The Influence of Genetics on Gamma-Herpesvirus Infections

Neneh Sallah

University of Cambridge
Wellcome Trust Sanger Institute

This dissertation is submitted for the degree of Doctor of Philosophy

September 2016

Murray Edwards College

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the contributions section within each chapter and/or specified in the text. It is not being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution. It does not exceed the prescribed word limit for the Faculty of Biology.

Neneh Sallah

September 2016

Abstract

Gamma-herpesviruses are double stranded DNA lymphotropic viruses that include Epstein-Barr Virus (EBV) and Kaposi's sarcoma herpesvirus (KSHV). They establish life-long infections in the human host and have been associated with a variety of malignant tumours. Upon exposure to an infectious pathogen, both host and pathogen genetic differences influence variation in an individual's immune response including potential disease outcome. Although both viruses have been studied extensively, genetic and environmental influences on susceptibility to infection in individuals and potential disease outcome as a result remain unclear. While EBV is nearly ubiquitous globally, KSHV displays striking geographic variation with highest prevalence in sub-Saharan Africa, particularly in Uganda. Thus, this thesis investigates how host and virus genetics influence pathogenesis in EBV and KSHV infections, particularly the contribution of human genetic variation, using the Ugandan General Population Cohort (GPC) as a study population. The GPC provides a phenotype rich dataset and the availability of human genomic data in a large subset of individuals provides the opportunity to investigate the genetics of infection.

In chapter 2, I characterised the seroprevalence of oncogenic viral infections, assessed the influence of co-infection on EBV and KSHV serological traits in the GPC, and assessed the genetic population structure and heritability of Immunoglobulin G (IgG) antibody response traits.

In chapters 3 and 4, I explored the influence of host genetic variation on EBV and KSHV IgG antibody levels respectively, as a proxy for infection and potential disease risk. I performed the first genome-wide association analysis of anti-EBV IgG traits and anti-KSHV IgG traits in Africa, using a combined approach including array genotyping, whole-genome sequencing and imputation to a panel with African sequence data to extensively capture genetic variation and aid locus discovery. For EBV infection, I identified novel loci and through trans-ethnic meta-analysis with a cohort of European ancestry I uncovered

distinct variants contributing to variation in immune responses in Uganda. For KSHV infection multiple putative candidate loci were identified with modest effect sizes potentially contributing to infection.

As Uganda sustains such high levels of KSHV seroprevalence compared to the rest of the world, in chapter 5, I also explored the viral genetic diversity of KSHV in the GPC by whole genome sequencing of viral DNA isolated from saliva of asymptomatic individuals, and analysed the population structure comparing it to published KSHV genomes from around the world. This analysis showed a greater appreciation of variation of genes in the central region of the genome, some of which are under positive selection, contributing to the clustering of genomes by geography, thus, suggesting the use of whole-genomes in KSHV viral characterisation.

Together, the findings described in this thesis reinforce the importance of conducting genetic studies of infectious disease in African populations to uncover functionally relevant loci associated with traits of interest, and independent of environmental factors. Furthermore, the ability to obtain both host and viral genomes from the same individuals, allows for a comprehensive assessment of factors underlying the course of infection. The development of African resources to fully capture genetic diversity across the continent and building of research capacity will be fundamental to facilitate large scale studies and uncover meaningful biological insights.

Acknowledgements

I would like to thank my supervisors Dr. Inês Barroso and Prof. Paul Kellam immensely for their unwavering support and continued guidance throughout my PhD. My research was possible as result of extensive collaboration and thus, I would like to acknowledge: Dr. Rob Newton and his team at the MRC/UVRI in Uganda for setting up and granting me access to the GPC, Dr. Manj Sandhu and his team at the Sanger for curating the genetic data and Dr. Denise Whitby and her team at the FNLCR in the US for providing the serological data for analyses. I would also like to thank the GPC study participants for their generosity. I thank the members of my thesis committee, Dr. Stacey Efstathiou and Dr. Arthur Kaser for their useful discussions.

Many thanks to those who've shared their expertise with me: Dr. Carl Anderson and Dr. Chris Franklin for introducing me to the analysis of human genetic data, UNIX and R; Dr. Eleanor Wheeler for all the useful Stats guidance; Dr. Anne Palser and Dr. Simon Watson for their guidance on virus genomic analysis. Thanks to all the members of teams 146 and 35, in particular, Carol Dunbar for all her admin support and always finding a way, and Fernando for all the good humour in N-333. Thanks to the Wellcome Trust Sanger Institute for generously funding my studies and the graduate programme for their support.

On a more personal note, I thank my many friends at Sanger, especially, Pinky, Carmen, Mia, Eva, Sophia, Michal, Tomi, Martin and the rest of PhD12, for sharing the peaks and troughs of research and writing-up, the motivational and therapeutic conversations and making the past four years an enjoyable experience! Thanks to the members of the "Murrays Breakfast Table" for the morning banter, ensuring a good start to every day!

Finally, I would like to thank my family, especially my parents, Zahra and O.G for their encouragement, constantly believing in me and being my inspiration. Thanking Alima for her invaluable company in Cambridge and keeping my life exciting! Huge thanks to Gibril, Jamil and Alpha for being a source of comfort and reminding me that there's more to life than 'work'! Thanks to Louis, for his patience and embarking on this journey with me. Thanks to Uncle Halim for delivering me safely to all my destinations and reminding me to relax. Thanks to all my 'Kusineras' for keeping my life interesting and my second mothers, Sheriffa and Granny for their encouragement! Thanks to my MRC Gambia family, in particular, Pa Tamba Ngom, Bouke de Jong, Harr Njai (R.I.P.) and Martin Antonio for taking me under their wings as a young, aspiring scientist.

This PhD is dedicated to my mother and all the great, persevering women in my family who've been my motivation and supported me endlessly! I thank God for keeping me determined and helping me overcome my challenges.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Full Name
1000G	1000 Genomes
AGVP	Africa Genome Variation Project
BL	Burkitt's lymphoma
CAEBV	Chronic active EBV
EAD	Early antigen D
EAF	Effect allele frequency
EBNA	EBV nuclear antigen
EBV	Epstein-Barr Virus
GPC	General Population Cohort
GWAS	Genome-wide association study
HAART	Highly active anti-retroviral therapy
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HHV	Human Herpesvirus
HIV	Human Immunodeficiency Virus
HL	Hodgkin's lymphoma
HLA	Human leukocyte antigen
IBD	Identity-by-Descent
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IM	Infectious mononucleosis
KICS	Kaposi's Sarcoma inflammatory cytokine syndrome
KS	Kaposi's Sarcoma
KSHV	Kaposi's sarcoma-associated herpesvirus
LANA	Latency-associated nuclear antigen
LCL	Lymphoblastoid cell line
LD	Linkage disequilibrium
LMM	Linear mixed model
LMP	Latency-associated membrane protein
MAF	Minor allele frequency
MCD	Multicentric castleman's disease
MFI	Mean fluorescence intensity
MHC	Major histocompatibility complex
NPC	Nasopharyngeal Carcinoma
OD	Optical density

OR	Odds ratio
ORF	Open reading frame
PBMCs	Peripheral blood mononuclear cells
PCA	Principal components analysis
PCR	Polymerase chain reaction
PEL	Primary effusion lymphoma
SNP	Single nucleotide polymorphism
UG2G	Uganda 2000 genomes
UGWAS	Uganda GWAS
VCA	Viral capsid antigen
WGS	Whole-genome sequencing