

>2.0 and A260/A230 ratio 2.0-2.2 represent the generally accepted range for high-quality RNA. Ns, non-significant.

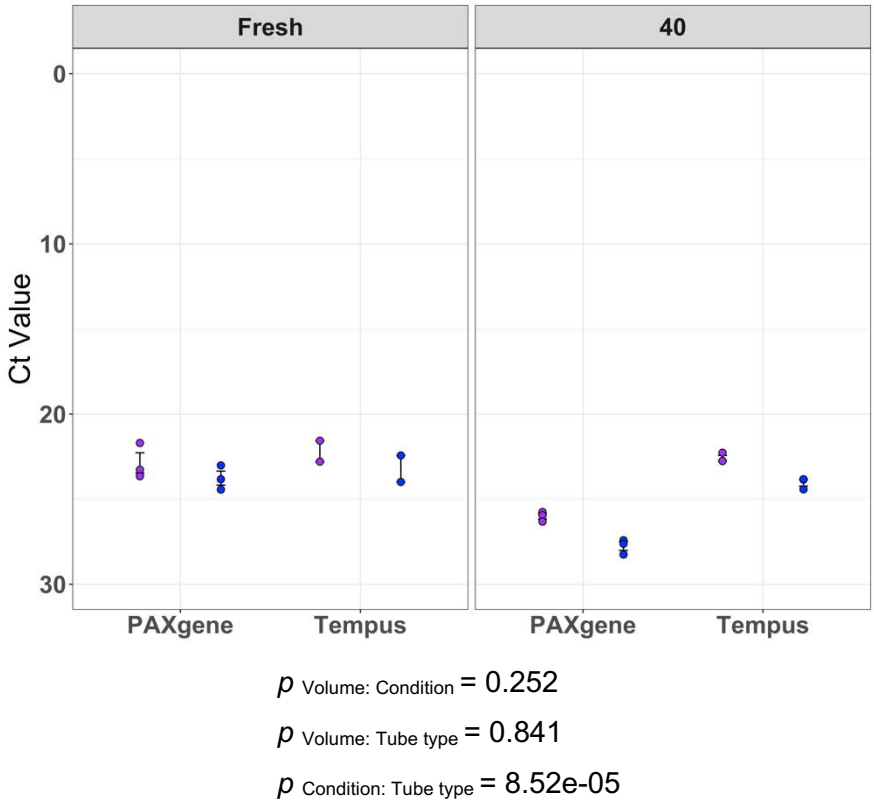


Figure S2: Reverse transcription at Full *vs.* Quarter concentration. The graph depicts RT-qPCR efficiency at reverse transcription using 5U (5units/µL; Quarter) reactions compared with 20U (Full) reactions (manufacturer's recommended protocol) for high-quality RNA (*i.e.*, freshly extracted) and low-quality RNA (*i.e.*, incubated 5 days at 40˚C) extracted from the two RNA stabilization systems (PAXgene® and Tempus™). Unpaired t-test results determined no differences between Full *vs.* Quarter reaction concentrations on Fresh extractions while Full strength of RT was significantly efficient compared to the Quarter concentration of RT when evaluating low-quality, degraded RNA. Two-Way ANOVA analysis indicated a significant interaction between tube type and their condition/RNA quality (Fresh or 5 days at 40˚C), indicating that the relationships between Ct value and reverse transcription concentration depend on tube type and RNA quality. $^{**}p< 0.01$, $^{*}p< 0.05$, ns – non-significant.

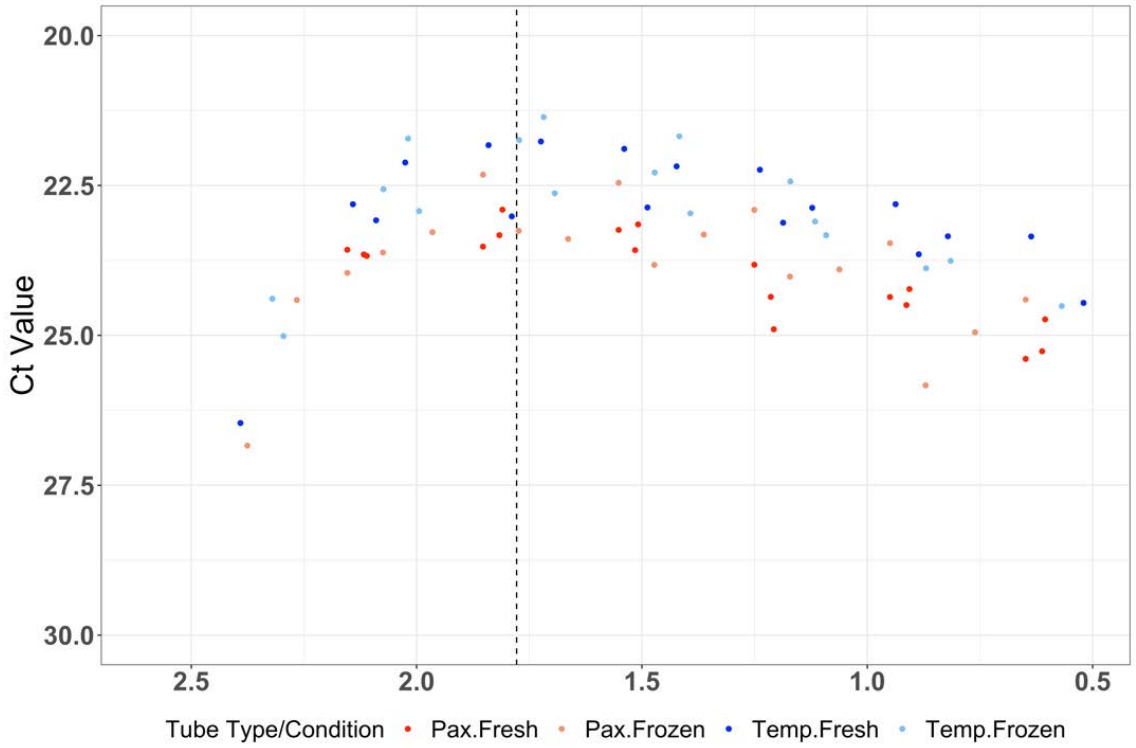


Figure S3: qPCR inhibitory behaviour at serially diluted (log_2 dilutions) total RNA samples measured with SDHA (short primers) and differences between tube types. Whole blood extracted at Fresh and Frozen conditions from the two different tube types (PAXgene® and Tempus™) were tested to investigate if there were any lingering inhibitory factors from PAXgene® or Tempus™ tubes A. Ct results generated at Log-linear amplification-capable sample dilutions exhibit a straight line (beyond the dashed vertical line, at 30ng/uL), while inhibited dilutions form a curved line (before the dashed vertical line, at lower dilutions >30ng/uL). Tube type-based inhibition would be expected if present data shifted across the x-axis, with inhibitory factors in tubes diluting out each titration, which is not observed here.

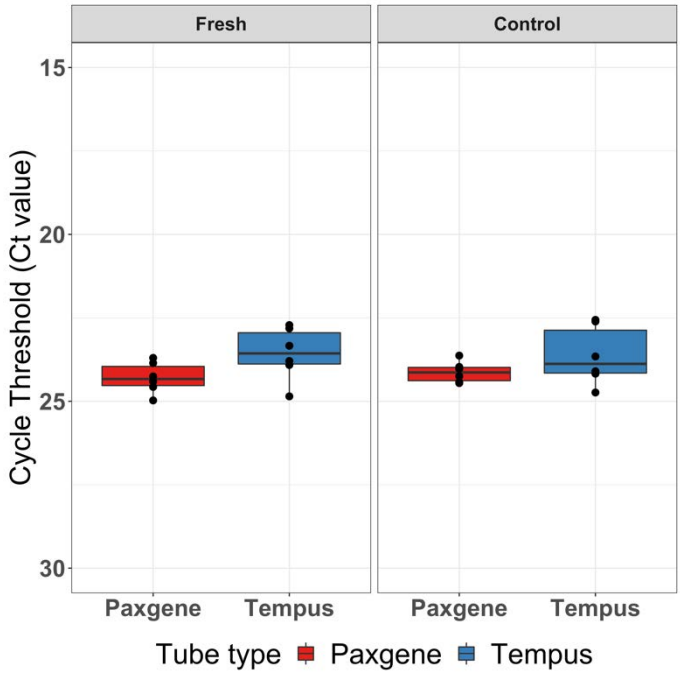


Figure S4: Cycle threshold (Ct) of the housekeeping gene, *SDHA* assessed for RNA samples extracted from PAXgene® and Tempus™ at 'Fresh' and 'Control' conditions. Box plots shows Ct values for *SDHA* with100-200 base pair (bp) (short-amplicon) for Fresh *vs.* Day1/Control; immediately frozen at -80°C for later extraction (matched n=6). Paired t-test results determined no overall differences between Fresh *vs.* Control conditions on Ct values.

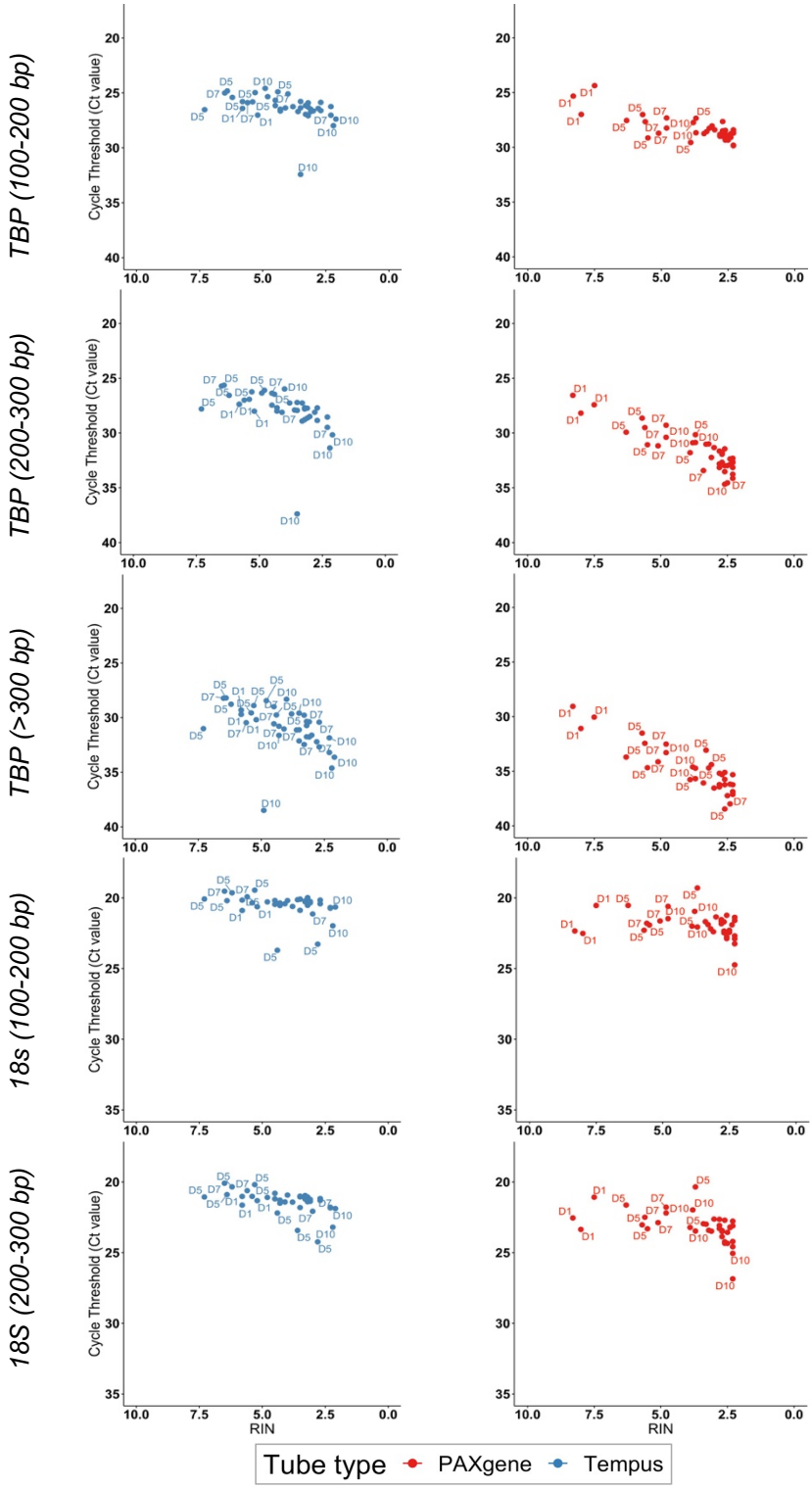


Figure S5: Cycle thresholds (Ct) in dependence on amplicon length and RNA integrity (RIN) for *TBP* and *18s*. The scatter plots show the Ct values in dependence on amplicon length and RNA integrity (RIN) for different amplicon lengths of *TBP* and *18s* when extracted from PAXgene® and Tempus™ at different storage lengths.

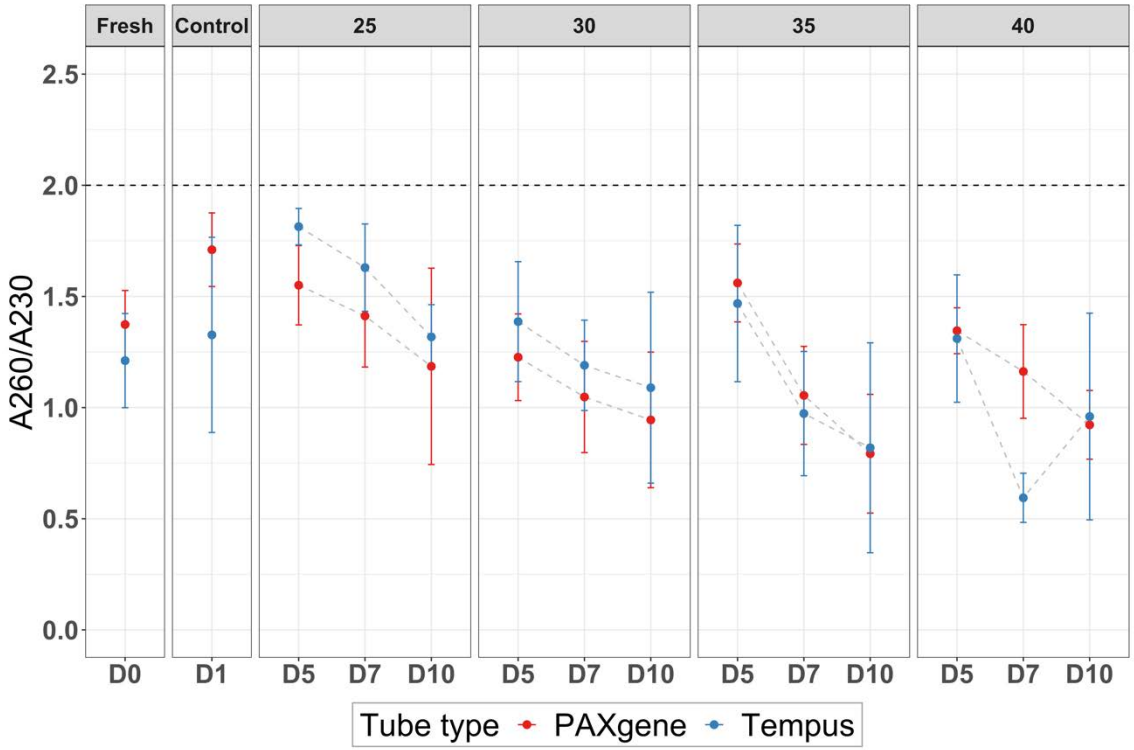


Figure S6 A260/230 ratios for PAXgene® and Tempus™ tubes blood collection systems under suboptimal tropical storage conditions. A260/A230 ratios determined spectrophotometrically across the different conditions [Fresh (n=8), Day1/Control; immediately frozen at -80°C for later extraction and various temperatures (25, 30, 35, 40°C) (matched subjects n=3)]. The dashed line indicates OD260/A230 = 2.0 for high-quality RNA.

Table S1 Summarised technical characteristics of RNA extraction methods from PAXgene® and Tempus™ Blood RNA Tubes.

Blood stabilization tube type	**PAXgene® Blood RNA Tubes**		**Tempus™ Blood RNA tubes**	
RNA isolation protocol	PAXgene® Blood RNA Kit	MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit, compatible with PAXgene® Blood RNA tubes	Tempus™ Spin RNA Isolation Kit	MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit, compatible with Tempus™ Blood RNA tubes
Blood volume	2.5mL	2.5mL	3mL	3mL
RNA isolation method	spin columns	magnetic beads	spin columns	magnetic beads
RNA processing method	manual	manual	manual	manual
Proteinase K treatment requirement	yes	yes	no	yes
DNase digestion implemented in Kit	yes	yes	no (optional)	yes
Elution volume (µL)	40	20-80	90	20-80
Expected RNA yield	>3µg	3-18µg	6-25µg	3-25µg

Table S2 List of primers used in this study for quantitative real-time PCR and their efficiencies.

Regression line fit to raw Cq values against log2 diluted cDNA concentrations. R^2 values were calculated using Pearson's correlation. Efficiencies were calculated as per MIQE guidelines (E = 10^((-1/slope)-1)) using the slope of the regression line previously described.

Gene	GenBank Accession	Primer Bank ID	Primer Sequence	Amplicon Length (bp)	Standard Curve linear equation	R^2	Efficiency (%)
TBP (Short)	NM_003194	285026518c1	**Forward** CCACTCACAGACTCTCACAAC **Reverse** CTGCGGTACAATCCCAGAACT	127	y = -3.203x + 23.385	0.995	105.21
TBP (Mid)	NM_003194	285026518c2	**Forward** CCACTCACAGACTCTCACAAC **Reverse** AATCAGTGCCGTGGTTCGTG	274	y = -3.0639x + 24.876	0.984	112.02
SDHA (Short)	NM_004168	156416002c1	**Forward** CAAACAGGAACCCGAGGTTTT **Reverse** CAGCTTGGTAACACATGCTGTAT	201	y = -3.2634x + 20.886	0.977	102.46
SDHA (Mid)	NM_004168	156416002c2	**Forward** CAAACAGGAACCCGAGGTTTT **Reverse** GGTGTCGTAGAAATGCCACCT	301	y = -3.4536x + 22.792	0.993	94.78
			Forward				

18s (Short)	NM_022551	14165467c1	GCGGCGGAAAATAGCCTTTG **Reverse** GATCACACGTTCCACCTCATC	139	y = -3.3503x + 17.163	0.996	98.83
18s (Mid)	NM_022551	14165467c2	**Forward** GCGGCGGAAAATAGCCTTTG **Reverse** GACCTGGCTGTATTTTCCATCC	229	y = -3.3656x + 17.715	0.993	98.21

Table S3 Multiple linear regression model outputs for RNA concentration, A260/A280 and RIN values at each temperature condition.

RNA concentration (ng/µL) - Spectrophotometer				
25˚C (adj-R^2 = 0.832 (*p* = 3.494e-05)				
	Estimate	Std. Error	Statistic	*p* value
(Intercept)	73.893	16.145	4.577	0.0006***
Tube type: Tempus	98.962	13.182	7.507	7.16e-06***
Days7	0.988	16.145	0.061	0.952
Days10	-15.822	16.145	-0.980	0.346
Subject B	-31.642	16.145	-1.960	0.074
Subject C	57.528	16.145	3.563	0.003900**
30˚C (adj-R^2 = 0.761 (*p* = 0.0003)				
(Intercept)	50.483	17.482	2.888	0.014*
Tube type: Tempus	69.487	14.274	4.868	0.0004***
Days7	-0.655	17.482	-0.037	0.971
Days10	3.330	17.482	0.190	0.852
Subject B	-34.077	17.482	-1.949	0.075
Subject C	67.992	17.482	3.889	0.002**
35˚C (adj-R^2 = 0.484 (*p* = 0.020)				
(Intercept)	79.980	23.110	3.460	0.005***
Tube type: Tempus	45.300	18.870	2.401	0.033*
Days7	-25.120	23.110	-1.087	0.298
Days10	-39.210	23.110	-1.696	0.116
Subject B	-30.830	23.110	-1.334	0.207
Subject C	49.340	23.110	2.135	0.054
40˚C (adj-R^2 = 0.630 (*p* = 0.003)				
(Intercept)	74.480	13.910	5.355	0.0002***
Tube type: Tempus	29.530	11.360	2.600	0.023*
Days7	-29.340	13.910	-2.109	0.057
Days10	-19.930	13.910	-1.433	0.177
Subject B	-38.470	13.910	-2.766	0.017*
Subject C	27.260	13.910	1.960	0.074

A260/A280				
25°C (adj-R^2 =-0.186 (p = 0.794)				
	Estimate	Std.error	Statistic	p value
(Intercept)	2.196	0.022	101.726	<2e-16***
Tube type: Tempus	-0.018	0.018	-1.020	0.328
Days7	-0.006	0.022	-0.294	0.774
Days10	-0.006	0.022	-0.263	0.797
Subject B	0.006	0.022	0.271	0.791
Subject C	-0.017	0.022	-0.781	0.450
30°C (adj-R^2 = 0.469 (p = 0.023)				
(Intercept)	2.204	0.020	111.518	<2e-16***
Tube type: Tempus	-0.053	0.016	-3.277	0.007**
Days7	-0.010	0.020	-0.506	0.622
Days10	-0.024	0.020	-1.197	0.254
Subject B	-0.054	0.020	-2.732	0.018*
Subject C	-0.017	0.020	-0.856	0.409
35°C (adj-R^2 = 0.372 (p = 0.055)				
(Intercept)	2.189	0.029	74.899	<2e-16***
Tube type: Tempus	-0.072	0.024	-3.022	0.011*
Days7	-0.026	0.029	-0.898	0.387
Days10	-0.018	0.029	-0.605	0.557
Subject B	-0.056	0.029	-1.916	0.079
Subject C	0.002	0.029	0.071	0.944
40°C (adj-R^2 = 0.706 (p = 0.001)				
(Intercept)	2.172	0.018	117.830	<2e-16***
Tube type: Tempus	-0.093	0.015	-6.157	4.89e-05***
Days7	-0.018	0.018	-0.968	0.352
Days10	-0.027	0.018	-1.469	0.167
Subject B	-0.020	0.018	-1.058	0.311
Subject C	0.024	0.018	1.307	0.216
log_2(RIN)				
25°C (adj-R^2 = 0.666 (p = 0.002)				
	Estimate	Std.error	Statistic	p value

(Intercept)	2.608	0.081	32.135	5.21e-13***
Tube type: Tempus	0.018	0.066	0.276	0.787
Days7	-0.241	0.081	-2.967	0.012*
Days10	-0.483	0.081	-5.947	6.75e-05***
Subject B	-0.025	0.081	-0.303	0.767
Subject C	-0.142	0.081	-1.746	0.106
30°C (adj-R^2 = 0.744 (p = 0.0004)				
(Intercept)	1.868	0.106	17.642	6e-10***
Tube type: Tempus	0.472	0.086	5.458	0.0001***
Days7	-0.111	0.106	-1.043	0.317
Days10	-0.474	0.106	-4.477	0.001***
Subject B	0.016	0.106	0.147	0.885
Subject C	-0.138	0.106	-1.306	0.216
35°C (adj-R^2 = 0.623 (p = 0.004)				
(Intercept)	1.737	0.138	12.630	2.73e-08***
Tube type: Tempus	0.496	0.112	4.416	0.001***
Days7	-0.383	0.138	-2.783	0.017*
Days10	-0.437	0.138	-3.177	0.008**
Subject B	-0.113	0.138	-0.823	0.427
Subject C	-0.170	0.138	-1.239	0.239
40°C (adj-R^2 = 0.579 (p = 0.01)				
(Intercept)	1.500	0.063	23.676	1.93e-11***
Tube type: Tempus	0.045	0.052	0.875	0.399
Days7	-0.179	0.063	-2.822	0.015*
Days10	-0.317	0.063	-5.007	0.0003***
Subject B	0.023	0.063	0.355	0.729
Subject C	-0.072	0.063	-1.135	0.278

Table S4 Multiple linear regression model for RNA concentration measured by Agilent 2100 Bioanalyzer

Agilent 2100 Bioanalyzer	**RNA concentration (ng/µL)**			
Explanatory variable	Estimate	Std. Error	t value	*p* value
(Intercept)	57.627	214.498	0.269	0.789
Tube type: Tempus	438.154	128.773	3.403	0.001**
Days	14.905	27.193	0.548	0.586
Temperature	1.462	6.413	0.228	0.82
Tube type Tempus: Days	-7.112	9.09	-0.782	0.437
Tube type Tempus: Temperature	-7.8	3.341	-2.335	0.023*
Days: Temperature	-0.507	0.813	-0.624	0.535
Adjusted R-squared: 0.603 (*p* value: 9.591e-12)				
25°C (adj-R^2 = 0.584 (*p* = 0.006)				
	Estimate	Std.error	Statistic	*p* value
(Intercept)	163.830	51.120	3.205	0.008***
Tube type: Tempus	182.220	41.740	4.366	0.001***
Days7	41.830	51.120	0.818	0.429
Days10	-15.670	51.120	-0.306	0.765
Subject B	-111.670	51.120	-2.184	0.050*
Subject C	29.330	51.120	0.574	0.005**
30°C (adj-R^2 = 0.724 (*p* = 0.001)				
(Intercept)	79.170	42.050	1.882	0.084
Tube type: Tempus	180.330	34.340	5.252	0.0002***
Days7	-34.170	42.050	-0.812	0.432
Days10	26.670	42.050	0.634	0.538
Subject B	-72.830	42.050	-1.732	0.109
Subject C	112.830	42.050	2.683	0.020
35°C (adj-R^2 = 0.451 (*p* = 0.027)				
(Intercept)	68.280	48.030	1.421	0.181
Tube type: Tempus	82.670	39.220	2.108	0.057
Days7	22.000	48.030	0.458	0.655
Days10	-82.330	48.030	-1.714	0.112

Subject B	22.000	48.030	0.458	0.655
Subject C	136.170	48.030	2.835	0.015*
40°C (adj-R^2 = 0.596 (*p* = 0.005)				
(Intercept)	86.000	35.557	2.419	0.032*
Tube type: Tempus	84.778	29.032	2.920	0.013*
Days7	-3.833	35.557	-0.108	0.916
Days10	-22.833	35.557	-0.642	0.533
Subject B	-38.167	35.557	-1.073	0.304
Subject C	118.500	35.557	3.333	0.006**

*** $p< 0.001$, ** $p< 0.01$, * $p< 0.05$ from separate multiple linear regression models with total RNA and RIN values as dependent variables and tube type, day and subjects as the independent variables for each temperature point.

Table S5: Multiple linear regression model outputs for Ct values obtained for 18s, SDHA and TBP with 100-200 bp (short-amplicon) and 200-300 bp (medium-amplicon) lengths at each temperature condition. *** p< 0.001, **p< 0.01, *p< 0.05 from separate multiple linear regression models with Ct values obtained for 18s, SDHA and TBP with 100-200 bp (short-amplicon) and 200-300 bp (medium-amplicon) lengths as dependent variable and tube type, days and subjects as the independent variables for each temperature point.

Primer	Temperature	25° C				30 °C				35° C				40° C			
		adj-R2 = 0.880 (*p* = 3.397e-06)				**adj-R2 = 0.271 (*p* = 0.09)**				**adj-R2 = 0.792 (*p* = 4.168e-05)**				**adj-R2 = 0.587 (*p* = 0.003)**			
	Coefficients:	Estimate	Std.error	t value	*p* value	Estimate	Std.error	t value	*p* value	Estimate	Std.error	t value	*p* value	Estimate	Std.error	t value	*p* value
***18s* short**	(Intercept)	20.919	0.280	74.760	< 2e-16***	20.909	1.077	19.406	5.55e-11***	21.422	0.386	55.518	< 2e-16***	20.910	0.800	26.137	1.27e-12***
	Tube type: Tempus	-1.365	0.140	-9.759	4.66e-07***	-1.337	0.527	-2.539	0.025*	-1.494	0.189	-7.918	2.5e-06***	-1.621	0.391	-4.144	0.001**
	Days	-0.007	0.035	-0.188	0.854	0.029	0.128	0.225	0.826	0.032	0.046	0.693	0.501	0.143	0.095	1.505	0.156
	SubjectB	0.831	0.175	4.762	0.0005***	0.689	0.645	1.068	0.305	0.543	0.231	2.350	0.035*	0.502	0.479	1.049	0.313
	SubjectC	0.970	0.175	5.559	0.0001***	1.261	0.645	1.954	0.073	0.315	0.231	1.363	0.196	1.395	0.479	2.913	0.012*
		adj-R2 = 0.909 (*p* = 6.477e-07)				**adj-R2 = 0.675 (*p* = 0.0007)**				**adj-R2 = 0.601 (*p* = 0.002)**				**adj-R2 = 0.69 (*p* = 0.0005)**			
***18s* medium**	(Intercept)	21.905	0.291	75.175	< 2e-16***	21.371	0.772	27.672	6.12e-13***	22.949	0.6376	35.9930	2.09e-14***	21.913	0.855	25.620	1.64e-12***
	Tube type: Tempus	-1.690	0.146	-11.598	7.06e-08***	-2.093	0.378	-5.543	9.50e-05***	-1.6184	0.3117	-5.1920	0.0002***	-2.212	0.418	-5.290	0.0001***
	Days	0.006	0.036	0.159	0.877	0.163	0.092	1.773	0.100	-0.0066	0.0758	-0.0870	0.9318	0.224	0.102	2.205	0.046*
	SubjectB	0.619	0.182	3.407	0.005**	0.635	0.462	1.372	0.193	0.3688	0.3818	0.9660	0.3516	0.468	0.512	0.915	0.377
	SubjectC	1.124	0.182	6.183	4.70e-05***	1.074	0.462	2.322	0.037*	0.6235	0.3818	1.6330	0.1264	1.502	0.512	2.932	0.012*
		adj-R2 = 0.819 (*p* = 1.707e-05)				**adj-R2 = 0.945 (*p* = 8.36e-09)**				**adj-R2 = 0.948 (*p* = 2.369e-08)**				**adj-R2 = 0.629 (*p* = 0.002)**			
***SDHA* short**	(Intercept)	25.042	0.886	28.276	4.64e-13***	25.827	0.522	49.465	3.45e-16***	26.632	0.432	61.611	< 2e-16***	25.034	1.274	19.655	4.73e-11***
	Tube type: Tempus	-3.330	0.433	-7.690	3.44e-06***	-4.278	0.255	-16.761	3.49e-10***	-3.374	0.205	-16.495	1.31e-09***	-2.934	0.623	-4.713	0.0004***
	Days	0.075	0.105	0.712	0.489	0.157	0.062	2.526	0.025*	0.099	0.051	1.930	0.078	0.383	0.152	2.525	0.025*
	SubjectB	0.478	0.530	0.901	0.384	0.400	0.313	1.280	0.223	0.009	0.241	0.037	0.971	-0.259	0.763	-0.339	0.740
	SubjectC	2.325	0.530	4.385	0.001***	0.877	0.313	2.805	0.015*	0.474	0.255	1.857	0.088	1.229	0.763	1.611	0.131
		adj-R2 = 0.888 (*p* = 8.167e-07)				**adj-R2 = 0.974 (*p* = 7.015e-11)**				**adj-R2 = 0.978 (*p* = 1.431e-10)**				**adj-R2 = 0.922 (*p* = 8.004e-08)**			
***SDHA* medium**	(Intercept)	26.787	0.709	37.808	1.11e-14***	28.07672	0.45044	62.332	< 2e-16***	29.006	0.414	70.050	< 2e-16***	28.390	0.806	35.230	2.75e-14***
	Tube type: Tempus	-3.722	0.346	-10.745	7.77e-08***	-5.39356	0.220	-24.493	2.91e-12***	-4.990	0.196	-25.468	8.17e-12***	-5.211	0.394	-13.227	6.45e-09***
	Days	0.136	0.084	1.619	0.129	0.25264	0.05358	4.715	0.0004***	0.212	0.049	4.334	0.001***	0.480	0.096	5.003	0.0002***
	SubjectB	0.780	0.424	1.838	0.089	0.6715	0.2697	2.49	0.03*	0.458	0.231	1.983	0.071	0.161	0.483	0.333	0.744
	SubjectC	1.901	0.424	4.480	0.001***	0.7805	0.2697	2.894	0.01*	0.621	0.244	2.540	0.026*	0.924	0.483	1.914	0.078
		adj-R2 = 0.904 (*p* = 3.009e-07)				**adj-R2 = 0.850 (*p* = 5.179e-06)**				**adj-R2 = 0.284 (*p* = 0.08)**				**adj-R2 = 0.877 (*p* = 1.463e-06)**			

	(Intercept)	27.418	0.425	64.460	< 2e-16***	27.595	0.617	44.757	1.26e-15***	26.971	1.300	20.744	2.39e-11***	28.346	0.436	65.036	< 2e-16***
	Tube type: Tempus	-2.454	0.208	-11.804	2.55e-08***	-2.947	0.301	-9.776	2.34e-07***	-1.490	0.636	-2.345	0.036*	-2.177	0.213	-10.219	1.4e-07***
	Days	0.031	0.051	0.603	0.557	0.143	0.073	1.944	0.074	0.205	0.155	1.326	0.208	0.101	0.052	1.951	0.073
	SubjectB	-0.001	0.255	-0.004	0.997	-0.116	0.369	-0.315	0.758	-0.481	0.779	-0.617	0.548	-0.708	0.261	-2.712	0.018*
	SubjectC	1.083	0.255	4.251	0.001***	0.260	0.369	0.704	0.494	0.946	0.779	1.215	0.246	0.348	0.261	1.332	0.206
TBP medium		**adj-R2 = 0.945 (p = 8.574e-09)**				**adj-R2 = 0.679 (p = 0.001)**				**adj-R2 = 0.539 (p = 0.006)**				**adj-R2 = 0.947 (p = 6.753e-09)**			
	(Intercept)	29.290	0.448	65.452	< 2e-16***	29.492	0.760	38.784	7.98e-15***	28.527	1.908	14.952	1.44e-09***	31.112	0.490	63.472	< 2e-16***
	Tube type: Tempus	-3.529	0.219	-16.133	5.6e-10***	-5.278	0.372	-14.197	2.72e-09***	-3.499	0.933	-3.752	0.002**	-3.824	0.240	-15.956	6.42e-10***
	Days	0.111	0.053	2.084	0.057	0.397	0.090	4.391	0.001***	0.479	0.227	2.112	0.055	0.260	0.058	4.459	0.001***
	SubjectB	-0.584	0.268	-2.180	0.048*	-0.669	0.455	-1.470	0.165	-0.508	1.142	-0.445	0.664	-1.128	0.293	-3.844	0.002**
	SubjectC	0.859	0.268	3.206	0.007**	0.489	0.455	1.074	0.302	1.999	1.142	1.749	0.104	0.461	0.293	1.571	0.140

Table S6: Two-way ANOVA results with multiple comparisons outputs for control and test conditions on PAXgene® and Tempus™ tube types on Ct values obtained for 18s, SDHA and TBP. **** p< 0.0001, *** p< 0.001, **p< 0.01, *p< 0.05, ns – non-significant

Bonferroni's multiple comparisons test	Predicted (LS) mean diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted *p* value
18s short (100-200bp) PAXgene®					
25 - 1 vs. 25 - 5	0.216	-1.394 to 1.826	No	ns	>0.999
25 - 1 vs. 25 - 7	0.455	-1.155 to 2.065	No	ns	>0.999
25 - 1 vs. 25 - 10	0.303	-1.307 to 1.913	No	ns	>0.999
25 - 1 vs. 30 - 5	0.734	-0.876 to 2.344	No	ns	>0.999
25 - 1 vs. 30 - 7	-0.058	-1.668 to 1.552	No	ns	>0.999
25 - 1 vs. 30 - 10	-0.600	-2.210 to 1.010	No	ns	>0.999
25 - 1 vs. 35 - 5	0.038	-1.572 to 1.648	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.384	-1.994 to 1.226	No	ns	>0.999
25 - 1 vs. 35 - 10	-0.088	-1.698 to 1.522	No	ns	>0.999
25 - 1 vs. 40- 5	-0.087	-1.697 to 1.523	No	ns	>0.999
25 - 1 vs. 40 - 7	-0.626	-2.236 to 0.984	No	ns	>0.999
25 - 1 vs. 40 - 10	-1.678	-3.288 to -0.068	Yes	*	0.035
Tempus™					
25 - 1 vs. 25 - 5	0.497	-1.113 to 2.107	No	ns	>0.999
25 - 1 vs. 25 - 7	0.371	-1.239 to 1.981	No	ns	>0.999
25 - 1 vs. 25 - 10	0.354	-1.446 to 2.154	No	ns	>0.999
25 - 1 vs. 30 - 5	-0.589	-2.199 to 1.021	No	ns	>0.999
25 - 1 vs. 30 - 7	0.458	-1.152 to 2.068	No	ns	>0.999
25 - 1 vs. 30 - 10	0.503	-1.107 to 2.113	No	ns	>0.999
25 - 1 vs. 35 - 5	0.288	-1.322 to 1.898	No	ns	>0.999
25 - 1 vs. 35 - 7	0.017	-1.593 to 1.627	No	ns	>0.999
25 - 1 vs. 35 - 10	0.025	-1.585 to 1.635	No	ns	>0.999
25 - 1 vs. 40- 5	-0.825	-2.435 to 0.785	No	ns	>0.999
25 - 1 vs. 40 - 7	0.123	-1.487 to 1.733	No	ns	>0.999
25 - 1 vs. 40 - 10	-0.543	-2.153 to 1.067	No	ns	>0.999
18s medium (200-300bp) PAXgene®					
25 - 1 vs. 25 - 5	-0.332	-3.247 to 2.583	No	ns	0.953
25 - 1 vs. 25 - 7	-0.059	-5.358 to 5.241	No	ns	>0.999

25 - 1 vs. 25 - 10	-0.220	-5.786 to 5.347	No	ns	>0.999
25 - 1 vs. 30 - 5	0.142	-5.013 to 5.297	No	ns	>0.999
25 - 1 vs. 30 - 7	-0.944	-8.888 to 6.999	No	ns	0.942
25 - 1 vs. 30 - 10	-1.627	-7.669 to 4.415	No	ns	0.489
25 - 1 vs. 35 - 5	-0.590	-10.91 to 9.726	No	ns	1.000
25 - 1 vs. 35 - 7	-1.093	-9.542 to 7.357	No	ns	0.917
25 - 1 vs. 35 - 10	-1.036	-13.51 to 11.44	No	ns	0.992
25 - 1 vs. 40- 5	-1.014	-10.91 to 8.883	No	ns	0.972
25 - 1 vs. 40 - 7	-1.567	-6.057 to 2.924	No	ns	0.336
25 - 1 vs. 40 - 10	-3.090	-7.650 to 1.470	No	ns	0.105
Tempus™					
25 - 1 vs. 25 - 5	0.575	-3.909 to 5.060	No	ns	0.919
25 - 1 vs. 25 - 7	0.458	-4.390 to 5.306	No	ns	0.983
25 - 1 vs. 25 - 10	0.115	-80.75 to 80.98	No	ns	>0.999
25 - 1 vs. 30 - 5	0.169	-10.41 to 10.75	No	ns	>0.999
25 - 1 vs. 30 - 7	0.397	-8.594 to 9.389	No	ns	>0.999
25 - 1 vs. 30 - 10	0.284	-2.694 to 3.262	No	ns	0.982
25 - 1 vs. 35 - 5	-0.646	-14.10 to 12.81	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.086	-7.871 to 7.699	No	ns	>0.999
25 - 1 vs. 35 - 10	-0.130	-5.603 to 5.343	No	ns	>0.999
25 - 1 vs. 40- 5	-0.962	-18.28 to 16.36	No	ns	>0.999
25 - 1 vs. 40 - 7	-0.093	-5.292 to 5.107	No	ns	>0.9999
25 - 1 vs. 40 - 10	-0.979	-10.28 to 8.320	No	ns	0.968
SDHA short (100-200bp) PAXgene®					
25 - 1 vs. 25 - 5	-2.491	-24.83 to 19.85	No	ns	>0.999
25 - 1 vs. 25 - 7	-2.014	-7.929 to 3.901	No	ns	0.412
25 - 1 vs. 25 - 10	-2.829	-7.956 to 2.298	No	ns	0.162
25 - 1 vs. 30 - 5	-2.863	-7.071 to 1.345	No	ns	0.107
25 - 1 vs. 30 - 7	-3.270	-6.294 to -0.246	Yes	*	0.043
25 - 1 vs. 30 - 10	-3.832	-4.927 to -2.738	Yes	**	0.004
25 - 1 vs. 35 - 5	-3.030	-5.341 to -0.718	Yes	*	0.029
25 - 1 vs. 35 - 7	-3.624	-6.454 to -0.794	Yes	*	0.031
25 - 1 vs. 35 - 10	-3.649	-9.402 to 2.104	No	ns	0.123
25 - 1 vs. 40- 5	-3.626	-7.994 to 0.7429	No	ns	0.072
25 - 1 vs. 40 - 7	-3.948	-4.529 to -3.367	Yes	**	0.001
25 - 1 vs. 40 - 10	-4.670	-6.432 to -2.908	Yes	**	0.007
Tempus™					

25 - 1 vs. 25 - 5	1.209	-9.472 to 11.890	No	ns	>0.999
25 - 1 vs. 25 - 7	0.926	-9.730 to 11.580	No	ns	>0.999
25 - 1 vs. 25 - 10	0.859	-14.830 to 16.540	No	ns	>0.999
25 - 1 vs. 30 - 5	1.503	-7.459 to 10.460	No	ns	>0.999
25 - 1 vs. 30 - 7	0.856	-9.513 to 11.230	No	ns	>0.999
25 - 1 vs. 30 - 10	0.850	-7.348 to 9.048	No	ns	>0.999
25 - 1 vs. 35 - 5	0.316	-3.551 to 4.184	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.039	-8.257 to 8.179	No	ns	>0.9999
25 - 1 vs. 35 - 10	0.083	-65.390 to 65.560	No	ns	>0.9999
25 - 1 vs. 40- 5	-0.179	-7.152 to 6.794	No	ns	>0.9999
25 - 1 vs. 40 - 7	-0.120	-7.101 to 6.860	No	ns	>0.9999
25 - 1 vs. 40 - 10	-2.802	-37.880 to 32.280	No	ns	>0.9999
SDHA medium (200-300bp) PAXgene®					
25 - 1 vs. 25 - 5	-3.096	-4.734 to -1.458	Yes	****	<0.0001
25 - 1 vs. 25 - 7	-2.592	-4.230 to -0.954	Yes	***	0.0002
25 - 1 vs. 25 - 10	-3.905	-5.543 to -2.267	Yes	****	<0.0001
25 - 1 vs. 30 - 5	-4.165	-5.803 to -2.527	Yes	****	<0.0001
25 - 1 vs. 30 - 7	-4.810	-6.448 to -3.172	Yes	****	<0.0001
25 - 1 vs. 30 - 10	-5.814	-7.452 to -4.176	Yes	****	<0.0001
25 - 1 vs. 35 - 5	-4.893	-6.531 to -3.255	Yes	****	<0.0001
25 - 1 vs. 35 - 7	-5.387	-7.025 to -3.749	Yes	****	<0.0001
25 - 1 vs. 35 - 10	-6.029	-7.667 to -4.391	Yes	****	<0.0001
25 - 1 vs. 40- 5	-5.667	-7.305 to -4.029	Yes	****	<0.0001
25 - 1 vs. 40 - 7	-6.650	-8.288 to -5.012	Yes	****	<0.0001
25 - 1 vs. 40 - 10	-8.037	-9.675 to -6.399	Yes	****	<0.0001
Tempus™					
25 - 1 vs. 25 - 5	1.039	-0.599 to 2.677	No	ns	0.747
25 - 1 vs. 25 - 7	0.336	-1.302 to 1.974	No	ns	>0.999
25 - 1 vs. 25 - 10	0.526	-1.112 to 2.164	No	ns	>0.999
25 - 1 vs. 30 - 5	0.976	-0.662 to 2.614	No	ns	0.950
25 - 1 vs. 30 - 7	0.641	-0.997 to 2.279	No	ns	>0.999
25 - 1 vs. 30 - 10	0.103	-1.535 to 1.741	No	ns	>0.999
25 - 1 vs. 35 - 5	0.260	-1.378 to 1.898	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.495	-2.133 to 1.143	No	ns	>0.999
25 - 1 vs. 35 - 10	-0.760	-2.587 to 1.066	No	ns	>0.999
25 - 1 vs. 40- 5	-0.601	-2.239 to 1.037	No	ns	>0.999

25 - 1 vs. 40 - 7	-0.851	-2.489 to 0.787	No	ns	>0.999
25 - 1 vs. 40 - 10	-2.941	-4.579 to -1.303	Yes	****	<0.0001
TBP short (100-200bp) PAXgene®					
25 - 1 vs. 25 - 5	-2.338	-4.725 to 0.049	No	ns	0.059
25 - 1 vs. 25 - 7	-2.337	-4.724 to 0.050	No	ns	0.059
25 - 1 vs. 25 - 10	-2.657	-5.044 to -0.270	Yes	*	0.021
25 - 1 vs. 30 - 5	-2.931	-5.318 to -0.544	Yes	**	0.008
25 - 1 vs. 30 - 7	-3.110	-5.497 to -0.723	Yes	**	0.005
25 - 1 vs. 30 - 10	-3.350	-5.737 to -0.963	Yes	**	0.002
25 - 1 vs. 35 - 5	-2.859	-5.246 to -0.472	Yes	*	0.011
25 - 1 vs. 35 - 7	-3.328	-5.715 to -0.941	Yes	**	0.002
25 - 1 vs. 35 - 10	-3.027	-5.414 to -0.640	Yes	**	0.006
25 - 1 vs. 40- 5	-3.337	-5.724 to -0.950	Yes	**	0.002
25 - 1 vs. 40 - 7	-3.178	-5.565 to -0.791	Yes	**	0.004
25 - 1 vs. 40 - 10	-3.713	-6.100 to -1.326	Yes	***	0.001
Tempus™					
25 - 1 vs. 25 - 5	1.060	-1.327 to 3.447	No	ns	>0.999
25 - 1 vs. 25 - 7	0.506	-1.881 to 2.893	No	ns	>0.999
25 - 1 vs. 25 - 10	1.020	-1.367 to 3.407	No	ns	>0.999
25 - 1 vs. 30 - 5	1.338	-1.049 to 3.725	No	ns	>0.999
25 - 1 vs. 30 - 7	0.402	-1.985 to 2.789	No	ns	>0.999
25 - 1 vs. 30 - 10	0.264	-2.123 to 2.651	No	ns	>0.999
25 - 1 vs. 35 - 5	-0.084	-2.471 to 2.303	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.172	-2.559 to 2.215	No	ns	>0.999
25 - 1 vs. 35 - 10	-1.934	-4.321 to 0.453	No	ns	0.204
25 - 1 vs. 40- 5	-0.251	-2.638 to 2.136	No	ns	>0.999
25 - 1 vs. 40 - 7	-0.096	-2.483 to 2.291	No	ns	>0.999
25 - 1 vs. 40 - 10	-0.796	-3.183 to 1.591	No	ns	>0.999
TBP medium (200-300bp) PAXgene®					
25 - 1 vs. 25 - 5	-2.490	-5.256 to 0.277	No	ns	0.106
25 - 1 vs. 25 - 7	-2.597	-5.364 to 0.169	No	ns	0.080
25 - 1 vs. 25 - 10	-3.327	-6.094 to -0.560	Yes	*	0.010
25 - 1 vs. 30 - 5	-3.600	-6.367 to -0.833	Yes	**	0.005
25 - 1 vs. 30 - 7	-4.976	-7.743 to -2.209	Yes	****	<0.0001
25 - 1 vs. 30 - 10	-6.284	-9.050 to -3.517	Yes	****	<0.0001

25 - 1 vs. 35 - 5	-4.347	-7.114 to -1.581	Yes	***	0.001
25 - 1 vs. 35 - 7	-5.561	-8.328 to -2.795	Yes	****	<0.0001
25 - 1 vs. 35 - 10	-5.537	-8.303 to -2.770	Yes	****	<0.0001
25 - 1 vs. 40- 5	-5.139	-7.906 to -2.373	Yes	****	<0.0001
25 - 1 vs. 40 - 7	-5.217	-7.984 to -2.451	Yes	****	<0.0001
25 - 1 vs. 40 - 10	-5.859	-8.626 to -3.093	Yes	****	<0.0001
Tempus™					
25 - 1 vs. 25 - 5	1.211	-1.556 to 3.977	No	ns	>0.999
25 - 1 vs. 25 - 7	0.644	-2.123 to 3.411	No	ns	>0.999
25 - 1 vs. 25 - 10	0.910	-1.857 to 3.676	No	ns	>0.999
25 - 1 vs. 30 - 5	1.320	-1.447 to 4.086	No	ns	>0.999
25 - 1 vs. 30 - 7	0.311	-2.456 to 3.078	No	ns	>0.999
25 - 1 vs. 30 - 10	-0.068	-2.834 to 2.699	No	ns	>0.999
25 - 1 vs. 35 - 5	-0.067	-2.834 to 2.699	No	ns	>0.999
25 - 1 vs. 35 - 7	-0.635	-3.401 to 2.132	No	ns	>0.999
25 - 1 vs. 35 - 10	-3.655	-6.422 to -0.888	Yes	**	0.004
25 - 1 vs. 40- 5	-0.621	-3.388 to 2.146	No	ns	>0.999
25 - 1 vs. 40 - 7	-1.094	-3.861 to 1.672	No	ns	>0.999
25 - 1 vs. 40 - 10	-2.440	-5.207 to 0.326	No	ns	0.121

Chapter 7
General Discussion

This chapter discusses the key findings of research presented in this thesis addressing the research gaps outlined in introductory Chapter 1. This concluding chapter summarizes the significant findings from the previous five chapters, describes the limitations and strengths of the investigations, and outlines recommendations for prospective research.

7 General Discussion

7.1 Overview

Epstein-Barr virus (EBV) and the bacterium *Mycobacterium tuberculosis* (MTB) can cause latent infections. EBV was recognized as the first human virus known to cause a wide range of lymphomas and epithelial tumours, including Burkitt lymphoma (BL), Hodgkin lymphoma (HL), nasopharyngeal carcinoma (NPC), gastric adenocarcinoma, natural killer (NK)/T-cell lymphoma (NKTCL), and post-transplant lymphoproliferative disorders (PTLDs) (1-3). EBV primarily infects epithelial cells of the oropharynx, replicates in B-lymphocytes, and establishes a lifelong persistent latent infection in memory B-cells with occasional reactivation (4-8). In certain impaired immune situations, latent EBV successively predisposes to malignant transformation via viral replication, gene expression and antigen activation, thereby promoting growth, proliferation, and survival of the cells (9). The virus establishes latent infection in B-cells as one of three latency types (Latency I, Latency II, or Latency III), expressing limited distinct sets of viral proteins to evade immune recognition (10). The distinct types of latency contribute to the pathogenesis of a particular tumour.

On the other hand, tuberculosis (TB) caused by MTB bacteria remains dormant or inactive in most individuals, especially in people with weakened immune systems. Latent tuberculosis (LTB) is a state of persistent immune responses to MTB with no evidence of clinically manifested active infection (11). Latent TB is an asymptomatic clinical state that is not transmissible, although it has a 5–10% risk of developing active TB infection by reactivation of MTB (12). Individuals with latent TB represent a reservoir for MTB, with a lifetime risk of developing active TB disease.

There is a crucial need to identify biomarkers for these latent infections in order to identify populations for targeted intervention to improve public health and limit the burden of these diseases. However, despite decades of research, effective and accurate laboratory diagnostics are lacking for any latent infections of humans. Moreover, available diagnostic tests cannot accurately discriminate between active and latent infections and predicting whether an individual with a latent infection will develop an active disease remains a major challenge (13, 14). A lack of understanding of immune responses to latent infections, and the role of specific antigens in this process, hinders the development of the required diagnostic or screening tests. To address this knowledge gap, the work presented in this thesis applied a proteome-wide multiplex approach and a high dimensional molecular profiling technique in combination with sophisticated computational analyses to human samples from various diseases including EBV

associated cancers (namely, natural killer/T-cell lymphoma (NKTCL) and classical Hodgkin lymphoma (cHL), EBV-associated lymphomas in the context of T-cell immunotherapy treatment for EBV-lymphomas, and tuberculosis, in order to inform immune signatures of disease and the associated aetiology.

In chapters 2, 3 and 4, a custom protein microarray representing the complete EBV proteome was used as a high throughput multiplex screening tool to evaluate both IgG and IgA antibody responses against 202 protein sequences representing the entire EBV proteome. An advanced analytical analysis pipeline was developed to deal with this complex high metric data and applied to characterize antibody responses in EBV-associated cancer models (namely, natural killer/T-cell lymphoma (NKTCL) and classical Hodgkin lymphoma (cHL)) and following T-cell immunotherapy treatment.

Findings from the published study described in Chapter 2 (15) demonstrated a distinct antibody profile for NKTCL compared to disease-free healthy controls in an East Asian study population. The humoral immune response to EBV in NKTCL patients has not been systematically studied so this study represented the first comprehensive characterization of EBV-directed humoral immunity in NKTCL. Data identified IgG antibody responses against six distinct EBV proteins as an antibody signature of NKTCL. IgG antibodies targeting EBNA3A were found to be a novel immune marker observed in NKTCL and have not been identified previously in B-cell or epithelial-origin tumours associated with EBV.

The second study utilized a case-control study on cHL to delineate the EBV status of cHL cases to demonstrate the possible use of EBV antibody biomarkers to predict the EBV status of cHL tumours in an East Asian study population and provide insights into the aetiological role of EBV in cHL pathogenesis (Chapter 3). These study findings demonstrated the generalizability of antibody markers previously reported in a European study population (16), supporting our hypothesis that the aetiology of EBV-positive cHL is similar across populations. This study identified 12 novel EBV-specific IgG antibodies that were significantly elevated in EBV-positive cHL cases compared to disease-free controls. BALF2-IgG was identified as the best predictor of the EBV status of the cHL cases in the current study population. In addition, the combination of two IgG markers (BdRF1 and BZLF1), together with patient demographics, accurately predicted the EBV status of cHL cases independent of geographic location and ethnic diversity of study populations.

The third study presented in this thesis used a cohort of EBV-positive lymphoma patients receiving autologous or third-party EBV-specific T-cells (EBVSTs) immunotherapy from Phase I clinical studies to characterize humoral responses to immunotherapy treatment

outcomes (Chapter 4). Little is known about humoral responses in cancer immunotherapies. Thus, understanding antibody responses during T-cell immunotherapy in patients who achieved a positive clinical response may provide insights into an exploitable spectrum of EBV antigen targets for immunotherapy. This study reported the first evaluation of proteome-wide microarray multiplex approach in the context of treatment outcomes. To our knowledge, this was the first study (and only study to date) that directly compared differentially expressed antibody profiles between responders and non-responders for EBVST immunotherapy. Results showed that the non-responders had elevated antibody responses post-treatment, whereas responders had decreased antibody responses. Six anti-EBV antibodies 3-IgA (BGLF3, BALF2, BBLF2/3) and 3-IgG (BGLF2, LF1, BGLF3) were identified that dramatically declined in the response group and were significantly elevated amongst non-responders at 3-months post-treatment as compared to pre-treatment. Chapter 3 results highlighted that the low levels of neutralizing antibodies in responders post-treatment play an important role in anti-tumour activity destroying tumour cells and alterations of hostile tumour microenvironment explaining inhibitory effects to suppress or anergize EBV-specific T-cell activity in non-responders with elevated antibody levels. Together, these results provide valuable insights into understanding humoral responses to EBVST immunotherapy, which are important for discovering and developing novel immunotherapy targets for treating EBV-associated lymphomas.

Collectively, the research presented in Chapters 2 to 4 provides valuable insights into the understanding of humoral responses to EBV latent infection associated EBV-lymphomas and in EBVST immunotherapy, which is important for the discovery, development and translation of biomarkers for latent infection associated disease diagnosis or targeted therapies.

This thesis also reports (chapter 5) the first study on whole-blood host transcriptomic changes associated with different TB infection states in Papua New Guinea (PNG). Balimo is a remote region of the Western Province of PNG with a high TB burden. Symptom-based clinical examinations predominantly diagnose TB in this region due to the limited diagnostic facilities available in the primary health care facility. Thus, the actual active or latent TB burden is unknown because of underdiagnosis or misdiagnosis due to other tropical infections in the area. Recent literature has highlighted host transcriptomic signatures as a promising platform for discovering biomarkers for TB diagnosis and progression (17-21). Host transcriptomic data presented in Chapter 5 identified putative gene signatures which appeared to distinguish the different TB infection states in PNG. This study revealed significantly differentially expressed gene (DEG) profiles in active pulmonary TB (PTB) and latent TB infection. Complement

genes; *C1QA*, *C1QB* and *C1QC* were among the highly upregulated genes in active PTB and these have been previously identified as biomarkers for active TB (22, 23). Also consistent with other reports of TB diagnosis and progression were the genes *SEPT4, ANKRD22*, *APOL4*, *PDCD1LG2*, and *CCRL2*. Together, our results suggest that the host transcriptomic profiles are heterogenous between different TB states and can be used to define and discriminate actual TB infections and inform biomarker discovery for TB diagnosis.

Molecular techniques such as high-throughput transcriptomics using RNA-seq as well as quantitative-PCR (qPCR) are highly dependent upon high-quality, high-quantity and intact RNA. However, collection and preservation of high-quality human samples in field situations, (such as the TB study using samples from Balimo reported in Chapter 5) can be challenging and potentially compromised due to limited resources, logistics of post-collection, processing delays and storage conditions. To investigate the potential impact of these factors, research presented in Chapter 6 assessed the relative performance of the two most widely used whole blood RNA collection systems, PAXgene® and Tempus™. The study analyzed optimal laboratory conditions as well as suboptimal conditions, including extended storage times and high storage temperatures effects on these blood preserving tubes simulating remote field study settings. The findings reported in this chapter showed that Tempus™ tubes maintained a slightly higher RNA quantity and integrity relative to PAXgene® tubes at suboptimal tropical conditions. Furthermore, Tempus™ tubes preferentially maintained the stability of mRNA transcripts tested by qPCR even when RNA quality decreased due to extended storage and higher tropical temperatures. However, both these tubes had similar RNA purities and preserved rRNA transcript equally. Altogether, findings from this chapter concluded that Tempus™ blood RNA collection tubes are preferable to use in suboptimal tropical conditions and in resource-limited field study settings for RNA-based studies.

7.2 Strengths of the research

The work presented in this thesis has numerous strengths. Firstly, the use of EBV-associated lymphomas enabled the analysis of differential humoral responses (both IgA and IgG) to EBV proteome in latency. The case-control studies presented in Chapters 2 and 3 reported the first comprehensive characterization of EBV-directed immunity in adults diagnosed with NKTCL and cHL, respectively, in Asia. Serological findings from our custom EBV protein-microarray results presented in Chapter 2 for putative cancer biomarkers (VCA and EBNA1) were internally validated using commercial ELISA assays to confirm our array-

based findings. Chapter 3 demonstrated the generalizability of antibody markers previously reported in a European study population (15) when evaluated in an East Asian study population using the same protein array approach. The successful replication of previous study findings highlighted the reproducibility of our protein microarray platform.

Additionally, the study presented in Chapter 4 demonstrated the first application of our EBV proteome-wide microarrays evaluating humoral responses in the context of treatment outcomes following EBVST immunotherapy. Another strength was the ability to access both pre-treatment and post-treatment samples from Phase I clinical trials to understand how antibody responses change during T-cell immunotherapy (Chapter 4). These study findings elucidated which EBV proteins are 'seen' by the immune system and are correlated with responders and non-responders to treatment, providing valuable insights into the novel EBV antigens facilitating successful targeted immunotherapy for EBV-lymphomas.

Also presented in this thesis was the first in-depth profiling of transcriptomic responses to different TB infection states in a high TB burden country (Chapter 5). The ability to employ RNA-seq as a high-throughput approach using whole-blood samples from a remote region of PNG enabled the profiling of transcriptional responses in both active and latent TB, and the identification of putative molecular signatures which could discriminate the different states of infection. This foundational study furthers our understanding of immune changes underlying TB infection in PNG, and provides a path to improve differential diagnosis of TB from other infections in this region.

Finally, data presented in Chapter 6 identified the most suitable blood RNA collection tube in suboptimal tropical conditions by comparing the most widely used commercially available blood RNA preserving systems (Tempus™ and PAXgene®) under suboptimal conditions. The effects of higher temperatures on Tempus™ and PAXgene® blood RNA stabilizing systems have not been previously studied. Thus, based on our findings on RNA quantity, quality, purity and transcript stability provide researchers with the choice of Tempus™ system for gene expression and molecular studies in rural and remote resource-limited settings where electricity and storage facilities are compromised.

7.3 Limitations of the research

There are several limitations of the research presented within this thesis. Despite the novelty and multiplexed nature of the EBV proteome-wide approach presented herein, the protein microarray has certain methodologic limitations. Firstly, our proteome array is not

designed to detect antibody responses against conformational epitopes, so this was an inherited technical limitation. Therefore, the information on post-translational processing associated with EBV-associated NKTCL and cHL is lacking. However, this phenomenon does not negatively impact for our study findings. Our custom proteome microarray results from previous studies have been validated using multiplex serology (24) and ELISA specifically tested for putative cancer biomarkers, including VCA, EBNA1, and EAd (15, 25). Further, study findings from Chapter 3 validated and tested the performance of this proteome array and generalizability of antibodies across two study populations

Secondly, our study findings from case-control studies from Chapters 2 and 3 are limited by their retrospective nature. Prospective study designs would be required to explore the alterations in anti-EBV antibody profiles prior to disease onset. However, the difficulty in conducting an adequately powered prospective study for a rare disease like NKTCL makes it unlikely that this limitation will be easily overcome in the future (26).

Thirdly, findings from Chapter 4 suggested that non-responders for the EBVST immunotherapy had elevated antibody responses, but these are initial study results which need to be replicated in larger patient cohorts from clinical immunotherapy trials. The analysis did not consider differences between immunotherapy protocols, tumour types, or EBV exposure in donors of third-party derived EBVSTs in clinical outcomes due to the limited sample size. Also, additional information on the history of infectious mononucleosis, education level, cigarette smoking, and other potential confounding factors for study subjects in Chapters 2-4 was unavailable.

Most of the study participants recruited in Chapter 5 were clinically diagnosed for TB based on the presented symptoms, with no laboratory confirmation available in their records. Additionally, we had only eight individuals with active PTB who had a positive culture and PCR for MTB. Thus, study findings reported herein require replication in a larger patient cohort with definitive TB diagnosis or in other publicly available datasets. Information on BCG vaccination status, HIV status, autoimmune diseases, exposure to other NTM species and tropical infections in these individuals is lacking on these study participants.

An overall limitation of our studies presented in Chapters 2-5 include the relatively limited sample sizes.

7.4 Conclusions and future directions

Overall, the research findings reported in this thesis advance our understanding of humoral responses to latent infections at the proteome-wide level, in particular of EBV latent infections in the context of EBV-associated lymphomas and EBV-specific T-cells immunotherapy. The biomarkers identified in the studies would be expected to have utility in diagnosis at the patient level and potentially for population-based screening to predict the likelihood of disease (for latent and EBV-associated cancers). The approach developed herein of identifying biomarkers for the diagnosis and screening of latent diseases, such as EBV-driven malignancies, could be also applied to other EBV associated tumours (*i.e.*, gastric carcinoma, diffuse large B-cell lymphoma). Additionally, the thesis took advantage of a different but synergistic platform using gene expression (transcriptomics) rather than protein expression (protein microarray) to extend our understanding of immune responses at the molecular level in different TB disease states, including latent TB. Taken together, the technological platforms presented in this doctoral work highlight the potential of utilizing custom protein-microarrays expressing target pathogen proteome as a multiplex tool, host transcriptomic analysis, and advanced analytical approaches including machine learning techniques for computational analyses to discover biomarkers of diseases caused by latent infections. The information gained by such studies would be expected to contribute to developing biomarkers effective for early diagnosis of individuals at risk of developing these diseases.

7.5 References

1. Hsu JL, Glaser SL. Epstein–Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. Critical Reviews in Oncology/Hematology. 2000;34(1):27-53.
2. Niedobitek G, Meru N, Delecluse HJ. Epstein-Barr virus infection and human malignancies. International Journal of Clinical and Experimental Pathology. 2001;82(3):149-70.
3. Epstein M, Barr Y, Achong B. A Second Virus-Carrying Tissue Culture Strain (EB2) of Lymphoblasts From Burkitt's Lymphoma. Pathologie et Biologie. 1964;12:1233.
4. Thorley-Lawson DA, Hawkins JB, Tracy SI, Shapiro M. The pathogenesis of Epstein–Barr virus persistent infection. Current opinion in virology. 2013;3(3):227-32.
5. Gerber P, Lucas S, Nonoyama M, Perlin E, Goldstein L. Oral excretion of Epstein-Barr virus by healthy subjects and patients with infectious mononucleosis. The Lancet. 1972;300(7785):988-9.
6. Sixbey JW, Nedrud JG, Raab-Traub N, Hanes RA, Pagano JS. Epstein–Barr virus replication in oropharyngeal epithelial cells. New England Journal of Medicine. 1984;310(19):1225-30.
7. Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA. EBV persistence in memory B cells in vivo. Immunity. 1998;9(3):395-404.
8. Laichalk LL, Hochberg D, Babcock GJ, Freeman RB, Thorley-Lawson DA. The dispersal of mucosal memory B cells: evidence from persistent EBV infection. Immunity. 2002;16(5):745-54.
9. Crombie JL, LaCasce AS. Epstein Barr Virus Associated B-Cell Lymphomas and Iatrogenic Lymphoproliferative Disorders. Frontiers in Oncology. 2019;9.
10. Ayee R, Ofori MEO, Wright E, Quaye O. Epstein Barr Virus Associated Lymphomas and Epithelia Cancers in Humans. Journal of Cancer. 2020;11(7):1737-50.
11. Mack U, Migliori G, Sester M, Rieder H, Ehlers S, Goletti D, et al. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. European Respiratory Journal. 2009;33(5):956-73.
12. Latent tuberculosis infection (LTBI) Fact sheet 2015: World Health Organization; 2016 [cited 2019 10 May]. Available from: https://www.who.int/tb/areas-of-work/preventive-care/ltbi/faqs/en/.

13. Bakkalci D, Jia Y, Winter JR, Lewis JE, Taylor GS, Stagg HR. Risk factors for Epstein Barr virus-associated cancers: a systematic review, critical appraisal, and mapping of the epidemiological evidence. Journal of Global Health. 2020;10(1):010405.
14. Carranza C, Pedraza-Sanchez S, de Oyarzabal-Mendez E, Torres M. Diagnosis for Latent Tuberculosis Infection: New Alternatives. Frontiers in Immunology. 2020;11.
15. Liu Z, Sarathkumara YD, Chan JKC, Kwong Y-L, Lam TH, Ip DKM, et al. Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma. Scientific Reports. 2021;11(1):23664.
16. Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. International Journal of Cancer. 2020;147(3):608-18.
17. Singhania A, Wilkinson RJ, Rodrigue M, Haldar P, O'Garra A. The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. Nature Immunology. 2018;19(11):1159-68.
18. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973.
19. Constantoulakis P, Filiou E, Rovina N, Chras G, Hamhougia A, Karabela S, et al. In vivo expression of innate immunity markers in patients with Mycobacterium tuberculosis infection. BMC infectious diseases. 2010;10(1):1-10.
20. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. Journal of molecular medicine. 2007;85(6):613-21.
21. Burel JG, Babor M, Pomaznoy M, Lindestam Arlehamn CS, Khan N, Sette A, et al. Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. Frontiers in Immunology. 2019;10(221).
22. Cai Y, Yang Q, Tang Y, Zhang M, Liu H, Zhang G, et al. Increased complement C1q level marks active disease in human tuberculosis. PLoS One. 2014;9(3):e92340.
23. Lubbers R, Sutherland JS, Goletti D, de Paus RA, van Moorsel CHM, Veltkamp M, et al. Complement Component C1q as Serum Biomarker to Detect Active Tuberculosis. Frontiers in Immunology. 2018;9:2427.
24. Simon J, Liu Z, Brenner N, Yu KJ, Hsu W-L, Wang C-P, et al. Validation of an Epstein-Barr Virus Antibody Risk Stratification Signature for Nasopharyngeal Carcinoma by Use of Multiplex Serology. Journal of Clinical Microbiology. 2020;58(5):e00077-20.

25. Coghill AE, Pfeiffer RM, Proietti C, Hsu W-L, Chien Y-C, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clinical Cancer Research. 2018;24(6):1305-14.
26. Aozasa K, Takakuwa T, Hongyo T, Yang W-I. Nasal NK/T-cell lymphoma: epidemiology and pathogenesis. International Journal of Hematology. 2008;87(2):110-7.

8 Appendices

Appendices contains the direct publications included in this thesis.

scientific reports

Check for updates

OPEN

Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma

Zhiwei Liu[1,11✉], Yomani D. Sarathkumara[2,11], John K. C. Chan[3,11], Yok-Lam Kwong[4,11], Tai Hing Lam[5,11], Dennis Kai Ming Ip[5,11], Brian C.-H. Chiu[6,11], Jun Xu[5], Yu-Chieh Su[7,8], Carla Proietti[2], Martha M. Cooper[2], Kelly J. Yu[1], Bryan Bassig[1], Raymond Liang[9], Wei Hu[1], Bu-Tian Ji[1], Anna E. Coghill[10], Ruth M. Pfeiffer[1], Allan Hildesheim[1,12], Nathaniel Rothman[1,12], Denise L. Doolan[2,12] & Qing Lan[1,12]

Extranodal natural killer/T-cell lymphoma (NKTCL) is an aggressive malignancy that has been etiologically linked to Epstein-Barr virus (EBV) infection, with EBV gene transcripts identified in almost all cases. However, the humoral immune response to EBV in NKTCL patients has not been well characterized. We examined the antibody response to EBV in plasma samples from 51 NKTCL cases and 154 controls from Hong Kong and Taiwan who were part of the multi-center, hospital-based AsiaLymph case–control study. The EBV-directed serological response was characterized using a protein microarray that measured IgG and IgA antibodies against 202 protein sequences representing the entire EBV proteome. We analyzed 157 IgG antibodies and 127 IgA antibodies that fulfilled quality control requirements. Associations between EBV serology and NKTCL status were disproportionately observed for IgG rather than IgA antibodies. Nine anti-EBV IgG responses were significantly elevated in NKTCL cases compared with controls and had $ORs_{highest\ vs.\ lowest\ tertile} > 6.0$ (Bonferroni-corrected P-values < 0.05). Among these nine elevated IgG responses in NKTCL patients, three IgG antibodies (all targeting EBNA3A) are novel and have not been observed for other EBV-associated tumors of B-cell or epithelial origin. IgG antibodies against EBNA1, which have consistently been elevated in other EBV-associated tumors, were not elevated in NKTCL cases. We characterize the antibody response against EBV for patients with NKTCL and identify IgG antibody responses against six distinct EBV proteins. Our findings suggest distinct serologic patterns of this NK/T-cell lymphoma compared with other EBV-associated tumors of B-cell or epithelial origin.

Abbreviations

BL	Burkitt lymphoma
CI	Confidence interval
CV	Coefficient of variation
EA	Early antigen
EBNA	Epstein-Barr nuclear antigen

[1]Division of Cancer Epidemiology and Genetics, 9609 Medical Center Drive, National Cancer Institute, Rockville, MD 20850, USA. [2]Centre for Molecular Therapeutics, Australian Institute of Tropical Health of Medicine, James Cook University, Cairns, Australia. [3]Department of Pathology, Queen Elizabeth Hospital, Hong Kong, SAR, China. [4]Queen Mary Hospital, The University of Hong Kong, Hong Kong, SAR, China. [5]School of Public Health, Faculty of Medicine, Li Ka Shing (LKS), The University of Hong Kong, Hong Kong, SAR, China. [6]Department of Public Health Sciences, University of Chicago, Chicago, USA. [7]Department of Medicine, School of Medicine, I-Shou University, Kaohsiung, Taiwan. [8]Division of Hematology-Oncology, Department of Internal Medicine, E-Da Hospital, Kaohsiung, Taiwan. [9]Hong Kong Sanatorium & Hospital, Hong Kong, SAR, China. [10]Cancer Epidemiology Program, Division of Population Sciences, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [11]These authors contributed equally: Zhiwei Liu, Yomani D. Sarathkumara, John K. C. Chan, Yok-Lam Kwong, Tai Hing Lam, Dennis Kai Ming Ip and Brian C.-H. Chiu. [12]These authors jointly supervised this work: Allan Hildesheim, Nathaniel Rothman, Denise L. Doolan and Qing Lan. ✉email: zhiwei.liu@nih.gov

EBV	Epstein-Barr virus
NKTCL	Extranodal natural killer/T-cell lymphoma
IQR	Interquartile range
HL	Hodgkin lymphoma
NPC	Nasopharyngeal carcinoma
OD	Optical density
sPLS-DA	Partial Least Squares Discriminant Analysis
VCA	Viral capsid antigen

Extranodal natural killer T-cell lymphoma (NKTCL; nasal type) is an aggressive malignancy that has been closely linked to infection with Epstein-Barr virus (EBV)[1]. Nearly all NKTCL is EBV positive, with EBV gene transcripts identified in almost 100% of NKTCL tumors[2,3]. EBV establishes lifelong latency in B cells in over 90% of adults worldwide but causes cancer in only a small fraction of infected individuals[4]. EBV-associated tumors include a subset of Hodgkin and non-Hodgkin lymphoma, as well as epithelial carcinomas of the nasopharynx and stomach[4]. Like EBV-positive Hodgkin lymphoma (HL) and nasopharyngeal carcinoma (NPC), EBV-infected cells in patients with NKTCL have been observed to express genes of latency I (EBNA1 and EBER1/2) or latency II (LMP1/2A/2B, EBNA1, and EBER1/2)[5,6]. However, the specific role of EBV in the pathogenesis of NKTCL is still poorly understood.

NK and T cells are typically not permissive of EBV infection and, consequently, EBV is not detected in NK or T cells in the blood of healthy carriers, and is only detected at low frequency in tonsillar NK or T cells[7]. A recent study suggested that EBV can infect mature peripheral T cells via binding of EBV glycoprotein gp350 to the cellular membrane protein CD21,[8] an established receptor for EBV infection of B-cells. However, EBV's role in NKTCL compared to B-cell lymphomas may differ following initial infection. It is possible that viral protein production is distinct following infection of T-cells, leading to different immune targets against which infected persons mount an antibody response. Study of the humoral (antibody) responses against EBV in patients with NKTCL, and comparison of these patterns to those observed in other EBV-associated cancers, could shed light on pathogenic mechanisms.

The humoral immune response to EBV in NKTCL patients is poorly characterized, with three case-only studies inclusive of a total of 155 patients reported to date[9–11]. Those studies found suggestive elevations in antibody levels against viral capsid antigen (VCA) and early antigen (EA) but not EBV nuclear antigen (EBNA). That pattern is distinct from that observed in other EBV-related cancers including nasopharyngeal carcinoma (NPC), Burkitt lymphoma (BL), and Hodgkin lymphoma (HL)[12–15]. In-depth, comprehensive characterization of serologic profiles that associate with NKTCL, and noting those that are distinct from other EBV-related cancers, could provide insight into the role of specific EBV proteins in the etiology of NKTCL. To investigate this, we utilized a multiplex technology targeting antibody responses to 202 peptide sequences representing the entire EBV proteome to comprehensively evaluate patterns of anti-EBV antibody responses in 205 adults from Hong Kong and Taiwan, including 51 NKTCL cases and 154 matched controls.

Results

Table 1 shows the distributions of demographic characteristic in 51 NKTCL cases and 154 matched controls from Hong Kong and Taiwan. Cases and controls had a similar sex, age, and study region distribution, reflective of the matched study design. Approximately two thirds of adults recruited were male, and 78.4% of cases (40/51) were recruited in Hong Kong.

NKTCL associations were disproportionately observed for IgG rather than IgA antibodies. Case–control comparisons of the mean standardized signal intensity for the 157 IgG and 127 IgA antibodies on the array revealed nominal ($P < 0.05$) elevations in 52 IgG antibodies but only six IgA antibodies (Fig. 1). Six anti-EBV IgG antibodies were significantly elevated in NKTCL cases compared to controls after adjustment for multiple testing ($P < 0.0002$; Fig. 1). Results from the remaining 46 anti-EBV IgG and six anti-EBV IgA antibodies that were nominally significantly elevated in NKTCL cases compared to controls ($P < 0.05$) are shown in Supplementary Table 1. Of note, we did not observe differences in anti-EBV EBNA1 IgG responses between NKTCL cases and controls (Supplementary Table 2).

Using logistic regression models with adjustment for sex, age, and study region, in addition to the six significant anti-EBV IgG antibodies mentioned above, we identified ab additional three IgG antibodies that were significantly elevated in NKTCL cases compared with controls. These elevations had at least a sixfold risk ($OR_{\text{highest vs. lowest tertile}} \geq 6.0$, Table 2). The most significant P value was observed for IgG antibody against latent protein EBNA3A (one of three variants shown in Fig. 2A). Accordingly, the strongest OR effect was observed for antibody against sequences representing EBNA3A (adjusted $OR_{\text{highest vs. lowest tertile}} = 16.33$, 95% confidence interval [CI]: 3.71 to 71.91, P-trend $= 1.6 \times 10^{-5}$), a protein expressed in latency IIb and III phases that has not been found to be strongly associated with other EBV-associated tumors of B-cell or epithelial origin[12–14]. Pronounced P values were also observed for the early lytic proteins BALF2 [EA(D)p138] (one representative variant shown in Fig. 2B) and BMRF1 [EA(D)p47 (one of two variants shown in Fig. 2C). Other IgG antibodies significantly and markedly elevated in NKTCL patients included those targeting antigens representing immediate early and late lytic proteins, BZLF1 [Zebra (Zta)], BVRF2 [VCAp40] and BPLF1 [Tegument protein] (Table 2 and Fig. 2D-2F).

We next examined the correlations between these nine highly differentially expressed anti-EBV IgG antibodies. Strong correlations were observed for antibodies targeting the same antigens (i.e., three variants for EBNA3A and two variants for BMRF1), with correlations ranging from 0.903 to 0.966. More modest correlation was observed between antibodies targeting different antigens, with correlations ranging from 0.313 to 0.747 (Supplementary Fig. 1). In a logistic regression model excluding 3 IgG antibodies (two antibodies against EBNA3A

Characteristics	NKTCL cases (N = 51)	Controls (N = 154)
Sex		
Male	34 (66.7)	102 (66.2)
Female	17 (33.3)	52 (33.8)
Age at diagnosis/selection (years)		
18–39	12 (23.5)	34 (22.1)
40–49	13 (25.5)	42 (27.3)
50–59	11 (21.6)	33 (21.4)
60–80	15 (29.4)	45 (29.2)
Region		
Hong Kong	40 (78.4)	123 (79.9)
Taiwan	11 (21.6)	31 (20.1)

Table 1. Characteristics of study population, by NK-T cell lymphoma (NKTCL) status in Hong Kong and Taiwan.

and one against BMRF1) that were highly correlated with antibodies targeting the same antigens, IgG antibodies against EBNA3A, BALF2, and BPLF1 retained statistical significance ($P < 0.05$).

In the sPLS-DA analysis, the top 10 anti-EBV IgG antibodies that were most informative for classifying NKTCL status were those targeting EBNA3A, BALF2, BRLF1, thymidine kinase (TK), BMRF1, and BZLF1 (Supplementary Fig. 2), largely consistent with the most significant antibodies defined using the t-test.

Results from ELISA assays confirmed our array-based findings. For example, we observed that VCA-IgG was significantly elevated among NKTCL cases compared with controls ($P = 5.2 \times 10^{-8}$, Supplementary Fig. 3A). There was suggestive evidence that VCA-IgA was also elevated ($P = 0.003$) but that association was not statistically significant after adjustment for multiple testing (Supplementary Fig. 3B). Antibodies against EBNA1 (both IgG and IgA) measured by ELISA were not elevated among NKTCL cases compared with controls (Supplementary Fig. 3C-D).

Finally, as an exploratory analysis, we leveraged genotyping data[16] from 94 controls included in the present study and observed that SNP rs9271588 (which maps to *HLA-DRB1*) was suggestively correlated with the most differentially expressed EBV-antibody EBNA3A-IgG ($P = 0.06$).

Discussion

This is, to our knowledge, the first study to comprehensively evaluate EBV-directed immunity in adults diagnosed with NKTCL in Asia. We investigated both IgG and IgA responses to each protein expressed in the EBV proteome. Profound differences in the anti-EBV antibody profile between NTCKL patients and matched controls were demonstrated, with significantly elevated IgG antibody responses against six distinct EBV proteins. Notably, the strongest NKTCL–EBV associations mapped to sequences representing EBNA3A (but not EBNA1), suggesting a possible role of this latent protein in disease pathogenesis.

In addition to IgG, we examined anti-EBV IgA antibodies in the context of NKTCL. IgA reflects recent exposure along mucosal surfaces such as the oral epithelium and has proven to be an informative biomarker for EBV-associated epithelial tumors (*e.g.*, nasopharyngeal carcinoma)[12,17]. However, IgA responses did not significantly differ between NKTCL patients and controls after correction for multiple testing. Although false negative findings cannot be entirely ruled out due to a modest sample size and relatively low activity of IgA antibodies, our findings may indicate that chronic reactivation or recent exposure to the virus at a mucosal site is less important in the pathogenesis of NKTCL.

The unique association of NKTCL with IgG antibodies against EBNA3A has not been previously reported[9,11,18]. Coghill et al. have reported an association between IgG antibodies against EBNA3A and BL in Africa, but the magnitude of association is smaller than the present study ($OR_{\text{highest vs. lowest tertile}} = 1.99$)[13]. Although EBV-encoded transcripts and proteins have been detected in patients with NKTCL[1,6,19–21], that expression pattern has generally been consistent with latency I or II infection, which is characterized by expression of EBNA-1, LMP-1, and LMP-2 genes but no other EBNA genes[1,5,6,20,21]. Therefore, the higher IgG antibody levels against EBNA3A observed in the current study might not be explained by high expression of EBNA3A gene in the tumor tissue. Instead, this observation could reflect a long-term systematic exposure to the upregulation of EBNA3A gene within circulating B cells infected with EBV, which could be an early event during the development of NKTCL.

In agreement with previously reported case-only studies that included 155 patients from the U.S. and China[9,11,18], we confirm elevations in NKTCL patients for IgG antibodies against sequences representing EBV EA, including EAD-p47 and -p138, EBV viral capsid BVRF2 (VCAp40), as well as virion production BPLF1 (tegument protein). Furthermore, in agreement with other epidemiological research[9–11], we report here elevated IgG antibodies against VCA and EA in patients with NKTCL, but no NKTCL associations with EBNA1. We also expand findings to the switch protein Zta (BZLF1), which has been associated with other EBV-associated malignancies [12–14] but not previously studied in the context of NKCTL. It is plausible that, again, systematic exposure to EBV, as indicated by elevations in IgG antibodies against EBV lytic proteins, potentially reflects an impaired T-cell response that allows virus to continue replication and spread from the B-cell compartment to NK/T-cells.

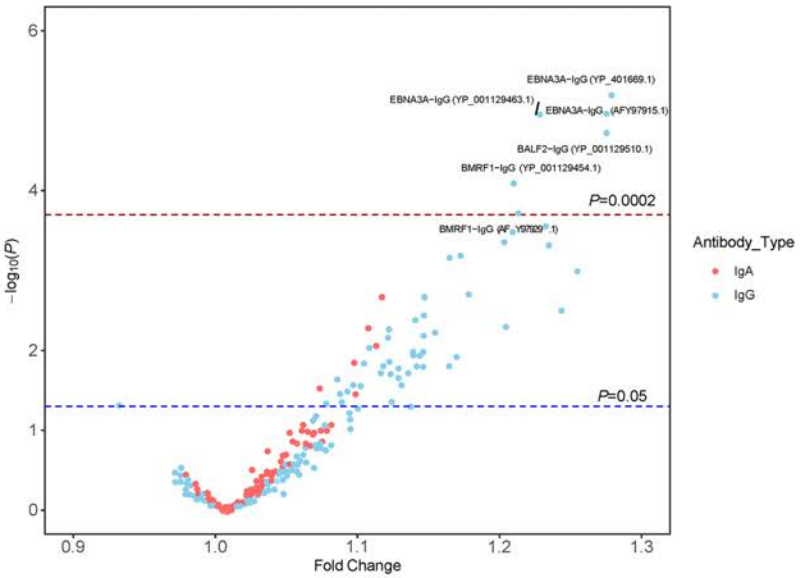


Figure 1. Case–control differences in the mean antibody response for 51 NK/T-cell lymphoma (NKTCL) cases versus 154 controls collected in Hong Kong and Taiwan. The x-axis displays the fold change (case vs. control ratio of standardized signal intensity) for all antibodies with CV ≤ 20%. The y-axis illustrates the p value corresponding to the t-test for a difference in standardized signal intensity between cases and controls. Six IgG-antibodies but no IgA antibodies were significantly elevated in NKTCL cases compared to controls at the $P < 0.0002$ (Bonferroni-corrected $P < 0.05$) threshold.

GWAS studies have implicated genetic susceptibility in NTCKL etiology, with signals consistently mapped to the HLA genes in the class II region[16,22]. In our study, we observed indicative evidence that genetic variation within the HLA class II region affected anti-EBV serologic immunity in controls. It is plausible that people with susceptible HLA variants might mount altered responses to EBV infection that predisposes to NKTCL development[23–26]. Future consortia-based efforts focusing on host genetic variants and anti-EBV antibodies would be required to explore the potential synergistic effects of HLA and EBV in the etiology of NKTCL.

Our results should be interpreted in light of certain methodologic limitations. First, our observations are based on data obtained from a case–control designed study so we are unable to determine whether alterations in anti-EBV antibody responses occurred prior to disease onset; i.e., predisposition to disease, as we have previously reported for other EBV-related tumors using the same EBV antibody array[12–14]. However, the difficulty of conducting an adequately powered prospective study for this rare disease makes it unlikely that this limitation will be easily overcome in the future[3]. Second, this is the only study to date examining the association between the proteome-wide anti-EBV antibody response and NKTCL, and we therefore lack an independent, external dataset for replication. Finally, this array was not designed to detect antibodies to conformational epitopes, which precluded us from examining NKTCL associations for selected transcripts that require glycosylation or other post-transcriptional modifications.

In conclusion, we characterize the antibody response against EBV for patients with NKTCL. Our findings suggest distinct serologic patterns of this NK/T-cell lymphoma compared with other EBV-associated tumors of B-cell or epithelial origin. This NKTCL–specific signature included pronounced differences in the immune response against six viral proteins involved in both latency and replication.

Methods

Study population. Plasma samples from 51 NKTCL cases and 154 control adults collected as part of the AsiaLymph, a multi-center hospital-based case–control study in Hong Kong and Taiwan conducted between 2012 and 2017, were selected for study. Eligible cases were aged between 18 and 79 years at diagnosis and living in the geographic region served by the partnering hospital at the time of cancer diagnosis. Cases with a prior history of lymphoma were ineligible. Blood and buccal cell collection were performed at the time of diagnosis and before receiving cancer therapy. Controls were drawn from patients seen at the same partnering hospital for diseases/conditions that were not associated with risk factors under study, including injuries and selected diseases of the circulatory, digestive, genitourinary, and central nervous system. Patients with a history of any

EBV protein and array sequence	Antibody type	t test *P*	NKTCL mean (SD)	Control mean (SD)	Fold change	NKTCL positivity	Control positivity	OR tertile 2 (95% CI)[b]	OR tertile 3 (95% CI)[b]	*P*-trend[c]
EBNA3A (YP_401669.1–80,382-82,877)	IgG	5.99×10^{-6}	1.76 (0.47)	1.38 (0.54)	1.27	96.1%	75.3%	2.44 (0.80–7.45)	6.59 (2.38–18.22)	6.51×10^{-5}
EBNA3A (AFY97915.1–80,252-82,747)	IgG	1.06×10^{-5}	1.68 (0.46)	1.32 (0.52)	1.27	94.1%	64.3%	4.79 (1.29–17.73)	11.14 (3.21–38.72)	1.84×10^{-5}
EBNA3A (YP_001129463.1–80,447-82,888)	IgG	1.08×10^{-5}	1.85 (0.44)	1.51 (0.48)	1.22	98.0%	88.3%	8.48 (1.83–39.22)	16.33 (3.71–71.91)	1.63×10^{-5}
BALF2 [EA(D)_p138] (YP_001129510.1–165,796-162,410–1)	IgG	1.79×10^{-5}	1.37 (0.39)	1.08 (0.41)	1.27	80.4%	52.6%	2.34 (0.75–7.28)	7.29 (2.60–20.43)	3.03×10^{-5}
BMRF1 [EA(D)_p47] (YP_001129454.1–67,745-68,959)	IgG	7.64×10^{-5}	1.78 (0.44)	1.48 (0.49)	1.20	96.1%	92.9%	2.52 (0.82–7.76)	6.83 (2.45–19.08)	5.70×10^{-5}
BMRF1 [EA(D)_p47] (AFY97929.1–67,486-68,700)	IgG	1.81×10^{-4}	1.67 (0.45)	1.38 (0.47)	1.21	94.1%	84.4%	2.88 (0.96–8.62)	6.32 (2.27–17.61)	1.60×10^{-4}
BZLF1 [Zebra (Zta)] (YP_001129467.1–91,697-91,197)	IgG	4.19×10^{-4}	1.49 (0.42)	1.24 (0.39)	1.20	96.1%	74.7%	4.85 (1.3–18.09)	11.13 (3.19–38.78)	1.99×10^{-5}
BVRF2 [VCAp40] (YP_001129501.1–136,465-138,282)	IgG	6.64×10^{-4}	1.74 (0.40)	1.50 (0.46)	1.17	100.0%	95.5%	2.92 (0.97–8.79)	6.75 (2.39–19.03)	1.19×10^{-4}
BPLF1 [Tegument protein] (CAA24839.1–71,527-62,078–2)	IgG	5.82×10^{-3}	1.93 (0.40)	1.73 (0.53)	1.11	98.0%	98.7%	2.20 (0.75–6.42)	6.00 (2.25–16.01)	1.17×10^{-4}

Table 2. OR and 95% CI for the association between anti-EBV antibody level and NK-T cell lymphoma (NKTCL) in Hong Kong and Taiwan[a]. Bold text is used to highlight the canonical EBV protein name. The remaining (non-bolded) text describes the sequence details of the array probe. CI, confidence interval. SD, standard deviation. [a]The table is ordered by t test *P* value (lowest to highest). [b]The odds of being a NKTCL case were calculated from a logistic regression model that included age group (18–39, 40–49, 50–59, 60–80 years), sex, region, and a three-level variable (tertiles) for anti-EBV antibody level. The tertiles were calculated using the underlying antibody distribution among disease-free controls. All ORs are expressed relative to the referent group of tertile 1 (lowest third of antibody distribution). [c]Two-sided *P* values for trend across marker categories were assessed with the Wald test using categorical values of the proteins with 1 degree of freedom.

lymphoma were not eligible for controls. Of all controls recruited in the two regions (N = 1496; 1119 from Hong Kong and 377 from Taiwan), we randomly selected 154 subjects who were frequency-matched to cases on sex, age (+/- 5 years), date of enrollment (within 3 months), and region (Hong Kong/Taiwan).

The study was approved by the institutional review boards at each participating site, and the US National Institutes of Health and US National Cancer Institute. Written informed consent was obtained from all participants. All laboratory testing was conducted under a protocol approved by James Cook University Human Research Ethics Committee. All methods were performed in accordance with the Declarations of Helsinki.

EBV protein microarray. The comprehensive EBV protein microarray chip used in this study has been described in detail previously[12,27]. Briefly, this microarray contains 202 protein sequences representing almost the entire EBV proteome, including 199 EBV protein sequences generated from five different EBV strains (AG876, Akata, B95-8, Mutu, and Raji) and three synthetic EBV peptides for which circulating antibodies are putative cancer biomarkers (VCAp18, EBNA1, and EAd p47). The 202 sequences represent each of the known open reading frames for EBV, as well as predicted splice variants of those open reading frames. Each of the protein sequences were cloned into the pXT7 expression vector, expressed using the *E. coli* cell-free protein system, and printed onto the microarray. Sequences include N-terminal 10×histidine (His) and C-terminal hemagglutinin (HA) tags for quality control and to confirm expression on the microarray. High coverage was achieved across the five prototypical EBV strains and ten Chinese strains, with > 97% of the predicted sequences from each strain represented on the microarray at > 99% homology. Four "noDNA" (no translated protein) spots were included to assess person-specific background.

Plasma samples from each of the study participants were tested on this EBV protein microarray as described previously[28]. Slides were scanned on an Axon GenePix 4300B (Molecular Devices, Australia); raw fluorescence intensities were corrected for spot-specific background; corrected data were transformed using variance stabilizing normalization (vsn) in Gmine[29]; and output was standardized to person-specific background (mean ± 1.5 SD of the four "no DNA" spots). Positivity was defined as a standardized signal intensity > 1.0. The standardized

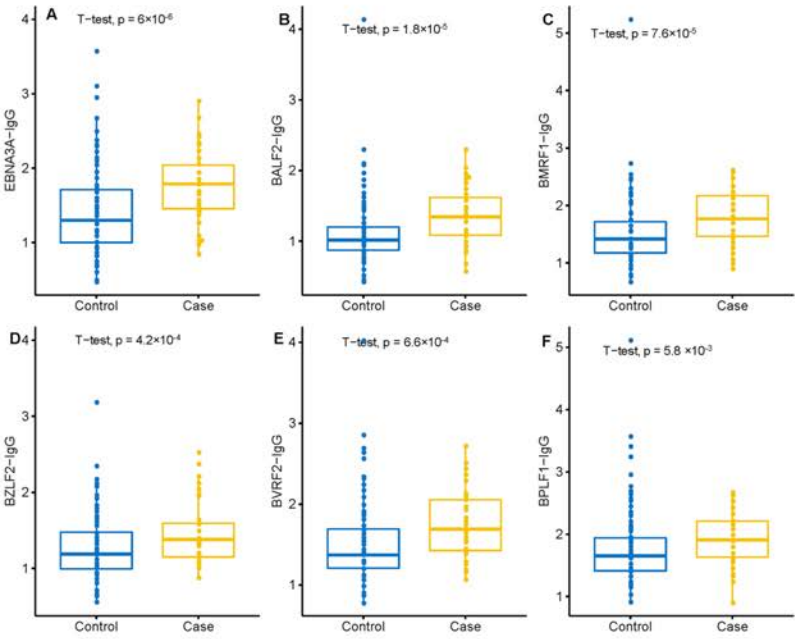


Figure 2. Signal intensity for the six significant anti-EBV IgG antibodies between NK/T-cell lymphoma (NKTCL) and controls, for (**A**) EBNA3A-IgG, (**B**) BALF2-IgG, (**C**) BMRF1-IgG, (**D**) BZLF2-IgG, (**E**) BVRF2-IgG, and (**F**) BPLF1-IgG. *P* values from the t-test are listed.

signal intensity for each spot was further grouped into three categories, with cutoffs for the categories defined using tertiles of the antibody distribution among the 154 controls.

Thirty-five samples were tested in duplicate, blinded to laboratory personnel, in order to assess assay reproducibility specific to this study population. The average coefficient of variation (CV) across the 202 EBV sequences was 16% [interquartile range (IQR), 14%–20%] for IgG antibody response and 19% (IQR, 16%–22%) for IgA antibody response, demonstrating a good reproducibly of our assay. We excluded 45 IgG and 75 IgA that had CVs > 20%, leaving a total of 157 IgG and 127 IgA antibodies for further analysis.

Antibody testing using ELISA kits. To internally validate the serological findings from the EBV microarray for putative cancer biomarkers, we utilized commercial ELISA assays to test for IgG and IgA antibodies against recombinant VCA and EBNA1; these two antigens have been extensively investigated in other EBV-related cancers[30]. ELISA assays were purchased from EUROIMMUN, Lübeck, Germany (IgG/IgA antibodies against VCA and IgG antibodies against EBNA1) and Zhongshan Biotech, Zhongshan, China (IgA antibodies against EBNA1)[31,32]. All samples were tested according to the manufacturers' instructions. Levels of antibodies were assessed by optical density (OD) values. Reference ODs (rODs) were obtained according to the manufacturers' instructions by dividing OD values by a reference control. The same thirty-five blinded duplicates tested by microarray were also tested by ELISA to assess assay reproducibility. The CVs for IgG antibodies against VCA and EBNA1 were 6.9% and 7.7%, respectively; for IgA antibodies, CVs against VCA and EBNA1 were 19.1% and 25.1%, respectively.

Statistical analysis. Differences in the mean standardized signal intensity between NKTCL patients and controls were assessed using an unpaired Student t test. Case–control differences were considered statistically significant at the $P < 0.0002$ threshold (equivalent to Bonferroni-corrected $P < 0.05$) to account for the number of comparisons. Odds ratios (ORs) quantifying the association between the three-level categorical variable for each antibody and NKTCL status were estimated using logistic regression models adjusted for sex, age group (18–39, 40–49, 50–59, 60–80 years), and region. In previous work, no sociodemographic or environmental factors were found to strongly and consistently correlate with elevated anti-EBV antibody responses other than smoking[27,33,34]; however, smoking was not associated with NKTCL in a previous study[35] and therefore was not included in our regression models. *P*-trend values were calculated from a model with each three-level antibody marker treated as an ordinal variable. Antibodies with *P*-trend < 0.0002 threshold (equivalent to Bonferroni-

corrected $P < 0.05$) were considered as statistical significance. For results from the ELISA assays, differences in the mean rOD between NKTCL patients and controls were assessed using an unpaired Student t test.

To identify the anti-EBV IgG antibodies that are most informative for distinguishing NKTL cases from controls, we employed sparse Partial Least Squares Discriminant Analysis (sPLS-DA), which was implemented using the splda function in MixOmics R package[36,37]. The sPLS-DA is a method for identifying the key variables of complex and sparse omics datasets that are associated with a biological outcome of interest and it has been shown to be successful with applications where the number of features far outnumber the number of samples[38] This procedure involves dimension reduction using Partial Least Squares regression (PLS) for discriminant analysis in combination with a Lasso penalization for feature selection. The number of features selected per component was optimized using tenfold cross validation repeated 5 times and the number associated with the lowest classification error rate was chosen for the final model. The final model was then applied to the entire dataset to obtain the most important anti-EBV IgG antibodies in distinguishing NKTCL cases from controls.

Amongst controls, we estimated the correlation between antibodies using Spearman correlation coefficients. We also evaluated whether previously reported NKTCL-associated genetic variants (i.e., rs13015714, mapped to *IL18RAP*, and rs9271588, mapped to *HLA-DRB1*)[16] were associated with the level of anti-EBV antibody response using linear regression models adjusted for sex, age group (18–39, 40–49, 50–59, 60–80 years), and region.

Data availability

For original data, please contact zhiwei.liu@nih.gov. Deidentified participant data can be shared.

Received: 15 August 2021; Accepted: 8 November 2021
Published online: 08 December 2021

References

1. Asano, N., Kato, S. & Nakamura, S. Epstein-Barr virus-associated natural killer/T-cell lymphomas. *Best Pract. Res. Clin. Haematol.* **26**(1), 15–21 (2013).
2. Kimura, H. EBV in T-/NK-Cell Tumorigenesis. *Adv. Exp. Med. Biol.* **1045**, 459–475 (2018).
3. Aozasa, K., Takakuwa, T., Hongyo, T. & Yang, W. I. Nasal NK/T-cell lymphoma: epidemiology and pathogenesis. *Int. J. Hematol.* **87**(2), 110–117 (2008).
4. 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**(12), 789–802 (2016).
5. Houldcroft, C. J. & Kellam, P. Host genetics of Epstein-Barr virus infection, latency and disease. *Rev. Med. Virol.* **25**(2), 71–84 (2015).
6. Xiong, J. *et al.* Genomic and transcriptomic characterization of natural killer T cell lymphoma. *Cancer Cell* **37**(3), 403–419 e406 (2020).
7. Hudnall, S. D. *et al.* Distribution and phenotype of Epstein-Barr virus-infected cells in human pharyngeal tonsils. *Mod. Pathol.* **18**(4), 519–527 (2005).
8. Smith, N. A., Coleman, C. B., Gewurz, B. E. & Rochford, R. CD21 (complement receptor 2) is the receptor for Epstein-Barr virus entry into T cells. *J. Virol.* **94**(11), 20 (2020).
9. Jones, J. F. *et al.* T-cell lymphomas containing Epstein-Barr viral DNA in patients with chronic Epstein-Barr virus infections. *N. Engl. J. Med.* **318**(12), 733–741 (1988).
10. Cao, S. M. *et al.* Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PLoS ONE* **6**(4), e19100 (2011).
11. Huang, Y. *et al.* Serum EBV EA-IgA and VCA-IgA antibodies can be used for risk group stratification and prognostic prediction in extranodal NK/T cell lymphoma: 24-year experience at a single institution. *Ann. Hematol.* **96**(8), 1331–1342 (2017).
12. Coghill, A. E. *et al.* Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. *Clin. Cancer Res.* **24**(6), 1305–1314 (2018).
13. Coghill, A. E. *et al.* The association between the comprehensive Epstein-Barr virus serologic profile and endemic Burkitt lymphoma. *Cancer Epidemiol. Biomark. Prev.* **29**(1), 57–62 (2020).
14. Liu, Z. *et al.* Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. *Int. J. Cancer.* **147**(3), 608–618 (2020).
15. Song, L. *et al.* Identification of anti-Epstein-Barr virus (EBV) antibody signature in EBV-associated gastric carcinoma. *Gastric Cancer* **2021**, 1–10 (2021).
16. Lin, G. W. *et al.* Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study in multiple populations. *Lancet Oncol.* **21**(2), 306–316 (2020).
17. Bhaduri-McIntosh, S. *et al.* Serum IgA antibodies to Epstein-Barr virus (EBV) early lytic antigens are present in primary EBV infection. *J. Infect. Dis.* **195**(4), 483–492 (2007).
18. Huang, Y. H., Wu, Q. L., Zong, Y. S., Feng, Y. F. & Hou, J. H. Nasopharyngeal extranodal NK/T-cell lymphoma, nasal type: Retrospective study of 18 consecutive cases in Guangzhou, China. *Int. J. Surg. Pathol.* **19**(1), 51–61 (2011).
19. Peng, R. J. *et al.* Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. *Leukemia* **33**(6), 1451–1462 (2019).
20. Minarovits, J. *et al.* Clonality, expression and methylation patterns of the Epstein-Barr virus genomes in lethal midline granulomas classified as peripheral angiocentric T cell lymphomas. *J. Gen. Virol.* **75**(Pt 1), 77–84 (1994).
21. Tao, Q., Ho, F. C., Loke, S. L. & Srivastava, G. Epstein-Barr virus is localized in the tumour cells of nasal lymphomas of NK, T or B cell type. *Int. J. Cancer.* **60**(3), 315–320 (1995).
22. Li, Z. *et al.* Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study. *Lancet Oncol.* **17**(9), 1240–1247 (2016).
23. Rubicz, R. *et al.* A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet.* **9**(1), e1003147 (2013).
24. Hammer, C. *et al.* Amino acid variation in HLA Class II proteins is a major determinant of humoral response to common viruses. *Am. J. Hum. Genet.* **97**(5), 738–743 (2015).
25. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* **10**(1), 59 (2018).
26. Sallah, N. *et al.* Distinct genetic architectures and environmental factors associate with host response to the gamma2-herpesvirus infections. *Nat. Commun.* **11**(1), 3849 (2020).

27. Liu, Z. *et al.* Patterns of interindividual variability in the antibody repertoire targeting proteins across the Epstein-Barr virus proteome. *J. Infect. Dis.* **217**(12), 1923–1931 (2018).
28. Lee, M. H. *et al.* Epstein-Barr virus-based nasopharyngeal carcinoma (NPC) risk prediction scores are elevated in NPC multiplex family members in Taiwan. *J. Infect. Dis.* **223**(3), 441–444 (2021).
29. Proietti, C. *et al.* Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) Web-Server. *Sci. Rep.* **6**, 38178 (2016).
30. Coghill, A. E. & Hildesheim, A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. *Am. J. Epidemiol.* **180**(7), 687–695 (2014).
31. Gan, Y. Y., Fones-Tan, A., Chan, S. H. & Gan, L. H. Epstein-Barr viral antigens used in the diagnosis of nasopharyngeal carcinoma. *J. Biomed. Sci.* **3**(3), 159–169 (1996).
32. Chen, M. R. *et al.* Use of bacterially expressed EBNA-1 protein cloned from a nasopharyngeal carcinoma (NPC) biopsy as a screening test for NPC patients. *J. Med. Virol.* **64**(1), 51–57 (2001).
33. Xu, F. H. *et al.* An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation. *J. Natl. Cancer Inst.* **104**(18), 1396–1410 (2012).
34. Pickard, A. *et al.* Epstein-Barr virus seroreactivity among unaffected individuals within high-risk nasopharyngeal carcinoma families in Taiwan. *Int. J. Cancer.* **111**(1), 117–123 (2004).
35. Xu, J. X. *et al.* Life-style and environmental factors in the development of nasal NK/T-cell lymphoma: A case-control study in East Asia. *Int. J. Cancer.* **120**(2), 406–410 (2007).
36. Le Cao, K. A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinform.* **12**, 253 (2011).
37. Rohart, F., Gautier, B., Singh, A. & Le Cao, K. A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**(11), e1005752 (2017).
38. Chung, D. & Keles, S. Sparse partial least squares classification for high dimensional data. *Stat. Appl. Genet. Mol. Biol.* **9**, Article 17 (2010).

Acknowledgements
This work was supported by the Intramural Research Program of National Cancer Institute (NCI), USA. D L Doolan is supported by a National Health and Medical Research Council of Australia (NHMRC) Principal Research Fellowship (#1137285). YDS is supported a Postgraduate Research Scholarship from James Cook University). We are grateful to the study subjects without whom this work would not be possible.

Author contributions
Z.L. designed research, analyzed and interpreted data, performed statistical analysis, and wrote the manuscript. Y.D.S., J.K.C.C., Y-L K., T.H.L., D.K.M.I., B.C.H.C., J.X., Y.C.S, M.M.C., K.J.Y., B.B., R.L., W.H., and B.T.J. collected data, and contributed vital new reagents or analytical tools. C.P. and R.M.P. analyzed and interpreted data. A.E.C. interpreted data and provided critical review. A.H., N.R., D.L.D., and Q.L. provided critical review and supervised the study.

Funding
Open Access funding provided by the National Institutes of Health (NIH). Intramural Research Program of National Cancer Institute (NCI), USA. D L Doolan is supported by a National Health and Medical Research Council of Australia (NHMRC) Principal Research Fellowship (#1137285). YDS is supported a Postgraduate Research Scholarship from James Cook University). We are grateful to the study subjects without whom this work would not be possible.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-02788-w.

Correspondence and requests for materials should be addressed to Z.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

International Journal of
Molecular Sciences

MDPI

Article

The Effect of Tropical Temperatures on the Quality of RNA Extracted from Stabilized Whole-Blood Samples

Yomani D. Sarathkumara [1], Daniel J. Browne [1], Ashton M. Kelly [1], David J. Pattinson [1], Catherine M. Rush [1,2], Jeffrey Warner [1,2], Carla Proietti [1] and Denise L. Doolan [1,*]

[1] Centre for Molecular Therapeutics, Australian Institute of Tropical Health of Medicine, James Cook University, Cairns, QLD 4878, Australia
[2] College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, QLD 4811, Australia
* Correspondence: denise.doolan@jcu.edu.au

Abstract: Whole-blood-derived transcriptional profiling is widely used in biomarker discovery, immunological research, and therapeutic development. Traditional molecular and high-throughput transcriptomic platforms, including molecular assays with quantitative PCR (qPCR) and RNA-sequencing (RNA-seq), are dependent upon high-quality and intact RNA. However, collecting high-quality RNA from field studies in remote tropical locations can be challenging due to resource restrictions and logistics of post-collection processing. The current study tested the relative performance of the two most widely used whole-blood RNA collection systems, PAXgene® and Tempus™, in optimal laboratory conditions as well as suboptimal conditions in tropical field sites, including the effects of extended storage times and high storage temperatures. We found that Tempus™ tubes maintained a slightly higher RNA quantity and integrity relative to PAXgene® tubes at suboptimal tropical conditions. Both PAXgene® and Tempus™ tubes gave similar RNA purity (A260/A280). Additionally, Tempus™ tubes preferentially maintained the stability of mRNA transcripts for two reference genes tested, *Succinate dehydrogenase complex, subunit A* (*SDHA*) and *TATA-box-binding protein* (*TBP*), even when RNA quality decreased with storage length and temperature. Both tube types preserved the rRNA transcript *18S ribosomal RNA* (*18S*) equally. Our results suggest that Tempus™ blood RNA collection tubes are preferable to PAXgene® for whole-blood collection in suboptimal tropical conditions for RNA-based studies in resource-limited settings.

Keywords: PAXgene®; Tempus™; blood RNA extraction; RT-qPCR; tropical; climate

Citation: Sarathkumara, Y.D.; Browne, D.J.; Kelly, A.M.; Pattinson, D.J.; Rush, C.M.; Warner, J.; Proietti, C.; Doolan, D.L. The Effect of Tropical Temperatures on the Quality of RNA Extracted from Stabilized Whole-Blood Samples. *Int. J. Mol. Sci.* **2022**, 23, 10609. https://doi.org/10.3390/ijms231810609

Academic Editor: Janusz Blasiak

Received: 15 August 2022
Accepted: 9 September 2022
Published: 13 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Gene expression profiles from whole-blood-derived RNA have proven useful as a molecular signature reflecting physiological and pathological changes in the body [1]. Immune signatures for various diseases and metabolic states, including biomarkers of disease, disease prognosis, or therapeutic efficacy, have been identified by blood transcriptional profiling [2–4]. Since blood transcript profiles are reproducible, cost-effective, and easy to implement, they can be rapidly translated into clinical practice [5]. High-quality and intact RNA is imperative for both traditional molecular diagnostics and high-throughput transcriptome sequencing techniques, such as quantitative PCR (qPCR) [6] and RNA-sequencing (RNA-seq) [7]. High-quality RNA can be readily obtained from fresh blood when processed immediately following collection. However, RNA quality can be adversely impacted by processing delays and complex logistical issues. Remote tropical locations are particularly challenging due to locality and resource restrictions [8]. Pre-extraction factors that influence gene expression or lead to RNA degradation include processing delays and storage conditions (i.e., duration and temperature) [9–12] since blood RNA is highly susceptible to enzymatic degradation and oxidative damage [13,14].

Commercially available blood collection systems with additives that stabilize RNA have been developed to address RNA stability, resulting in significantly enhanced quantity and quality of RNA extracted from whole blood [15,16]. The two most widely used blood RNA stabilizing systems are PAXgene® and Tempus™ [17,18]. These systems contain proprietary solutions that lyse cells, inactivate RNases, and minimize changes in gene expression [19]. Those systems are designed to facilitate long-term whole-blood storage at low temperatures (−80 °C), removing the necessity of isolating RNA immediately post-collection and allowing batched processing [8,20]. However, each system has different stabilization efficiencies, thus impacting the resultant transcriptional profiles [21]. Several studies have investigated the relative performance of PAXgene® and Tempus™ in specific suboptimal conditions [17,20]. There is, however, an unmet need to identify the optimal RNA stabilization system for use in suboptimal tropical conditions, including high temperatures and extended times before storage at −80 °C.

In this study, we compared the effects of extended storage times and high temperatures simulating suboptimal tropical conditions before freezing of whole-blood RNA stabilized in either the PAXgene® or Tempus™ systems.

2. Results

2.1. Magnetic-Bead and Spin-Column-Based RNA Purification Systems Extracted Equivalent Whole-Blood RNA Quantity, Quality, and Purity

The quantity, quality, and purity of RNA extracted were evaluated using two common RNA isolation systems [18]: spin-column-based (PAXgene® QIAGEN/BD; Tempus™ Applied Biosystems) or magnetic-bead-based (MagMAX™ Life Technologies; compatible with either PAXgene® or Tempus™ collection tube) systems. RNA concentration (ng/µL) and purity (A260/A280 and A260/A230) were evaluated by spectroscopic quantification using NanoPhotometer® N60 (Implen, München, Germany). RNA integrity number (RIN) was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA yield (ng) was normalized to whole-blood volume collected in each blood RNA collection system (i.e., 2.5 mL in PAXgene® vs. 3 mL in Tempus™). All RNA samples were extracted fresh post-collection for the comparison of RNA isolation protocols.

We found the RNA isolation protocol did not affect the amount of total RNA extracted, the RNA quality, or the purity of RNA (normalized total RNA: $p = 0.875$, RIN: $p = 0.124$, A260/A280: $p = 0.101$, A260/A230: $p = 0.318$, MagMAX™ vs. columns, unpaired t-test) (Figure S1). Additionally, all samples had RIN of >7.0 and an A260/A280 ratio between 1.98–2.15, suggesting recovery of high-quality and high-purity RNA from both RNA isolation systems. A260/230 ratios were used as a secondary measurement detecting presence of residual phenol/ethanol, salts, and carbohydrates that can affect RNA quality. We observed differences between column vs. MagMAX™ extraction methods although differences were not significant. Comparatively, better values were obtained with Tempus™ column extractions. However, RNA yields obtained from the Tempus™ tubes were significantly lower than those obtained with PAXgene® tubes ($p = 0.008$, unpaired t-test), suggesting that the PAXgene® tubes gave higher RNA yields than Tempus™ when extracted in optimal laboratory conditions (Figure S1). Given the high quality of the extracted RNA and the advantages of extracting in a 96-well-plate format [6], magnetic bead-based (MagMAX™) RNA purification system compatible with PAXgene® and Tempus™ systems were used for the rest of the extractions.

2.2. A Higher Quantity of RNA Was Obtained Using Tempus™ Blood RNA Tubes in Suboptimal Tropical Conditions

Next, we compared the quantity, quality, and purity of RNA extracted from whole-blood samples stored in suboptimal tropical conditions collected in PAXgene® and Tempus™ tubes. To simulate suboptimal tropical conditions, matched whole-blood samples were collected in either PAXgene® or Tempus™ tubes and stored at different temperatures (25, 30, 35, or 40 °C) for different lengths of time (0, 1, 5, 7, or 10 days) before final storage at −80 °C for later extraction. These samples were compared to matched samples immediately frozen

at −80 °C for later extraction (D1/Control) or unmatched samples collected and processed post collection immediately in optimal laboratory conditions (D0/Fresh). The tube type had no effect on RNA quantity (i.e., normalized RNA concentration) when extracted in Fresh ($p = 0.065$, paired t-test) or Control ($p = 0.274$, paired *t*-test) conditions (Figure 1A). However, we found that tube type significantly affected normalized RNA concentration ($p < 0.0001$; Table 1, Model 1) in samples subjected to suboptimal tropical conditions.

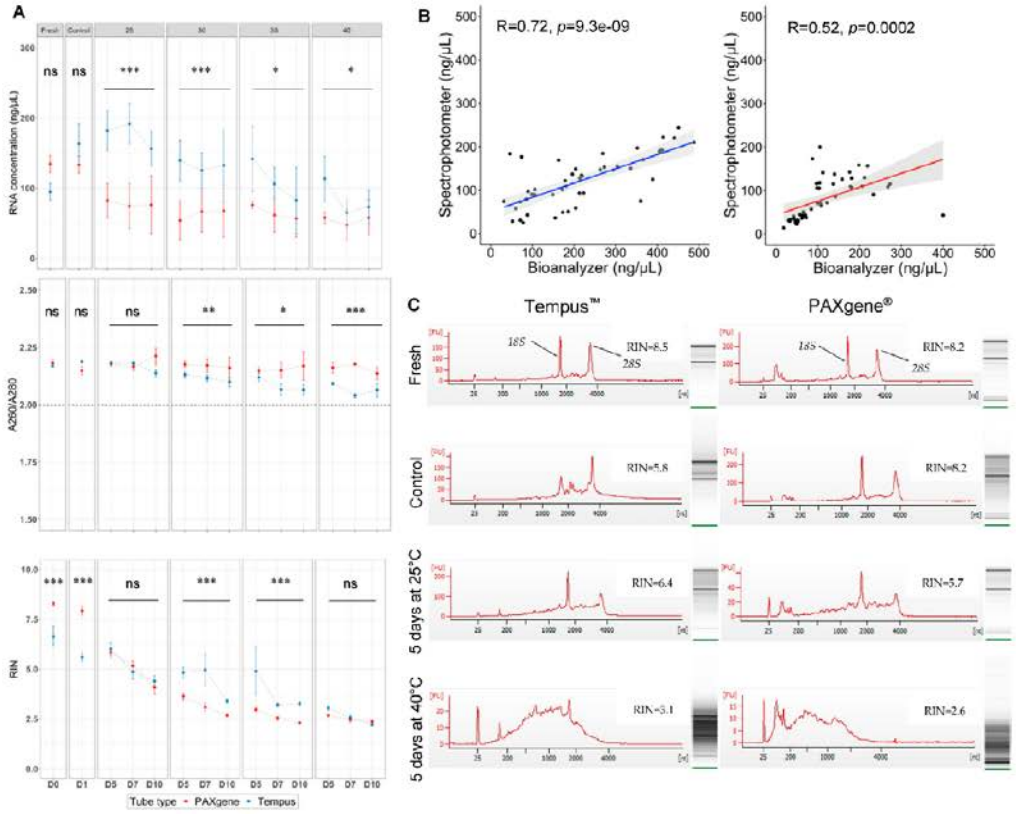


Figure 1. Comparison RNA concentration, purity, and integrity between PAXgene® and Tempus™ blood collection systems under suboptimal tropical storage conditions. (**A**) Mean RNA yield normalized to whole-blood volume (ng/µL), A260/A280 ratios determined spectrophotometrically, and RNA integrity number (RIN) across the different conditions (Fresh ($n = 8$), Day1/Control; immediately frozen at −80 °C for later extraction and various temperatures (25, 30, 35, 40 °C) (matched subjects $n = 3$)). (**B**) Scatter plots revealed correlations between normalized RNA concentration (ng/uL) by Spectrophotometer or Bioanalyzer for PAXgene® (left panel) and Tempus™ (right panel). Pearson's correlation assessed correlations between variables. (**C**) Electropherograms showed two distinct peaks (*28S* and *18S*), gel images showed two bands comprising the *28S* and *18S* from high-quality RNA, and smears indicated RNA degradation. The uncropped non-quantitative gel per *n* is shown; *18S* and *28S* peaks of sample #3 (Fresh) and #2 (Control, 5 days at 25 °C and 5 days at 40 °C). Note the different scales in Figure 1C. The dashed line indicates OD260/A280 = 2.0 for high-quality RNA. Effect of tube type in the linear regression models for each temperature point are indicated; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns, non-significant; RIN, RNA integrity number.

Table 1. Multiple linear regression models. Model 1 on normalized RNA concentration (ng/µL), Model 2 on A260/A280 ratios, and Model 3 on RIN (expressed as $\log_2$ RIN).

Model 1 = RNA Concentration (ng/µL)				
Explanatory variable	Estimate	Std. Error	*t* value	*p*-value
(Intercept)	66.230	78.853	0.84	0.404
Tube type: Tempus™	254.894	47.339	5.384	1.15×10^{-6} ***
Days	5.049	9.997	0.505	0.615
Temperature	−0.106	2.358	−0.045	0.964
Tube-type Tempus™: Days	−5.859	3.342	−1.753	0.084
Tube-type Tempus™: Temperature	−4.650	1.228	−3.786	0.0003 *
Days: Temperature	−0.170	0.299	−0.57	0.571
Adjusted R-squared: 0.764 (*p* value: $<2.2 \times 10^{-16}$)				
Model 2 = A260/A280				
Explanatory variable	Estimate	Std. Error	t value	*p*-value
(Intercept)	2.181	0.102	21.466	$<2 \times 10^{-16}$ ***
Tube type: Tempus™	0.161	0.061	2.638	0.011 *
Days	0.008	0.013	0.606	0.547
Temperature	0.000	0.003	−0.085	0.933
Tube-type Tempus™: Days	−0.008	0.004	−1.951	0.055
Tube-type Tempus™: Temperature	−0.005	0.002	−3.075	0.003 **
Days: Temperature	0.000	0.000	−0.567	0.572
Adjusted R-squared:0.565 (*p*-value: 1.458×10^{-10})				
Model 3 = $\log_2$(RIN)				
Explanatory variable	Estimate	Std. Error	t value	*p*-value
(Intercept)	4.905	0.572	8.581	3.45×10^{-12} ***
Tube type: Tempus™	0.338	0.343	0.985	0.328
Days	−0.148	0.072	−2.047	0.045 *
Temperature	−0.082	0.017	−4.777	1.10×10^{-5} ***
Tube-type Tempus™: Days	−0.020	0.024	−0.836	0.407
Tube-type Tempus™: Temperature	0.002	0.009	0.236	0.814
Days: Temperature	0.002	0.002	1.057	0.295
Adjusted R-squared: 0.797 (*p* value: $<2.2 \times 10^{-16}$)				

Independent multiple linear regression models: Model 1 on normalized RNA concentration, Model 2 on A260/A280 ratios, and Model 3 on $\log_2$(RIN) values as dependent variable and tube types, days, temperatures, and subjects as the independent variables. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

To investigate these data further, we applied three multiple linear regression models to evaluate the effects of explanatory variables (i.e., tube type, storage times, temperature, and biological subject) on normalized RNA concentrations, A260/A280 ratios, and RIN values (Table 1). Model 1 explained 76% of the variation in extracted RNA quantity ($R^2 = 0.764$, $p < 2.2 \times 10^{-16}$). Tempus™ tubes had a significant effect on the RNA yield ($p = 1.15 \times 10^{-6}$), and temperature variation in tube type significantly impacted RNA yield ($p = 0.0003$; Table 1, Model 1). Additionally, we applied separate linear models to each temperature condition to explore the effect of the temperature on the tube type (Table S3). We found that the concentration of RNA extracted from whole-blood collected in Tempus™ tubes was significantly greater than for PAXgene® tubes at all evaluated temperatures ($p = 7.16 \times 10^{-6}$ (25 °C); $p = 0.0004$ (30 °C); $p = 0.033$ (35 °C); $p = 0.023$ (40 °C), Table S3). As a secondary measurement of RNA concentration, we obtained the RNA concentration readings from Agilent Bioanalyzer. Similarly, Tempus™ tubes gave higher RNA yields when measured with Agilent Bioanalyzer ($p = 0.001$; Table S4). Interestingly, RNA concentration measurements by spectrophotometer and bioanalyzer were more strongly correlated in Tempus™ than in PAXgene® tubes ($p = 0.0002$, $R^2 = 0.270$ (PAXgene®); $p = 9.3 \times 10^{-9}$,

R^2 = 0.518 (Tempus™)); Figure 1B), suggesting that tube-specific contents influence concentration measurements.

When considering RNA purity, all extracted RNA samples had A260/A280 ratios >2 regardless of tube type, indicating a high purity under all test conditions (Figure 1A, middle panel). There was no difference in A260/A280 ratios between tube type for RNA extracted from fresh and control RNA samples ((p = 0.480, (Fresh); p = 0.111, (Control), paired t-test). However, at tropical storage and temperature conditions, the tube type (p = 0.011) and the storage temperature on tube type (p = 0.003) had significant effects on A260/A280 ratios as per the multiple linear regression model (Table 1, Model 2). We used A260/230 ratio as a secondary measurement of RNA purity although with low quantities, as ng/µL RNA A260/230 ratios are highly variable A260/230 ratios (Figure S6). A260/230 ratios are irrelevant for RNA quality and only describe dissolved solvent concentration relative to RNA. Linear models further indicated that the higher temperatures decreased purity as evidenced by A260/A280 ratio (30 °C (p = 0.007), 35 °C (p = 0.011) and 40 °C ($p = 4.89 \times 10^{-5}$), Table S3). Taken together, these data demonstrated that higher RNA yields are extracted from Tempus™ blood tubes compared to PAXgene® tubes in suboptimal tropical conditions. These data also showed that RNA yields significantly decreased with increasing temperature in both PAXgene® and Tempus™ tubes.

2.3. A Higher Quality of RNA Was Obtained Using PAXgene® Tubes in Optimal Laboratory Conditions

We determined if high RNA quality was preserved using PAXgene® or Tempus™ tubes in suboptimal tropical conditions and compared the RNA quality with samples extracted in optimal laboratory conditions. RIN values declined over time and temperature irrespective of the whole-blood collection tube (Figure 1A). PAXgene® had significantly higher RIN values in fresh (p = 0.013, paired t-test) and control conditions (p = 0.001, paired t-test) compared to Tempus™. The electropherograms showed comparable results for different conditions applied on PAXgene® and Tempus™ systems (Figure 1C). Ribosomal RNA bands were clearly visible in fresh extractions and control samples. Most RNA eluates stored at room temperature (25 °C) for 5–7 days obtained RIN values around 5–6 with visible *18S* and *28S* bands. In contrast, RNA stored at 40 °C did not show distinct rRNA banding. These data demonstrated that higher-quality RNA was obtained with PAXgene® tubes (compared to Tempus™) when RNA was extracted post collection immediately or when samples were maintained at the optimal storage conditions as recommended by the manufacturers.

2.4. The Highest Quality RNA Was Obtained Using Tempus™ Blood RNA Tubes in Suboptimal Tropical Conditions

A multiple linear regression model was built to explore the effects of suboptimal tropical conditions on RNA integrity to determine the effect of tube type, storage temperature, and storage time on RNA integrity (measured by RIN values) (Table 1, Model 3). Model 3 explained approximately 79.7% (R^2 = 0.797) of the variation in RIN values. RIN values decreased significantly over time (p = 0.045), decreasing by 0.862 per day (Model 3, days estimate = −0.148; $e^{-0.148}$ = 0.862). RIN values also decreased significantly with the increasing of temperature ($p < 0.0001$), decreasing by 0.921 (Model 3, temperature estimate = −0.082; $e^{-0.082}$ = 0.921) for each degree Celsius (°C) increase in temperature. These data demonstrated that RNA integrity significantly decreased with the length of storage time and temperature.

The impact of storage time on RNA integrity (RIN values) was also analysed at each temperature for each tube type (Table S3). RIN values were significantly higher in Tempus™ than PAXgene® tubes at 30 °C (p = 0.0001) and 35 °C (p = 0.001). However, no significant differences in RIN at 25 °C (p = 0.787) or 40 °C (p = 0.399) was found for tube types (Figure 1A). In agreement with previously published studies [9,16], these results demonstrated that RNA integrity is temperature-sensitive, and that both tube types

produced low-quality RNA at increased storage times and temperatures. Nevertheless, our data suggest that Tempus™ tubes may provide better RNA integrity (higher RIN values) under certain suboptimal tropical conditions compared to PAXgene® tubes.

2.5. *Tempus™ Tubes Maintain mRNA Integrity across Suboptimal Tropical Conditions*

In order to validate our RNA quality measurements, we quantified mRNA and rRNA extracted from PAXgene® and Tempus™ tubes using RT-qPCR. We tested the relative mRNA abundance of two human reference genes, *Succinate dehydrogenase complex, subunit A* (*SDHA*) and *TATA-box-binding protein* (*TBP*), and one rRNA transcript, *18S ribosomal RNA* (*18S*). The RNA concentration of all samples was normalized at pre-cDNA synthesis (i.e., at 30 ng/µL). Hence, an increasing cycle threshold (Ct) value indicated a decreasing relative transcript quality rather than abundance [22]. We tested RT-qPCR primer sets designed to amplify different-sized fragments of the same target gene (i.e., amplicons between 100–300 bp) and differences in the relative RNA quality (e.g., increased Ct values) would be expected to be intensified when assaying genes with primers amplifying larger amplicons.

There was no significant difference in the mean Ct values between the tube types for Control samples from smaller amplicons ($p = ns$, paired *t*-test: D1/Control, 100–200 bp) or with larger amplicons ($p = ns$, paired *t*-test: D1/Control, 200–300 bp, Figure 2). For example, mean Ct values obtained for larger amplicons (200–300 bp) tested in PAXgene® tubes (**22.32** (*18S*), **25.48** (*SDHA*), **27.39** (*TBP*)) and in Tempus™ tubes (**21.33** (*18S*), **25.59** (*SDHA*), **27.59** (*TBP*)) were largely consistent. In addition, we found no statistically significant difference between matched fresh and control samples between PAXgene® and Tempus™ tubes (Figure S4). However, the Ct value varied significantly at tropical storage conditions across all three tested genes across short (100–200 bp) and medium amplicons (200–300 bp) (Figure 2). Tempus™ tubes maintained significantly higher transcript stability, as indicated by lower Ct values obtained for three tested genes compared to PAXgene® tubes at suboptimal tropical conditions (Table S5).

Multiple comparison testing found that the *18S* rRNA Ct values were not statistically significantly influenced by incubation temperature or duration when rRNA was collected in either PAXgene® or Tempus™ tubes. In contrast, mRNA (*SDHA* and *TBP*) collected in PAXgene® tubes were significantly impacted by storage time and temperature compared to Tempus™ tubes (Table S6).

We showed that the RNA degraded samples, as indicated by the decreasing level of RIN, had higher Ct values (Figure 3A). The Ct shifted towards higher cycle numbers for *SDHA* with larger amplicons than short- and medium-length amplicons, which was much more evident in PAXgene® tubes than in Tempus™ (Figure 3A). These results indicated that relative overall stability in terms of mRNA expression levels was maintained in Tempus™ compared to PAXgene® tubes. A similar relationship between RIN and Ct values was observed for *TBP* (Figure S5). However, as clearly shown in Figure S5, both tube types had Ct < 30 for all product lengths for *18S*, suggesting both PAXgene® and Tempus™ tubes preserved rRNA at suboptimal tropical conditions. As indicative of decreasing relative transcript quality, increasing Ct values were validated by correlating change in Ct values with RIN. We considered the change in Ct values (ΔCt) as the difference between samples collected under suboptimal tropical conditions and the mean of the control samples. Strong statistically significant correlations were found between ΔCt and RIN for all tested genes (Figure 3B). These negative correlations indicated that with the decreasing RIN values, the ΔCt of 200–300bp amplicons increased, thus validating the use of RT-qPCR to assess the quality of the RNA. Taken together, these data demonstrate that Tempus™ collection tubes better maintain mRNA stability in suboptimal tropical conditions even despite a decreasing RIN.

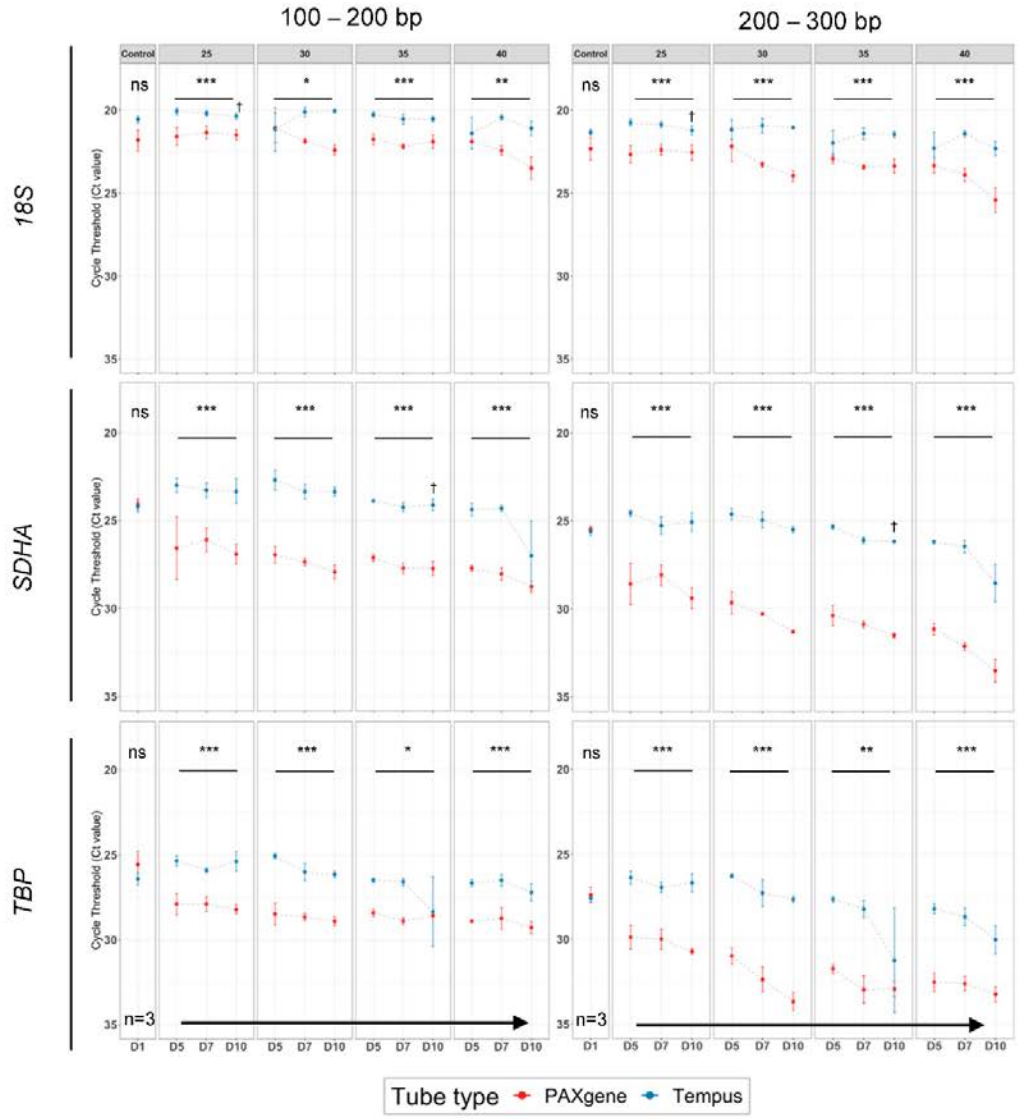


Figure 2. Cycle thresholds (Ct) of housekeeping genes assessing RNA from PAXgene® and Tempus™ stored at different temperatures and times. The mean cycle threshold (Ct) values for *18S*, *SDHA*, and *TBP* with 100–200 base pair (bp) (short-amplicon) or 200–300 bp (medium-amplicon) lengths across different conditions (control and at multiple storage temperatures and days (matched $n = 3$)). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; ns, non-significant; †, data from only two observations were potentially available due to LOD. Blue, Tempus™ Blood RNA tubes; red, PAXgene® Blood RNA tubes.

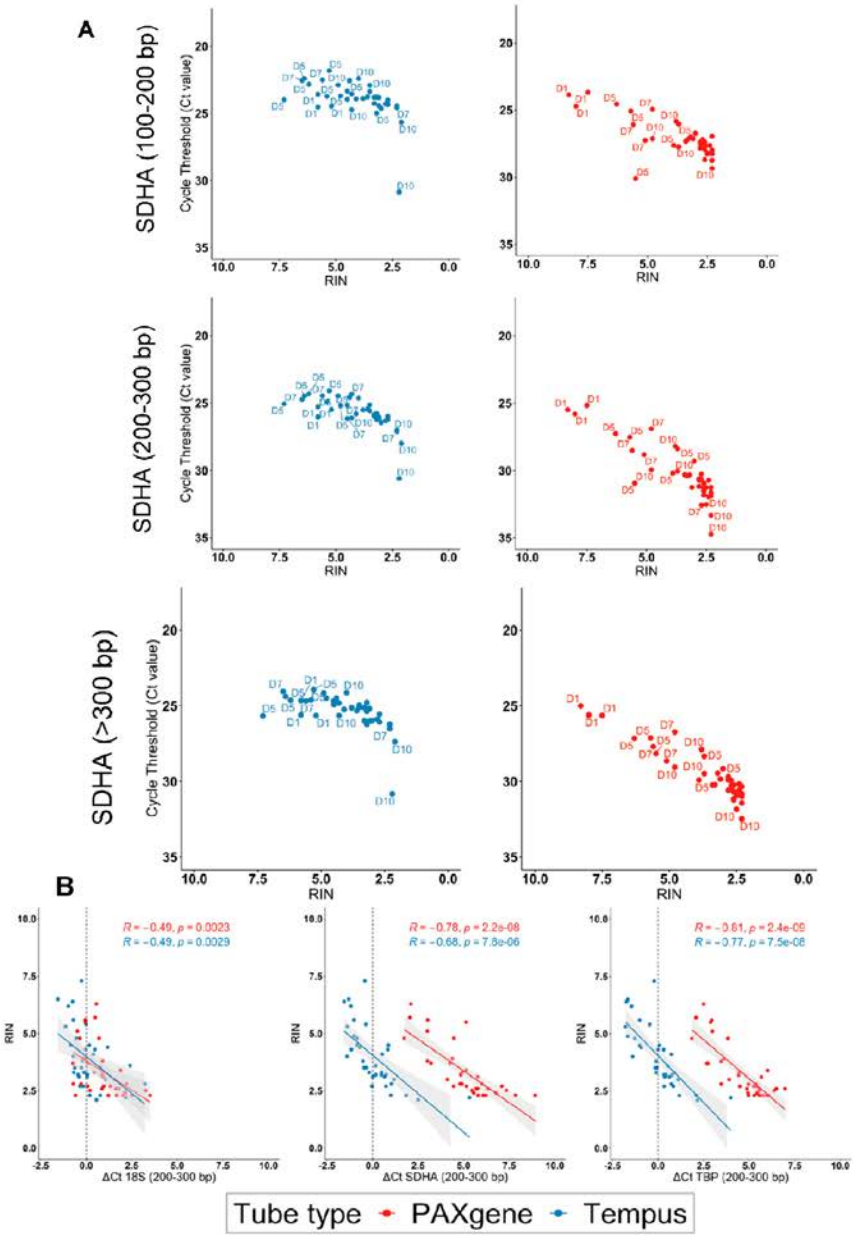


Figure 3. Cycle thresholds (Ct) in dependence on amplicon length and RNA integrity (RIN). (**A**) Scatter plots show the Ct values in dependence on amplicon length and RNA integrity (RIN) for different lengths of *SDHA* amplicons differences for each tube type. (**B**) Spearman correlation of the ΔCt values with RIN value (RNA quality) for medium-length amplicons (200–300 bp) of *18S* (**left**), *SDHA* (**middle**), and *TBP* (**right**). Blue, Tempus™ Blood RNA tubes; red, PAXgene® Blood RNA tubes.

To test if the presence of PCR inhibitors, which are often co-extracted from whole blood (e.g., haemoglobin, lactoferrin, anticoagulants, etc.) [23,24], could have contributed to these results, RT-qPCR was performed on a $\log_2$ serial dilution of undiluted extraction eluent. We considered that a trendline gradient of Ct values relative to the dilution greater than −3.3 (i.e., E′ < 100%) was indicative of the presence of PCR inhibitors [23]. There was no apparent effect of inhibitors in both PAXgene® or Tempus™ tubes when the samples were diluted below 60 ng/uL (Figure S3). These data demonstrated that our findings were unlikely to be a consequence of inhibitors present in the RT-qPCR reaction.

In summary, our data showed that Tempus™ tubes maintained a higher RNA quantity and integrity comparatively to PAXgene® tubes when RNA is stored in suboptimal tropical conditions. Furthermore, Tempus™ tubes maintained stability of mRNA in conditions where RNA samples were heavily degraded as indicated by RIN. Taken together, this study establishes that the Tempus™ blood RNA collection system resulted in a better quality of RNA and enhanced stability of mRNA when whole-blood samples are stored under suboptimal tropical conditions.

3. Discussion

Gene expression profiling with molecular techniques such as RT-qPCR and next-generation sequencing requires high-quality intact RNA. It is well-established that the pre-analytical variables in blood sample collection and processing have profound effects on RNA quality that may consequently introduce substantial technical bias for molecular analysis [11,25]. Pre-analytical handling of blood samples and storage can be challenging in tropical remote field study settings where freezing at −80 °C immediately post-collection may not be an option. Here, we evaluated PAXgene® and Tempus™ blood RNA stabilization tubes for preserving RNA quantity, purity, quality, and gene transcript stability at suboptimal tropical conditions.

According to the respective manufacturers, PAXgene® blood RNA tubes effectively stabilize RNA for up to three days at room temperature, five days at 2–8 °C, and up to 11 years at −20 °C or −70 °C, whilst Tempus™ blood RNA tubes stabilize RNA for up to five days at room temperature and at least a week at 4 °C or −80 °C for long-term storage [26,27]. Duale et al. showed that the RNA yield, quality, and integrity were stable up to six years of storage at −80 °C in Tempus™ blood RNA tubes [8].

This study determined the impact of warm tropical temperatures (25, 30, 35, and 40 °C) and prolonged storage times (0, 5, 7, and 10 days) on total RNA yield, purity, quality, and transcript stability of the two most widely used commercially available blood RNA stabilizing systems, PAXgene® and Tempus™. These conditions were selected to simulate conditions in field sites in tropical or subtropical regions for a 10-day field trip, representing a challenging situation for preserving RNA. The performances of commercially available kits with columns (spin-column-based) and MagMAX™ (magnetic-beads-based) protocols were used to extract RNA from blood collected in PAXgene® and Tempus™ tubes. The total RNA yield, RNA integrity (RIN), and purity were used as performance measures.

In experiment A, both columns vs. MagMAX™ extraction methods had no significant differences in normalized total RNA yields, RIN, and A260/280 ratios in PAXgene® and Tempus™ tubes. However, total RNA obtained from Tempus™ tubes was significantly lower than PAXgene® tubes. We observed comparable average OD 260/280 ratios and RIN > 7 for column and MagMAX™ extraction methods for blood collected in PAXgene® and Tempus™ tubes. One limitation of experiment A was that only two biological subjects were evaluated for Tempus™ column vs. MagMAX™ extractions due to the unavailability of Tempus™ blood collection tubes of the same batch. An A260/A280 ratio between 1.8 to 2.2 indicates highly purified RNA with minimum DNA contamination [28]. These data demonstrated that a similar quantity, quality, and purity of RNA could be obtained using either spin-column or magnetic-bead RNA purifications when purifying RNA from whole-blood collected in either PAXgene® or Tempus™ systems. Higher RIN value indicates better RNA integrity, and RIN values above seven are considered ideal for high-throughput

downstream applications [29,30]. However, RNA samples with a RIN of five have been used in gene expression studies [14,31]. MagMAX™ RNA extractions were used as the method of RNA purification for the rest of the study.

Both PAXgene® and Tempus™ tubes gave similar RNA yields and purity (A260/A280) when the samples were extracted fresh or post freezing, whilst RIN values were significantly higher in blood samples extracted from PAXgene® than Tempus™. The reason for this difference is not yet known. Overall RIN values were much lower in control samples compared to freshly extracted RNA. A previous study investigating the impact of storage duration (24, 32, and 40 h) and storage temperatures (24 °C, 4 °C, and −80 °C) of whole blood collected in heparin tubes on the qualities of DNA and RNA showed that RNA integrity declined dramatically when the samples were frozen [9]. Freezing blood samples will lead to irreversible cellular damages, causing osmotic and ice injuries of red blood cells due to water crystallization [32]. Activated intracellular enzymes such as RNases can be released upon thawing, resulting in RNA degradation [33].

The linear regression model on normalized RNA concentrations indicated that the Tempus™ tubes result in a higher RNA yield than PAXgene®, while we demonstrated that the tube type and temperature significantly affect RNA yields. However, similar RNA purities were obtained from both tube types. A similar study by Duale et al., comparing PAXgene® vs. Tempus™ tubes stored for 0, 2, 5, and 7 days at RT (~22 °C) and then stored at −80 °C until extraction, showed that RNA yields collected in the Tempus™ tubes were consistently higher than PAXgene® tubes. However, RNA quality (average 260/280 ratios and RIN values) was similar in both systems [17]. Consistent with our study, other studies have reported that higher RNA yields were obtained with Tempus™ tubes compared to PAXgene® tubes [34–37]. However, we observed low A260/230 ratios (less than 2.0) for both tube types in our study. This ratio is decreased in the presence of residual phenol, salts, and carbohydrates that can affect the accuracy of downstream application and used as a secondary measurement for RNA purity [38]. Historically, low A260/A230 ratios are reportedly attributed to the high salt content of the elution buffers contained in PAXgene® extraction kits [16,17,34,35] and as well as in Tempus™ extraction kits [16,22,39].

Our results indicated a gradual decrease of RNA quality in terms of RIN values over storage duration and increased temperatures in both tube types. However, RNA extracted from Tempus™ tubes had improved RINs compared to PAXgene® at suboptimal tropical conditions. RNA integrity was influenced mainly by increased storage temperatures at higher temperatures. It has been well-documented that RNA molecules are sensitive to physical degradation due to high temperatures [40]. However, the effects of higher temperatures on Tempus™ and PAXgene® blood RNA stabilizing systems have not been previously studied. Our data suggested that good quality RNA (average RIN >5) can be obtained in both tube types when samples are kept at 25 °C room temperature for up to 5 days of storage duration. Overall, we demonstrated that satisfactory amounts of good-quality RNA can be achieved using blood RNA stabilizing systems in warm tropical temperatures (25–30 °C) and at storage times up to a week. When the storage temperatures are above 30 °C, the RNA quality drops significantly and may not be adequate in downstream applications. RT-qPCR data further demonstrated that PAXgene® tubes do not preserve mRNA with the same efficiency as Tempus™ tubes, but both tubes equally preserved rRNA from degradation in suboptimal tropical conditions.

In most low-resource settings, microscopy and serological assays such as ELISAs remain the standard methods for diagnosis of tropical infections, especially in low-income and middle-income countries, despite limited sensitivity and specificity. More sensitive molecular methods have potential to inform disease, diagnosis, and treatment and to facilitate field-based intervention and biobanking studies (i.e., large-scale field trials). This report provides important information to facilitate such studies. In particular, our data show that the Tempus™ blood RNA collection system resulted in higher-quality RNA and maintained more consistent stability of mRNA when whole-blood samples were stored under suboptimal tropical conditions as compared with the PAXgene® system.

4. Materials and Methods

4.1. Sample Collection

Whole blood was collected from healthy adult volunteer donors into Tempus™ (3 mL) Blood RNA tubes (Applied Biosystems, Foster City, CA, USA) or PAXgene® (2.5 mL) Blood RNA tubes (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland) according to manufacturer's instructions. Briefly, whole blood was collected directly into each tube by standard venepuncture and immediately shaken vigorously for 10 s to ensure that the stabilizing reagent made uniform contact with the sample as per the manufacturer's instructions.

4.2. Experimental Design

Experiment A compared RNA yields, purity, and integrity from MagMAX™ extractions using spin columns or magnetic beads (Figure 4, top panel). All blood samples collected for experiment A were kept for two hours at 25 °C room temperature (RT) after collection and freshly extracted (i.e., no storage at −80 °C). Experiment B evaluated RNA yields, purity, and integrity in whole-blood samples stored at different temperatures (25, 30, 35, or 40 °C) and storage times (0, 5, 7, or 10 days; Figure 4). All samples collected for experiment B were frozen at −20 °C overnight then transferred to −80 °C until RNA extraction with magnetic-beads-based MagMAX™. All blood samples stored in PAXgene® tubes were thawed for 2 h at room temperature, whereas Tempus™ tubes were thawed for 30 min on ice prior RNA isolation.

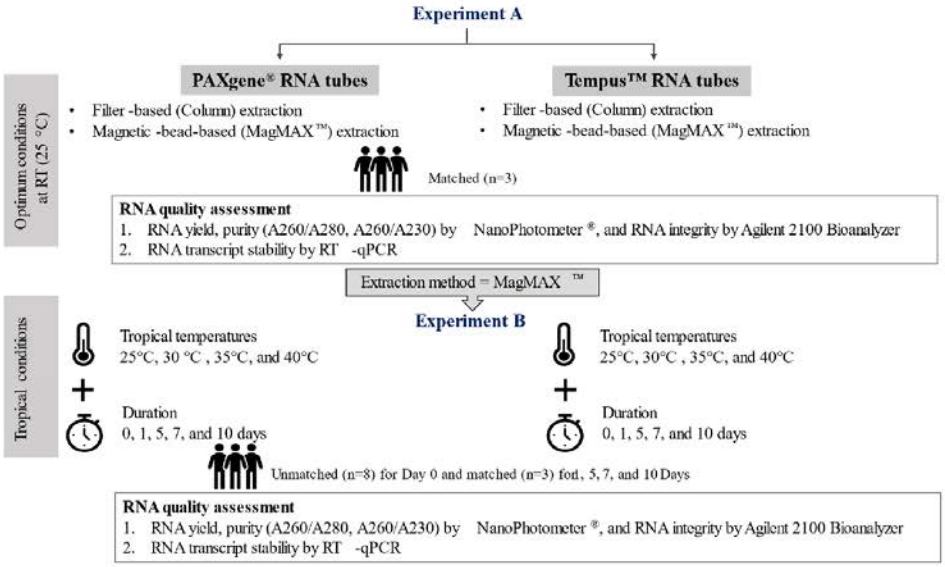


Figure 4. Experimental design. **Experiment A.** Comparison of total RNA yields, purity, and integrity using spin column-based and magnetic bead-based kits for the two types of blood stabilization systems (PAXgene® Blood RNA Tubes and Tempus™ Blood RNA Tubes). All RNA samples were extracted fresh post collection. **Experiment B.** Systematic testing of different temperatures (25, 30, 35, or 40 °C) and storage times (0, 1, 5, 7, or 10 days) were immediately frozen at −80 °C for later extraction with matched samples (n = 3). Samples immediately frozen at −80 °C for later extraction is referred to "D1/Control". Unmatched samples (n = 8) collected in PAXgene® or Tempus™ RNA stabilizing systems and processed post collection immediately at optimal laboratory conditions (D0/Fresh).

4.3. RNA Extraction

4.3.1. Column-Based RNA Purification

Total RNA from whole blood collected in PAXgene® tubes was extracted according to manufacturer's instructions using PAXgene® Blood RNA Kit (PreAnalytiX, QIAGEN/BD, Hombrechtikon, Switzerland), which included DNase I treatment. Total RNA was eluted in 40 µL elution buffer. According to the manufacturer's instructions, total RNA from blood collected in Tempus™ tubes was extracted using the Tempus™ Spin RNA Isolation Kit (Applied Biosystems, CA, USA). RNA was eluted in 90 µL of elution solution. DNase treatment was an optional step in the Tempus™ column extraction system and therefore not included, as genomic DNA contamination using this procedure is minimal (less than 0.05% by weight) according to the manufacturer's specifications.

4.3.2. Magnetic-Bead-Based RNA Purification (MagMAX™)

Total RNA from whole blood was extracted from PAXgene® Blood RNA Tubes using MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit (Life Technologies, CA, USA) according to the manufacturer's protocol, including a TURBO™ DNase and protease step. RNA was isolated using MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit, compatible with Tempus™ Blood RNA tubes (Life Technologies, CA, USA) following the manufacturer's protocol with TURBO™ DNase treatment. All extracted RNA samples were stored at −80 °C. The technical characteristics of each extraction method are summarised in Table S1.

4.4. RNA Yield, Purity and Integrity

RNA concentration (ng/µL), A260/A280, and A260/230 ratios to indicate RNA purity were measured by spectroscopic quantification using NanoPhotometer® N60 (Implen, München, Germany). RNA integrity was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the Eukaryote Total RNA Nano assay, complementing RNA 6000 NanoChip kit (Agilent Technologies, CA, USA), following the manufacturer's instructions. The RNA integrity number (RIN) was calculated by the Agilent 2100 Expert software (Version B.02.10.SI764, Agilent). The RIN ranges from 1 to 10; an RIN of fully intact RNA is 10, and an RIN of completely degraded RNA is 1.

4.5. Reverse Transcription Quantitative PCR (RT-qPCR)

4.5.1. Reverse Transcription

All reverse transcription (RT) reactions were conducted using the SuperScript IV™ First-Strand Synthesis System™ (ThermoFisher Scientific, Waltham, MA, USA). All samples were primed with 37.5 ng of random hexamers and 10 mM dNTPs at 65 °C for 5 min and then 4 °C for 1 min. Reverse transcription was then performed using the SuperScript IV™ reverse-transcriptase (SSIV) for 10 min at 23 °C, 10 min at 50 °C, and 10 min at 85 °C. SSIV concentration was assessed at 20 U (20 units/µL RNA) reactions compared with 5 U reactions (Figure S2) as previously described [41]. All subsequent RNA samples were reverse transcribed at 30 ng/µL (Figure S3) using 5 U reactions in 15 µL total volume reactions for test conditions. All cDNA samples were stored at 4 °C.

4.5.2. Quantitative PCR (qPCR)

qPCR was run with 5 µL total reaction volume using SsoAdvanced SYBR® SuperMix (Bio-Rad, Hercules, CA, USA), which facilitates excellent reaction efficiencies [6]. All reactions contained 0.5 µM of desalt-grade primers (Sigma-Aldrich) with 0.75 ng/uL sample cDNA. Each sample was run in technical triplicate replicate, followed by a melt curve analysis to ensure primer specificity. Primers used for the RT-qPCR assays were sourced from Primer Bank™ [42] (Table S2). Primer efficiencies were calculated as per MIQE guidelines [43] as previously published from cDNA standards [6]. Reaction efficiency was calculated from $\log_2$ dilutions of pooled cDNA from 1×10^6 unstimulated PBMCs. The PCR cycling program included an enzyme activation step at 95 °C for 2 min and then

40 cycles of annealing and extension at 95 °C for 15 s and 60 °C for 30 s, respectively. The cycle threshold (Ct) value was set to 0.3 ΔRN, and a pooled cDNA positive control was included across all plates to ensure reproducibility. qPCR was performed using the QuantStudio 5 Real-Time PCR system running QuantStudio Design and Analysis Software (v1.5.1, Applied Biosystems).

4.6. Statistical Analysis

Statistical analyses were performed using R statistical software (https://www.r-project.org/ (accessed on 24 November 2021, RStudio Inc., Boston, MA, USA, Version 1.4.1103). Unpaired *t*-tests were used to compare the data from spin-column-based and magnetic-bead-based MagMAX™ RNA extractions to determine any significant difference between RNA isolation systems for total RNA yield, A260/A280 ratios, and RNA integrity. Statistical significance was defined using p-values < 0.05. Normal distribution of data and normality of residuals were evaluated using the Shapiro–Wilks test.

The average Ct values for each replicates/triplicate and targets with Ct-values > 35 or undefined were considered beyond the limit of detection (LOD) and removed from the analysis [17]. Paired *t*-tests determined the differences of RNA yields normalized to input whole blood volume, A260/A280 ratios, RIN values, and Ct values for short and medium lengths of all housekeeping genes between the two RNA stabilization systems (PAXgene® and Tempus™) for RNA extracted on fresh and control conditions. Whole-blood samples processed post collection immediately at optimal laboratory conditions without −80 °C storage are referred to as D0/Fresh. Samples frozen immediately at −80 °C for later extraction are referred to as D1/Control.

Multiple linear regression models were fit to investigate the overall relationship of independent variables at experimental conditions (i.e., tube type, temperature, and storage time) on normalized RNA yields, A260/A280 ratios, RIN values, or Ct values. To further explore the significant interaction between temperatures and tube type, separate analyses were performed at each temperature. To compare the changes of Ct values on control and test conditions on PAXgene® and Tempus™ tube types, we performed a two-way analysis of variance with multiple comparisons.

5. Conclusions

We conclude that collection of whole blood samples in Tempus™ tubes is the preferred system of choice for gene expression and molecular studies in rural and remote resource-limited settings where electricity and storage facilities are compromised. Our findings are especially relevant to research on RNA biology, which could help future directions of diagnosis, treatment, and interventions against diseases that are prevalent in tropical countries, including neglected tropical diseases.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms231810609/s1.

Author Contributions: Y.D.S., D.J.B., C.P. and D.L.D., designed the study; Y.D.S. recruited volunteers, drafted the manuscript, prepared figures, and analysed the data; Y.D.S. prepared whole-blood lysates, isolated RNA, and performed bioanalyzer analysis of RNA and quantification; D.J.B. and A.M.K. designed primers, performed RT-qPCR and the gene expression analysis, and helped prepare the manuscript; C.P., D.J.P., C.M.R. and J.W. contributed to data analysis and interpreting the results in the context and reviewed the manuscript. All authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Health and Medical Research Council of Australia (NHMRC Program Grant #1069466). Y.D.S. is supported by a Postgraduate Research Scholarship from James Cook University. A.M.K. is supported by a Scholarship from the Australian Institute of Tropical Health and Medicine. D.J.B. is supported by a James Cook University Prestige Research Training Program Stipend. D.L.D. is supported by an NHMRC Principal Research Fellowship (#1137285).

Institutional Review Board Statement: Protocols for obtaining volunteer blood samples were reviewed and approved by the James Cook University Human Research Ethics Committee (H7886). All participants provided written informed consent. All experiments were performed in compliance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: We are grateful to the volunteers who provided samples for this study and thank Yide Wong for his support as a phlebotomist.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mohr, S.; Liew, C.-C. The peripheral-blood transcriptome: New insights into disease and risk assessment. *Trends Mol. Med.* **2007**, *13*, 422–432. [CrossRef]
2. Zaas, A.K.; Chen, M.; Varkey, J.; Veldman, T.; Hero, A.O., 3rd; Lucas, J.; Huang, Y.; Turner, R.; Gilbert, A.; Lambkin-Williams, R.; et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* **2009**, *6*, 207–217. [CrossRef]
3. Holcomb, Z.E.; Tsalik, E.L.; Woods, C.W.; McClain, M.T. Host-Based Peripheral Blood Gene Expression Analysis for Diagnosis of Infectious Diseases. *J. Clin. Microbiol.* **2017**, *55*, 360–368. [CrossRef]
4. Gliddon, H.D.; Herberg, J.A.; Levin, M.; Kaforou, M. Genome-wide host RNA signatures of infectious diseases: Discovery and clinical translation. *Immunology* **2018**, *153*, 171–178. [CrossRef]
5. Gautam, A.; Donohue, D.; Hoke, A.; Miller, S.A.; Srinivasan, S.; Sowe, B.; Detwiler, L.; Lynch, J.; Levangie, M.; Hammamieh, R.; et al. Investigating gene expression profiles of whole blood and peripheral blood mononuclear cells using multiple collection and processing methods. *PLoS ONE* **2019**, *14*, e0225137. [CrossRef]
6. Browne, D.J.; Brady, J.L.; Waardenberg, A.J.; Loiseau, C.; Doolan, D.L. An Analytically and Diagnostically Sensitive RNA Extraction and RT-qPCR Protocol for Peripheral Blood Mononuclear Cells. *Front. Immunol.* **2020**, *11*, 402. [CrossRef]
7. Kukurba, K.R.; Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.* **2015**, *2015*, 951–969. [CrossRef]
8. Duale, N.; Lipkin, W.I.; Briese, T.; Aarem, J.; Rønningen, K.S.; Aas, K.K.; Magnus, P.; Harbak, K.; Susser, E.; Brunborg, G. Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank—Evaluation of RNA quality and stability. *BMC Res. Notes* **2014**, *7*, 633. [CrossRef]
9. Huang, L.-H.; Lin, P.-H.; Tsai, K.-W.; Wang, L.-J.; Huang, Y.-H.; Kuo, H.-C.; Li, S.-C. The effects of storage temperature and duration of blood samples on DNA and RNA qualities. *PLoS ONE* **2017**, *12*, e0184692. [CrossRef]
10. Malentacchi, F.; Pizzamiglio, S.; Wyrich, R.; Verderio, P.; Ciniselli, C.; Pazzagli, M.; Gelmini, S. Effects of Transport and Storage Conditions on Gene Expression in Blood Samples. *Biopreserv. Biobank.* **2016**, *14*, 122–128. [CrossRef]
11. Malentacchi, F.; Pazzagli, M.; Simi, L.; Orlando, C.; Wyrich, R.; Günther, K.; Verderio, P.; Pizzamiglio, S.; Ciniselli, C.M.; Zhang, H.; et al. SPIDIA-RNA: Second External Quality Assessment for the Pre-Analytical Phase of Blood Samples Used for RNA Based Analyses. *PLoS ONE* **2014**, *9*, e112293. [CrossRef] [PubMed]
12. Qin, J.; Williams, T.L.; Fernando, M.R. A novel blood collection device stabilizes cell-free RNA in blood during sample shipping and storage. *BMC Res. Notes* **2013**, *6*, 380. [CrossRef]
13. Donohue, D.E.; Gautam, A.; Miller, S.A.; Srinivasan, S.; Abu-Amara, D.; Campbell, R.; Marmar, C.R.; Hammamieh, R.; Jett, M. Gene expression profiling of whole blood: A comparative assessment of RNA-stabilizing collection methods. *PLoS ONE* **2019**, *14*, e0223065. [CrossRef]
14. Opitz, L.; Salinas-Riester, G.; Grade, M.; Jung, K.; Jo, P.; Emons, G.; Ghadimi, B.M.; Beißbarth, T.; Gaedcke, J. Impact of RNA degradation on gene expression profiling. *BMC Med. Genom.* **2010**, *3*, 36. [CrossRef]
15. Langelaan, M.L.P.; Dylus, J.; Bock, E.; Jongen, B.; Mertens, A.A.M.; Raijmakers, M.T.M. Improved pre-analytical process for RNA isolation from whole blood samples. *Ned. Tijdschr. Klin. Chemie Lab.* **2014**, *39*, 164–165.
16. Feddersen, S.; Bastholt, L.; Pedersen, S.M. Stabilization of circulating thyroglobulin mRNA transcripts in patients treated for differentiated thyroid carcinoma. *Ann. Clin. Biochem.* **2017**, *54*, 558–566. [CrossRef]
17. Duale, N.; Brunborg, G.; Ronningen, K.S.; Briese, T.; Aarem, J.; Aas, K.K.; Magnus, P.; Stoltenberg, C.; Susser, E.; Lipkin, I.W. Human blood RNA stabilization in samples collected and transported for a large biobank. *BMC Res. Notes* **2012**, *5*, 510. [CrossRef]
18. Aarem, J.; Brunborg, G.; Aas, K.K.; Harbak, K.; Taipale, M.M.; Magnus, P.; Knudsen, G.P.; Duale, N. Comparison of blood RNA isolation methods from samples stabilized in Tempus tubes and stored at a large human biobank. *BMC Res. Notes* **2016**, *9*, 430. [CrossRef]
19. Yip, L.; Fuhlbrigge, R.; Atkinson, M.A.; Fathman, C.G. Impact of blood collection and processing on peripheral blood gene expression profiling in type 1 diabetes. *BMC Genom.* **2017**, *18*, 636. [CrossRef]

20. Weber, D.G.; Casjens, S.; Rozynek, P.; Lehnert, M.; Zilch-Schöneweis, S.; Bryk, O.; Taeger, D.; Gomolka, M.; Kreuzer, M.; Otten, H.; et al. Assessment of mRNA and microRNA Stabilization in Peripheral Human Blood for Multicenter Studies and Biobanks. *Biomark. Insights* **2010**, *5*, 95–102. [CrossRef]
21. Menke, A.; Rex-Haffner, M.; Klengel, T.; Binder, E.B.; Mehta, D. Peripheral blood gene expression: It all boils down to the RNA collection tubes. *BMC Res. Notes* **2012**, *5*, 1. [CrossRef]
22. Eikmans, M.; Rekers, N.V.; Anholts, J.D.H.; Heidt, S.; Claas, F.H.J. Blood cell mRNAs and microRNAs: Optimized protocols for extraction and preservation. *Blood* **2013**, *121*, e81–e89. [CrossRef] [PubMed]
23. Schrader, C.; Schielke, A.; Ellerbroek, L.; Johne, R. PCR inhibitors—occurrence, properties and removal. *J. Appl. Microbiol.* **2012**, *113*, 1014–1026. [CrossRef]
24. Al-Soud, W.A.; Rådström, P. Purification and characterization of PCR-inhibitory components in blood cells. *J. Clin. Microbiol.* **2001**, *39*, 485–493. [CrossRef]
25. Zhang, H.; Korenková, V.; Sjöback, R.; Švec, D.; Björkman, J.; Kruhøffer, M.; Verderio, P.; Pizzamiglio, S.; Ciniselli, C.M.; Wyrich, R.; et al. Biomarkers for Monitoring Pre-Analytical Quality Variation of mRNA in Blood Samples. *PLoS ONE* **2014**, *9*, e111644. [CrossRef]
26. PAXgene Blood RNA System. Technical Note [Internet]. 2018. Available online: https://www.preanalytix.com/storage/download/_ProductResources_/TechnicalNotes/PROM-7266-002_BD-7969_TN_Blood_RNA_System_RNA_stability_over_11_years_storage_1118_WW_WEB.pdf (accessed on 24 March 2021).
27. Applied Biosystems™ Tempus™ Blood RNA Tube [Internet]. Available online: https://www.thermofisher.com/order/catalog/product/4342792#/4342792 (accessed on 24 March 2021).
28. Mlcochova, H.; Hezova, R.; Stanik, M.; Slaby, O. Urine microRNAs as potential noninvasive biomarkers in urologic cancers. *Urol. Oncol. Semin. Orig. Investig.* **2014**, *32*, 41.e1–41.e9. [CrossRef]
29. Sheng, Q.; Vickers, K.; Zhao, S.; Wang, J.; Samuels, D.C.; Koues, O.; Shyr, Y.; Guo, Y. Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief. Funct. Genom.* **2017**, *16*, 194–204. [CrossRef]
30. Puchta, M.; Boczkowska, M.; Groszyk, J. Low RIN Value for RNA-Seq Library Construction from Long-Term Stored Seeds: A Case Study of Barley Seeds. *Genes* **2020**, *11*, 1190. [CrossRef]
31. Shen, Y.; Li, R.; Tian, F.; Chen, Z.; Lu, N.; Bai, Y.; Ge, Q.; Lu, Z. Impact of RNA integrity and blood sample storage conditions on the gene expression analysis. *Onco. Targets. Ther.* **2018**, *11*, 3573–3581. [CrossRef]
32. Yang, J.; Sui, X.; Wen, C.; Pan, C.; Zhu, Y.; Zhang, J.; Zhang, L. A hemocompatible cryoprotectant inspired by freezing-tolerant plants. *Colloids Surf. B Biointerfaces* **2019**, *176*, 106–114. [CrossRef]
33. Koh, E.J.; Yu, S.Y.; Kim, S.H.; Kim, S.J.; Lee, E.I.; Hwang, S.Y. Understanding Confounding Effects of Blood Handling Strategies on RNA Quality and Transcriptomic Alteration Using RNA Sequencing. *Biochip J.* **2021**, *15*, 187–194. [CrossRef]
34. Asare, A.L.; Kolchinsky, S.A.; Gao, Z.; Wang, R.; Raddassi, K.; Bourcier, K.; Seyfert-Margolis, V. Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation. *BMC Genom.* **2008**, *9*, 474. [CrossRef]
35. Häntzsch, M.; Tolios, A.; Beutner, F.; Nagel, D.; Thiery, J.; Teupser, D.; Holdt, L.M. Comparison of whole blood RNA preservation tubes and novel generation RNA extraction kits for analysis of mRNA and miRNA profiles. *PLoS ONE* **2014**, *9*, e113298. [CrossRef] [PubMed]
36. Matheson, L.A.; Duong, T.T.; Rosenberg, A.M.; Yeung, R.S.M. Assessment of sample collection and storage methods for multicenter immunologic research in children. *J. Immunol. Methods* **2008**, *339*, 82–89. [CrossRef]
37. Nikula, T.; Mykkänen, J.; Simell, O.; Lahesmaa, R. Genome-wide comparison of two RNA-stabilizing reagents for transcriptional profiling of peripheral blood. *Transl. Res.* **2013**, *161*, 181–188. [CrossRef]
38. Roy, D.; Tomo, S.; Modi, A.; Purohit, P.; Sharma, P. Optimising total RNA quality and quantity by phenol-chloroform extraction method from human visceral adipose tissue: A standardisation study. *MethodsX* **2020**, *7*, 101113. [CrossRef]
39. Carrillo-Ávila, J.A.; de la Puente, R.; Catalina, P.; Rejón, J.D.; Espín-Vallejo, L.; Valdivieso, V.; Aguilar-Quesada, R. Evaluation of RNA purification methods by using different blood stabilization tubes: Identification of key features for epidemiological studies. *BMC Res. Notes* **2020**, *13*, 77. [CrossRef]
40. Sidova, M.; Tomankova, S.; Abaffy, P.; Kubista, M.; Sindelka, R. Effects of post-mortem and physical degradation on RNA integrity and quality. *Biomol. Detect. Quantif.* **2015**, *5*, 3–9. [CrossRef]
41. Browne, D.J.; Kelly, A.M.; Brady, J.L.; Doolan, D.L. A high-throughput screening (HTS) RT-qPCR assay for quantifying surrogate markers of immunity from PBMCs. *Front. Immunol.* **2022**, *13*, 962220. [CrossRef]
42. Spandidos, A.; Wang, X.; Wang, H.; Seed, B. PrimerBank: A resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res.* **2009**, *38*, 792–799. [CrossRef]
43. Bustin, S.A.; Benes, V.; Garson, J.A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.W.; Shipley, G.L.; et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55*, 611–622. [CrossRef]