An investigation of genetic factors in Ebola virus disease

Nathalie Emma MacDermott

Section of Paediatrics, Division of Infectious Diseases, Department of Medicine
Imperial College London

Copyright declaration

Statement of own work/Declaration of originality

The following thesis contains research that I have conducted over the last five years, which has been supervised by Prof Michael Levin at the Department of Medicine, Division of Infectious Diseases, Imperial College London and Dr Timothy Brooks, at the Rare and Imported Pathogen Laboratory, Public Health England, Porton Down. It is submitted for the degree of Doctor of Philosophy. Any sources, figures or data that were not generated in this project are stated in the text. The exome sequencing work described in chapter 5 was undertaken by the Oxford Wellcome Centre for Human Genomics, and the preliminary analysis was conducted by Dr Evangelos Bellos. The genotyping described in chapter 6 was undertaken by Illumina Fast-track Services, California, and the preliminary analysis was conducted by Dr Clive Hoggart. The data and work presented here are not yet published, but several manuscripts are in preparation.

Nathalie MacDermott

London, June 2021

Figure 0-1 'Makontha, a ghost community'.

The entire community of Makontha, Marampa Kingdom, Port Loko District (estimated population 50-60 people) died of Ebola virus disease during the 2014-2016 epidemic in Sierra Leone.

The community remains uninhabited.

'The only thing necessary for the triumph of evil is for good men to do nothing.'

John Stuart Mill/Rev Charles F Aked (paraphrase)

For the people of Sierra Leone, Liberia and Guinea

For all those whose lives have been altered by Ebola virus disease

Abstract

Introduction

The West African Ebola epidemic was the largest Ebola epidemic to date with over 28,000 cases. The large number of cases permitted assessment of different disease phenotypes and outcomes of Ebola virus disease (EVD). Given the variety of disease phenotypes in EVD, a genetic predisposition to disease phenotype and outcome was hypothesised.

Methods

Samples from 325 deceased patients and 174 surviving patients were provided through the Sierra Leone Ministry of Health-Public Health England Ebola Biobank. Additionally, 1021 household contacts, 1004 community controls and 504 Ebola survivors were recruited in Sierra Leone. Participants provided a saliva sample for DNA extraction and an oral fluid sample for anti-EBOV IgG antibodies. Exome sequencing was undertaken on 250 extreme phenotype cases and genome wide genotyping was undertaken on 2153 Ebola patients, household contacts and community controls. Data analysis of the exome data included within family segregation studies, gene burden testing and pathway analysis. The genotyped data was interrogated through a genome wide association study comparing deceased and surviving cases.

Results

Of the household contacts, 3.5% were positive for anti-EBOV IgG. Seropositivity correlated with risk exposure level, with the highest risk level demonstrating seropositivity rates of 15.6%. Ebola survivors with more severe acute disease demonstrated lower levels of anti-EBOV IgG antibodies (p=0.01), as did those with more severe post-Ebola syndrome, although this was not significant. Exome sequencing revealed multiple protective mutations within cholesterol metabolism pathways. A key finding was a protective variant in the PCSK9 gene (p=0.002). Preliminary GWAS analysis of deceased versus surviving Ebola patients identified a genome wide significant (p=2.9×10^{-8}) SNP in the Carbonic Anhydrase 5a gene.

Conclusions

The study established different extreme phenotypes of EVD, including highly exposed antibody negative and asymptomatic antibody positive individuals. Genetic factors affect both susceptibility to and severity of Ebola virus disease; with rare deleterious mutations in genes within cholesterol metabolism pathways, and common polymorphisms determining outcome of exposure to Ebola virus.

Acknowledgements

First and foremost, I would like to thank the study participants for having the courage and strength to participate in the study, as well as the Government of Sierra Leone for recognising the study's importance and granting it ethical approval. I would also like to thank all those who have made this research possible. This includes:

- My supervisors, Prof Mike Levin and Dr Tim Brooks, for their support, guidance and assistance throughout the project

- Dr Evangelos Bellos and Dr Clive Hoggart for their assistance with the analysis of the exome sequencing data and genome wide association data respectively.

- Sam Alberman for his assistance with a large amount of the laboratory work, both in Sierra Leone and the UK.

- The field team, Daniel Sesay, Osman Kamara, Alim Billo Bah, Santigie Dobson Kanu, Cecilia Touray and David Ama Collier, for their perserverance and long hours in the field recruiting participants.

- The University of Makeni for their support with the field work and granting me access to the Infectious Diseases Research laboratory at the University of Makeni. Included within this I would like to thank Dr Raoul Emeric Wadoum for his support with the laboratory work and his friendship, Prof Ian Goodfellow for developing the laboratory and permitting me to use it, and Mercy Mwaura for coordinating monetary transfer for the payment of the field team and purchasing of lab and fieldwork equipment.

- Prof Judith Glynn and Dr Hilary Bower at the London School of Hygiene and Tropical Medicine for their support and guidance. For permitting me to utilise their exposure risk assessment and modify their participant questionnaires, as well as assisting with the commercialisation of the anti-EBOV IgG capture assay.

- Prof Richard Tedder, Dr Dahn Samuel, Dr Steve Dicks and team at Public Health England for developing the oral fluid anti-EBOV IgG capture assay and granting me access to it. I would also like to thank Kalon Diagnostics Ltd for commercialising the assay and making it available within a rapid time frame.

- Dr Vasiliki Ioanna Mitravela for her perseverance in drawing pedigrees and entering data into the database.

- Dr Sharron Stubbs for her support and diligence in ordering all of the equipment and helping to prepare it for shipment to Sierra Leone.

Rafi the kitten in the community and approximately 2 weeks later at my house in Makeni. March/April 2017.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
£	British pound
%	Percentage
+ve	Positive
-ve	Negative
°C	Degrees Celsius
°F	Degrees Fahrenheit
Le	Leones (Sierra Leone currency)
ml	Millilitre
mph	Miles per hour
ng	Nanogram
ul	Microlitre
ABCA	ATP-binding cassette transporter A gene
AI	Antibody Index
AP	Adipocyte protein
APO	Apolipoprotein
ARSB	Arylsulfatase B gene
ATM	Automated teller machine/cash machine
BEND6	BEN Domain Containing 6 protein coding gene
BSc	Bachelor of Science degree
BSL	Biological safety level
BRC	British Research Council
BWA	Burrows-Wheeler Alignment
CA5A	Carbonic anhydrase 5a
CC	Community Control
CDC	Centers for Disease Control (United States)
CES	Carboxylesterase
CI	Confidence Interval
CLT	Clathrin light chain
CT	Cycle threshold
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of Congo
EBOV	Ebola virus
EBV	Epstein Barr virus
DfID	Department for International Development
EIA	Enzyme immunosorbent assay
ELISA	Enzyme linked immunosorbent assay
ELWA	Eternal Love Winning Africa (hospital in Monrovia)
ETC	Ebola treatment centre
ETU	Ebola treatment unit
EVD	Ebola virus disease
FC	Free cholesterol
GATK	Genome Analysis Toolkit
GOAL	Irish NGO

GP	Glycoprotein
GWAS	Genome wide association study
HC	Household contact
HIV	Human immune-deficiency virus
HLA	Human leucocyte antigen
HMG-CoA reductase	3-hydroxy-3-methyl-glutaryl-coenzyme A reductase
HRC	Haplotype reference consortium
ID	Identity
IgG	Immunoglobulin G
IGHI	Institute of Global Health Innovation
IMC	International Medical Corps
Inc	Incorporated
INSERM	Institut National de la Sante et de la Recherche Medicale
IQR	Inter-quartile range
ISSF	Institutional Strategic Support Fund
ITMAT	Institute for Translational Medicine and Therapeutics
IV	Intra-venous
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LDLRAP	Low density lipoprotein receptor-related protein-associated protein
LIPA	Lipase A
LIPE	Lipase E
LIPM	Lipase M
LPIN	Lipin protein gene
LRRC20	Leucine Rich Repeat Containing 20 protein coding gene
LSHTM	London School of Hygiene and Tropical Medicine
LSR	Lipolysis-stimulated lipoprotein receptor
Ltd	Limited
MOHSL	Ministry of Health Sierra Leone
MSF	Medecins sans Frontieres
MX	Interferon induced GTP binding protein gene
NCEH	Neutral cholesterol ester hydrolase
NGO	Non-governmental organisation
nm	Nanometre
NP	Ebola virus nucleoprotein
NPC	Niemann Pick C
OD	Optical density
OG-500/575	Oragene 500/575 saliva collection device
OR	Odds ratio
PC	Principle component
PCA	Principle component analysis
PCR	Polymerase chain reaction
PCSK	Proprotein convertase subtilisin-kexin
PD	Patient died
PES	Post Ebola Syndrome
PHE	Public Health England
PPE	Personal Protective Equipment
PS	Patient survived
QC	Quality control
QT	Quart (United states volume measurement)
RAF	Royal Airforce

RNA	Ribonucleic acid
ROC	Receiver operating characteristic curve
RTN4	Reticulon 4 protein coding gene
RT-PCR	Reverse transcription polymerase chain reaction
SLC25	Family of nuclear encoded transporters
SLESRC	Sierra Leone Ethics and Scientific Review Committee
SNP	Single nucleotide polymorphism
SOAT	Sterol-O-acyltransferase
SQLE	Squalene epoxidase
SS	Survivor syndrome
STARD	StAR related lipid transfer domain
TLR	Toll like receptor
TMB	3,3',5,5'-tetramethylbenzidine
UK	United Kingdom
UNICEF	United Nations Children's Fund
VEP	Variant Effect Predictor
VHF	Viral haemorrhagic fever
VLDL	Very low density lipoprotein
VP24	Ebola virus membrane associated protein
VP30	Ebola virus hexameric zinc-finger protein
VP35	Ebola virus polymerase co-factor
VP40	Ebola virus matrix protein

Summary of the thesis

This thesis describes a program of work I undertook over five years to address the question of whether genetic factors play a role in both susceptibility and outcome of EVD. The scientific question of the contribution of genetics to EVD, and my decision to focus a PhD on this disease arose from my personal experience of the Ebola epidemic in Liberia, clinical observations of differences in disease phenotype and the evolving literature demonstrating the role of a host's genetic makeup in their response to infections.

Chapter 1 explains the background to this thesis, including a brief introduction to Ebola virus disease, a review of the evidence for genetic susceptibility to infectious diseases as well as a literature review of genetic susceptibility to viral infections and Ebola virus disease. It also describes the differences in disease phenotype observed during the clinical management of patients with EVD in Liberia and provides the background to the situation in West Africa at the time the study took place.

Chapter 2 provides an overview of the methods and study design, including reasons for the study design used and how the participant numbers required for the study were identified. It describes obtaining access to samples from the Ministry of Health of Sierra Leone and Public Health England biobank and accessing clinical data from Ebola treatment facilities. It also discusses the initial planning phase to recruit participants to the study from communities in Sierra Leone and provides a brief overview of the laboratory work undertaken and techniques used.

Chapter 3 explains in detail the participant categories identified for recruitment and the methods used to recruit the target participant numbers from communities in Sierra Leone. It then describes the data collected and provides a descriptive analysis of the participants recruited to each participant category. This includes clinical data for the biobank samples and the types of persisting symptoms experienced by Ebola survivors.

Chapter 4 discusses the methods used to identify which study participants demonstrated antibodies to Ebola virus. It then analyses the community seropositivity data, comparing differences between communities and adults and children. Lastly it compares the rates of seropositivity in relation to the level of exposure to Ebola described by the participants. The data from this chapter identifies the Ebola

virus exposure phenotypes of 'resistant to disease' and 'resistant to infection', which later inform part of the genetic analysis.

Chapter 5 provides a brief description of the methods used to determine Ebola antibody levels in Ebola survivors and explores the relationship between Ebola antibody level and the prevalence of Post-Ebola Syndrome (PES) among recruited Ebola survivors. It highlights potential associations between antibody level, age and the severity of PES, as well as exploring these associations in relation to different types of acute disease and different symptoms of PES.

Chapter 6 explains the methodology behind the exome sequencing, definition of extreme phenotypes and samples selected for exome sequencing. It presents the key findings of the preliminary exome sequencing analysis, along with a brief discussion of these findings, their meaning and potential impact.

Chapter 7 discusses the methodology behind the genome wide association study and the samples selected for the GWAS. It describes the key finding of the preliminary SNP analysis, including a brief discussion of the relevance of this finding, its meaning and potential impact.

Chapter 8 summarises the findings of the study and discusses the impact these findings may have in relation to our understanding of EVD and infectious disease pathogenesis. It also details the work that will be taken forward to confirm the results and expound them further. Lastly it discusses the limitations of the study and how the study findings will be communicated to both the scientific community and the study participants.

Chapter 1 Introduction

1.1 Ebola virus disease: an overview

Ebola virus disease (EVD) is a viral infection caused by the filovirus *Ebolavirus[1, 2]*. It is known to cause a severe, potentially haemorrhagic, illness if it infects humans, with mortality rates varying from 25-90%[3-7]. There are six species of Ebola virus, four of which are known to be pathogenic in humans. The species associated with the most known outbreaks in humans is *Zaire ebolavirus*, but there have also been several outbreaks of *Sudan ebolavirus*, and two significant outbreaks of *Bundibugyo ebolavirus[8-12]*. There has only been one known human case of *Tai Forest ebolavirus* and *Reston ebolavirus* is not known to cause disease in humans, only pigs. A new species of Ebola virus was recently discovered in Sierra Leone and has been named *Bombali ebolavirus[13]*.

Fruit bats are believed to be the natural reservoir of Ebola virus, but do not show signs of disease. The intermediate hosts, such as non-human primates and forest antelope, who likely contract the disease from coming into contact with bats or the fruit they have fed off, do develop a severe haemorrhagic illness and invariably succumb to the disease[1, 3]. It is this intermediate host that is the most common source of viral entry into a human host, usually through hunting and preparation of bush meat, although direct contact between bats and humans is also a known mechanism[14, 15] . The viral entry to a human host through either the reservoir or the intermediate host is often a sporadic event, however once the virus enters a human host, it then spreads readily from person to person through contact with sick or deceased individuals or their bodily fluids[15-17]. It has been suggested that recombination of viruses is possible when there is mixing of different bat populations amid recrudescent disease or during periods of increased viral shedding[14, 15]. This could well result in more species of Ebola appearing in the future and an increasing incidence of sporadic events that result in human infection, with resulting outbreaks or epidemics. Recently a new species of Ebola virus was identified in bats in Sierra Leone, named *Bombali ebolavirus* after the district in which it was identified[13]. It is unclear whether Ebola Bombali infects humans and causes symptomatic disease or disease outbreaks, but studies are commencing to determine if there is serological evidence of this in populations in the region[13].

Ebola virus disease is classified as a viral haemorrhagic fever due to the virus' ability to cause endothelial damage and a coagulopathy which may manifest with haemorrhage from various bodily orifices. Despite this, only a modest proportion of patients may demonstrate haemorrhagic signs, with reported rates of approximately 10-30%[6, 7, 18-22]. The Ebola virus is classified as a class 4 pathogen due to

the high risk of contracting the disease when handling human or animal specimens[16, 17] and the high rates of mortality associated with infection.

1.1.1 Epidemiology of Ebola outbreaks

The Ebola virus was first identified in 1976 during an outbreak of haemorrhagic fever in the Democratic Republic of Congo (DRC), this was caused by what is now idenitfed as *Zaire ebolavirus* (EBOV). Since then three Ebola virus species (Zaire, Sudan and Bundibugyo) have been responsible for sporadic outbreaks of haemorrhagic fever in a handful of countries in central Africa, with the majority of epidemics in DRC and Uganda[23]. Prior to 2014 there had been no known outbreaks of Ebola in West Africa, and the largest Ebola epidemic had not exceeded 450 cases. Figure 1-1 demonstrates the Ebola epidemics that have occurred from 1976 up to 2018.
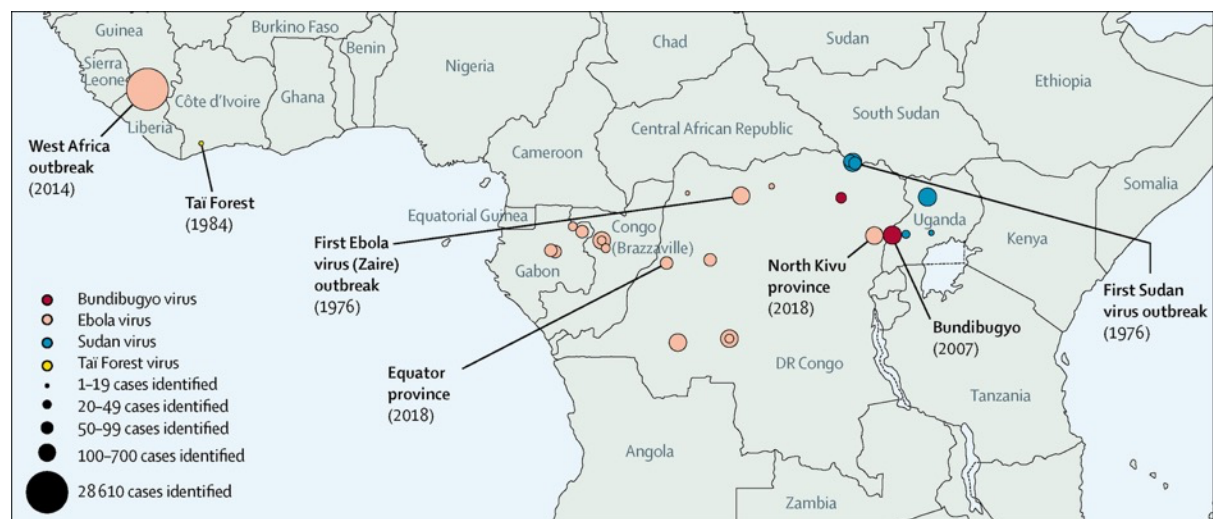
Figure 1-1 Outbreaks of Ebola virus on the African continent from 1976 to 2018.[23]

1.2 The West African Ebola epidemic

The Ebola virus epidemic in West Africa from 2013-2016 was the largest Ebola epidemic in history resulting in over 28,000 cases of the disease and over 11,000 deaths[24]. It predominantly affected the nations of Guinea, Liberia and Sierra Leone, but also spread to a number of other nations, including autochthonous transmission in Nigeria, Mali and the United States. This epidemic was caused by *Zaire ebolavirus*, specifically the Makona strain of *Zaire ebolavirus*[23].

1.2.1 Cultural context

The three most affected nations of the West African Ebola epidemic are all low income nations, ranking near the bottom of the UN Human Development Index[25]. Both Liberia and Sierra Leone were also

recovering from horrific civil wars, which ended less than twelve years before the epidemic began[26, 27]. Both wars affected large regions of each country and their capital cities, resulting in destruction of much of the functioning infrastructure. Subsequently both Liberia and Sierra Leone have an extremely limited health infrastructure and ability to respond to outbreak situations. Similarly, their populations, traumatised by over ten years of civil war and familiar with the corruption that pervades both political systems, are very untrusting of their governments. These factors compounded the epidemic situation in both countries, with numerous rumours of malintent by governments and aid workers. These rumours resulted in an unwillingness of people to seek testing or treatment at treatment facilities, and at times outbursts of violence towards health care workers and community engagement workers. All of this compounded the response to the epidemic, but also added to the complexities of conducting research in such a setting. Entering such communities to request their cooperation in a study on Ebola was not straightforward and required significant planning to ensure the safety of the field team and success of the recruitment.

There are multiple traditional cultural influences that contributed to the spread of the epidemic in Sierra Leone and its neighbours, Liberia and Guinea. One of these was the cultural importance of funeral rites, and a belief that the send off of a deceased loved one through their funeral was essential to the departure of their soul. Funeral practices included washing the deceased's body, using the water that was used to wash the body to wash oneself or drink, and also touching or kissing the body as a sign of respect to the person who had died. All of these features enhanced the spread of Ebola[15, 28-31], as has been similarly experienced in previous, albeit smaller and more contained, Ebola outbreaks in Central Africa[32-34]. Another important component in the spread of disease was the development of more robust roads, the mobility of the population and their ability to travel long distances, including to capital cities, with comparative ease when compared to previously. The cultural context in Sierra Leone is discussed further below as it played an important role in undertaking the fieldwork for this study.

1.2.2 Traditional hierarchy and ethnicities in Sierra Leone

Sierra Leone is a nation that hosts multiple different ethnicities. While the predominant ethnicities are Temne and Mende, each contributing to approximately one third of the population, the remaining third is contributed to by thirteen other ethnicities[35]. Different ethnicities predominate in different areas of Sierra Leone (Figure 1-2), however there is now a significant amount of inter-mixing of ethnicities in different communities, particularly in urban areas due to rural to urban migration[36]. The rural areas in which this study took place were predominantly Temne or Limba communities, but minority

ethnicities within communities sometimes adopt the Temne ethnicity as their own, even though this may not be their true genetic ethnicity.

Figure 1-2 Distribution of major ethnicities in Sierra Leone (image in public domain).

The traditional leadership hierarchy within communities varies depending on the dominant ethnicity in the region, however the structure is usually that of a Paramount Chief who has responsibility for a large region which hosts multiple sections. This is usually an inherited position but can sometimes be an elected position by the community members. Under the Paramount Chief are multiple Section Chiefs who are responsible for a number of communities (approximately 5-10) in a smaller region and who answer to the Paramount Chief. These positions may be inherited or appointed by the Paramount Chief. Under the Section Chief is the Local Chief who is the chief to one specific community and answers to the Section Chief. This position may be inherited or elected depending on the community context. This hierarchy predominantly applies to rural areas, while urban areas vary in their structure, there is always a headman or headwoman over each community.

1.2.3 Clinical features of Ebola virus disease

Ebola virus disease (EVD) is the name attributed to the clinical syndrome caused by Ebola virus. The early features of EVD tend to be common to most infectious diseases and subsequently are often missed outside the context of an epidemic or by healthcare workers naïve to the risk in their settings[6, 18-21, 37-42]. The early and late features of the disease are shown in Table 1-1 below and the case definition used during the epidemic in West Africa is shown in Figure 1-3.

Early symptoms/signs	Late symptoms/signs
Fever	Shock
Headache	Comatose/unresponsive
Myalgia/arthralgia	Seizures
Weakness/fatigue	Bleeding
Anorexia	From IV line sites
Abdominal pain	From nose/mouth/ears
Nausea and vomiting	Melaena
Diarrhoea	Haematemesis
Hiccoughs	Haematuria
Sore throat	Metrorrhagia/Menorrhagia
Cough or respiratory symptoms	Sub-conjunctival haemorrhage
Conjunctivitis	Bleeding from eyes
	Jaundice and deranged liver enzymes
	Reduced urine output/anuria
	Renal impairment
	Oedema

Table 1-1 Early and late symptoms and signs associated with EVD[6, 43-45].

Suspect case:

1. History of contact with a known case and one of:
   a. A fever of >37.5 ∘C/99.5 ∘F
   b. Three or more symptoms
   c. Unexplained bleeding/miscarriage

   OR

2. Any patient presenting with a fever and 3 symptoms during the epidemic/from an endemic area

Probable case:

1. Any suspected case evaluated by a clinician and deemed to be highly likely of having EVD infection

   OR

2. A deceased suspect case (where it was not possible to collect specimens for laboratory confirmation) that has an epidemiological link with a confirmed case.

Confirmed case:

1. Laboratory confirmed case on the basis of positive detection of viral RNA by real time, reverse transcription-polymerase chain reaction (RT-PCR).

Figure 1-3 The Ebola case definition used during the epidemic in West Africa[5-7, 22, 43].

As mentioned earlier, EVD has a very high mortality varying from 25-90% depending on the virus species and the context of the outbreak[2, 5, 6, 45]. *Zaire ebolavirus* has the highest mortality rates varying from 40-90%[5, 6, 18, 40, 45]. Mortality rates vary with age, with extremes of age having higher mortality rates (age 0-5 years and greater than 60 years)[6, 7, 22, 46]. Teenagers appear to have the lowest mortality rates at approximately 40-45%, with all other age groups having relatively similar mortality rates[22]. Following on from the West African epidemic there are now several factors known to predispose to worse outcome in the context of EVD. These include time from symptom onset to presentation at a treatment facility, viral load and organ impairment[5, 18, 21, 37, 39, 40]. While haemorrhagic disease is reported in only 10-30% of patients, it is also suggestive of a poor outcome[6, 18, 21, 37]. It is likely co-morbidities contribute to worse outcome, but this data was not well reported during the epidemic and so this remains unclear. However there is conflicting information in relation

to malaria co-infection, with one large study documenting malaria as protective in the context of Ebola[47] (although it excluded children under five years of age who are typically more likely to have malaria) and another study documenting malaria co-infection as detrimental[48, 49]. Similarly, early, good supportive care with adequate replacement of fluid losses and management of electrolyte disturbances is known to improve outcome[39-41].

Prior to the West African epidemic post-Ebola sequelae in Ebola survivors were not well understood, although there were reports of eye problems such as uveitis in some survivors and headaches and joint pain in others[12, 50, 51]. There was also the suggestion that male Ebola survivors might continue to secrete Ebola virus in semen for some time after surviving, although the longest determined period of time was 91 days[16, 51]. Due to the scale of the West African epidemic there are now thousands of Ebola survivors who can be studied. Subsequently it has been determined that a condition exists which has been named Post-Ebola Syndrome (PES) and is present in a significant proportion of Ebola survivors. This syndrome appears to include eye disease, largely uveitis (sometimes associated with cataract formation or blindness); joint problems; neurological problems and possibly fertility problems[52-59]. A paper published during the writing of this thesis, now sheds some light on the number of Ebola survivors experiencing symptoms consistent with PES and what these symptoms constitute[60]. The symptoms described included urinary frequency, headaches, muscle pain, memory loss and uveitis, with prevalence of symptoms ranging from 14.7% to 47.6%[60]. It has also been determined that male Ebola survivors have persistence of virus in seminal fluids and that in some contexts this is live virus that can transmit infection to previously uninfected individuals[17, 61-66]. While the persistence of virus in semen appears common within the first few months from symptom onset, this becomes less frequent with time since symptom onset. In a small number of survivors, however, virus persistence has been demonstrated even at 18 months after symptom onset[67].

1.2.3.1 Clinical experience which led to the thesis topic

Through direct personal experience of working in an Ebola treatment facility based at the ELWA mission hospital, just outside Monrovia, Liberia in July of 2014, it was noticed that there were different ways in which patients appeared to succumb to EVD. These 'modes of death' seemed to fall into three specific groups, the data of which has since been published in the American Journal of Tropical Medicine and Hygiene[18] (Figure 1-4). It was also observed that there were patients with clear histories of exposure, some with positive test results, who never appeared to become unwell. This led to the question of why different patients had different clinical syndromes in response to exposure to the same strain of virus.
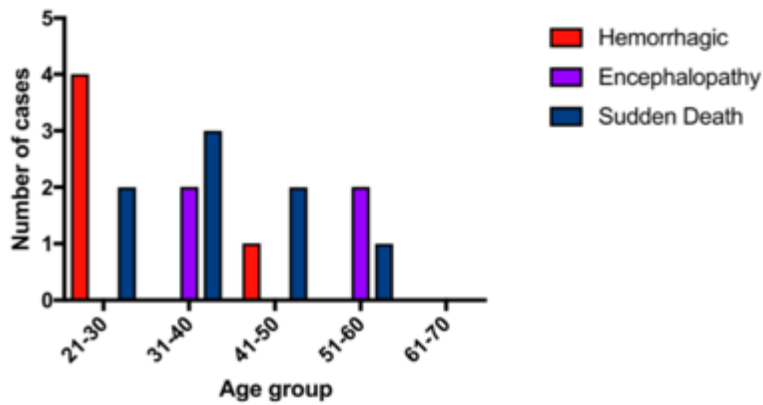
FIGURE 4. Relationship of age to mode of death in Ebola virus disease. This figure appears in color at www.ajtmh.org.
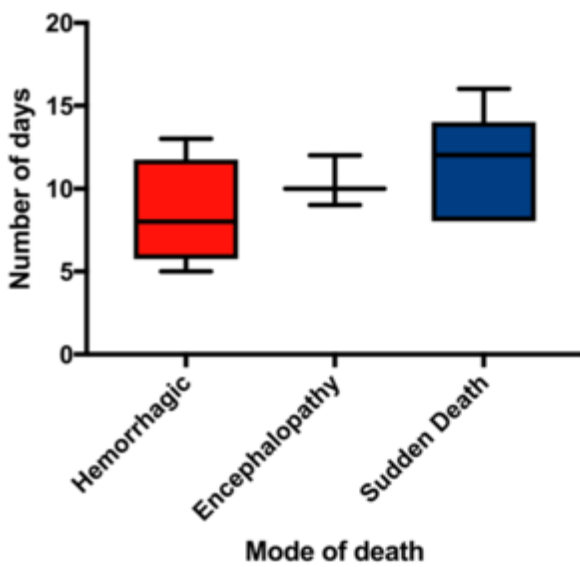
FIGURE 2. Time to death from symptom onset compared with mode of death in Ebola virus disease. This figure appears in color at www.ajtmh.org.
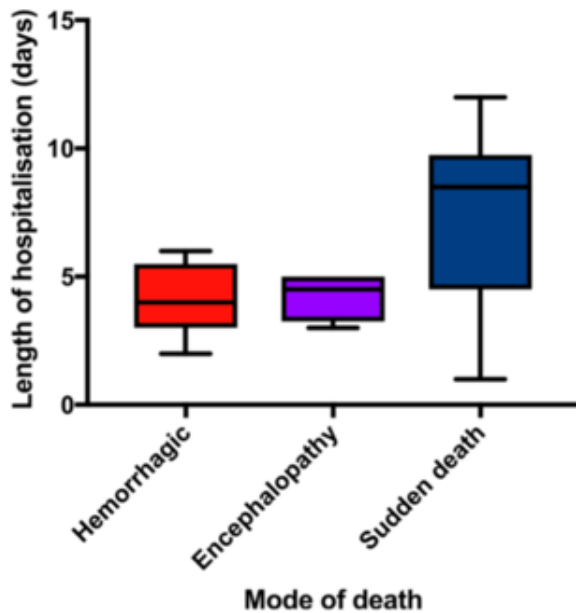
FIGURE 3. Duration of hospitalization by mode of death. This figure appears in color at www.ajtmh.org.

Figure 1-4 Graphs showing variation in mode of death by age, time to death and length of hospitalisation based on clinical data from the ELWA2 Ebola treatment facility, Monrovia, Liberia, June-August 2014. As published in the American Journal of Tropical Medicine (Mobula et al, 2018)[18].

As the epidemic progressed and larger numbers of people survived the disease, it also became apparent that Ebola survivors had different rates of recovery and development of post-Ebola syndrome[53, 55, 60]. This included those who survived with no ongoing problems and those who developed post-Ebola syndrome, which varied from mild, intermittent joint pain through to the development of neurological sequelae or severe inflammatory eye disease resulting in cataract formation or blindness[53, 54, 56, 58]. A group of survivors also demonstrated virus persistence in various bodily fluids, most notably semen, intra-ocular fluid and cerebrospinal fluid[56, 58, 61]. In the case of male survivors, the presence of virus in semen appeared to be intermittently expressed but could persist for as long as 18 months if not longer[67]. In several cases there was transmission of Ebola from a surviving male to a previously unaffected woman[62, 66, 67]. This further emphasised the possibility of genetic determinants that impact on the body's ability to contain and clear the virus.

1.2.4 Pathogenesis

The Ebola virus (EBOV) has a lipid bilayer that is embedded with glycoprotein (Figure 1-5). It uses its envelope glycoprotein to bind to the host cell surface and fuse with the host cell membrane. It also has a secretory glycoprotein, which it uses as an immune evasion mechanism by shedding it. This secreted glycoprotein mops up host neutralising antibody and also binds to neutrophils preventing their

activation[68-74]. Further components that constitute Ebola virus virions and their function are shown in Table 1-2.

Ebola virus component	Function
Lipid bilayer	Virus coating, fuses with host cell membrane
Glycoprotein (GP)	Envelope GP - Binding to host cell surface receptor, host cell membrane fusion Secretory GP – mops up host neutralising antibody
VP40 matrix protein	Important in virus particle assembly and budding
VP24 membrane associated protein	Blocks interferon alpha, beta and gamma signalling pathways preventing destruction by host cells
VP35 polymerase co-factor	Essential role in viral RNA transcription Suppresses host dendritic cell maturation Inhibits induction of interferon alpha & beta
NP - nucleoprotein	Encapsidates the viral genome protecting it from nucleases and host innate immune response
VP30 Hexameric zinc-finger protein	Viral transcriptional activator
Single stranded negative sense RNA	Essential for viral replication

Table 1-2 Components of the Ebola virus virion[71, 75-78].

Figure 1-5 Ebola virus structure showing glycoprotein coated lipid bi-layer, viral proteins and single stranded, negative sense RNA. Used with permission from Emory University Department of Visual Medical Education[79].

EBOV utilises several tactics to evade the immune system in addition to the role of its envelope and shed glycoprotein described above. These include infection of monocytes, dendritic cells and phagocytes, which disrupts antigen trafficking and cytokine production; reduced activation but enhanced stimulation of T cells, ultimately triggering early T cell death; and increased production of pro-inflammatory cytokines with reduced production of anti-inflammatory cytokines[1, 80, 81]. EBOV also infects and depletes natural killer cells[81]. This impairment of the immune response ultimately results in unchecked viral replication resulting in an increasing viral load. This results in organ failure through the increasing inflammatory response and the direct cytopathic effect of the virus, which has a particular affinity for hepatocytes, renal cells and endothelial cells[3, 8, 39, 81]. If the host immune system is unable to gain the upper hand, ultimately multi-organ failure and death ensues.

Over the last decade there has been increasing understanding of the role of host lipids in sustaining the lifecycle of many viruses[82]. This can commence from the point of entry to the cell where viruses may fuse with the cell's lipid bilayer membrane, or bind to lipids and lipid receptors at the cell membrane and utilise them to gain entry to the cell, such as Hepatitis C virus utilising the LDL receptor as part of its process of gaining entry to host cells[82-84]. HIV-1 is also known to require lipid as a co-factor for viral entry[82, 83]. Viruses may also use host lipids within cells for the purposes of lipid membrane formation around newly formed viral nucleic acid or for remodelling of such membranes; providing energy for the purposes of replication; and for virion assembly and budding[82, 83]. Host lipids also appear to have a role in influencing the infectivity of newly released virions[82, 83]. As a lipid envelope virus, EBOV is known to fuse with host cell lipid bilayer membranes and it has also been implicated in utilising the NPC1 receptor (a cholesterol transport receptor) to exit the endosome into the cell cytoplasm where it can replicate[85-87]. Filoviruses have also been identified, among others, as being susceptible to modifications in cholesterol metabolism which may restrict their replication[82, 83].

1.2.4.1 The NPC1 receptor

There have been suggestions based on other viral diseases that some RNA viruses utilise lipid or cholesterol metabolism pathways to invade host cells and enter the cytoplasm where they can replicate[82, 83]. One of the receptors known to be utilised by Hepatitis C and HIV is the Nieman Pick C1 receptor[84, 88]. Based on these findings Carette et al (2011) investigated whether Ebola virus utilises the Nieman Pick C1 (NPC1) receptor to exit the endosome into the cell cytoplasm[85]. The NPC1 receptor is a cholesterol transport receptor. Absence of this receptor results in the lysosomal storage disease Nieman Pick type C, which is invariably fatal in late childhood/early adulthood[89].

Several initial cellular based studies identified that EBOV utilises the NPC1 receptor to enter the cell cytoplasm[85, 86, 90]. A further study using mouse models sought to demonstrate a difference in clinical phenotype when comparing wild type, receptor null and heterozygote NPC1 mice[87]. The receptor null mice did not become infected with EBOV, the heterozygotes became infected but had a 90% survival rate and the wild type mice became infected, experienced a more severe clinical disease and had mortality rates of 90%[87]. This is also supported by two studies that identified a mutation in the EBOV glycoprotein at the NPC1 binding site during the West African epidemic, which caused it to increase its infectivity in humans[69, 73]. A further study assessing the binding location of EBOV to the NPC1 receptor identified that EBOV binds to the C domain of the receptor whereas cholesterol is thought to bind at a transmembrane domain (Figure 1-6)[77]. Therefore, EBOV binding is independent of the NPC1 receptor's cholesterol transport function.
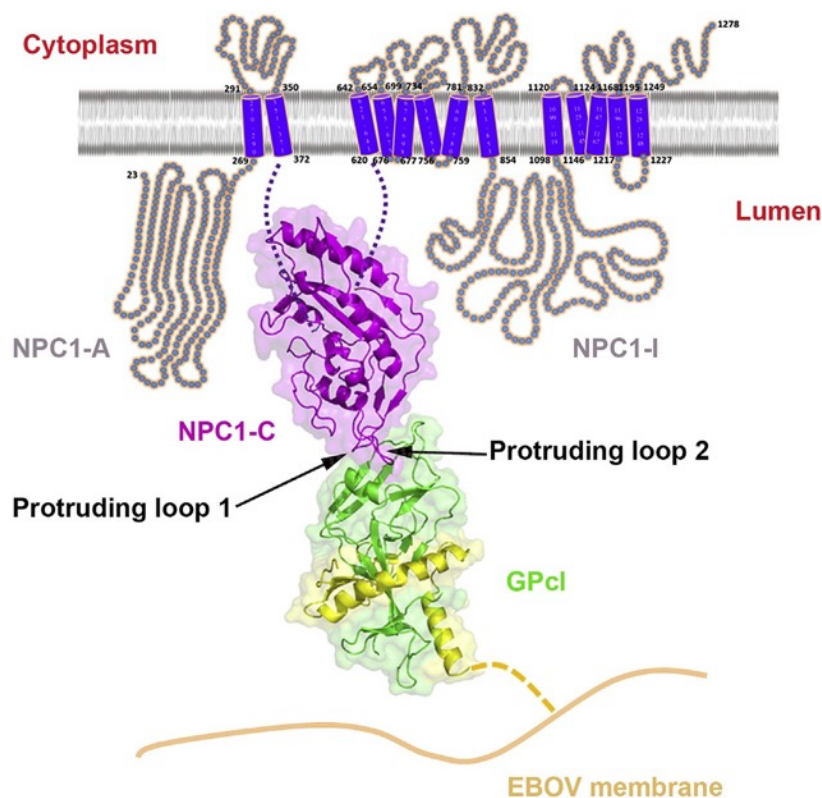
Figure 1-6 EBOV GP binding to the C luminal domain of the NPC-1 receptor. There are additional A and I luminal domains, as well as transmembrane and N-terminal domains which Ebola does not bind to. Mutation of domain C resulted in either failure of EBOV to bind to the receptor or reduced binding of EBOV to the receptor depending on the type of mutation. Taken from Wang et al (2016)[77].

1.3 Evidence of a genetic contribution to infectious diseases

Over the last two decades there has been increasing understanding of the importance of genetics in both susceptibility to infection with specific pathogens (eg. HIV) and susceptibility to a variety of different disease states and poor outcomes from certain pathogens (eg. HIV, Hepatitis C, TB)[91]. While initial suggestion of a host genetic role in the outcome from infectious diseases stemmed from the identification of genetic variants with a causal relationship to primary immune deficiencies, this progressed to studies of siblings/identical twins and identification of specific monogenic traits contributing to an increased risk from certain infections[92]. Over the last 20 years the study of genetic susceptibility to specific infections has progressed greatly to encompass large scale genome wide association studies and whole genome/exome sequencing. These studies have demonstrated the significant role of host genetic variants in infection with, and outcome from, infectious diseases; and importantly have contributed greatly to the understanding of the pathogenesis of different infectious agents[91, 92].

1.3.1 Approaches to identifying the genetic basis of disease

Genetic contributions to disease can be established by a range of approaches. Early approaches used analysis of multi-case families through linkage studies to identify the chromosomal location of candidate genes[93]. More recently assessment of exomes or whole genomes has become possible as a result of the technological advances in genome sequencing[94]. While these approaches are powerful methods to identify rare Mendelian variants, they are not applicable to identifying the role of common polymorphisms[95, 96]. These polymorphisms are better identified through case control studies using either candidate gene approaches or genome wide association studies. Figure 1-7 below depicts the differences between whole exome sequencing and GWAS and which analysis is more appropriate to identify the type of mutation being sought[95, 96].

Whole exome sequencing vs Genome Wide Association Studies
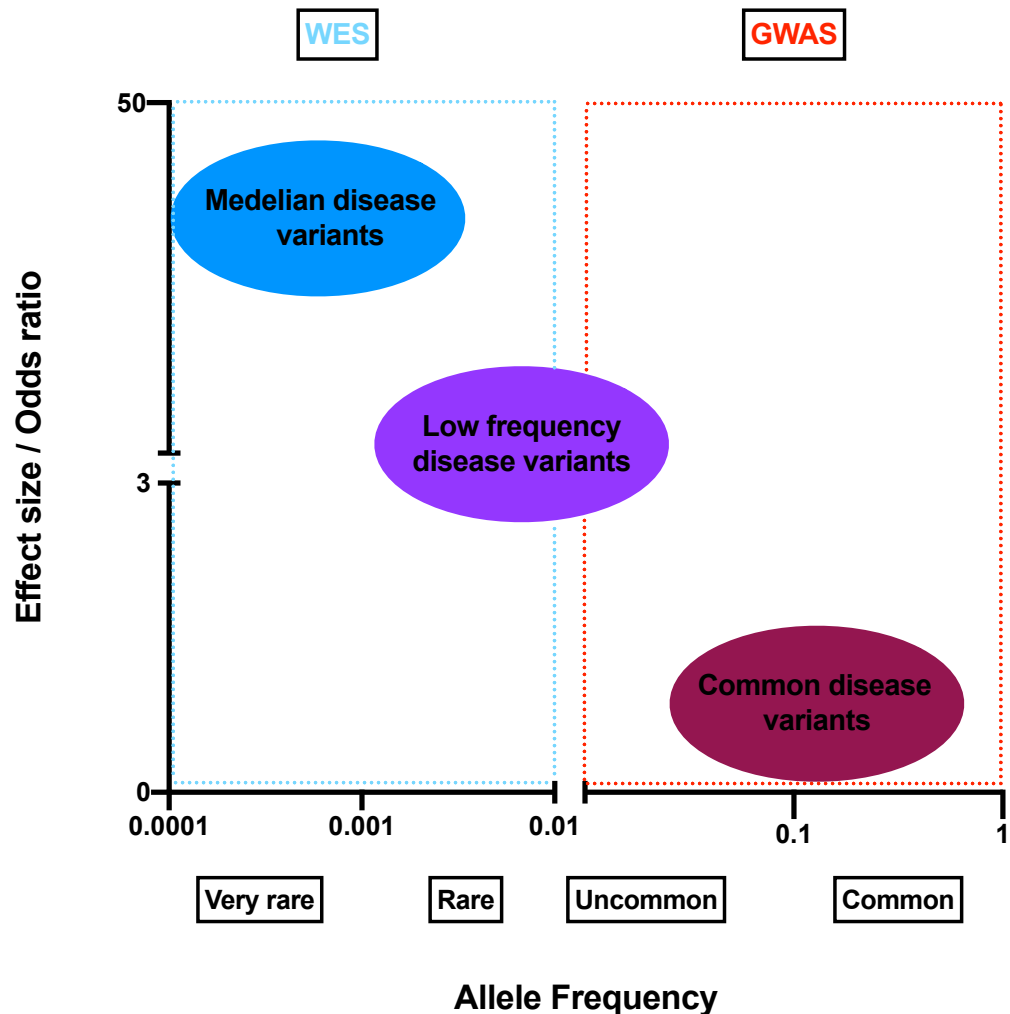
Figure 1-7 A comparison of whole exome sequencing and genome wide association studies. WES = Whole exome sequencing, GWAS = Genome wide association study. Modified from Manolio et al 2009[96].

There are advantages and disadvantages to both GWAS and whole exome sequencing. GWAS are more likely to identify common mutations affecting susceptibility or disease severity but require large groups for comparison to achieve statistical significance, and caution to ensure the finding is not an artefact of imbalance between the cases and control population[96, 97]. Exome sequencing does not require such large groups for comparison, but is more likely to identify rare mendelian mutations, which may only be applicable to a very small subset of the population[95, 96]. While utilising a combination of the two facilitates maximum use of genetic data, differences in study design are required to ensure optimal, well phenotyped study cohorts that are suitable for analysis through GWAS and exome sequencing.

1.3.2 Genome wide association studies and exome sequencing in infectious diseases

As genetic sequencing and genotyping technologies have become increasingly accessible over the last decade there has been a move towards greater understanding of the host response to severe infection, and why some individuals may suffer a more severe disease phenotype and outcome when exposed to the same pathogen compared with others who only experience a mild disease phenotype or complete protection[93, 98, 99]. Genetic factors play an important role in response to infection[93], with genome wide association (GWAS), and study of familial cases having identified genetic variants underlying susceptibility, severity and outcome of many infections. Early studies on Human Immune-deficiency Virus (HIV), viral hepatitis and meningococcal disease revealed that specific genetic mutations may increase susceptibility to a more severe disease phenotype or may provide some degree of protection from a severe phenotype[100-112]. A summary of some of the key gene variants identified in increasing susceptibility to specific infections or providing protection to certain infections are provided in Table 1-3 and Table 1-4 below. Some studies have also indicated host genetic mutations that are predictive of response to specific therapies in certain infectious diseases, such as HIV and Hepatitis B and C, which have enabled patient targeted treatments with resulting improved outcomes[101-103, 105, 106].

Gene	Infectious disease susceptibility	Impact
Properdin Complement Factor H Complement Factor D	Neisseria infection	Failure of the membrane attack complex in pathogen killing
APOL1	Trypanosomiasis	Reduced trypanosome lysis
IFN-gamma receptor	Mycobacterial disease	Inadequate interferon gamma response to mycobacteria causing reduced pathogen killing
IL-17A/IL-17RA/IL-17RC IL-17F	Chronic mucocutaneous candidiasis	Reduced pathogen killing
CARD9 deficiency	Invasive dermatophytic disease	Unclear
SAP XIAP	X-linked lymphoproliferation in EBV	Failure of CD8 cytotoxic T-cells to kill EBV infected B-cells
IRAK4 MYD88	Invasive pneumococcal disease	Abnormal functioning of TLR and IL-1R pathways causing impaired killing of pneumococci.
TLR3 TRIF	Herpes simplex encephalitis	Impairment of TLR-3 control of IFN alpha and beta production in cortical

UNC93B		neurons/oligodendrocytes causing impaired viral clearance/killing
IRF7	Severe influenza	Impairment of IRF7- dependent amplification of antiviral interferons causing impaired pathogen killing
MDA5	Severe rhinovirus infection	Impaired recognition of rhinovirus and reduced interferon induction causing reduced pathogen killing
RNA polymerase III	Severe varicella-zoster infection	Poor interferon induction and reduced interferon production resulting in reduced pathogen killing
TIRAP	Invasive staphylococcal disease	Impaired signalling for pathogen killing

Table 1-3 A summary of gene variants identified in increased susceptibility or severity to specific infectious diseases[93, 98]. Those listed in red highlight the single stranded RNA viruses.

Gene	Infectious disease protectivity	Impact
DARC	Plasmodium vivax malaria	Resistance to erythrocyte infection with P Vivax
CCR5	HIV-1	Protection against infection with CCR5-tropic HIV-1
FUT2	Norovirus Rotavirus	Resistance to infection
CASP12	Severe sepsis	Resistance to septic shock

Table 1-4 A summary of gene variants identified to be protective in the context of specific infectious diseases[93, 98]. Those listed in red highlight the single stranded RNA viruses.

As demonstrated by the examples given above, studies of genetic variants in susceptibility to infectious disease not only provide better understanding of the host immune response and the pathogenesis of specific infectious diseases, they may also provide insights into new therapies[103, 106]. In some cases, such therapies may already be licensed for other conditions and therefore may have passed safety trials, if not efficacy trials, for the specific infection. For instance, given the evolving information of the use of host lipids by viruses in both pathogenesis and increased infectivity it is a possibility that therapies available for hyperlipidaemia may play a role in the treatment of severe viral infections[82, 83, 113]. There is some recent suggestion that patients with severe pneumonia may benefit from

treatment with statins, although the current evidence remains limited and the results of large, randomised control trials are awaited[114, 115].

The infections highlighted in red in Table 1-3 and Table 1-4 are caused by single stranded RNA viruses. This is to underscore that while large scale genetic studies have not been conducted in relation to EVD, there is previous demonstration of specific gene mutations which predispose either to susceptibility to infection with single stranded RNA viruses or susceptibility to severe disease with single stranded RNA viruses. Hepatitis C is a specific example of a single stranded RNA virus in which there has been a significant amount of research to determine underlying genetic predisposition to outcome from infection. This was in part due to the clinical observation that approximately 25% of patients achieve spontaneous clearance of Hepatitis C, while 75% progress to chronic infection[91]. GWAS identified that HLA class II may be associated with spontaneous clearance of Hepatitis C virus and polymorphisms in IFNL3 have been identified as being associated with improved outcome and viral clearance following treatment with pegylated interferon and ribavirin[91].

1.4 Literature supportive of genetic susceptibility to EVD

The contribution of genetics to EVD is suggested by the marked variability in outcome among apparently healthy individuals. This includes those who are highly exposed but never develop the disease; asymptomatic and minimally symptomatic cases, and a relatively fixed mortality rate across different outbreaks of the same EBOV species. There is limited data on genetic susceptibility to EVD specifically however, several different papers are suggestive that factors other than exposure, co-morbidity and variations in viral genetic sequence contribute to outcome in EVD[5, 18, 21, 39, 116].

1.3.1 Evidence of different disease phenotypes and asymptomatic infection

There are several early papers describing asymptomatic infection with Ebola virus, however they have been criticised for potential cross-reactivity in the assays used to determine Ebola virus antibodies[117-119]. The earliest Ebola virus sero-surveys in West Africa were conducted in the 1980's in two different communities in Liberia. Both studies describe a higher incidence of antibodies to Ebola virus when compared to other haemorrhagic fever viruses[117, 119]. While one study was conducted in the context of a Lassa fever outbreak, the second study was conducted outside of an outbreak setting. This study determined that antibodies to Ebola virus were more prevalent in the population than antibodies to Lassa fever, despite it being well known even then that Liberia is subject to sporadic Lassa fever outbreaks[119]. As no known outbreaks of EVD had occurred in Liberia at this time, it is suggestive that

among some communities there was the presence of asymptomatic or minimally symptomatic infection with EBOV.

More recently sero-surveys have been conducted in West African communities that were affected by Ebola virus. Two of these studies utilised an oral fluid antibody assay to detect Ebola IgG, one was conducted in Sierra Leone and the second in the community of Meliandou in Guinea, where the West African outbreak is believed to have commenced[120, 121]. Both studies determined that asymptomatic or minimally symptomatic infection, while not common, does exist (4.30% and 3.57% respectively). They also determined that different levels of exposure pre-disposed to higher risk of infection, but that there were those within communities with very high exposure who were seronegative[120, 122]. The presence of asymptomatic or minimally symptomatic infection suggests that some individuals may carry protective mutations that prevent them from becoming unwell with EVD (disease severity genes); that others can be highly exposed but remain antibody negative suggests some individuals carry protective traits that prevent them becoming infected with EBOV (disease susceptibility genes).

Variations in clinical disease syndromes and prevalence of Post-Ebola syndrome (PES) in Ebola survivors raises questions about why there are disparities. While there may be predisposing factors such as viral load during acute disease, length of illness and others, it is unclear why there are different disease phenotypes. That some survivors develop severe uveitis and others do not[53, 60], and some male survivors have virus persistence in seminal fluid and others do not[60, 61, 63], suggests there are differences in host response to the virus with some individuals more able to clear the virus than others. Again, these different phenotypes are suggestive of genetic differences that may be protective or detrimental in the presence of Ebola virus infection.

1.3.2 Role of human leucocyte antigen (HLA) B in outcome of EVD

The earliest suggestion of a genetic susceptibility to disease outcome from EVD is a study from the Ebola epidemic in Uganda (2000-2001), which found different HLA types were associated with either fatal or non-fatal disease outcomes[123]. This study identified that HLA alleles B67 and B15 were associated with fatal outcomes and HLA alleles B07 and B14 were associated with survival[123]. To date this and a study looking at natural killer cell receptor mutations[124] are the only genetic studies of host susceptibility to EVD, although a relatively recent paper from Guinea demonstrates different transcriptomic signatures in EVD[125].

1.3.3 Transcriptomic signatures in acute EVD

While different from genomics, transcriptomics can highlight variations in upregulation and down regulation of genes which occur during an active disease process and which may highlight differences between individuals who go on to experience different disease outcomes. A recent publication from a small group of EVD patients in Guinea studied differences in transcriptomic signatures during acute EVD in patients who survived the disease versus those who died[125]. While there was a predominance of deceased patients when compared with survivors (n=88 vs n=24), the authors were still able to demonstrate a difference between the groups. Deceased patients demonstrated greater up regulation of acute phase responses, interferon signalling and albumin and fibrinogen genes. The authors conclude that they believe they would be able to predict patient outcome on the basis of a gene panel they developed[125]. That transcriptomic signatures vary between EVD patients and can be predictive of outcome suggests a difference in the host response to infection, and potentially a genetic predisposition to one outcome or another.

1.3.4 The NPC1 gene

As mentioned earlier the role of the NPC1 receptor has been explored in the role of pathogenesis in EVD, with evidence from a study using mouse models that wild type mice experienced a more severe clinical disease and had mortality rates of 90%, when compared with heterozygous or receptor null mice[87]. Humans born with variants in the NPC1 gene that alter its cholesterol transport function have a condition called Nieman Pick C disease, which is invariably fatal by early adulthood. The pathogenesis of Nieman Pick C disease is based on the failure of the NPC1 receptor to transport lipids resulting in accumulation of sphingolipids in lysosomes[89]. It is unlikely a variant in NPC1 that affects its cholesterol transport function would be genetically preserved due to the adverse consequences, however a variant in NPC1 at the site that Ebola binds (a separate location on the receptor to where cholesterol binds, see Figure 1-6) would not alter the lipid transport function of the receptor. On this basis, a variant in this region may potentially be preferentially selected in regions where EBOV infection is present.

Similarly, further studies have demonstrated that mutations in other genes related to endosomal-lysosomal fusion may play a role in inhibiting or reducing EBOV entry to the cell cytoplasm. Of particular note were the six sub-units of the homotypic fusion and vacuole protein-sorting (HOPS) complex (VPS11, VPS16, VPS18, VPS33A, VPS39 and VPS41), which mediates fusion of endosomes and lysosomes and plays a role in endosome maturation. After NPC-1, VPS33A and VPS11 were particularly enriched in an EBOV resistant cell population[85, 126]. These studies support a genetic predisposition to both

infection with and survival from Ebola virus disease on the basis of the NPC1 gene[127] and related genes involved in endosomal-lysosomal fusion.

On these grounds, and the findings of the role of lipid pathway genes in the pathogenesis of other RNA viruses, it was postulated that there may be specific gene variants in lipid pathways that may result in susceptibility to or protection from infection with EBOV. Due to the absence of any previous studies of human genetics and EVD indicating a causal role of specific candidate genes in either susceptibility to infection or susceptibility to severe disease, it was not assumed that it would be possible to undertake candidate gene analysis, however based on the above data it was felt important to utilise DNA obtained through this study not just for genotyping, but also for whole exome sequencing. The analytical approach to the latter would initially require a broad look to determine if there were any specific genes that appeared to impact on susceptibility to infection or susceptibility to severe disease, and which may be implicated as possible candidate genes. It was also considered that initial analysis might highlight several genes that clustered within one of the above lipid pathways. In the event of either of the above situations it would then be possible to refine the analysis and undertake either a candidate gene analysis or a specific pathway analysis.

The advantages of utilising a candidate gene approach are that it is a relatively quick process once all the data is available and it may provide weight to potential genes already identified through a broader analytical approach, but which may have small effect sizes. It can however be negatively impacted by a lack of statistical power and can be subject to a relatively high rate of false positives, with failure to replicate findings in further cohorts[128]. Pathway analysis helps to overcome the issues of lack of statistical power and small effect sizes by looking at a group of genes from certain pathways which may demonstrate a high mutational burden within one disease phenotype when compared with another. Alone these genes may not be statistically significant or may demonstrate a small effect size so may not appear to impact the disease phenotype, however when identified together the effect sizes are aggregated and therefore may demonstrate a significant relationship to disease phenotype. This process is dependent though on robust understanding of the pathways involved[128].

1.4 Research proposal and hypotheses

The Paediatric department at Imperial College London has an ongoing research program on the genetic basis of infectious diseases including tuberculosis, meningococcal disease, and other bacterial and viral infections, as well as Kawasaki disease. This research has shown a role for both Mendelian severe mutations, and common polymorphisms in a range of infectious diseases[93]. Following discussion with

Prof Levin a proposal to establish a cohort of EVD deceased cases, survivors, household contacts and unexposed controls was developed in order to investigate the hypotheses (Figure 1-8):

1. That differences in outcome of EBOV exposure have a genetic basis.

2. That genetic differences in response to EBOV might arise from either rare mendelian mutations, common polymorphisms, or a combination of both genetic effects.

3. That identification of rare variants influencing EVD outcome in families might point to the pathways in which common variants may also act to influence susceptibility and severity of disease.

4. That identification of the genes underlying severity of disease and resistance to infection might help to identify new pathways for therapeutic intervention.

This proposal received initial funding from the Wellcome ISSF and the Institute of Global Health Innovation at Imperial College in the form of a one year fellowship to conduct pilot work. This pilot work resulted in the award of a Wellcome Clinical Research Training Fellowship funding a further three years in order to investigate the hypotheses illustrated below in Figure 1-8. The remainder of this thesis will describe the program of work undertaken to establish the cohorts listed above and in turn begin to address each of the hypotheses.

Potential locations of genetic variation that could reduce susceptibility or severity of disease in EVD
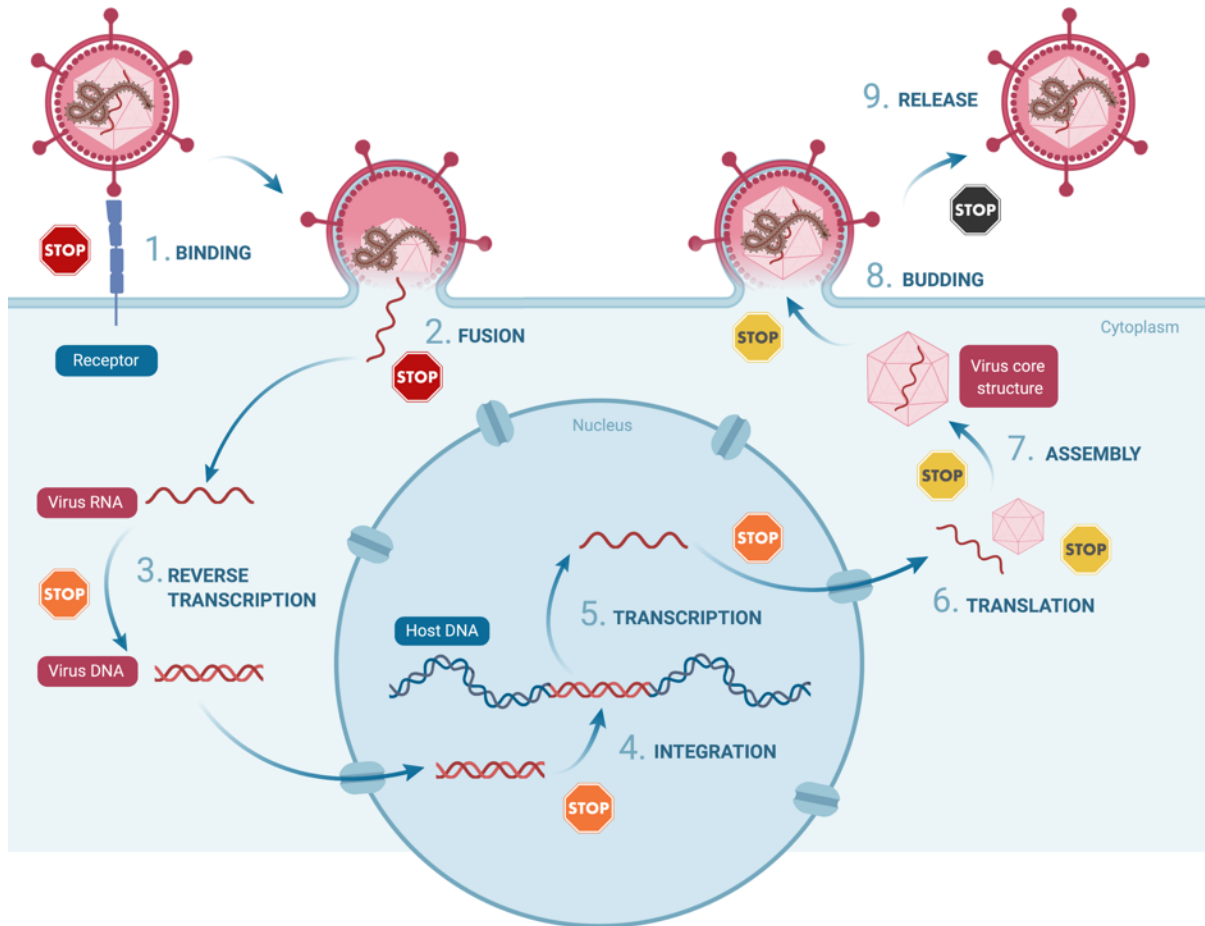
Figure 1-8 Locations of possible host genetic variation in the EBOV host cell entry and replication cycle which may be involved in susceptibility to infection with EBOV or severity of clinical disease. The red STOP signs are proposed as locations where variants may interfere with EBOV host cell entry resulting in an individual with high level exposure not becoming infected. The orange STOP signs are proposed as locations where variants may interfere with viral replication resulting in an individual becoming infected with EBOV but not experiencing significant symptoms due to reduced production or activity of new virions. The yellow STOP signs are proposed as locations where variants may interfere with proper formation of new virions resulting in an individual becoming infected with EBOV and experiencing clinical symptoms of EVD, but which may be of a milder nature than in a host without such variants. The black STOP sign is proposed as the point at which the host inflammatory response may alter the outcome of EVD infection; this is where a new, fully formed virion has been produced in an individual infected with EBOV, therefore the clinical response to EBOV infection is now dependent on the innate and humoral host immune response which determines the severity of clinical symptoms experienced and the ultimate outcome of disease.

Chapter 2 Materials and methods overview

This chapter provides an overview of the methods used in the study. A detailed description of the fieldwork undertaken to recruit the participant cohorts and the challenges faced in carrying out the study in a nation with very limited research infrastructure can be found in Appendix A. A summary of the specific methods used is found in Table 2-1 below. More detailed descriptions of specific methods are also found in chapters three through seven.

Methods used	Purpose	Genre of methodology
Funding and ethical approval	Approval of study	Good clinical practice
MOHSL-PHE Ebola biobank sample access	Access a cohort of EVD deceased participants	Scientific collaboration
Tracing outcome data	Phenotyping of participant categories	Epidemiological
Determining outcome based on sample CT value	Define participant categories	Epidemiological
Determining recruitment sample size	Power the study	Statistical
Developing an international collaboration	Power the study	Scientific collaboration
Developing a relationship with University of Makeni	Facilitate field and lab work in Sierra Leone	Scientific collaboration
Identifying fieldwork cohorts	Phenotyping of participant categories	Epidemiological
Risk exposure stratification	Phenotyping of participant categories	Scientific collaboration & Epidemiological
Fieldwork	Participant recruitment	Epidemiological & anthropological
Sample handling and processing	Optimising recruited samples	Laboratory
Enzyme linked immunosorbent assay	Determining Ebola antibody status	Laboratory
Comparison with LSHTM samples	Rule out Ebola survivor sero-conversion	Laboratory & Scientific collaboration

Sample shipment to the UK	DNA extraction and storage of samples	Logistics & Laboratory
DNA extraction	Host DNA for sequencing	Laboratory
DNA concentration and plating	DNA optimisation for sequencing	Laboratory
Exome sequencing and genotyping	Genetic data for analysis	Laboratory
Antibody data analysis	Understanding community disease transmission/classifying participant categories	Statistics, epidemiological & phenotyping
Bioinformatic data analysis	Understanding genetic determinants of disease phenotype	Bioinformatics

Table 2-1 Summary of methods used.

2.1 Early planning

Initial funding of the study was committed to in November 2014, through the award of a one year Wellcome Institutional Strategic Support Fund (ISSF) Ebola Fellowship at Imperial College London, funded by both the ISSF and the Institute for Global Health Innovation, commencing from March 2015. Contact was made with Dr Tim Brooks at Public Health England (PHE), who oversaw the PHE diagnostic labs in Sierra Leone. With the permission of the government of Sierra Leone, Dr Brooks and PHE established an Ebola Biobank at Porton Down (the 'Sierra Leone Ministry of Health and Sanitation and Public Health England Ebola biobank' from this point on referred to as the 'Ebola biobank') to store aliquots of diagnostic samples from Ebola patients in Sierra Leone. This contact led to a successful application to access samples for a deceased and survivor cohort through the Ebola biobank, facilitating the study being based in Sierra Leone.

2.1.1 Initial assessments, ethical approval and funding

In March of 2015 a successful preliminary trip to Sierra Leone led to contact with members of the Ebola task force and governmental actors. A further trip in July/August of 2015 led to preliminary assessments of where the field and laboratory work could be based, as well as a successful application for ethical approval from the Sierra Leone Ethics and Scientific Review Committee (SLESRC) in Freetown. Ethical approval in Sierra Leone was followed by ethical approval from the Imperial College Research Ethics

Committee in January 2016. Shortly prior to this further funding for the study had been secured through a successful application for a three year Wellcome Clinical Research Training Fellowship, commencing from March 2016.

2.2 Study design

The initial study design was to conduct a genome wide association study. This was decided on the basis of looking for common variants affecting disease severity, and also because it was unclear at the outset whether it would be possible to determine extreme phenotypes that differed in disease susceptibility. On this basis the power calculation shown in section 2.3 was calculated for the purpose of powering a GWAS. The recruitment of participants to the study, with the exception of samples accessed through the Ebola biobank, was conducted in Ebola affected communities in Sierra Leone. Participants were recruited as households, either affected (household contacts) or unaffected (community controls). This is explained in greater detail in chapter 3. By recruiting entire households in affected communities, it was possible to ensure recruitment of adequate participant numbers, however it did mean there would be high levels of relatedness among participants. In communities in rural areas of Sierra Leone there is often a degree of relatedness among the majority of the community even if participants are not identified as being from the same household. While ideally for a GWAS unrelated individuals are sought, due to the circumstances of the recruitment and the communities in which the recruitment was being conducted, this was not possible. An inventory of each household was conducted during recruitment to determine the degree of relatedness of individuals for exclusion from later analysis if necessary.

2.3 Powering the study

Determining the power of a genome wide association study (GWAS) when there is limited previous research on the topic is essentially a hypothetical exercise as the allele frequency within the population must be assumed to make the calculation. As it is unclear specifically what allele is being looked for, an assumption must be made of the potential allele frequency of an unknown allele within the study population. Similarly, the rates of linkage disequilibrium (LD) among the population studied must be applied, in this case LD rates were not known for the specific population and had to be assumed based on data from another West African population. Using the Genetic Power Calculator[129] for the primary outcome measure of deceased versus surviving patients, a relatively high allele frequency of 35% was assumed for the calculation with a prevalence of 60% among the population on the basis of relatively consistent mortality rates of 50-60% of an infected, symptomatic population. While linkage disequilibrium rates are lower for African populations a study of the Yoruba population in Nigeria[130, 131] demonstrated linkage disequilibrium rates of approximately 0.2. With the assumption of

comparing 600 fatal and 600 surviving cases, the study was powered at >80% for a significant association. These assumptions and the power calculation are shown in Table 2-2 and Table 2-3.

Case-control parameters	
Number of cases	600
Number of controls	600
High risk allele frequency (A)	0.35
Prevalence	0.6
Genotype relative risk Aa	50
Genotype relative risk AA	25
Genotypic risk for aa (baseline)	0.02287

Linkage disequilibrium statistics	
Linkage Disequilibrium (D')	0.021
Linkage disequilibrium (r-sq)	0.01279
Haplotype frequency (AB)	0.2532
Haplotype frequency (Ab)	0.09677
Haplotype frequency (aB)	0.3968
Haplotype frequency (ab)	0.2532

Table 2-2 Genetic power calculator parameters demonstrating the assumed allele frequencies, effect size and level of linkage disequilibrium for the power calculation. Due to the high mortality rates and high infection rates among those in contact with EBOV a relatively high allele frequency and effect size was assumed. Linkage disequilibrium rates are known to be lower for African populations[131], a study[130] of the Yoruba population in Nigeria demonstrated linkage disequilibrium rates of approximately 0.2.

Alpha	Power	N cases for 80% power
0.1	0.8948	441
0.05	0.8258	560
0.01	0.6262	834
0.001	0.3472	1220

Table 2-3 Results of the power calculation using the Genetic Power Calculator (GPC)[129]. Results shown for the Case-control statistics: allelic 1 df test (B versus b).

It was initially unclear if it would be possible to obtain 600 fatal cases and if there would be sufficient enough DNA extracted from all samples. Therefore, in order to adequately power the study researchers conducting similar studies in Liberia and Guinea were contacted. These researchers included Prof Laurent Abel (INSERM, Paris) and Prof Anavaj Sakuntabhai (Institut Pasteur) for studies in Guinea and

Prof Pardis Sabeti (Broad Institute, Harvard University) for studies in Liberia and the Eastern part of Sierra Leone. It was agreed between all parties to provide validation cohorts for significant results in individual analyses and to conduct a meta-analysis of genotyped data from all three nations to ensure a well powered GWAS. This brought together approximately 10,000 samples, of which roughly 1200 were deceased samples. It must be emphasised that all parties had the relevant approvals to access these samples and conduct this research from the relevant countries involved, as well as ethical approval from these countries and their own institutions.

2.4 Sample access and tracing outcome data

The Ebola biobank was created at Porton Down at the end of 2015, and a formal application process to access samples implemented. A successful application to access samples was submitted in April of 2016. To determine the most appropriate samples to access from the biobank, outcome data was required. While data such as lab sample ID, sample volume, PCR cycle-threshold value (CT value) for Ebola virus and the age and sex of the patient was available, there was no outcome or clinical data for the samples. This data was, however, essential to determining how to categorise the samples for the genomic analysis.

To determine outcome for the samples, anonymised clinical data was obtained from several of the non-governmental organisations (NGOs) who had operated Ebola treatment facilities in Sierra Leone. These included the International Medical Corps (IMC), GOAL and Save the Children. Despite significant challenges in linking clinical data to samples in the Ebola biobank it was possible to determine outcome for 286 plasma samples (147 survived, 139 deceased patients). The CT values for these samples were also included in the biobank data set. The CT value is the cycle threshold value from the diagnostic real time PCR test that was used to diagnose a patient with EVD during the epidemic. Real time PCR identifies the presence of the virus in the patient's sample by detecting the viral RNA and amplifying it through multiple cycles of replication. The cycle threshold is the point at which, when crossed, a PCR test would be called as a positive result. It is representative of the number of PCR cycles that need to be completed before the viral RNA is identified. The more viral RNA in the sample, the less PCR cycles need to be completed before sufficient fluorescence is emitted for the positive threshold to be crossed. Therefore, the lower the CT value the higher the viral load. Using the CT values for the 'outcome identified' plasma samples from the biobank data set it was possible to determine a CT value cut off to establish with >98% certainty that a patient had died (Figure 2-1). The ROC curve used to determine this cut-off is shown in Figure 2-2, where using the Youden score to optimise the specificity at >98% for a true 'deceased' case led to a CT value cut-off of <18 being established.
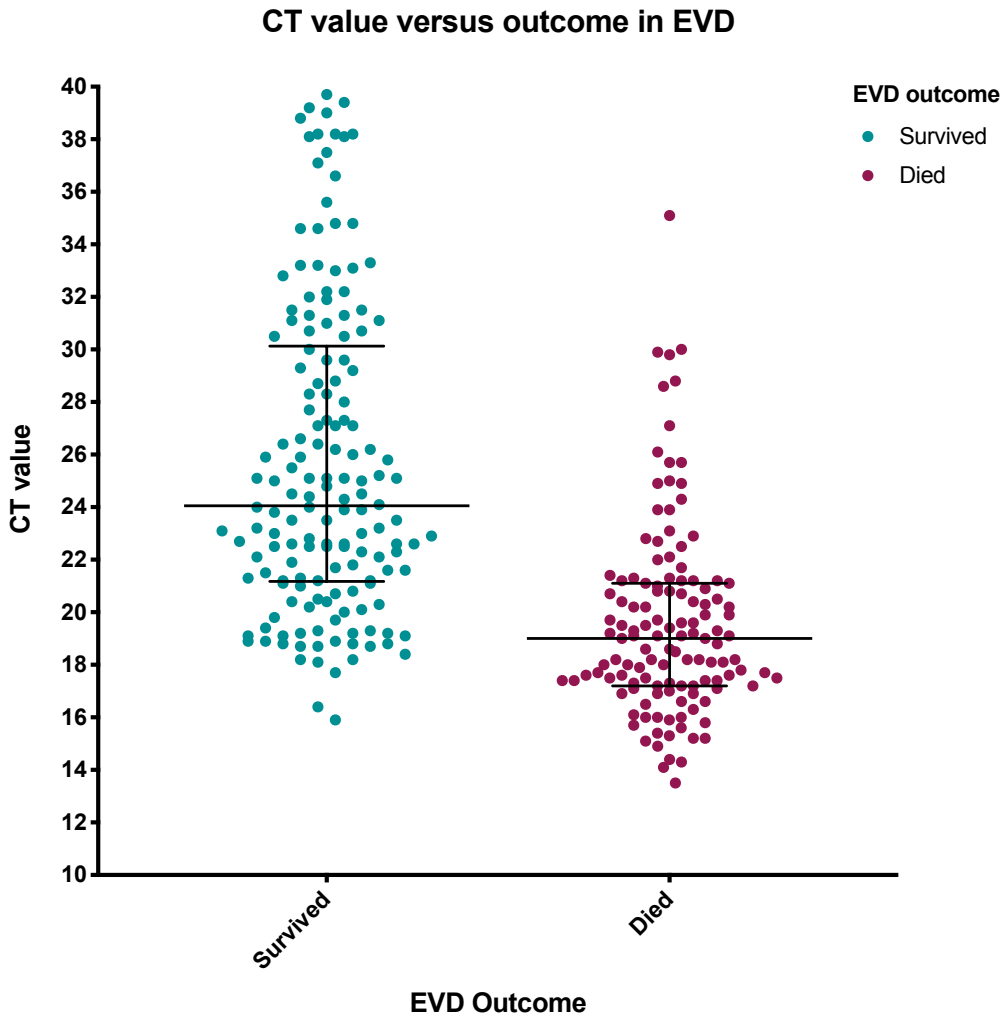
Figure 2-1 CT value of EVD patients who survived and died displaying median and inter-quartile range. Only two of the known survivors demonstrated CT values of less than 18.

ROC curve of CT value versus EVD outcome
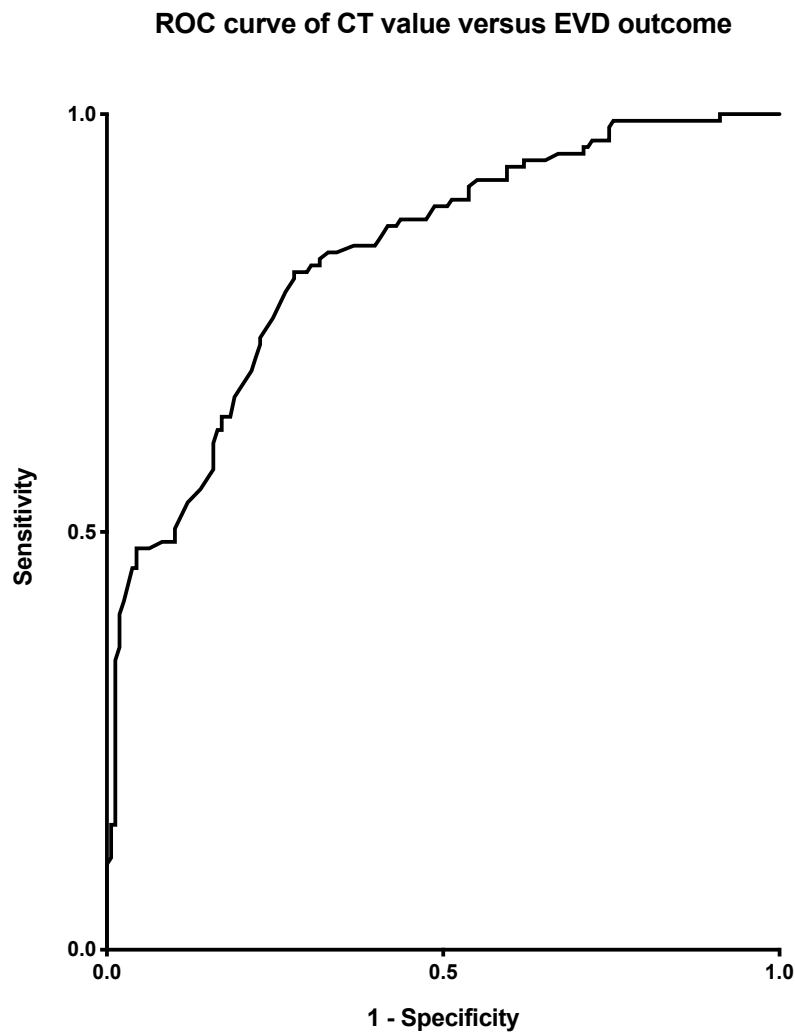
Figure 2-2 ROC curve of CT value as a predictor of death in EVD patients. Using this ROC curve, the Youden score was used to optimise the specificity at >98% for a true 'deceased' case, this identified a CT value cut-off of <18 for the assumption that a case with unknown outcome was deceased.

Using this cut off enabled the identification of an 'assumed deceased' participant group and therefore a further 94 'deceased' plasma samples in the biobank. Spun supernatant from swab samples was also stored in the biobank. As swabs were largely only taken from deceased individuals in the community, it was assumed the swab samples were from deceased individuals. There were 119 swabs with a sufficient supernatant volume available in the biobank.

Due to limitations in samples in the Ebola biobank, access was granted to up to 500 samples/application from the biobank. While the application was provisionally approved in April 2016, formal approval by the Ebola biobank committee occurred in August 2017. Due to the limitations in samples that could be accessed, samples from deceased individuals were prioritised and an additional fieldwork arm to recruit Ebola survivors developed. The samples accessed from the Ebola biobank are shown below in Table

2-4. While 499 samples from the biobank were applied for, of which 352 were deceased or assumed deceased and 147 were Ebola survivors, the samples that were received totalled 492 as seven samples (five deceased and two survivor) could not be located.

Outcome	Plasma sample	Swab supernatant sample	Total
Deceased	139	0	139
Assumed deceased	94	119	213
Survived	147	0	147
			499

Table 2-4 Sample types/outcomes applied for from the Ebola biobank.

2.5 Planning the fieldwork

2.5.1 Identifying a laboratory in Sierra Leone

In 2015 an infectious disease research laboratory had been developed by Prof Ian Goodfellow at the University of Makeni in Sierra Leone. Following discussions with the Vice-chancellor of the University, a successful application to partner with the laboratory to conduct the field and Sierra Leone based laboratory work was made. This relationship made it possible to store samples that required freezing, heat treat saliva samples for DNA extraction to neutralise any pathogens prior to shipment to the UK and conduct the Ebola antibody Enzyme Immunosorbent Assay (EIA) in country. The lab also acted as a base for the storage of equipment for the fieldwork and training of the field team.

2.5.2 Fieldwork details

A detailed description of the field and laboratory work undertaken is contained in appendices A-I, which describe the methods used, the challenges faced, and how these were overcome. Chapter three describes a summary of the fieldwork and the recruitment categories used. Recruitment of participants took place in several different communities in order to control for ethnicity, this was based on approximate location data traced from the Ebola biobank samples. By identifying the communities where the biobank samples originated from, it was possible to crudely control for ethnicity by recruiting the household contacts, community controls and Ebola survivors from the same communities and regions.

In order to reassure communities and improve the chances of participant recruitment a non-invasive saliva collection kit manufactured by DNA Genotek (Oragene OG-500 or OG-575, for young children)

was identified. These kits provide consistent, adequate yields of DNA from saliva for downstream sequencing (as per manufacturers guidelines, see Appendix B). To assist with phenotyping the study participants it was necessary to undertake Ebola IgG antibody studies on all participants. An oral fluid Ebola IgG capture assay designed by Prof Richard Tedder and his team at PHE, but now in manufacture by Kalon Diagnostics Ltd was selected (this assay is described further below). This assay identifies the presence of IgG antibodies to Ebola, which confirms if a person has been previously infected with the virus. It can therefore be used as a test of previous exposure, as opposed to the RT-PCR which was used to identify acute infection in those who were unwell at the time of testing.

2.6 Laboratory work in Sierra Leone

As mentioned above, the laboratory work in Sierra Leone was conducted at the Infectious disease research laboratory at the University of Makeni. All laboratory work was conducted by myself, with the assistance of Dr Raoul Emeric Wadoum (oral fluid extraction from swabs) and Mr Samuel Alberman (anti-EBOV IgG capture EIA).

2.6.1 Sample handling and processing

At the time of collection all samples were sealed in their specific receptacle and placed in a sealed specimen bag. The *Oragene* saliva samples were then placed in a heavy duty storm case to protect them in transit from the field to the laboratory, the oral fluid samples were placed in a heavy duty cooler containing ice to keep them at approximately 4-8 °C. Once at the laboratory, the oral fluid samples were immediately placed in a -20 °C freezer. The *Oragene* saliva samples were heat treated in a water bath at 60 °C for sixty minutes, the outer receptacle was then dried and stored in a designated, labelled *Oragene* tube storage box which could hold 36 tubes. These boxes were then stored in a large outer plastic container with lid until all samples had been collected and shipment to the UK could be arranged in the heavy duty storm cases.

At the end of each week the oral fluid swabs were defrosted. Using the level 2 biological safety cabinet in the BSL 2 laboratory, the samples were eluted in a viral transport medium (produced by PHE) through a mechanism of frothing the swab in viral transport medium in the original receptacle and then wringing out the swab and aliquoting the fluid into 2ml *Sarstedt* tubes. These *Sarstedt* tubes were then frozen at -20 °C until the EIA was conducted several months later. The swab was discarded at this stage. All standard operating procedures are in Appendix J.

2.6.2 Enzyme Linked Immunosorbent Assay (EIA)

The Anti-EBOV IgG Capture EIA used was specific to EBOV glycoprotein antigen, as such it only identifies IgG to EBOV glycoprotein, and not to other viral proteins. The EIA was formulated based on the EBOV Mayinga glycoprotein antigen and has been demonstrated to have a sensitivity of >95% and a specificity of >99%[120, 132]. This work is described in more detail in chapter 4.

2.6.3 Sample shipment

All the saliva samples were shipped to the UK in September 2017 for DNA extraction, following a protracted process of obtaining permissions and signatures (detailed in appendix B). All permissions were successfully obtained, and the samples arrived at Imperial College in September 2017.

2.7 Laboratory work in the UK

The DNA extraction from the saliva samples was conducted at Imperial College London once the samples arrived from Sierra Leone. Following this, quantification and concentrating of the DNA and plating of the samples for exome sequencing and genotyping was undertaken. These samples were then sent to the Oxford Wellcome Genomics Centre and Illumina Fasttrack Services in California, respectively. These latter aspects are described in more detail in chapters six and seven. With the exception of DNA extraction from the biobank samples which required a level 4 biological safety laboratory, all laboratory work described here was conducted by myself with the assistance of Mr Samuel Alberman.

2.7.1 DNA extraction

Using the *Qiagen QIAsymphony* automated extractor, which is the only automated extractor with a validated protocol for DNA extraction from *Oragene* saliva samples, all samples were extracted. The *QIAsymphony* uses a magnetic bead extraction method to extract DNA. The validated protocol for host DNA extraction from *Oragene* saliva samples utilises the *Qiagen QIASymphony* DNA Midi kit and extracts from 1ml of saliva for OG-500 samples and 350ul of saliva for OG-575 samples. The elution volume can be varied from 50ul to 200ul. DNA extraction commenced once all specific extraction protocols had been uploaded to the instrument in November 2017.

As the samples had been heat treated at 60 $^{\circ}$C for 60 minutes prior to leaving Sierra Leone, they were deemed safe to handle at BSL 2. Due to small amounts of debris in the samples it was necessary to aliquot the saliva sample from the original receptacle into separate barcode labelled storage tubes,

avoiding any obvious particles of debris. A 96 sample extraction took approximately seven hours. These samples were then quantified and stored, to be plated at a later date, which is described in detail in chapters six and seven.

2.7.2 Biobank samples

Formal approval to access samples from the Ebola biobank was granted in August 2017. As the samples were infected with Ebola virus the DNA extraction needed to take place in a BSL4 lab. This extraction was therefore conducted by PHE at Porton Down and the extracted DNA from the biobank samples arrived in March 2018.

Unfortunately, only plasma (and a few serum samples) had been stored at the PHE labs during the epidemic, the cell pellet had been discarded. Spun supernatant from swabs of deceased patients had also been stored. While there was concern that neither of these sample types were optimal for a good DNA yield, a trial run of mouth swabs and serum in healthy controls at the end of 2015 yielded sufficient quantities of DNA, with serum samples largely in excess of 50ng/ul and swab samples largely in excess of 30ng/ul. It was hypothesised that due to the endothelial damage most EVD patients are known to succumb to, hopefully there would be a significant amount of cellular material in the serum of deceased patients, which would enhance the DNA yields.

Through discussions with *Qiagen* it had been determined that the most suitable kit for potentially low yield samples was the *Qiagen QIAmp DNA Minikit*. This kit is suitable for host DNA extraction from 200ul of sample. This kit was provided to PHE for the extraction of the Ebola biobank samples. When the samples arrived at Imperial at the end of March 2018, while the majority had sufficient yields, some of the lower yield samples were too dilute as the final elution volume of 200ul had been too large. While 499 samples had been requested, 492 samples were provided as seven samples could not be located.

Following the DNA extraction, all samples were quantified, concentrated if concentrations were too low and then at a later stage plated for shipment to either Oxford for the exome sequencing or California for the genotyping. This process is described further in chapters six and seven.

2.7.3 Accessing funds for exome sequencing and genotyping

Funding for the genotyping and exome sequencing was made available through the award of an Imperial BRC Institute of Translational Medicine and Therapeutics (ITMAT) 'Push for Impact' award in April 2018. This funding covered the cost of the exome sequencing. The genotyping costs were covered

in part by other grants to Prof Levin and in part by Prof Sabeti, who covered the cost of a significant number of the arrays and genotyping costs.

2.8 Analysis and Statistics

Descriptive statistics of the participants recruited to the study, as well as analysis of the antibody, exome sequencing and genotyping results is described in the relevant chapters. Included with this are the specific methods related to that chapter, an interpretation of the results and a discussion of the meaning of these results in the context of EVD. The statistical analysis of the descriptive statistics and antibody results included basic descriptive statistics using means, medians, geometric means and odds ratios. P-values were calculated for both parametric and non-parametric data using t-tests, analysis of variance (ANOVA), Wilcoxon rank tests, Kruskal-Wallis tests and chi-squared tests. Correlations were determined using regression analysis and logistic regression where a binary outcome was present for one of the comparators. All statistics and graphs were conducted using GraphPad Prism 8 (*GraphPad Software Inc.*), Stata 14.2 (*Statacorp LP*), and 'R Studio' 1.1 (*RStudio Inc*); images were created using 'biorender' (*www.biorender.com*).

Chapter 3 Study participant categories

3.1 Introduction

This chapter describes the fieldwork undertaken to recruit participants to the study in the categories of household contacts, community controls and Ebola survivors. These participants were prospectively recruited to the study and are referred to as 'recruited participants' from this point forward. The deceased cohort and a small number of Ebola survivors were accessed through the Ebola biobank as mentioned in chapter two, these samples are referred to as 'biobank participants' from now on. As mentioned in chapter two, the study was initially designed to conduct a GWAS with recruitment of large groups of individuals in the categories of 'household contacts', 'community controls' and 'Ebola survivors'. That study participants were related to each other was unavoidable as Ebola is a disease which spreads among close contacts within households, the vast majority of whom are related to each other to some degree.

A key step in undertaking a genetic study to identify differences in host response to EBOV infection is to ensure well defined disease phenotypes. As mentioned in chapter one, there were different expressions of disease following exposure to EBOV. While the patients who were diagnosed with EVD and had a positive EBOV PCR result (biobank participants) are relatively easy to define in terms of outcome of either survived or died, those that were exposed but did not become unwell or did not seek health care are an unknown classification or phenotype. On these grounds it was important to determine whether the participants recruited to the study through the fieldwork were exposed to the virus and, if they had been exposed, whether they had antibodies to EBOV. This information helps to define the study participants into further categories for the later genetic analysis.

In this chapter the definitions used to identify and characterise these groups are described, along with details of the epidemiological and phenotypic characteristics of each group and the number of participants per group. While antibody status was utilised to further define the groups presented here, this data is presented in more detail in chapters 4 and 5.

3.2 Methods

3.2.1 Genetics study design

The initial study design was based on undertaking a genome wide association study comparing Ebola deceased versus Ebola survivors, as well as highly exposed contacts that were either antibody negative or antibody positive but had not experienced significant symptoms. The latter two categories were unknown at the time of participant recruitment, and it was unclear what the level of seroprevalence to Ebola antibodies would be within affected communities. As participant recruitment progressed it became apparent that within groups there were significant 'sub-categories', particularly in relation to Ebola survivors and the post-Ebola symptoms they experienced. Therefore, it was decided that a small subset of samples would be suitable for exome sequencing in addition to GWAS.

The initial GWAS study design aimed to recruit approximately 600 Ebola deceased, 600 Ebola survivors, 1000 Ebola contacts and 1000 controls (an ethnicity control). This was based on the power calculation described in chapter 2. During the participant recruitment it became apparent that accessing 600 survivor samples and 600 deceased samples from the biobank would not be possible, on this basis deceased samples were prioritised from the biobank and additional participant recruitment was undertaken in Sierra Leone to recruit Ebola survivors. This decision also permitted the opportunity to explore differences in PES symptomatology experienced by Ebola survivors.

While optimally related participants would be avoided in GWAS participant recruitment this was not feasible due to the sample sizes required and the relatedness of persons within rural communities in Sierra Leone. Therefore, it was decided to prioritise obtaining the appropriate sample size and later exclude participants based on genetic relatedness. This latter aspect was particularly important as despite obtaining an inventory for all household contacts, many of the reported relationships were inaccurate, particularly in relation to paternity.

The process of recruitment of participants to the study and obtaining samples through the biobank was based on answering the study hypotheses described in chapter 1. In order to do this, it was important to appropriately categorise and phenotype the study participants. Chapters 3, 4 and 5 describe the process of categorising participants and include an analysis of data obtained through this process. This data was analysed in more detail than simply identifying a rough disease phenotype in order to determine if there were more specific disease phenotypes within each category that were as yet unknown ('sub-categories'), and which would warrant further exploration from an epidemiological perspective, but also from a genomic perspective if sample sizes permitted.

3.2.2 Identifying participant categories

As explained in chapter 1, different clinical syndromes and disease phenotypes were identified in people exposed to EBOV. These clinical syndromes are highlighted in Figure 3-1.
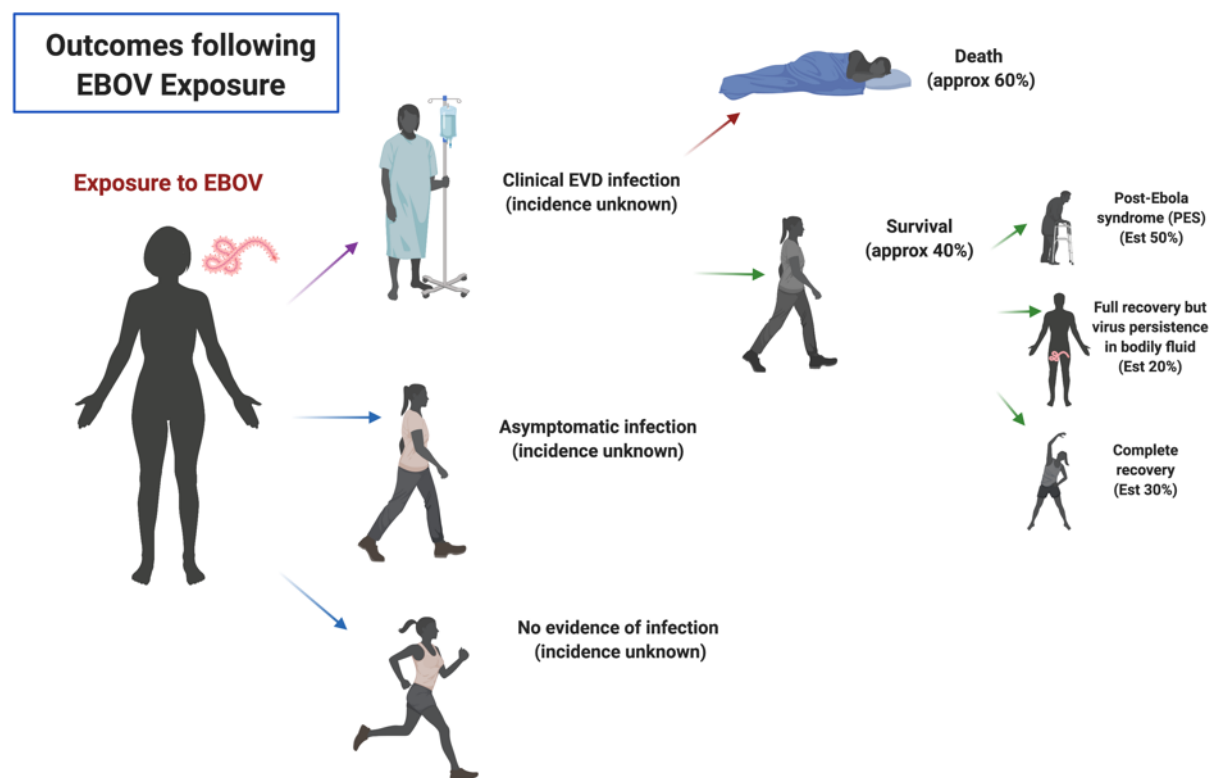
Figure 3-1 Outcomes following EVD exposure flowchart, developed to determine participant categories and approximate recruitment numbers required for the study.

At the time of deciding on recruitment categories and numbers it was unclear what the likely seropositive rates among asymptomatic/minimally symptomatic household contacts would be and what the rates of post-Ebola problems were in Ebola survivors. Assumptions based on personal clinical experience and some early studies[12, 50] were made for the latter which are shown in Figure 3-1.

3.2.3 Exposure risk stratification

In order to stratify the study participants in terms of their exposure to EBOV a risk stratification was used. This risk stratification had been published and demonstrated in two separate studies to correlate well with antibody status among Ebola contacts who had not been formally diagnosed with the disease during the epidemic[120, 122]. Recently a further study has demonstrated that the concepts in the risk stratification are robust, with contacts exposed to deceased individuals and those having contact with

the bodily fluids of an infected person having the highest risk of being seropositive[133]. The risk stratification is shown in Figure 3-2 and was used with permission for the participant recruitment. Based on this risk stratification and the data published in the aforementioned studies[120-122, 133], participants were further categorised into three groups – high risk exposure (risk levels 1-3), moderate risk exposure (risk levels 4-6) and low level exposure (risk levels 7-8) for the purposes of the genetic analysis.

Individual questionnaire: Part B: To be completed after discussion with household.

Record the type of contact the person had with Ebola cases in the household or elsewhere including deceased possible Ebola cases while they were ill, or with the body after death. Ask probing questions if necessary to fill in gaps or check responses.   1 = contact   0 = no contact

Participant ID sticker				
Only the highest level of contact needs to be indicated	In household	Outside household		
Level 1: Contact with the body / prepared the body				
Level 2: Direct contact with body fluids eg. blood, diarrhoea, vomit, urine. Or a baby who was breastfed by an EV+ woman				
Level 3: Direct close contact with wet case (a person with diarrhoea/vomiting/ bleeding) eg helped dress, embraced, carried, helped care, shared bed, or mother breastfeeding an EV+ child				
Level 4: Direct close contact with dry case (eg helped dress, embraced, carried, helped care, shared bed)				
Level 5: Indirect close contact with wet case (eg washed clothes, bed linen)				
Level 6: Indirect close contact with dry case (eg washed clothes, bed linen)				
Level 7: Minimal contact (eg shared meals, shared utensils)				
Level 8: No actual contact known (eg kept distance once person was symptomatic)				

Record any details that may help define the level. Record the reason for the highest level given – i.e. what made you decide that the person had contact at that level. Keep all notes with forms for the household

Figure 3-2 The risk stratification used for recruitment (used with permission from Prof Glynn, LSHTM). Risk levels 1-3 were defined as high risk exposure, risk levels 4-6 as moderate risk exposure and risk levels 7-8 as low level exposure for the purposes of the genetic analysis.

3.2.4 Participant category definitions

The following groups were recruited as study participants through the fieldwork:

1. *Household contacts* - anybody living in a house at the same time that another household member was symptomatic with EVD, but who were not diagnosed with EVD themselves.

a. *Resistant to disease* – household contact with a positive antibody result for anti-EBOV IgG

b. *Resistant to infection* - household contact with high risk exposure (level 1-3) and a negative antibody result for anti-EBOV IgG

2. *Community controls* - anybody living within an affected community who did not have a symptomatic EVD household member in their house, and who was not diagnosed with EVD themself.

a. *Resistant to disease* - a person in any risk exposure level with a positive antibody result for anti-EBOV IgG

3. *Ebola survivors* - a person who suffered an acute infection with EVD and survived; had a positive Ebola viral PCR test result and is able to present a survivor certificate from an Ebola treatment facility or is verified by the chair of the survivors' committee who has previously witnessed the certificate.

a. *Ebola survivors with PES* - Ebola survivors who have ongoing symptoms that meet the definition of Post-Ebola Syndrome.

b. *Ebola survivors without PES* - Ebola survivors who have had no ongoing problems since surviving the disease

c. *Anti-EBOV IgG negative Ebola survivors* - Ebola survivors who have a negative antibody result to Ebola

The samples accessed through the Ebola biobank were classified as follows:

1. *Patient deceased* - a person with a positive Ebola viral PCR test who is recorded as deceased within Ebola treatment facility data records or is assumed to have died based on viraemia (see Figure 2-1 above) or the sample was a swab taken from a dead body.

2. *Patient survived* - a person with a positive Ebola viral PCR test and is recorded as survived within Ebola treatment facility data records.

In order to adequately power the study an ambitious recruitment target was set as demonstrated in Table 3-1 below.

Recruitment category	Target recruitment number
Household Contacts	1000
Community Controls	1000
Ebola Survivors	500

Table 3-1 Target recruitment numbers by category.

3.2.5 Conducting the fieldwork

An extensive description of the fieldwork is contained in appendix A, which describes in more detail the methods used, the challenges faced, and how these were overcome. Recruitment of participants took place in several different communities in order to control for ethnicity. Ethnicity data was not stored by treatment facilities during the epidemic, so instead communities were identified for recruitment based on the location from which the biobank participants came from. The regions participants were recruited from are shown in Figure 3-3 below by participant category.
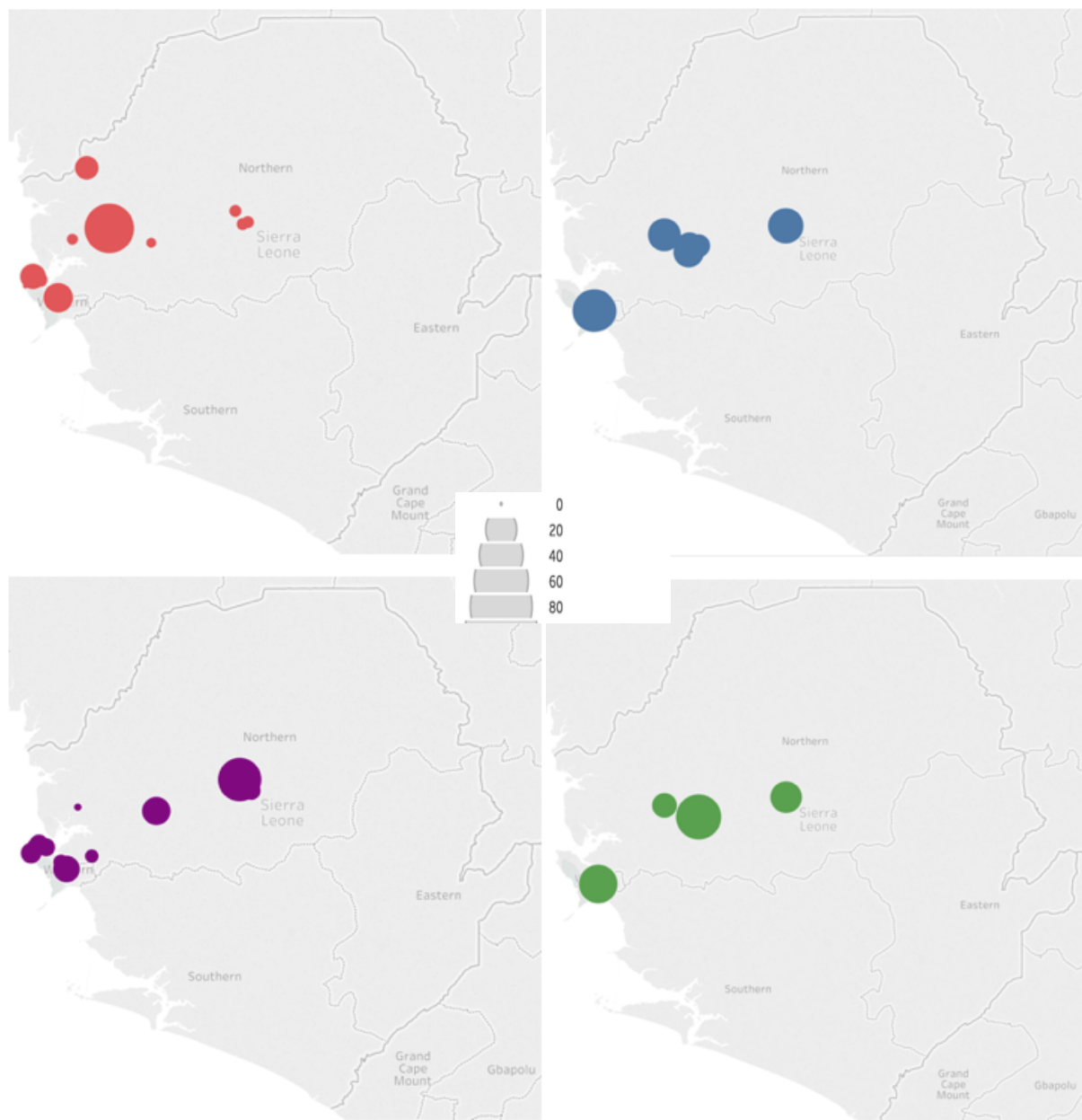
Figure 3-3 Maps of Sierra Leone demonstrating participant recruitment location and distribution. Biobank samples = Red, Household contacts = Blue, Ebola survivors = Purple, and Community Controls = Green.

To improve community participation in the study the need for invasive sampling was minimised by utilising a saliva collection kit manufactured by *DNA Genotek* (*Oragene OG-500* for adults and children over 3 years of age, *OG-575* for children under 3 years). These kits provide consistent, adequate yields of DNA from saliva for downstream sequencing (as per manufacturers guidelines, see Appendix D). To assist with phenotyping the recruited participants, Ebola IgG antibody studies were also undertaken. An oral fluid Ebola IgG capture assay manufactured by *Kalon Diagnostics Ltd* was utilised (this assay is described further below).

The communities targeted for recruitment had all been significantly affected by Ebola and were very sensitive to research studies on this topic. On these grounds the strategy below was used to enhance recruitment (Figure 3-4).
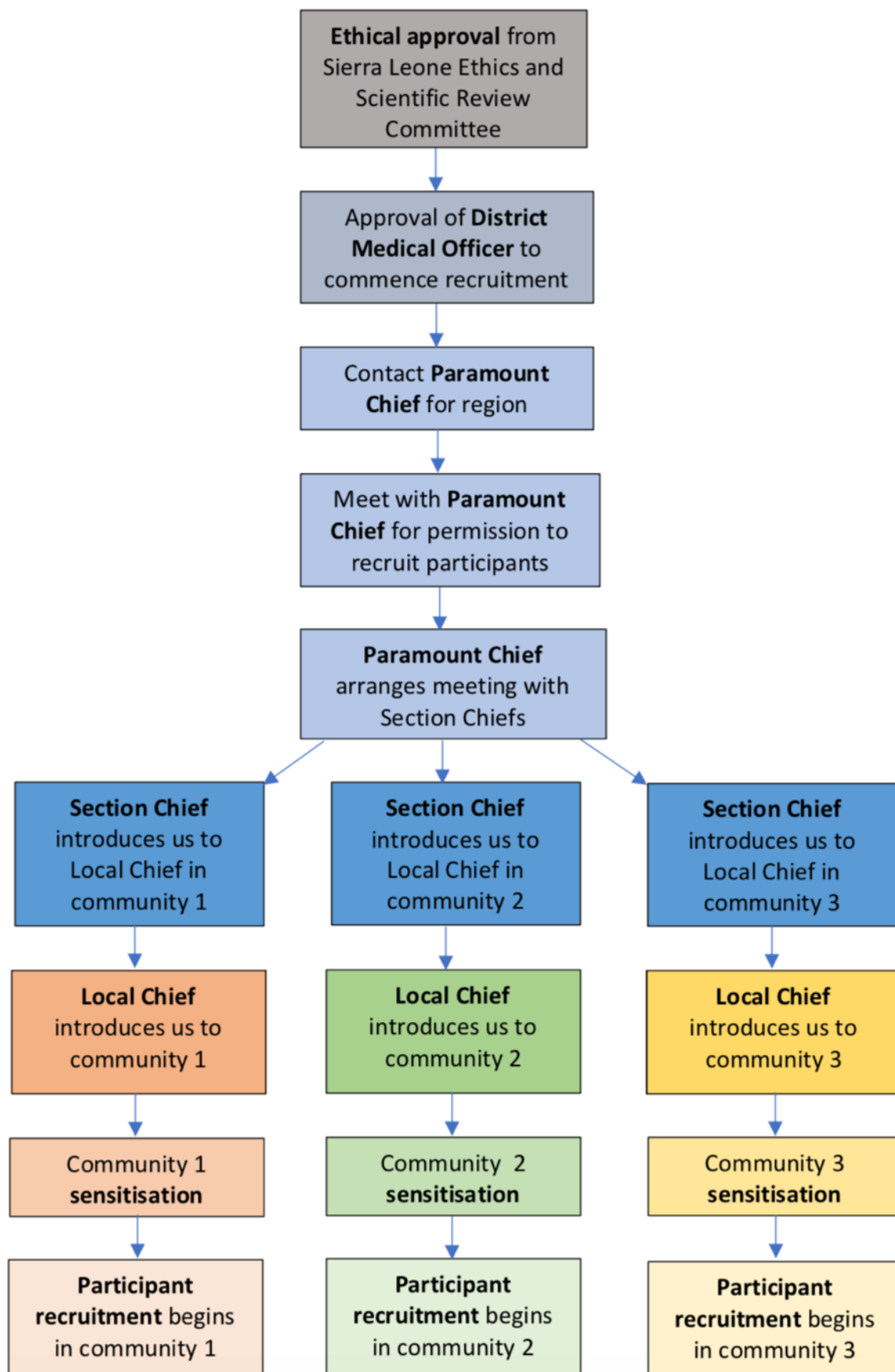
Figure 3-4 Flow chart of the community engagement and sensitisation process.

Data collection was conducted using handwritten paper forms (see Appendix G, H and I), while this was not as efficient for data entry as an electronic method, it was the most feasible method due to limitations in internet access. The forms for recruitment of the household contacts (Appendix G) and community controls (Appendix H) were used with permission from Prof Glynn, following their use in an earlier study seeking to capture similar data[120]. The forms for the recruitment of the Ebola survivors (Appendix I) were developed by myself based on data that it was felt essential to capture for the genomic analysis, as well as the growing understanding of what symptoms constituted post-Ebola syndrome. Ebola survivors were recruited using a different strategy as additional approval was required from the local Ebola survivors' committee. With the support of the local survivors' committee, information was disseminated about the study to the Ebola survivors within their locality, who were then invited to a central location for recruitment. Ebola survivors were recruited individually, rather than in families, however documentation was made of any related survivors. Data was also obtained in relation to any ongoing problems or symptoms of post-Ebola syndrome that Ebola survivors suffered and was categorised in relation to the frequency and severity of the ongoing symptoms, as shown in *Figure 3-5*.

22.	Since you survived have you had any continuing problems?	Yes		
		No		
23.	If yes please describe the problems			
24.	*[When person has finished describing symptoms in their own words, ask them if they had any symptoms on this list that they have not mentioned, if yes tick the box to the right and insert the relevant number to clarify their symptoms]*	Eye disease *[1. Uveitis, 2. Cataracts, 3. Blindness]*		
		Neurological problems *[1. Seizures, 2. Recurrent headaches (>2/week), 3. Persistent sleep disturbance (>4 nights/week), 4. Specific paralysis, 5. Other (describe)]*		
		Deafness		
		Joint problems *[1. Occasional joint pain, 2. Regular joint pain limiting daily activities (>3 days/week), 3. Recurrent joint swelling (>2 episodes)]*		
		Significant weight change *[1. Weight loss, 2. Weight gain]*		
		Recurrent infections >1 every 2 months *[1. Malaria, 2. Infections requiring antibiotics, 3. Both]*		
		Fertility problems *[1. Problems having sexual relations (men), 2. Problems conceiving a baby (men and women)], 3. Recurrent miscarriage (women)]*		

Figure 3-5 A section of the Ebola survivor questionnaire demonstrating the categorisation of ongoing symptoms/PES. Uveitis was further stratified based on the length of treatment with steroid eye drops: mild = <1 month, moderate = 1-3 months, severe = >3months. These categories were later stratified to 'No PES', 'PES' and 'Severe PES', as discussed in chapter 5.

3.2.6 Statistical analysis

The statistics described in this chapter are basic descriptive statistics including percentages, means with 95% confidence intervals and odds ratios with 95% confidence intervals. P-values were calculated using t-tests, Kruskal-Wallis and Chi-squared tests, and logistic regression, with a p-value <0.05 denoting significance. All statistics were calculated in Stata 14.2 (*Statacorp LP*) and 'R Studio' 1.1 (*RStudio Inc*); graphs were created using GraphPad Prism version 8.0.1 (*GraphPad Software Inc*). Images were created using 'biorender' (*www.biorender.com*).

3.3 Data analysis

3.3.1 Analysis categories

As described earlier there were three categories of recruited participants and two categories of biobank participants.

The recruited participants exceeded the target recruitment numbers in all categories and are listed in Table 3-2 below:

Category	Target recruitment no.	No. recruited	No. analysed
Household contacts	1000	1021	1021
Community controls	1000	1004	1004
Ebola survivors	500	504	448

Table 3-2 Numbers of participants recruited and analysed.

The biobank participants accessed through the MOHSL-PHE Ebola biobank are shown in Table 3-3 with the numbers received and analysed:

Category	Sample no. requested	Sample no. received	No. analysed
Ebola patient deceased	352	347	Varies with available data
Ebola patient survived	147	145	Varies with available data

Table 3-3 Numbers of participant samples received from MOHSL-PHE Ebola biobank and number analysed.

The descriptive features of the recruited participant categories and biobank participant categories are discussed in this chapter. Chapters 4 and 5 discuss the antibody analysis for the recruited participants, conducted in order to refine the phenotyping of participants for the genomic analysis. The genomic analysis is described in chapters 6 and 7. Not all of the Ebola survivors were analysed due to difficulties in categorising 56 of the Ebola survivors who were negative for anti-EBOV IgG in their oral fluid samples, this is discussed further in chapter 5.

A total of 499 samples were requested from the biobank, but 492 were received due to seven missing samples. Antibody testing was not conducted on the biobank participants as they all had confirmed EVD and were PCR positive for Ebola virus, therefore their disease phenotype for the genomic analysis was already known. Significant amounts of data were missing for the biobank samples; therefore, the descriptive features of these categories are limited to the data that was available, which varied with the relevant parameter.

3.3.2 Descriptive features of the groups

The following basic demographics were observed for each of the categories:

	Recruited participant categories			Biobank participant categories	
	Household contacts *(%)*	Community controls *(%)*	Ebola survivors *(%)*	Ebola patient deceased *(%)*	Ebola patient survived *(%)*
Age group					
0-5 years	281 *(28)*	342 *(34)*	16 *(4)*	38 *(17)*	3 *(4)*
6-11 years	199 *(19)*	180 *(18)*	37 *(8)*	16 *(7)*	9 *(12)*
12-18 years	156 *(15)*	121 *(12)*	58 *(13)*	21 *(9)*	12 *(16)*
19-39 years	262 *(26)*	229 *(23)*	266 *(59)*	77 *(33)*	39 *(52)*
40-60 years	88 *(9)*	98 *(10)*	65 *(15)*	53 *(23)*	11 *(15)*
60+ years	35 *(3)*	33 *(3)*	6 *(1)*	25 *(11)*	1 *(1)*
Missing	0	1	0	12	2
Sex					
Male	455 *(45)*	425 *(42)*	175 *(39)*	33 *(45)*	13 *(46)*
Female	566 *(55)*	579 *(58)*	273 *(61)*	40 *(55)*	15 *(54)*
Missing	0	0	0	169	49
Ethnicity					
Fula	11 *(1)*	11 *(1)*	16 *(4)*	NA	NA
Kissi	4 *(<0.5)*	0	0	NA	NA
Kono	16 *(2)*	13 *(1)*	5 *(1)*	NA	NA
Krio	4 *(<0.5)*	6 *(0.5)*	2 *(0.5)*	NA	NA
Kuranko	12 *(1)*	11 *(1)*	5 *(1)*	NA	NA
Limba	194 *(19)*	170 *(17)*	62 *(14)*	NA	NA
Loko	24 *(2)*	15 *(2)*	9 *(2)*	NA	NA
Mandingo	12 *(1)*	20 *(2)*	7 *(2)*	NA	NA
Mende	55 *(5)*	68 *(7)*	21 *(5)*	NA	NA
Sherbro	12 *(1)*	4 *(<0.5)*	3 *(1)*	NA	NA
Susu	16 *(2)*	13 *(1)*	7 *(2)*	NA	NA
Temne	654 *(64)*	668 *(67)*	294 *(66)*	NA	NA
Yalunka	0	0	1 *(<0.5)*	NA	NA
Loko/Mende	1 *(<0.5)*	0	0	NA	NA
Mandingo/Kono	0	0	1 *(<0.5)*	NA	NA
Mandingo/Susu	0	0	1 *(<0.5)*	NA	NA
Temne/Fula	0	0	1 *(<0.5)*	NA	NA
Temne/Limba	5 *(0.5)*	0	7 *(2)*	NA	NA
Temne/Loko	0	0	1 *(<0.5)*	NA	NA
Temne/Mandingo	0	1 *(<0.5)*	1 *(<0.5)*	NA	NA
Temne/Susu	0	0	1 *(<0.5)*	NA	NA
Missing	1	4	3	NA	NA

Table 3-4 Basic demographics of the participant categories (NA = not available).

Age for all the recruited participants was reduced by two years to be representative of age at the time of the epidemic. As can be seen from Table 3-4 and Figure 3-6, the distribution of participants in different age groups was similar between the household contact and community control categories, with a median age of 13 and 11 respectively; but differed from the two survivor categories which were similar to each other with median ages of 25 and 26. The distribution of ages in the deceased category

diverged from the four other categories with a median age of 30. A non-parametric Kruskal-Wallis test demonstrates there were significant age differences between these five groups with a p-value < 0.001.

Age distribution for participant categories
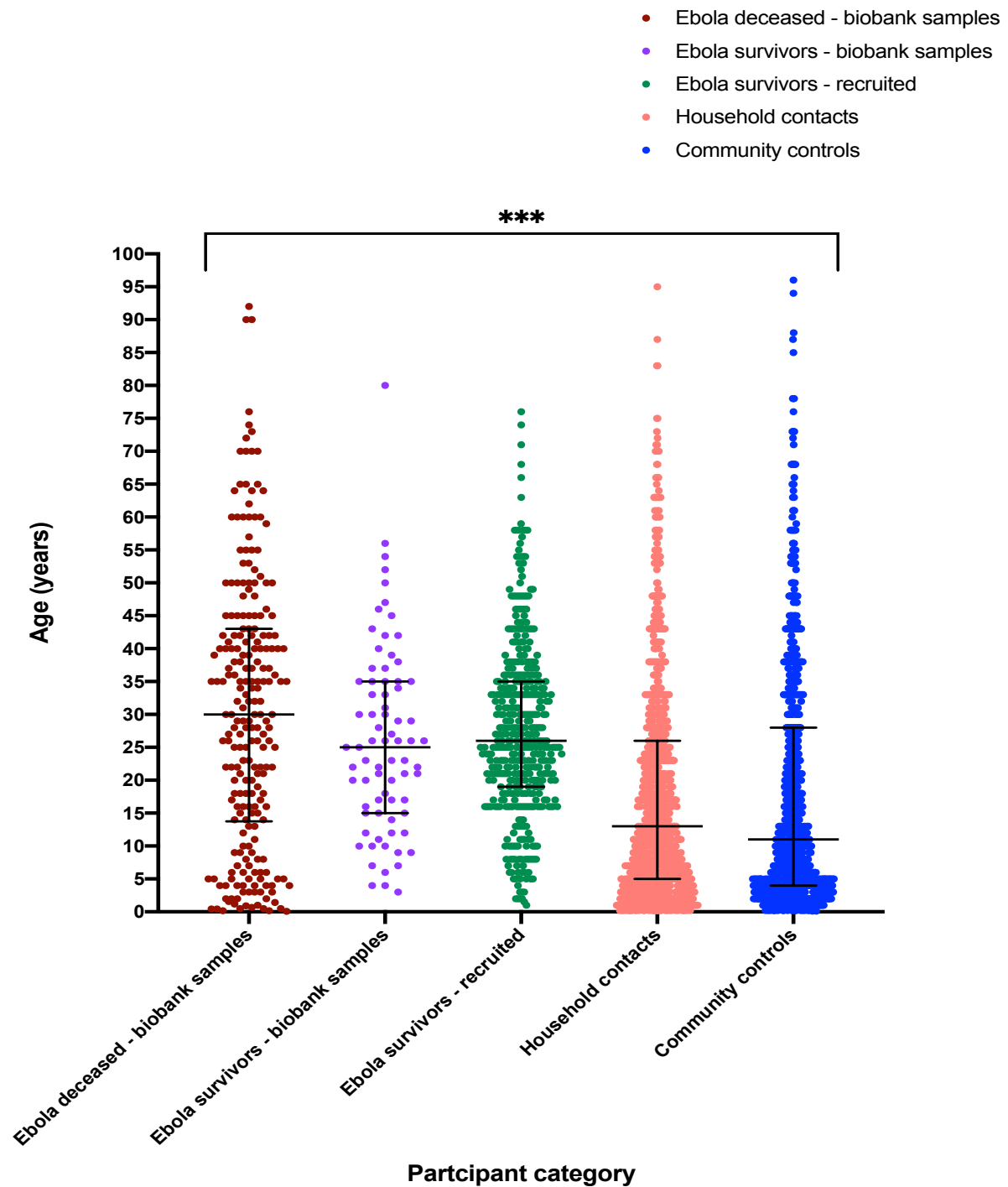
Figure 3-6 Age distribution across participant categories with error bars displaying median and interquartile range. *** significant at p<0.001.

Where the data is available it demonstrates there was no significant variation in the proportion of males and females between the categories (Table 3-4, Figure 3-7), with a p-value of 0.371. In each category there was a slight predominance of females.

Sex distribution of study participants
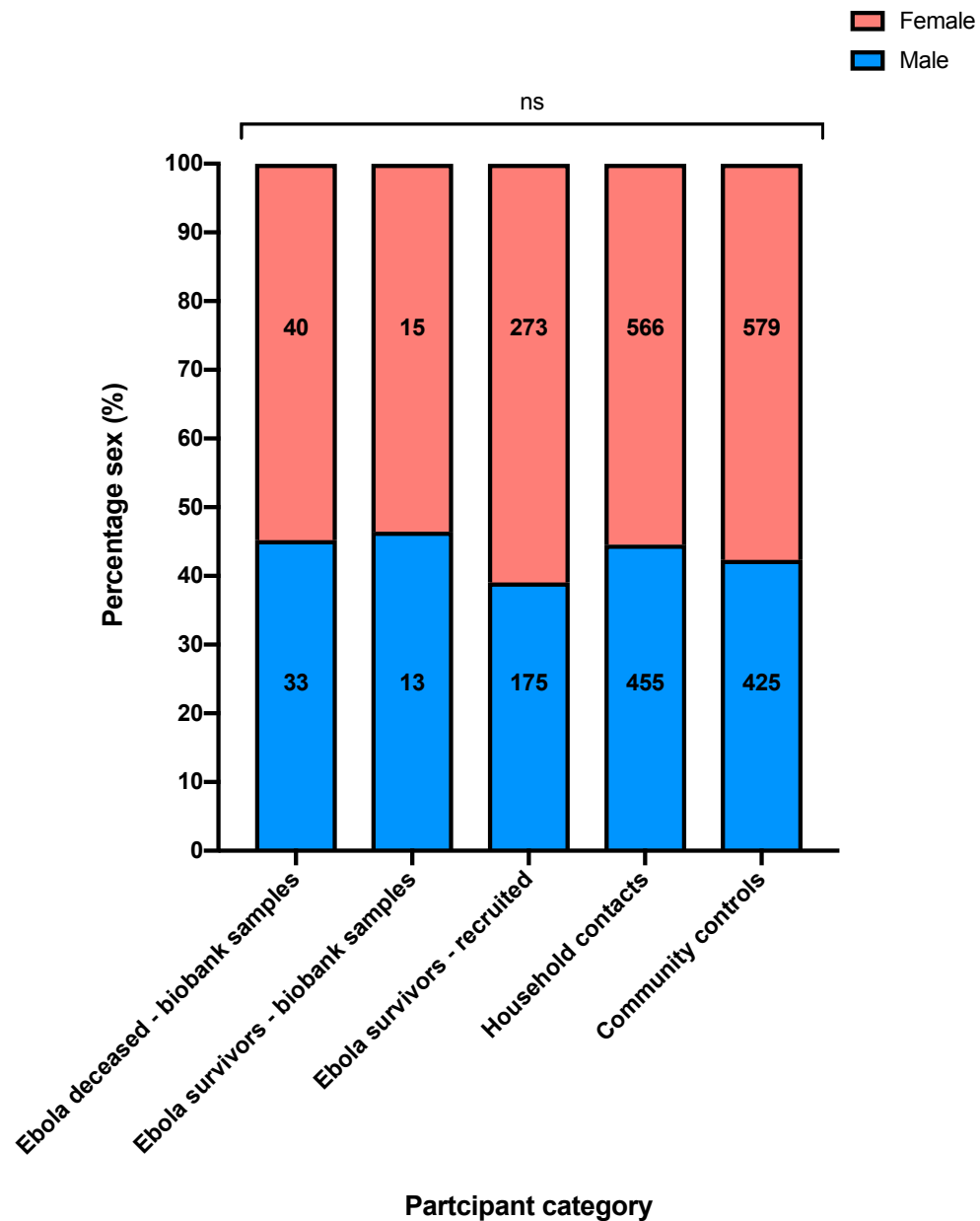
Figure 3-7 Sex distribution across participant categories, with numbers of participants of each sex listed within the bars, ns=not significant.

Ethnicity data was not available for the biobank samples; however, the recruited participants were recruited from similar communities to the locations that the biobank participants derived from. Subsequently the participant samples originated from roughly the same regions of the country. There was little variation in distribution of ethnicity across recruited participant categories (Figure 3-8).

Household Contact - ethnicity distribution

Legend:
- Fula
- Kissi
- Kono
- Krio
- Kuranko
- Limba
- Loko
- Mandingo
- Mende
- Sherbro
- Susu
- Temne
- Temne/Limba
- Temne/Loko
- Missing
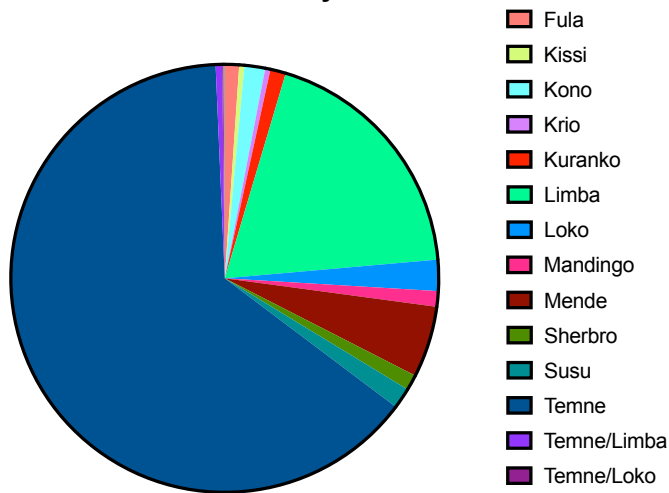
Total=1021

Community Control - ethnicity distribution

Legend:
- Fula
- Kono
- Krio
- Kuranko
- Loko
- Limba
- Mandingo
- Mende
- Sherbro
- Susu
- Temne
- Temne/Mandingo
- Missing
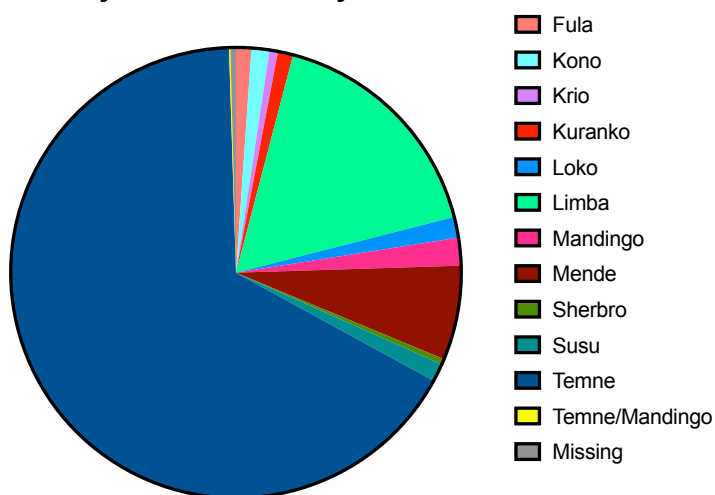
Total=1004

Ebola survivors - ethnicity distribution

Legend:
- Fula
- Kono
- Krio
- Kuranko
- Limba
- Loko
- Mandingo
- Mende
- Sherbro
- Susu
- Temne
- Yalunka
- Mandingo/Kono
- Mandingo/Susu
- Temne/Fula
- Temne/Limba
- Temne/Loko
- Temne/Mandingo
- Temne/Susu
- Missing
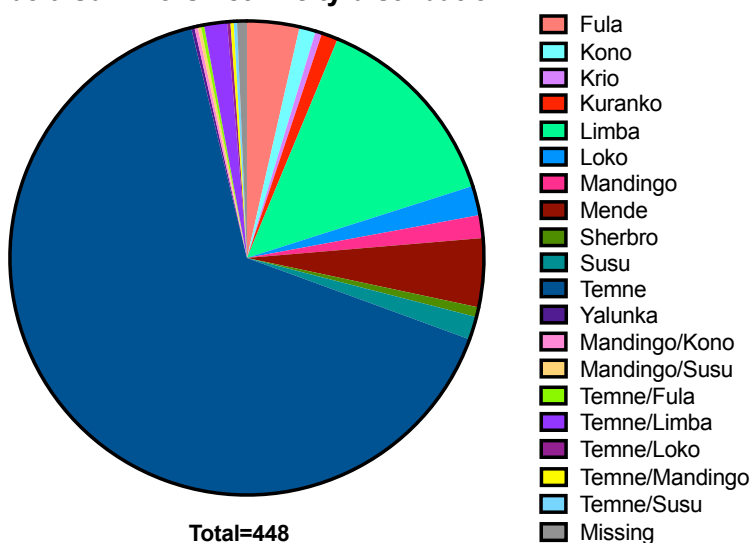
Total=448

Figure 3-8 Ethnicity distribution across recruited participant categories.

3.3.3 Descriptive epidemiology of rural and urban communities

Data was collected on the type of community that participants lived in during the West African Ebola epidemic.

	Recruited participant categories			Biobank participant categories	
	Household contacts (%)	Community controls (%)	Ebola survivors (%)	Ebola deceased (%)	Ebola survivors (%)
Community type					
By participant no.					
Rural	558 *(55)*	567 *(56)*	102 *(23)*	33 *(17)*	3 *(4)*
Urban	463 *(45)*	437 *(44)*	346 *(77)*	160 *(83)*	66 *(96)*
Missing	0	0	0	49	8
By household no.					
Rural	86 *(45)*	105 *(54)*	NA	NA	NA
Urban	104 *(55)*	88 *(46)*	NA	NA	NA

Table 3-5 Community type by participant category (NA = not applicable).

There is minimal variation in the rural-urban distribution of participants in the household contact and community control categories (Table 3-5). The community type for the recruited Ebola survivors was predominantly urban, as was the case for the biobank participants, although the data from the two biobank participant groups may be inaccurate, which is discussed further below. A graphical representation of this data is shown in Figure 3-9. A chi-squared test demonstrated a significant difference in the distribution of rural and urban participants across these five categories, p<0.001.

Rural-Urban distribution of participant categories
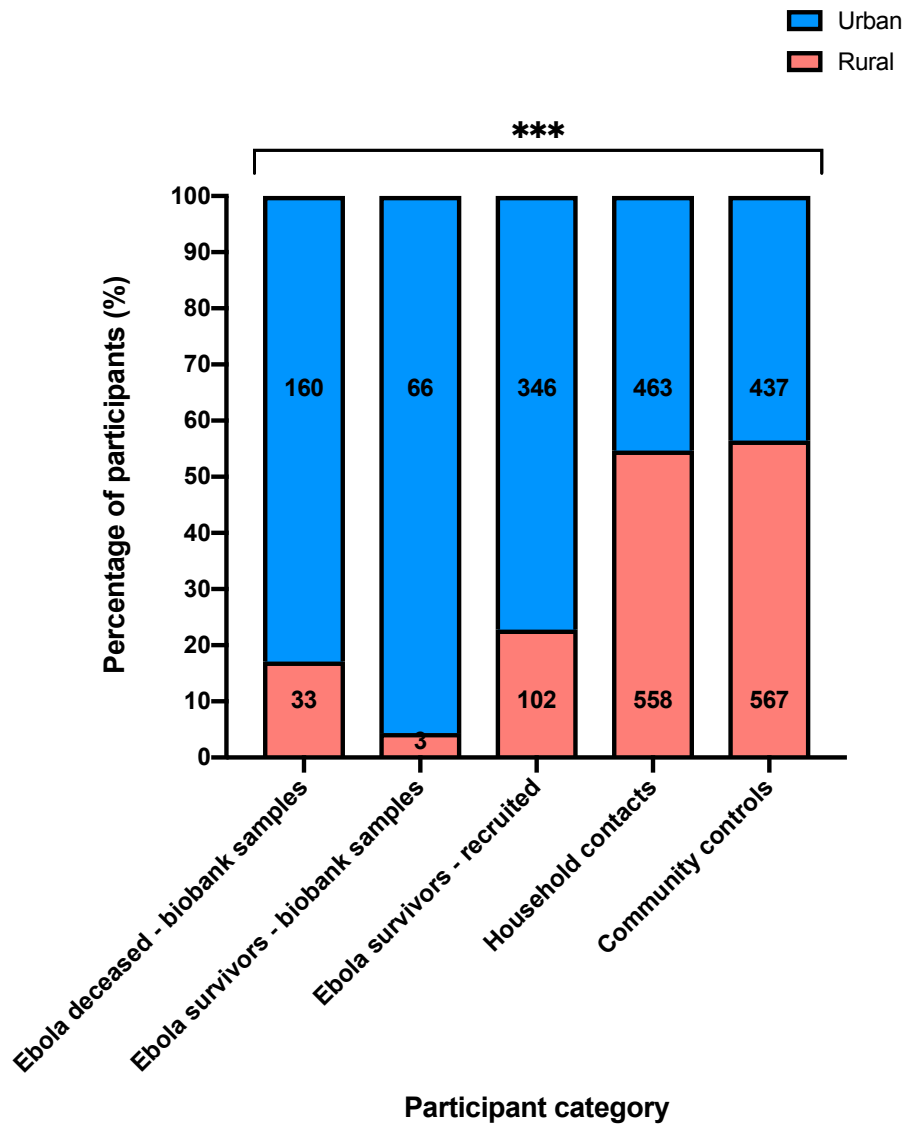
Figure 3-9 Distribution of participants by category and community type, with numbers of participants from each community type listed within the bars. ***significant at p<0.001.

3.3.4 Variations in household circumstances

As recruitment of household contacts and community controls was done by recruitment of entire households, it was important to determine if there were any variations between households which may have altered their chances of being affected by EVD. Household status was assessed on the following parameters:

- Number of rooms in the household
- Size of household at the time of epidemic (children and adults)
- Number of rooms shared with another household

- Number of miscarriages, stillbirths or neonatal deaths

3.3.4.1 Variations in household size

Table 3-6 lists the differences in mean values of parameters that varied within households.

	Household contacts (CI)	Community controls (CI)
Mean no. people/household		
Rural	11.5 (10.2-12.8)**	9.2 (8.1-10.2)**
Urban	9.7 (8.6-10.7)#	10.0 (9.0-11.0)#
Total	10.6 (9.6-11.3)#	9.6 (8.8-10.2)#
Mean no. rooms/household		
Rural	5.5 (4.8-6.0)#	4.7 (4.3-5.0)#
Urban	4.1 (3.6-4.4)#	4.0 (3.6-4.3)#
Total	4.7 (4.3-5.0)#	4.4 (4.0-4.6)#
Mean no. rooms shared/household		
Rural	0.39 (0.21-.056)#	0.47 (0.27-0.64)**
Urban	0.38 (0.23-0.53)#	0.17 (0.05-0.27)**
Total	0.39 (0.27-0.49)#	0.33 (0.21-0.44)#
No. miscarriages/stillbirths/neonatal deaths		
Total	22/310**	7/269**
Total excl EVD +ve mothers	14/302#	7/269#

*Table 3-6 Variations in household size and infant deaths at the time of the EVD epidemic for household contact and community control households. Denominator for 'No. miscarriages/stillbirths/neonatal deaths' is the number of women of childbearing age (13-45 years) in that category #=not significant, *=significant at p<0.05, **=significant at p<0.01, ***=significant at p<0.001*

As can be seen in Table 3-6 there were a greater number of people per household among the household contacts compared with the community controls. This appears to be largely related to household contacts from rural areas, which is more clearly depicted in Figure 3-10 below. While the difference in the number of people per household between the household contacts and community controls was not significant (p=0.087), nor was the difference between rural and urban households in both groups combined (p=0.492), the difference between the number of people per household in rural household contacts was significantly higher than rural community controls (p=0.006). Similarly, while there was little overall variation in the mean number of rooms per household, there was a larger number of rooms per household for household contacts in rural areas. There was little variation in the number of rooms a household shared with another household between rural and urban communities in the household contacts and overall between the household contacts and community controls, but there did appear to be more community control households sharing rooms in rural areas compared with urban areas (p=0.004).
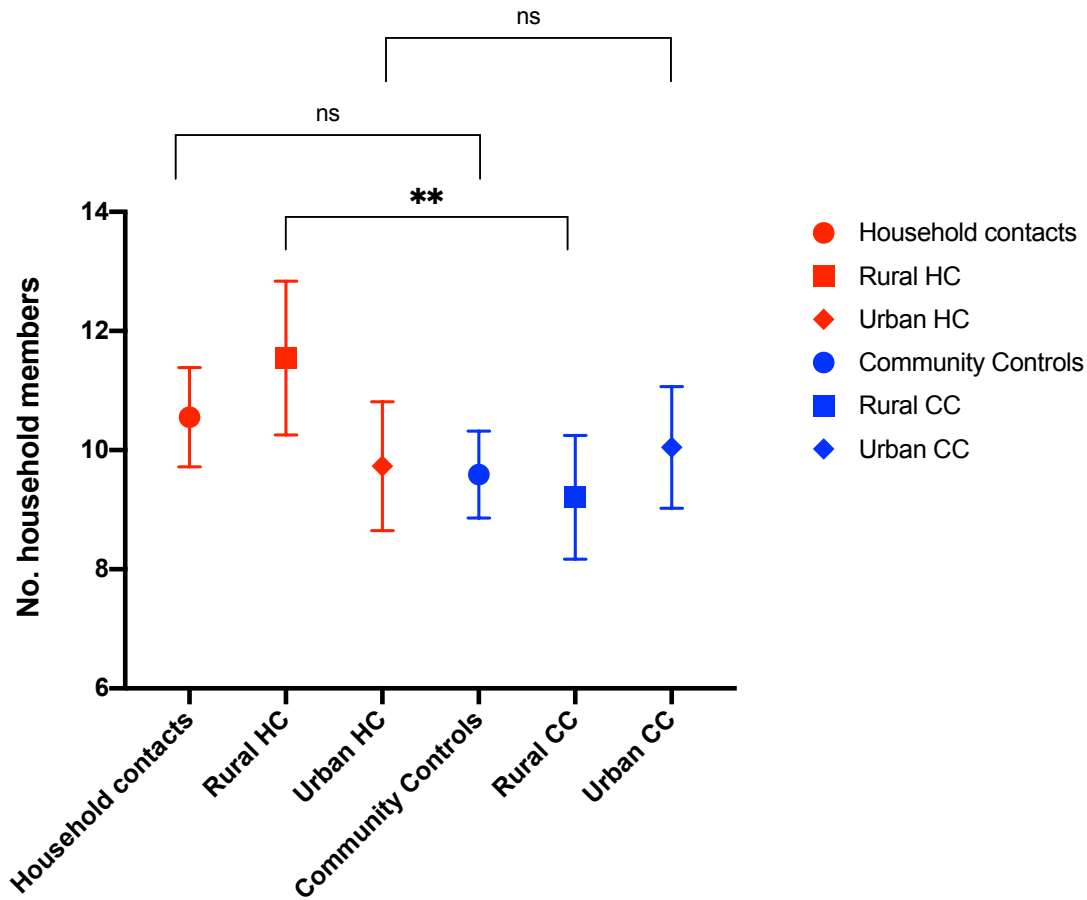
*Figure 3-10 Variations in mean household size (with 95% CI) for household contact and community control households in urban and rural areas. ns = not significant, **significant at p<0.01.*

An interesting variation between household contact and community control households was the number of miscarriages, stillbirths and neonatal deaths in these categories. Even when excluding mothers with known EVD in the categories there was still twice the rate of pregnancy loss and neonatal death in household contacts compared to community controls (Figure 3-11). The odds ratio for a woman of childbearing age (ages 13 to 45) living in a household with an EVD case, suffering a pregnancy loss or neonatal death was 3.0 (p=0.01, 95% CI 1.2-7.1) when compared with women of childbearing age in control households. When excluding mothers with known EVD from this analysis the odds ratio falls to 1.8 but is no longer significant with p=0.19 (95% CI 0.73-4.6).

Miscarriages, stillbirths and neonatal deaths among household contact and community control women of childbearing age (13-45 years)
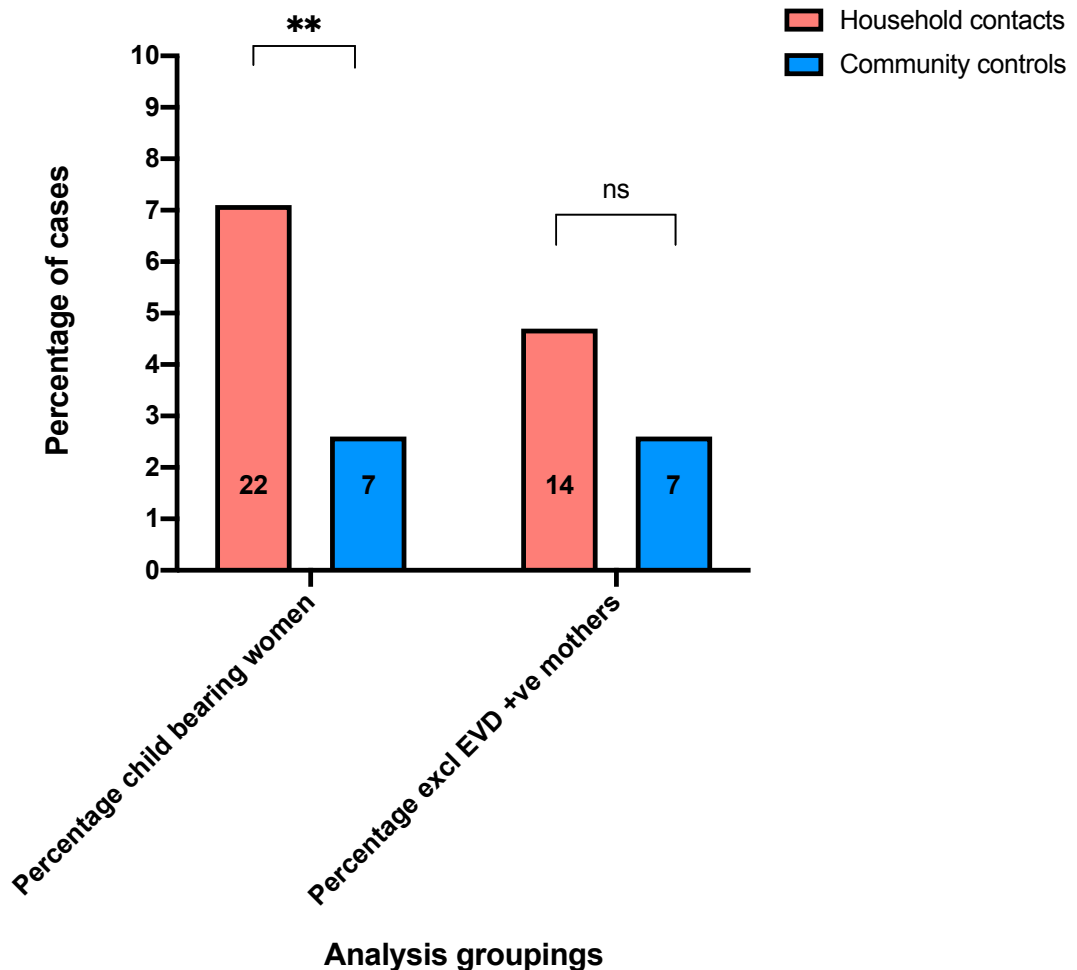
Figure 3-11 Total numbers of miscarriages, stillbirths and neonatal deaths during the West African EVD epidemic in household contacts and community control households. The first set of bars represent the total number of pregnancy losses, the second set excludes those pregnancy losses in women who were known to be EVD infected. **significant at p<0.01, ns = not significant.

3.3.4.2 Variations in access to water and sanitation

While access to water and sanitation did not vary significantly between household contact and community control households, it did vary between rural and urban communities, with rural communities experiencing more limited access to clean water, soap and a personal or communal latrine.

3.3.5 Phenotypic characteristics of acute infection

In an effort to define disease phenotypes for the genomic analysis, data was collected from recruited Ebola survivors on the types of symptoms they experienced during their acute disease. Where available this data was also obtained for the biobank participants from treatment facility records. Regrettably the data stored by different facilities varied and some was not made available. However, below the data that was obtained is described and the similarities and differences between Ebola survivors and Ebola deceased biobank participants highlighted.

3.3.5.1 Clinical symptoms in EVD patients

While data was available for all the recruited Ebola survivors, data was only available for approximately one-third of the biobank Ebola survivor and deceased categories. Due to difficulties accessing very specific symptoms, the symptoms were categorised into three groups, which were deemed to be relatively reliably reported. These were:

1. Wet symptoms (to include any symptoms of vomiting or diarrhoea)
2. Haemorrhagic disease (any form of abnormal bleeding from any orifice)
3. Profuse haemorrhagic disease (abnormal bleeding from three or more bodily sites)

Any participant not experiencing any of the above symptoms was defined as experiencing 'dry symptoms'. The results for both recruited Ebola survivors and biobank survivors (data combined) and biobank deceased patients is displayed below in graphical format (Figure 3-12):

Clinical symptoms in participants infected with EVD
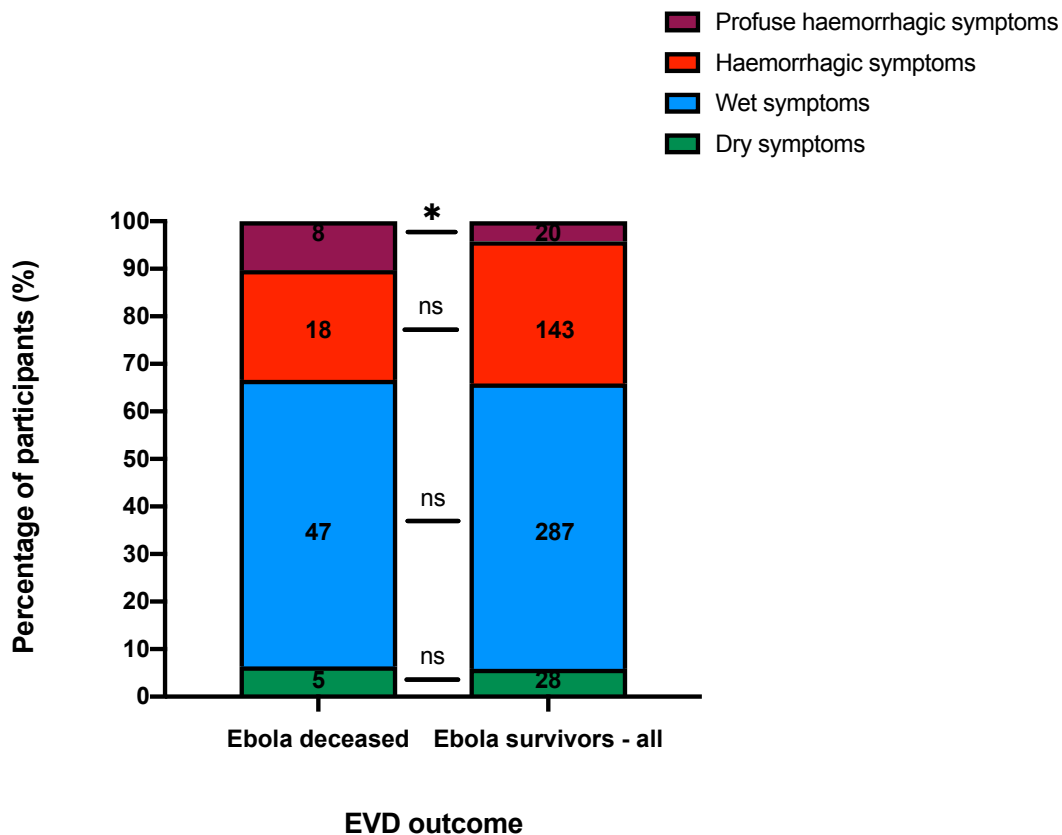
*Figure 3-12 Clinical symptoms in participants infected with EVD with numbers of participants suffering the symptom type listed within the bars. ns = not significant, *significant at p<0.05.*

It is clear from the above figure (Figure 3-12) that there was little difference between groups in terms of experiencing dry and wet symptoms. In excess of 90% of participants infected with Ebola experienced either diarrhoea, vomiting or bleeding, or all of these symptoms. There is not a significant difference in rates of haemorrhagic disease between the deceased and surviving participants (p=0.22), however there is a significantly increased rate of profuse haemorrhagic disease in the deceased participants. The odds of death if profuse haemorrhagic disease was experienced was 2.6 (p=0.028, 95% CI=1.1-6.7).

3.3.5.2 Malaria co-infection

Data on co-morbid conditions was not available for the biobank participants. Reporting of previous medical problems was included on the survivor questionnaire for the prospectively recruited Ebola survivors, but revealed little significant information due to limitations in healthcare provision in Sierra Leone. Data was however available for the biobank participants on malaria co-infection during their acute EVD disease. This information is presented below (Figure 3-13) for biobank deceased and

surviving participants and demonstrates that significantly more deceased participants had malaria co-infection than surviving participants (OR=9.0, p=0.003, 95% CI=2.1-38.9). The age range for those with malaria was from 1 year to 73 years, with half of participants with malaria being 12 years or under, nine participants were under 5 years. The two surviving participants were aged 12 and 39 years.

*Figure 3-13 Malaria co-infection in Ebola survivors and deceased participants in the biobank participant categories, with numbers of participants suffering from malaria listed within the bars. **significant at p<0.01.*

3.3.6 Ebola survivors and PES

At the time of conducting this study it was unclear what the prevalence of Ebola survivors with ongoing symptoms and Post-Ebola Syndrome (PES) was. In order to determine if there are genetic differences between Ebola survivors who have ongoing problems versus those who don't, this was included within the recruitment questionnaire for Ebola survivors as shown earlier in *Figure 3-5*.

3.3.6.1 Prevalence of PES among recruited Ebola survivors

Due to concerns regarding reporting reliability of Ebola survivors, symptoms were stratified by severity and for preliminary analysis divided into three groups: no PES, PES, and severe PES. Severe PES was defined as frequent symptoms with life limiting effects. These symptoms are listed in Table 3-7. The pie chart below (Figure 3-14) demonstrates the prevalence of these categories among the Ebola survivors that were recruited, with 68% of Ebola survivors suffering from PES or severe PES.

Organ system	Sequelae included within the 'Severe PES' category
Eye	Cataracts/blindness
	Severe uveitis (treatment with eye drops >3 months)
Neurological	Complex neurological sequelae (eg. Paralysis/new onset seizure disorder)
	Persistent sleep disturbance (>4 nights/week)
	Recurrent headaches (>2 days/week)
Musculoskeletal	Frequent, debilitating joint pain (>3 days/week)
Reproductive	Impotence
	Difficulty conceiving

Table 3-7 Sequelae included within the 'Severe PES' category for recruited Ebola survivors.

Ebola survivors - prevalence of PES
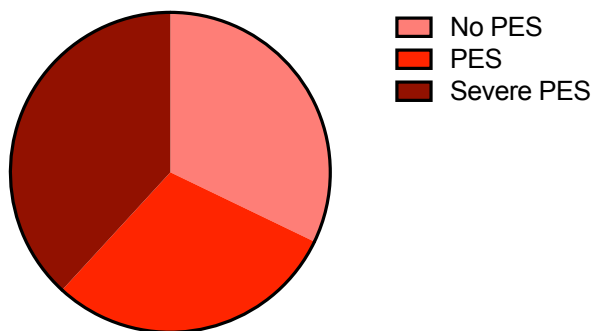
Total=448

Figure 3-14 Prevalence of PES and severe PES among recruited Ebola survivors.

The bar chart below (Figure 3-15) shows the prevalence of specific individual symptoms broken down by severity. A significant proportion of survivors with PES experienced more than one symptom, with 39% of those with PES experiencing two or more symptoms.
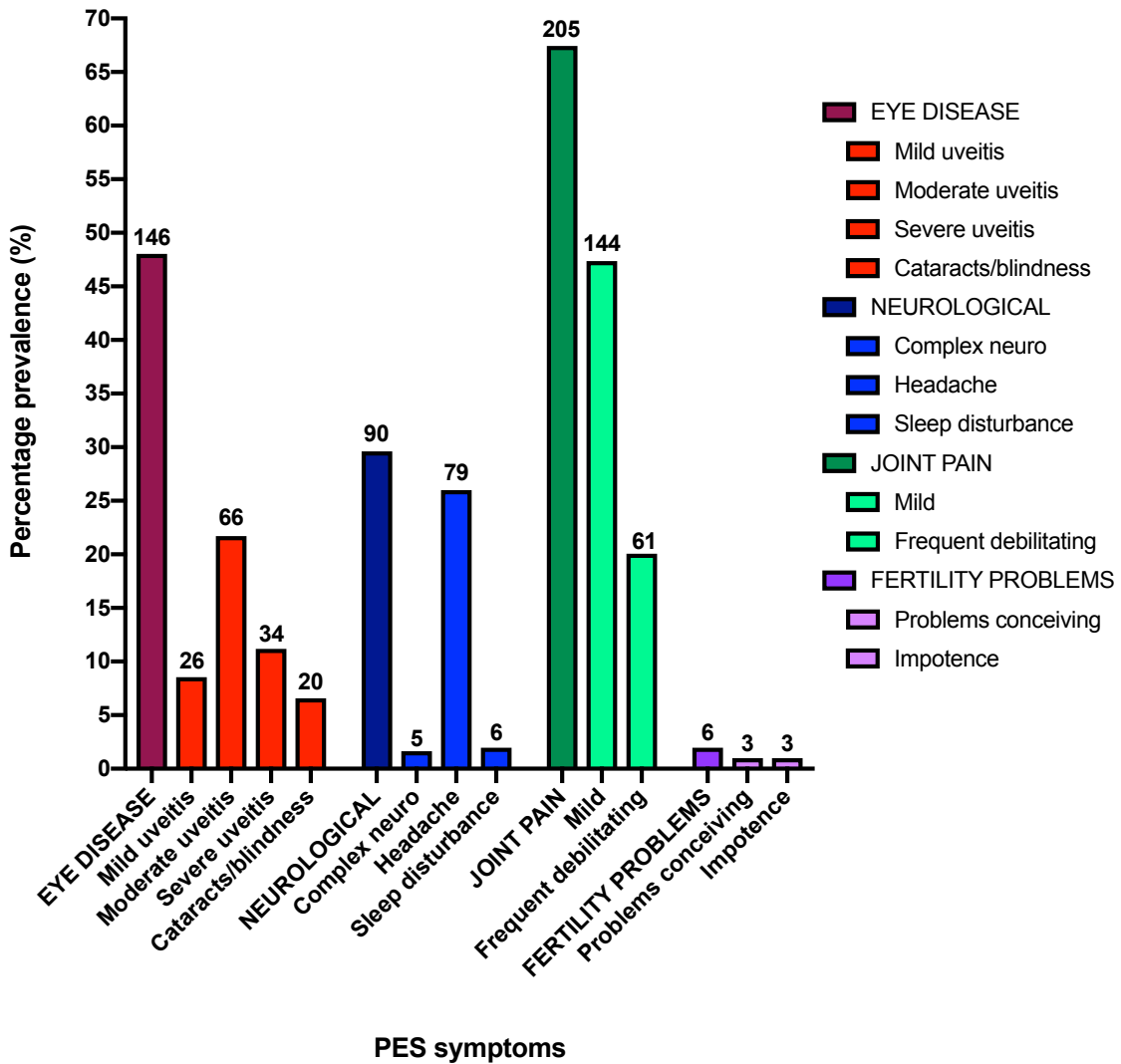
Figure 3-15 Prevalence of symptoms among recruited Ebola survivors with PES, with numbers of participants suffering the symptom type listed above the bars.

3.3.6.2 Variations in prevalence of PES

The prevalence of PES and severe PES among recruited Ebola survivors varied with age, sex and ethnicity (severe PES only) as shown in Table 3-8, Table 3-9 and Table 3-10 below.

		Any PES (mild/severe) (n=304)	No PES (n=144)
Mean age in years (standard deviation)		31.71 (12.98)	24.56 (12.42)
Male sex (%)		104 (34.2)	71 (49.3)
Urban community (%)		235 (77.3)	111 (77.1)
Ethnicity (%)	*Temne (n=294)*	190 (62.5)	104 (72.2)
	Limba (n=62)	47 (15.5)	15 (10.4)
	Other/mixed (n=92)	67 (22.0)	25 (17.4)

Table 3-8 Descriptive statistics comparing Ebola survivors with mild/severe PES and Ebola survivors with no PES

		Severe PES (n=173)	No severe PES (n=275)
Mean age in years (standard deviation)		31.99 (13.60)	27.79 (12.73)
Male sex (%)		56 (32.4)	119 (43.3)
Urban community (%)		130 (75.1)	216 (78.5)
Ethnicity (%)	*Temne (n=294)*	97 (56.1)	197 (71.6)
	Limba (n=62)	30 (17.3)	32 (11.6)
	Other/mixed (n=92)	46 (26.6)	46 (16.7)

Table 3-9 Descriptive statistics comparing Ebola survivors with severe PES and Ebola survivors without severe PES

	Variable		Unadjusted Odds ratio (95% CI)	p-value	Adjusted Odds ratio (95% CI)	p-value
PES						
	Age (per 10 years)		1.59 (1.34-1.90)	<0.001	1.61 (1.34-1.92)	<0.001
	Female sex		1.87 (1.25-2.80)	0.002	1.80 (1.18-2.74)	0.006
	Rural community		0.99 (0.62-1.58)	0.959	0.87 (0.52-1.46)	0.599
	Ethnicity	*Limba*	1.72 (0.91-3.22)	0.092	1.90 (0.98-3.66)	0.056
		Other/mixed	1.47 (0.87-2.46)	0.147	1.32 (0.76-2.30)	0.322
Severe PES						
	Age (per 10 years)		1.27 (1.10-1.48)	0.001	1.26 (1.08-1.46)	0.003
	Female sex		1.59 (1.07-2.37)	0.022	1.53 (1.01-2.30)	0.043
	Rural community		1.21 (0.77-1.90)	0.404	1.13 (0.70-1.85)	0.597
	Ethnicity	*Limba*	1.90 (1.09-3.31)	0.023	1.89 (1.07-3.34)	0.028
		Other/mixed	2.03 (1.26-3.27)	0.003	2.09 (1.26-3.44)	0.004

Table 3-10 Analysis of variables associated with mild/severe PES and severe PES alone using logistic regression.

The prevalence of PES among recruited Ebola survivors increased with age (Figure 3-16), with just under half of child and teenage survivors describing ongoing symptoms, versus over 70% of adult survivors. The odds ratio for PES in adult survivors when compared with child survivors was 3.0 (p<0.001, 95% CI=1.9-4.7) and when using logistic regression the odds of PES increased by 61% for each 10 years of age gained (OR 1.61, CI 1.34-1.92, P<0.001). The proportion of those with PES that were defined as

severe PES varied from 50 to 70% among age groups, and the prevalence of severe PES also increased with age (Figure 3-16). The odds ratio for severe PES among adult survivors compared with child survivors was 1.6 (p=0.04, 95% CI=1.03-2.6) and when using logistic regression the odds of severe PES increased by 26% for each 10 years of age gained (OR 1.26, CI 1.08-1.46, P=0.003).
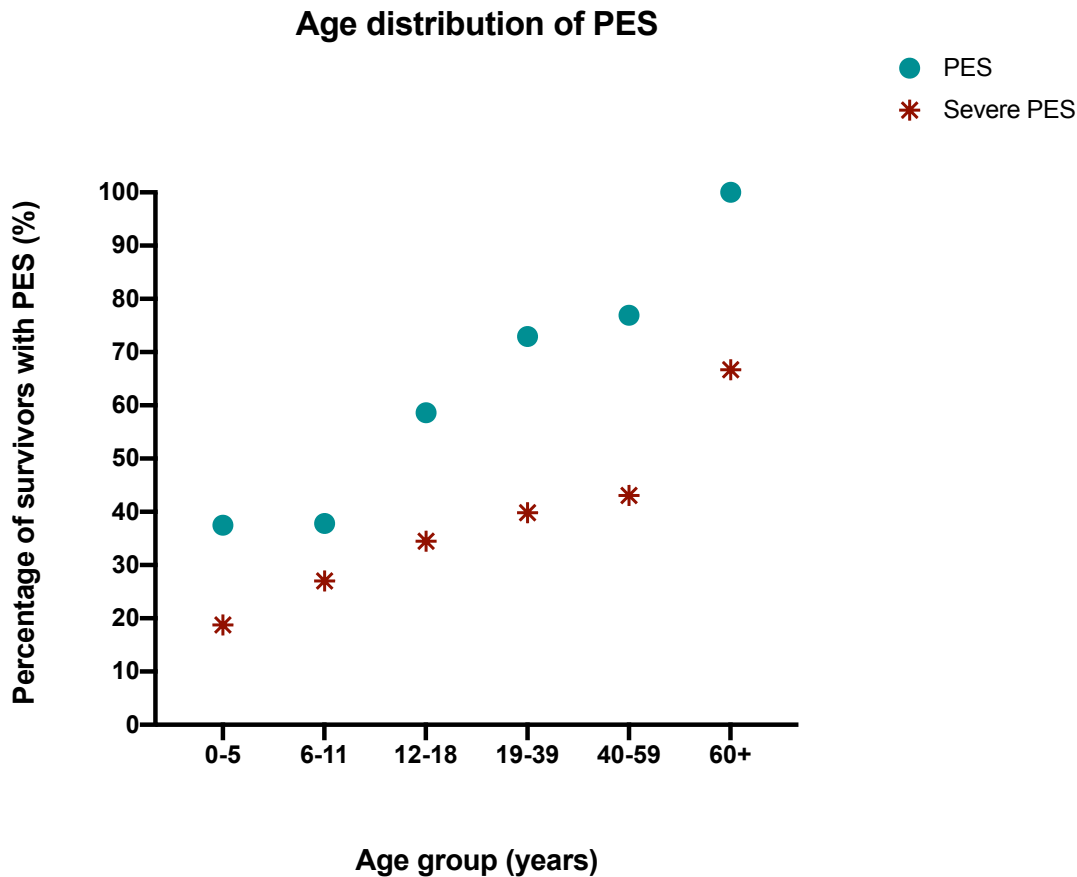
Figure 3-16 Percentage prevalence of PES and severe PES by age group.

Sex also demonstrated variation in prevalence of PES and severe PES with a higher proportion of women experienced both PES (OR 1.80, CI 1.18-2.74, P=0.006) and severe PES (OR 1.53, CI 1.01-2.30, P=0.043) than men. Interestingly the prevalence of severe PES varied with ethnicity, with both the Limba ethnicity (OR 1.89, CI 1.07-3.34, p=0.028) and all other 'minor' ethnicities (OR 2.09, CI 1.26-3.44, p=0.004) experiencing a higher prevalence of PES than the Temne ethnicity. There was no variation in the prevalence of PES by community type with 68% of recruited survivors in both types of community reporting symptoms of PES. There was however a slightly higher proportion of recruited survivors from rural areas who described symptoms of severe PES when compared with recruited survivors from urban areas (41.2% vs 37.3%), although this was not statistically significant (p=0.48).

3.3.6.3 PES resolution

The majority of recruited survivors reported that their symptoms were improving, with over 85% reporting they had resolved or were improving (Figure 3-17). Only one survivor reported their symptoms as worsening, which was an older survivor with worsening sight in an eye that had already been diagnosed as 'blind'. For those survivors who reported their symptoms as 'resolved' the symptom duration had been anywhere from 2 weeks to 6 months, however most were unable to recall their symptom duration.
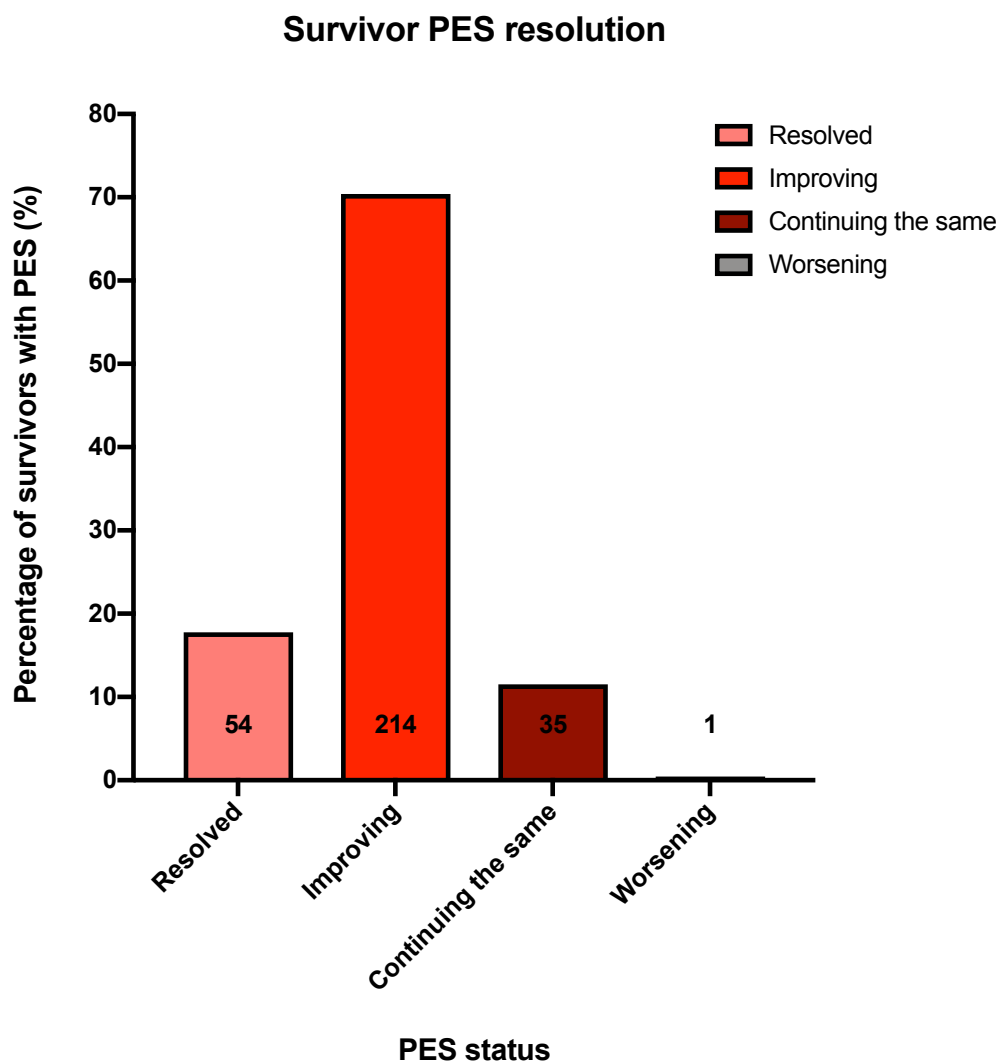
Figure 3-17 Recruited survivors reporting of the status of their PES symptoms, with numbers of participants within each group listed within/above the bars.

3.3.6.4 Ebola virus persistence in survivor's bodily fluids

Male survivors over the age of 16 years of age were offered semen testing through government programmes to determine if Ebola virus continued to be present in their semen following their survival

from Ebola[64]. The time period survivors reported being offered semen testing varied from 12 to 24 months after the onset of symptoms of the acute disease process. Out of 141 males aged over 16 years in the recruited Ebola survivor category, 86.5% underwent virus persistence testing on at least one occasion. Table 3-11, below, indicates the results of the semen testing.

		Number	Percentage (%)
Semen tested			
	Yes	122	86.5
	No	19	13.5
Semen testing result			% of those tested
	Positive	6	4.9
	Indeterminate	1	0.82
	Negative	98	80.3
	Did not receive result	17	13.9
Months semen remained positive for			
	15	1	0.8
	16	1	0.8
	17	0	0
	18	1	0.8
	19	3	2.5

Table 3-11 Results of semen testing for Ebola virus in recruited male survivors over 16 years of age.

Concerningly, it appears that 2.5% of the Ebola survivors who were tested continued to have Ebola virus present in their semen up until 19 months following the onset of symptoms of their acute EVD. All recruited survivors were asked if they had been included in any other testing of bodily fluid for Ebola virus, such as intra-ocular fluid, tears, cerebro-spinal fluid or vaginal secretions. None of the survivors recruited had been enrolled in any other virus persistence screening programme.

3.4 Discussion

3.4.1 Summary of results

3.4.1.1 Descriptive features of the groups

While it was not possible to directly control for age, sex and ethnicity across the participant categories, it is reassuring that there were limited differences between these demographics across the groups. The interpretation of any differences in sex distribution in the biobank participant categories is difficult due to there being a significant amount of missing data, however the limited data available shows a slight female preponderance similar to that of the other participant categories. The distribution of ages was consistent in the household contact and community control categories, with a preponderance of children under the age of 18 years. In regard to the survivor groups the age distribution changes, with

a higher proportion of participants in the young to middle age adult range, this distribution changes yet again in the deceased biobank participant category with a greater array across all age groups. There was minimal variation in ethnicity across the recruited participant categories, however this data was not available for the biobank participants.

3.4.1.2 Descriptive epidemiology of rural and urban communities

There is little variation in participant numbers from rural and urban communities in the household contacts and community controls, however there was a predominance of recruited Ebola survivors from urban areas. The data reported for the biobank participant categories is based on that collected by the Ebola treatment facilities at the time of patient admission, hence a significant amount is missing. However, similar to the recruited survivors an urban predominance is also demonstrated.

There were differences in the numbers of households recruited in rural and urban areas. In the household contacts, more *households* were recruited in urban areas, but more *participants* were recruited in rural areas due to greater household size. This was not as clearly demonstrated in the community controls, despite household contact and community control households being recruited from the same communities. It is apparent from the data that the household size in household contacts was generally larger than that of community controls, this appeared to be predominantly due to household size in rural areas.

There was no variation between household contact and community control households in relation to access to water and sanitation, although overall rural communities had more limited access to clean water, soap and a personal or communal latrine. There was however significant variation in pregnancy loss and neonatal death between household contact and community control households, with household contact households experiencing twice the rate of pregnancy loss and neonatal death than community control households (p=0.01). This difference was maintained even when EVD positive mothers were excluded from the analysis, however the result was no longer significant (p=0.19).

3.4.1.3 Phenotypic characteristics in acute infection

While it was difficult to obtain clinical data from the biobank participant samples stored during the epidemic, it was possible to ascertain certain clinical syndromes. It is clear that deceased participants were more likely to have profuse haemorrhagic symptoms (OR=2.8, p=0.018), and that malaria co-infection is detrimental in the context of acute EVD with an odds ratio of 9.0 for a fatal outcome (p=0.003, 95% CI=2.1-38.9).

3.4.1.4 Ebola survivors and PES

The rates of survivors with Post-Ebola syndrome were higher than initially anticipated, with over two-thirds of survivors suffering from ongoing problems. The range of symptoms experienced by survivors generally fit into four organ systems, namely eye disease, joint disease, neurological disease or reproductive disease. The predominant complaints were of eye or joint disease, with a lower prevalence of neurological and fertility problems. The most severe forms of post-Ebola symptoms are those which result in ongoing life-limiting problems, and include loss of eye sight, joint pain limiting work possibilities or activities of daily living, neurological sequelae resulting in chronic headaches, persistent sleep disturbance, seizure disorders and reduced mobility or inability to reproduce. Reassuringly however over 85% of survivors described these symptoms as either improving or resolved at approximately 2 years after their acute disease.

PES varied with age with adults more likely to suffer from PES than children and a significant positive correlation between age and prevalence of both PES ($p < 0.001$) and severe PES ($p < 0.01$). There was also an increased prevalence of PES in women compared with men ($p < 0.01$) and an increased prevalence of severe PES among non-Temne ethnicities ($p < 0.05$). A small proportion (2.5%) of male Ebola survivors over the age of 16 had Ebola virus persistence in their semen up to 19 months after the onset of symptoms for their acute disease.

3.4.2 Interpretation of results

3.4.2.1 Descriptive features of the groups

The similarity in sex distribution between household contacts and community controls likely reflects that there was a slight preponderance of women affected by EVD in West Africa, approximately in the ratio of 60% female to 40% male[42]. This is reflected in the deceased and both survivor categories in this study. It is also likely that there was a slight preponderance of women recruited as women were more willing to engage in the study than men. In families divided over their decision to participate in the study, it was often the female members of the family who participated when their male relatives refused. While it may also be assumed that the men were away from home at their farms during recruitment, this was not actually what was experienced. In rural areas the women may equally have been away on farms or selling at the market on market day, and in urban areas there were also a large number of female led households where there was a mother and her children, but the father was not present.

The distribution of ages was consistent in the household contact and community control categories, with a preponderance of children under the age of 18 years. This is representative of the population demographic that is observed in Sierra Leone. In the survivor groups the age distribution changes with a predominance of young to middle age adults, which is probably representative of the distribution of those more likely to be infected with EVD (the young to middle age adult population)[6, 7, 24, 134] and the age groups which had the highest mortality rates from EVD (children under the age of 5 years and adults over the age of 60 years)[7, 22]. This also explains why the distribution in the deceased category is more spread across the age groups, because while more of the young to middle age adult population were infected with EVD resulting in a higher number of deceased patients in these age groups, the mortality rates were higher in the extremes of age and so proportionally there are higher numbers of participants from these age groups in the deceased group when compared with the survivor groups.

3.4.2.2 Descriptive epidemiology of rural and urban communities

The reason for the collection of data in rural and urban communities was for the purposes of understanding transmission dynamics in communities when comparing household contacts and community controls and understanding variations in seropositivity (discussed in chapter 4). That there is little variation in participant numbers from rural and urban communities is consistent with the recruitment strategy used. It is interesting that there was a predominance of recruited Ebola survivors from urban areas. This may reflect the recruitment strategy, as for survivor recruitment the local survivor committee's meeting location was utilised, these tended to be in urban centres. This meant longer distances for survivors from rural areas to travel, but also logistical challenges in communicating with survivors from more rural areas, which may have resulted in them being under-represented in the study. This urban distribution, however, may also represent what was seen during the epidemic, which was that due to higher population density in urban areas, once the disease entered an urban setting it tended to spread relatively rapidly. As such it was likely to infect more people within a given area than in rural areas where the population density is not as high[24, 122].

The data reported for the biobank participant categories is based on that collected by the Ebola treatment facilities at the time of admission, due to the extenuating circumstances at the time a significant amount is missing. The data that is reported may also be less accurate, however the urban predominance seen in the recruited Ebola survivor category is also demonstrated here, and as mentioned above, to some extent may represent what was witnessed during the epidemic in terms of disease spread in urban areas.

Another reason for assessing the distribution of rural and urban areas was to determine if there were differences in household circumstances within communities. As can be seen there were differences in the numbers of households recruited in rural and urban areas. In the household contacts, more *households* were recruited in urban areas, but more *participants* in rural areas. This is representative of the size of households in rural areas, where some households had upwards of twenty people living within one building, and consistent with the finding that the household size in household contact households was larger than that of community control households, which was predominantly related to household size in rural areas. In urban areas the household sizes were smaller. This difference in household size was not as clearly demonstrated in the community controls, with similar proportions of participants from households in rural and urban areas. This suggests that in rural areas where the population density is not as great as in urban areas, household size was related to whether a household was affected by EVD or not. It would make sense that larger households in rural areas were more at risk of having diseased individuals in their household simply due to the fact that there were on average more people in the household[122]. The presence of larger numbers of people would increase the chance that a household member may be exposed to or become infected with EVD. However, it may also have to do with the density of people within the household and thus an increased risk of transmission within the household. As described above, there appear to be more confirmed EVD cases from urban areas, which is likely to be related to population density in urban areas compared with rural areas. Potentially the population density of a household in rural areas was therefore also a risk factor for a household being affected.

There was no variation between household contact and community control households in relation to access to water and sanitation, although overall rural communities had more limited access to clean water, soap and a personal or communal latrine. This is in large part due to a lack of running water and soap but is also related to higher rates of poverty in rural areas, where a small fee must often be paid to access water from the bore hole. The variation in pregnancy loss and neonatal death between household contact and community control households is interesting to note. Particularly when infants of EVD positive mothers were excluded from the counts. It could be suggestive of possible subclinical infection occurring in pregnant women in affected households resulting in pregnancy loss or neonatal death of an infected infant, but without any other apparent symptoms in the mother.

3.4.2.3 Phenotypic characteristics in acute infection
While it was difficult to obtain clinical data from the biobank participant samples stored during the epidemic, it was possible to ascertain certain clinical syndromes in order to compare groups, but also

to be able to differentiate participants by disease phenotype for genetic analysis purposes. It is clear that deceased participants were more likely to have profuse haemorrhagic symptoms (OR=2.8, p=0.018). This makes sense, as in an environment where intensive care and blood products to address coagulopathy are not readily available, a profusely haemorrhagic disease phenotype is more likely to be fatal. Interestingly though, the overall number of participants where haemorrhagic disease was reported is somewhat higher than the average previously reported from clinical data during the epidemic[5-7, 21, 22, 39], which was in the region of 10%. This figure was low compared to what was observed at some treatment facilities during the epidemic and the figures of approximately 30% described here are more representative of what some organisations operating treatment facilities reported[18]. It is likely that milder haemorrhagic symptoms were under-reported during the epidemic, as data collection was very challenging and symptoms such as melaena may not have been reported unless a relatively detailed history was obtained from the patient, which was not always possible.

The data on malaria co-infection is very interesting. It would seem obvious that malaria co-infection with Ebola would likely pre-dispose to a worse outcome, however a paper published in 2016 suggested that malaria appeared protective in the context of Ebola[47]. This has since been refuted by a further publication from the West African epidemic[49]. The data described here certainly lends itself to the latter suggestion that malaria co-infection is detrimental in the context of acute EVD with an odds ratio of 9.0 for a fatal outcome in the presence of malaria co-infection (p=0.003, 95% CI=2.1-38.9). While the possibility of re-activation of an asymptomatic parasitaemia in adults with severe Ebola should be considered, given that over half of the participants that had malaria and succumbed to their malaria/EVD co-infection were under the age of 12, it would seem unlikely that this accounts for the outcomes observed in this dataset.

3.4.2.4 Ebola survivors and PES

The rates of survivors with Post-Ebola symptoms are higher than initially anticipated, but not unexpected as further data has emerged on the prevalence of PES in Ebola survivors[60]. With over two-thirds of survivors suffering from ongoing problems, this suggests a significant problem which needs to be addressed during and following any Ebola epidemic. While the predominant complaints were of eye or joint disease, the impact of neurological and fertility problems should not be downplayed due to their lower prevalence, as they have a significant impact on an individual's wellbeing. The most severe forms of post-Ebola problems are those which result in ongoing life-limiting problems, such as loss of eye sight, joint pain limiting work possibilities and significant neurological sequelae. While it is reassuring that the majority of survivors described their symptoms as improving or resolved approximately 2 years after their acute disease, there is some evidence that access to prompt

treatment of sequelae is likely to limit more severe forms of PES developing, particularly in relation to eye disease and loss of sight[53, 57]. It is therefore imperative that clinics are established to monitor and provide treatment to survivors with PES from the beginning of an epidemic.

Variations in PES with age and duration of hospital admission are interesting, but not unexpected. It is known that children tend to recover more quickly from viral infections with less associated problems[135]. In the case of EVD it is known that the extremes of age are more likely to have a fatal outcome, which in part accounts for the low numbers of survivors recruited in the under 5 year age group and the over 60 year age group. However, the data is clear, those children and infants who do survive are less likely to have persisting problems when compared with their older counter parts. The increased prevalence of PES in women compared with men may warrant further exploration, although may be related to women being more willing to participate in research studies or more forthcoming with information related to their health. There is the possibility however that it could relate to the underlying disease process or the initial viral exposure dose, but it is not within the remit of this study to investigate this. The increased prevalence of severe PES among non-Temne ethnicities is noteworthy and may be important to consider during the genomic analysis.

It is concerning that 2.5% of the male Ebola survivors over the age of 16 had Ebola virus persistence in their semen up to 19 months after the onset of symptoms for their acute disease. While it has previously been debated whether live virus capable of causing infection is actually what is identified in these tests, it has now been demonstrated that sexual transmission can occur[62, 66, 67] and has occurred even at 18 months after symptom onset of EVD[67]. These results shed further light, alongside the growing body of evidence, that sexually active male Ebola survivors need to be advised not to engage in unprotected sexual intercourse until they have had a minimum of two negative PCR tests for Ebola virus in semen, separated by at least one month[59, 63]. Ideally this should be three tests, as there is now evidence for variable expression of virus in semen, with the possibility of a negative test one month and a positive test the next[60, 61].

Although transmission is likely rare, given the severe nature and high fatality rates of EVD it is a public health imperative that sexual transmission be prevented. However, this must be balanced with what is practically achievable given the numerous challenges of trying to follow up male Ebola survivors in the middle of an epidemic in regions with limited health infrastructure. While clear advice regarding using condoms may be provided at discharge from treatment facilities, the challenges of compliance with this have clearly been demonstrated through the decades of sexual health promotion in relation to the

HIV/AIDS pandemic[136, 137]. While many male survivors may intend to comply with advice and testing schedules, unless these testing facilities are discrete and easily accessible, and condoms are readily provided free of charge, compliance is likely to be limited, particularly in rural areas. It is also very important that any testing service does not add additional public stigma to an already stigmatised Ebola survivor population[64].

3.4.3 Limitations

The main limitations of the aspects of the study contained within this chapter pertain to missing data, largely in relation to the biobank participants, and the manner with which survivor participants were recruited to the study.

3.4.3.1 Missing data

There is a lot of missing data on sex for the biobank participant categories, as well as participant community of origin and clinical details. Unfortunately specific information was not consistently reported on the lab sample submission forms and could not always be traced from the treatment facility datasets. Despite this there are similar male to female ratios reported for the biobank participants as for the recruited Ebola survivors, and similarly while the community data on the biobank participants may be inaccurate, it still followed a similar patter to that seen among the recruited survivors, with an urban community predominance. Although some clinical data was missing for the biobank participants, it was still possible to categorise most participants in terms of disease severity and thereby demonstrate differences between those who died and those who survived in relation to symptom severity.

3.4.3.2 Participant recruitment

The survivor recruitment was conducted at either the meeting location or the survivor clinic of the local Ebola Survivor's Association. These venues were located in urban centres, which may have resulted in a bias of recruitment of survivors from urban regions and an under-representation of survivors from rural areas. This may relate to messages pertaining to recruitment to studies not reaching survivors in rural areas and also the logistical challenge of survivors in rural areas travelling to the urban centre for the purposes of recruitment. However, in order to reach survivors in rural communities, some rural communities were visited and sensitised accordingly, and the survivor communication network that was in place through the survivors' association and community Ebola survivor representatives was utilised. A higher transport supplement was also paid to those participants travelling from rural areas.

Similarly, it is possible that women were over-represented in all recruited participant categories due to a greater willingness to engage with research studies than their male counterparts from the same communities. This was witnessed both in rural communities and when recruiting Ebola survivors. There may also have been a greater willingness of women to express health related concerns when recruiting Ebola survivors resulting in an under-representation of men in the survivor PES category. However, given that the male to female ratios in all participant categories, including the biobank participant categories, were quite similar it is reassuring that there was not a significant recruitment bias.

3.4.4 Wider interpretation of results

It is reassuring, particularly from the perspective of the genetic analysis that there was minimal variation in ethnicity across the recruited participant categories. This indicates that the recruitment strategy of controlling for ethnicity by matching by community was relatively successful. It also suggests that the ethnicity distribution of the two biobank participant categories is likely similar to the recruited participant categories given the majority of the biobank participants came from the same region. This has implications for the genomic analysis that is described in later chapters.

With over two-thirds of survivors suffering from ongoing problems, this suggests a significant problem which needs to be addressed during and following any Ebola epidemic. The evidence that access to prompt treatment may also limit more severe forms of PES developing, particularly in relation to eye disease and loss of sight[53, 57], makes it essential that Ebola survivor clinics are established to monitor and provide treatment to Ebola survivors with ongoing problems from the beginning of an epidemic. Similarly given that there is now proven sexual transmission of Ebola even up to 18 months following surviving the acute infection, it is imperative that screening of semen from sexually active male Ebola survivors be facilitated from the start of any epidemic in order to prevent sexually transmitted cases of EBOV infection.

The variations that the above data have demonstrated in relation to outcome among EVD survivors is not only interesting from a practical, clinical perspective, but also lends further evidence to the case for host genetic variations in ability to contain and eliminate the virus. Further dividing the recruited survivors into categories based on severity of ongoing problems helps to further refine the disease phenotypes for the genetic analysis and provides additional sub-categories for comparison within the exome sequencing analysis.

Chapter 4 Ebola IgG seropositivity among affected communities

4.1 Introduction

This chapter describes the results of Ebola IgG antibody studies conducted on the recruited household contacts and community controls. As antibody levels are an indication of previous infection, they provide a tool to explore how often symptomatic and asymptomatic infection occurred in household contacts of EVD patients and within affected communities. Seropositivity (the presence of antibodies in serum, although in this case used synonymously for those with antibodies present in oral fluid) has previously been shown to be related to EBOV exposure level[120, 122], therefore this chapter first describes the use of exposure level to stratify household contacts in terms of presumed quantity of exposure, and also to describe Ebola survivor's possible modes of infection. It then relates the exposure level to IgG seropositivity for household contacts and community controls. IgG seropositivity and antibody levels for Ebola survivors are described in chapter 5.

The EBOV exposure level was determined using an exposure risk stratification which is fully described in chapter three and demonstrated in Figure 3-2. In summary the risk stratification describes eight different risk levels, with risk level 1 referring to the highest level of risk of exposure to Ebola virus and risk level 8 referring to the lowest level of risk of exposure to Ebola virus. Participants were further categorised into three groups – high risk exposure (risk levels 1-3), moderate risk exposure (risk levels 4-6) and low level exposure (risk levels 7-8).

The use of antibody studies to determine previous Ebola infection in undiagnosed individuals following the epidemic in Sierra Leone is important with regards to genetic studies as it helps to identify a category of participants who had either asymptomatic or subclinical EVD, which was not diagnosed during the epidemic. This category provides evidence of an extreme genetic phenotype as clearly such individuals were able to contain the disease without expressing significant symptoms of it, these have been termed 'disease resistant' individuals. This suggests a highly effective immune response to EVD. Such studies also help to identify a category of study participants who were highly exposed to EBOV during the epidemic but show no evidence of previous infection with the virus as they are negative for anti-EBOV IgG. That these participants were so highly exposed (often contact with multiple dead bodies or bodily fluids of EVD patients) but did not become infected suggests they are 'resistant to infection' and are able to contain the virus prior to reaching the stage of generating a humoral immune response

that produces antibodies. These highly exposed but antibody negative individuals represent another extreme phenotype that warrants study from a genetic perspective. At the time of undertaking participant recruitment to the study it was unclear that these categories existed, and what their prevalence was in affected communities. Subsequently some data on this has been published[120, 121, 133] but with smaller numbers than described here. It was therefore both interesting and important to analyse this data from an epidemiological perspective, not only to further understand the seroprevalence in communities and relationship to exposure and symptomatology, but also to observe if the findings of this study were consistent with other recently published studies.

The data discussed in this chapter relates to seropositivity within communities, this denotes the proportion of the community that are anti-EBOV IgG positive and is expressed as a percentage. The following chapter describes the seropositivity within survivors in order to classify them as true Ebola survivors who have clear evidence of previous infection with EBOV. It also looks at the antibody level (or index) within survivors and compares this across different parameters, particularly the prevalence and severity of post-Ebola syndrome. The antibody index as described below is a measure of the quantity of antibody detected in saliva samples from each individual and reflects indirectly the concentration of antibody in the person's plasma.

4.2 Methods

The methods of participant recruitment of household contacts, community controls and Ebola survivors are described briefly in chapter three and in more detail in Appendix A. The methods of antibody detection and determining seropositivity and antibody level/index are described below.

4.2.1 Enzyme Immunosorbent Assay (EIA)

As mentioned in chapter 3, in order to improve community participation in the study, non-invasive sampling methods were needed. The sampling method used to determine seropositivity and antibody level was an oral fluid Anti-EBOV IgG Capture Enzyme Immunosorbent Assay (EIA) developed by Prof Richard Tedder and colleagues at PHE. An example of the oral fluid collection device is shown in Figure 4-1.

Figure 4-1 A member of the field team, Cecilia Touray, demonstrates the oral fluid collection device (a blue sponge that is rubbed around the gums for 90 seconds) to community members during sensitisation of one of the communities in Lunsar district, Sierra Leone, January 2017.

The assay is specific to EBOV glycoprotein antigen, as such it will only identify IgG to EBOV glycoprotein, and not to other viral proteins (Figure 4-2). The EIA was formulated based on the EBOV Mayinga glycoprotein antigen and has been demonstrated to have a sensitivity of >95% and a specificity of >99%[120, 132]. The coefficient of variation is a measure of the variation of the assay reaction across an individual EIA plate and also between EIA plates. This figure demonstrates the consistency of the reaction within different wells of the plate and across plates when different plates are run in tandem in the same environmental conditions. For the anti-EBOV EIA on oral fluid samples the coefficient of variation has been demonstrated in previous studies conducted in Sierra Leone (utilising the same equipment and similar laboratories) to be 8% within plates and 17.9% between plates, both of which are within accepted limits [132].

EBOV IgG Capture Assay

Legend

Anti-human IgG

Human IgG in sample

Bound Ebola glycoprotein antigen

Substrate

Anti-human IgG coated EIA plate well
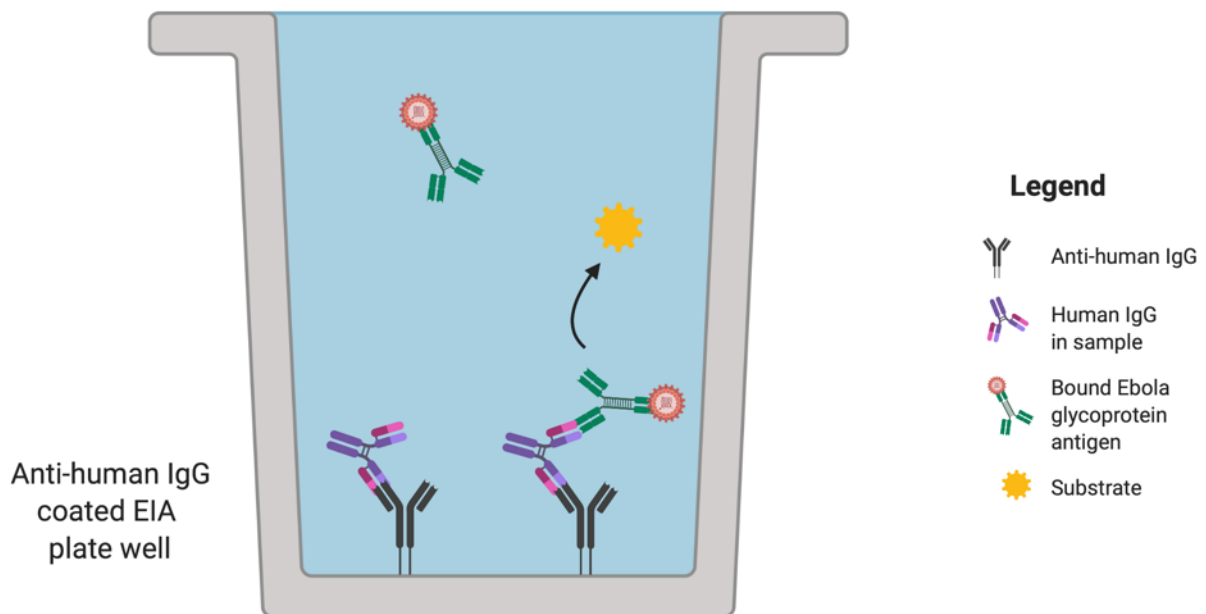
Figure 4-2 Diagram of the EBOV IgG capture assay. The diagram demonstrates a well of the anti-human IgG coated EIA plate. The human IgG in the sample binds to the anti-human IgG coating the well when the sample is added to the well. The bound Ebola glycoprotein antigen is then added to the well which binds only to human IgG that recognises Ebola antigen (which is only present if a participant has been previously infected with Ebola). If the bound Ebola glycoprotein antigen binds with human IgG in the sample, it releases TMB substrate which results in a colour change of the solution in the well. This colour change is then detected by a plate reader which measures the optical density of the well.

The oral fluid had already been eluted in viral transport medium by washing the oral fluid devices with 1ml of PHE viral transport medium, which contained 3-10% newborn calf serum, antibacterial and antifungal reagents and 0.05% Tween 20. The eluted samples were then stored in a -20 °C freezer until recruitment was completed and the ELISA could be performed on all samples together over a period of a few weeks, in order to reduce variability in conditions in which the EIA was being conducted. Duplicates of the positive control, negative control and the cut-off calibrator were added to the first six wells of the 96 well plate, coated with anti-human IgG antibody, at a volume of 100ul; 100ul of the oral fluid samples were added to the remaining 90 wells of the plate. Following a 60 minute incubation at 37 °C the wells were washed using a plate washer (*BioTek*, ELx50 TS Washer), with four cycles of 350ul wash fluid added to each well. Then 100ul of EBOV tracer (bound Ebola glycoprotein antigen) was added to each well and incubated at 37 °C for 120 minutes. The plate was then washed in the plate washer following the same cycle as above, and 100ul of TMB substrate added to each well. The plate

was then protected from light and incubated at room temperature for 30 minutes. Stop solution was then added to each well and the optical densities read in a plate reader (*Thermo Scientific*, Multiskan FC Microplate Photometer) at 450nm, with a reference wavelength of 620nm. All samples were handled in a BSL level 2 laboratory and samples were always handled and plated within a class 2 biological safety cabinet. Once the samples were ready to be read by the plate reader they were transported covered to the BSL level 1 laboratory where the plate reader was installed.

The plate layout for the EIA is shown in Appendix C. While ideally all plates would have included a mixture of household contacts, controls and survivors, it was decided to use separate plates for the survivor samples in order to prevent cross-contamination of contact and control samples resulting in false positive results. Given the consistency in performance of the assay and the limited variation within and between plates, this was felt to be the optimal plate layout.

If a negative result was obtained for any of the Ebola survivor samples, the sample was repeated to confirm the result. In the household contacts and community controls, if there was a positive result the EIA was repeated to confirm the result. The optical density cut-off was calculated based on the results of the 504 Ebola survivors and 1004 community control samples that were recruited as part of this study. On the basis that the survivor samples should be positive, and the community control samples negative this is a standard method of determining the cut off for a positive result. Using the results from these samples, a ROC curve was created which enabled the cut-off optical density for calling a positive EIA result to be defined using the Youden score (Figure 4-3 and Figure 4-4). The Youden score identifies the point of maximum sensitivity and specificity, which ensured a specificity of 98.6% and a sensitivity of 88.7% for an optical density cut-off of 0.088. In these circumstances the specificity was prioritised over the sensitivity to avoid false positives, which were perceived to potentially have a more significant impact on later genomic analysis, particularly given the substantially lower number of positive samples in the household contacts and community controls than negative samples. With regards to the survivor samples, any negative samples were excluded from the genomic analysis, so false negatives would not have impacted this analysis.

ROC curve: ROC of ELISA optical densities (survivors vs controls)
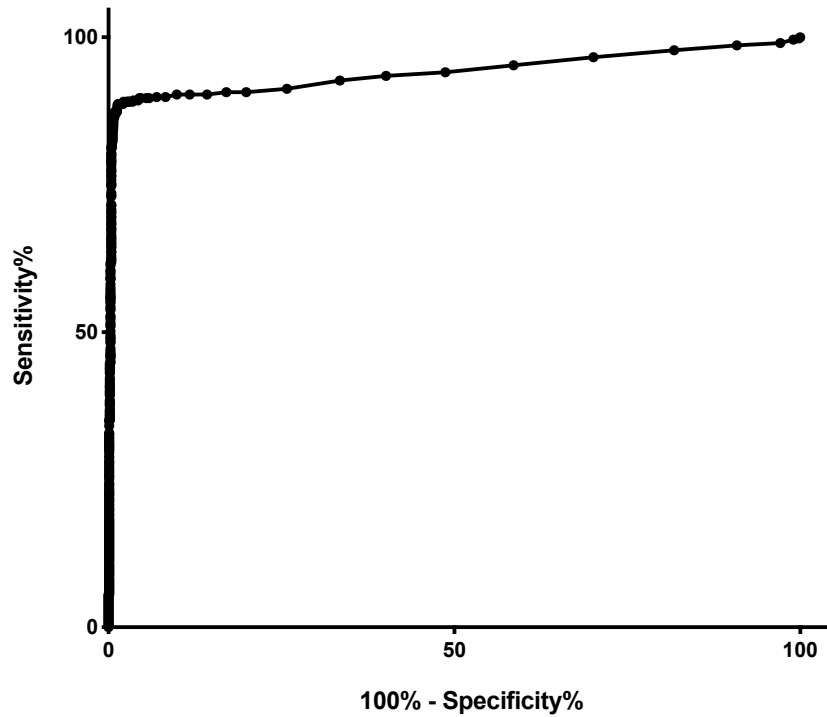
Figure 4-3 ROC curve of the EIA optical densities of community controls and Ebola survivors, used to determine the optical density cut-off for a positive EIA result.
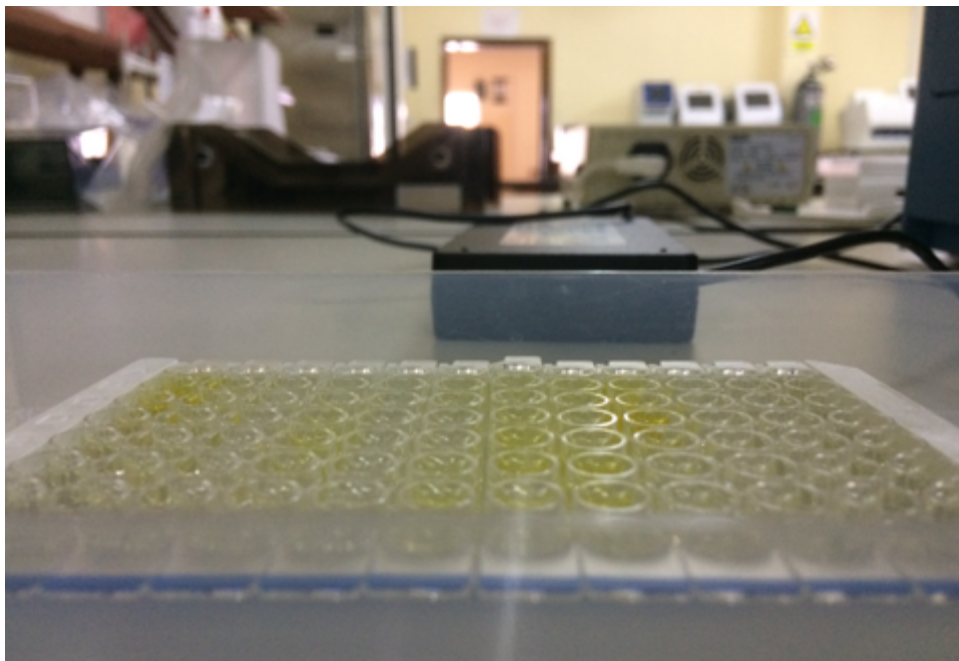
Figure 4-4 An Ebola survivor EIA plate showing variations in the colour change across wells thus demonstrating different amounts of Ebola glycoprotein antigen binding to the human IgG in the wells resulting in the release of different amounts of substrate in each well. Photo taken at the University of Makeni Infectious Diseases Research laboratory in Makeni, Bombali District, Sierra Leone. August 2017. With thanks to Samuel Alberman for the photo.

The EIA kit came with a reagent referred to as a cut-off calibrator. This reagent was included within two of the wells on each EIA plate and was used to control for over or under-reactivity of plates. The cut-off calibrator was used to calculate an antibody index for each sample within the plate based on the sample's optical density. The cut-off calibrator for each plate was adjusted on the basis of the data above demonstrating an optimal optical density cut-off of 0.088 for a positive sample in this study. Each sample's optical density result was then divided by the adjusted cut-off calibrator for that plate to determine the antibody index. The following antibody index parameters were used to call a result:

Antibody index	Result
>1	Positive
0.95-0.99	Indeterminate
<0.95	Negative

Table 4-1 Antibody index result interpretation.

Although the antibody index is not a concentration of antibody per se, it is related to the optical density of the sample, which reflects the concentration of antibody in the sample. While the relationship of optical density to antibody concentration is not linear, it is accepted that the larger the antibody index, the higher the optical density and therefore the higher the concentration of antibody. On this basis the antibody index value can be utilised to compare differences between groups in those who were antibody positive, for instance differences in antibody level between Ebola survivors with PES and those without PES (shown in chapter 5).

4.2.2 Statistics

The statistics described in this chapter are basic descriptive statistics including medians, percentages and odds ratios with 95% confidence intervals. P-values were calculated using t-tests and Wilcoxon rank sum tests with a p-value <0.05 denoting significance. Linear regression and Spearman rank tests were used to determine the significance of correlations across a set of parameters and logistic regression to determine odds ratios of significant correlations. All statistics were calculated in Stata 14.2 (*Statacorp LP*) and 'R Studio' 1.1 (*RStudio Inc*); graphs were created using GraphPad Prism version 8.0.1 (*GraphPad Software Inc*). Images were created using 'biorender' (*www.biorender.com*).

4.3 Analysis

4.3.1 Stratifying for exposure to Ebola – risk exposure level

Risk exposure levels, as described in the methods (Figure 3-2), were recorded for all household contacts and community control participants. As expected, the vast majority of community controls were risk

level 8, with over 99% reporting they had 'no known contact' with an EVD infected individual. Among household contacts all risk exposure levels were accounted for, which varied both with community type (Figure 4-5, p<0.001) and age (Figure 4-6, p<0.001). For comparison purposes risk exposure level was broken down into three categories: high risk exposure (levels 1-3), moderate risk exposure (levels 4-6) and low risk exposure (levels 7-8). The odds of high risk exposure in urban communities was 3.3 times that of rural communities (p<0.001, 95% CI=2.5-4.4).

Risk exposure levels for household contacts in rural and urban communities
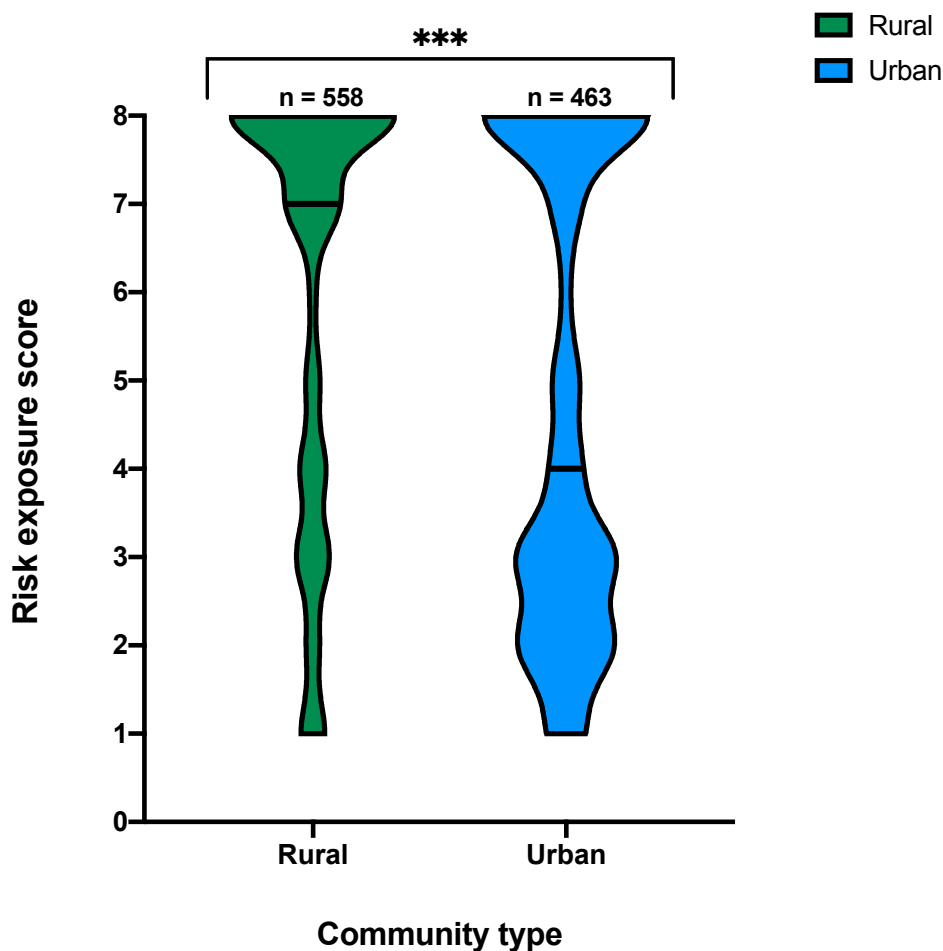
Figure 4-5 Risk exposure levels among rural and urban household contacts with the median score reflected by the horizontal black line, and participant numbers listed above each violin. The width of the violin is representative of the proportion of participants at the relevant risk exposure level. ***p<0.001.

The graph below (Figure 4-6) shows that overall children had a lower risk of exposure, with high risk exposure increasing with age. The odds of adults having high risk exposure compared to children and teenagers was 2.9 (p<0.001, 95% CI=2.2-3.8). Using linear regression, a significant inverse correlation

105

was demonstrated for risk exposure level with age (where *lower risk levels* correspond with *higher risk of exposure*), p<0.001.

Distribution of risk exposure score by age group
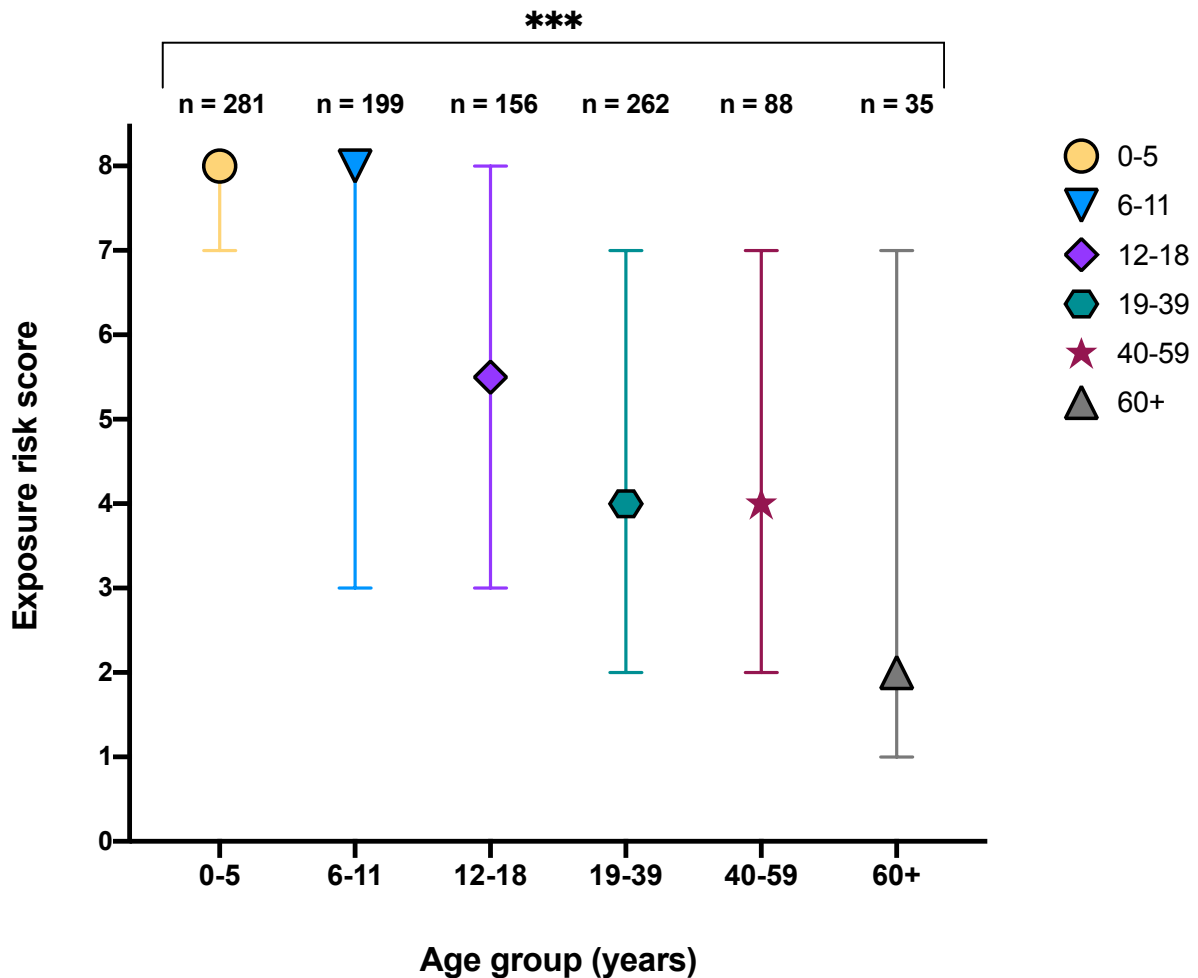
*Figure 4-6 Distribution of risk exposure score by age group with bars indicating the median and interquartile range, numbers of participants are shown above each plot. Risk level one represents the highest level of exposure to Ebola (contact with a dead body) and risk level 8 the lowest level of exposure to Ebola (no known contact).***significantly correlated at p<0.001.*

In order to corroborate the significance of these findings in relation to high risk exposure and determine if there were any confounding factors, unadjusted and adjusted odds ratios were calculated using logistic regression and are shown in Table 4-2 below. As can be seen in the table below the findings in relation to risk exposure and age were unaffected by confounding factors and held true with older ages more likely to be highly exposed to Ebola among household contacts. The odds of high risk exposure increased by 3% with each year of age gained (OR=1.03, p<0.001). Similarly participants living in urban communities were much more likely to be highly exposed to Ebola than their rural counterparts with

adjusted odds of 4.38 (CI 3.21-5.98, p<0.001). Reviewing the adjusted odds ratios, there were no significant gender differences in relation to high risk exposure, but the non-Temne, non-limba ethnicities demonstrated a lower risk of exposure than their Temne and Limba counterparts (OR 0.52, CI 0.35-0.77, p=0.001).

	Variable		Unadjusted Odds ratio (95% CI)	p-value		Adjusted Odds ratio (95% CI)	p-value
High risk exposure							
	Age (years)		1.03 (1.02-1.04)	<0.001		1.03 (1.02-1.04)	<0.001
	Male sex		0.75 (0.57-0.98)	0.032		0.89 (0.67-1.19)	0.429
	Urban community		3.33 (2.53-4.38)	<0.001		4.38 (3.21-5.98)	<0.001
	Ethnicity	Limba	0.69 (0.49-0.99)	0.046		0.70 (0.48-1.04)	0.081
		Other/mixed	1.01 (0.71-1.43)	0.969		0.52 (0.35-0.77)	0.001

Table 4-2 Adjusted and unadjusted odds ratios for variables associated with high risk exposure to Ebola virus among household contacts.

4.3.2 Reported mode of contact in Ebola survivors

Over 95% of the prospectively recruited Ebola survivors knew how they had become infected with Ebola virus (Figure 4-7). Just under 70% reported their mode of contact as caring for a family member. A further 14% described their mode of contact as either direct contact or sharing a bed with an individual who was unwell with EVD. Less than 1% described their mode of contact as attending a funeral.

Survivor mode of contact

Legend:
- Cared for 1st degree relative
- Cared for extended family
- Cared for partner/spouse
- Shared bed with relative
- Cared for friend/neighbour
- Unknown
- Direct contact with relative
- Healthcare worker exposure
- Direct contact with friend/neighbour
- Attended funeral
- Delivered infant
- Breastfed infant
- Shared bed with partner/spouse

Total=448

Figure 4-7 Mode of contact reported by Ebola survivors resulting in them becoming infected with EVD.

4.3.3 Anti-EBOV IgG seropositivity in household contacts

The overall rate of seropositivity in household contacts was 3.5%. This varied based on risk exposure level, age and whether a participant had experienced symptoms consistent with the Ebola case definition. The unadjusted and adjusted odds ratios for participant characteristics which may vary with seropositivity are shown in Table 4-3 below.

	Variable		Unadjusted Odds ratio (95% CI)	p-value	Adjusted Odds ratio (95% CI)	p-value
Anti-EBOV IgG Seropositive						
	High risk exposure		3.42 (1.73-6.78)	<0.001	2.76 (1.32-5.77)	0.007
	Age (years)		1.02 (1.01-1.04)	0.004	1.02 (1.00-1.04)	0.033
	Male sex		1.12 (0.57-2.18)	0.744	1.27 (0.64-2.51)	0.491
	Urban community		1.53 (0.78-2.99)	0.213	1.20 (0.57-2.54)	0.633
	Ethnicity	*Limba*	0.53 (0.18-1.54)	0.243	0.60 (0.20-1.75)	0.347
		Other/mixed	1.06 (0.45-2.50)	0.892	1.04 (0.42-2.58)	0.940

Table 4-3 Adjusted and unadjusted odds ratios for participant characteristics that may impact on anti-EBOV IgG seropositivity in household contacts.

4.3.3.1 Seropositivity by risk exposure level

The risk exposure stratification used in this study had already been demonstrated to correlate with seropositivity[120]. The results described here confirm that seropositivity was related to risk of exposure, with 15.6% of participants in risk exposure level 1 being seropositive for Ebola antibodies, and 6.9% of high risk exposure (levels 1-3) participants being seropositive (Figure 4-8). This was versus 1.9% of moderate and low exposure participants (levels 4-8) being seropositive. The odds ratio for seropositivity in high exposure participants versus low and moderate exposure participants was 2.76 (p=0.007, 95% CI 1.32-5.77) and the odds of seropositivity in participants in risk exposure level 1 compared to the aggregate of risk levels 2 to 8 was 7.1 (p<0.001, 95% CI=3.4-14.8).

Percentage seropositivity by risk level in household contacts
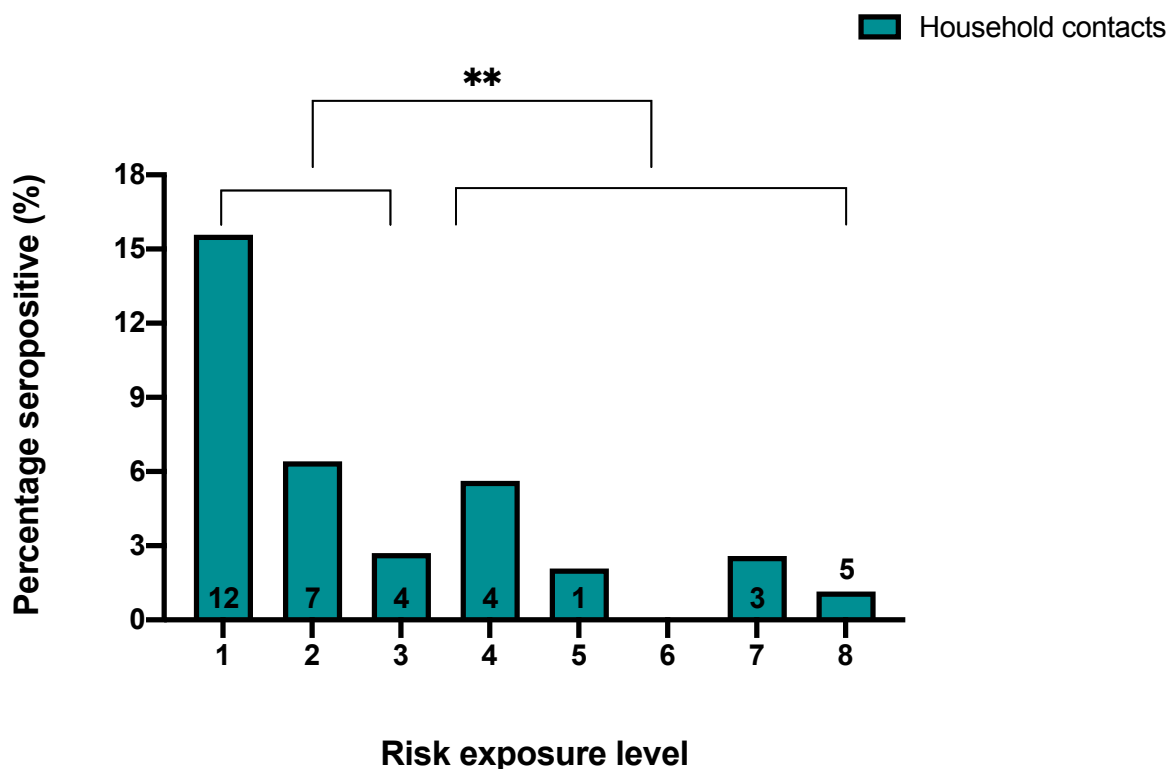
Figure 4-8 Rates of seropositivity by risk exposure level where risk level one represents the highest level of exposure to Ebola (contact with a dead body) and risk level 8 the lowest level of exposure to Ebola (no known contact). Numbers of seropositive participants are shown within/above the bars. **significant at p<0.01 for rates of seropositivity in high risk exposure participants (levels 1-3) versus moderate (levels 4-6) and low risk exposure participants (levels 7 & 8).

4.3.3.2 Variations in seropositivity by age and community type

In household contacts seropositivity rates were higher in older age groups and increased with age, with the exception of a very low level of seropositivity in the 6-11 year age group. Overall seropositivity rates were 5.7% for adults and 2.2% for children (p=0.003). Using logistic regression, the adjusted odds ratio was 1.02 (p=0.033) for seropositivity with increasing age in years. This corresponds to a 2% increase in the odds of being seropositive for each year of age gained and was only partly confounded by older ages experiencing higher risks of exposure (unadjusted p=0.004 vs adjusted p=0.030). Seropositivity levels also appeared higher in urban areas compared with rural areas with 15.4% of participants in the 40-59 year age group in urban areas seropositive versus none in the same group in rural areas. Seropositivity in urban areas was 1.5 times that of rural areas (4.3% vs 2.9%), although this difference was not statistically significant when controlling for other factors such as rates of high risk exposure (p=0.63). Differences in rates of seropositivity between urban and rural children and adults are shown

in Table 4-4 and Figure 4-9 below. When comparing adults in urban areas to all other household contacts the odds of being seropositive were 3.9 times higher (OR 3.85, p<0.001, 95% CI 1.94-7.64).

Community type	Adult seropositivity (%)	Child seropositivity (%)
Urban	8.9***	1.7#
Rural	3.2 #	2.6#

*Table 4-4 Household contact seropositivity in adults and children in rural and urban communities. #=not significant, *=significant at p<0.05, **=significant at p<0.01, ***=significant at p<0.001*
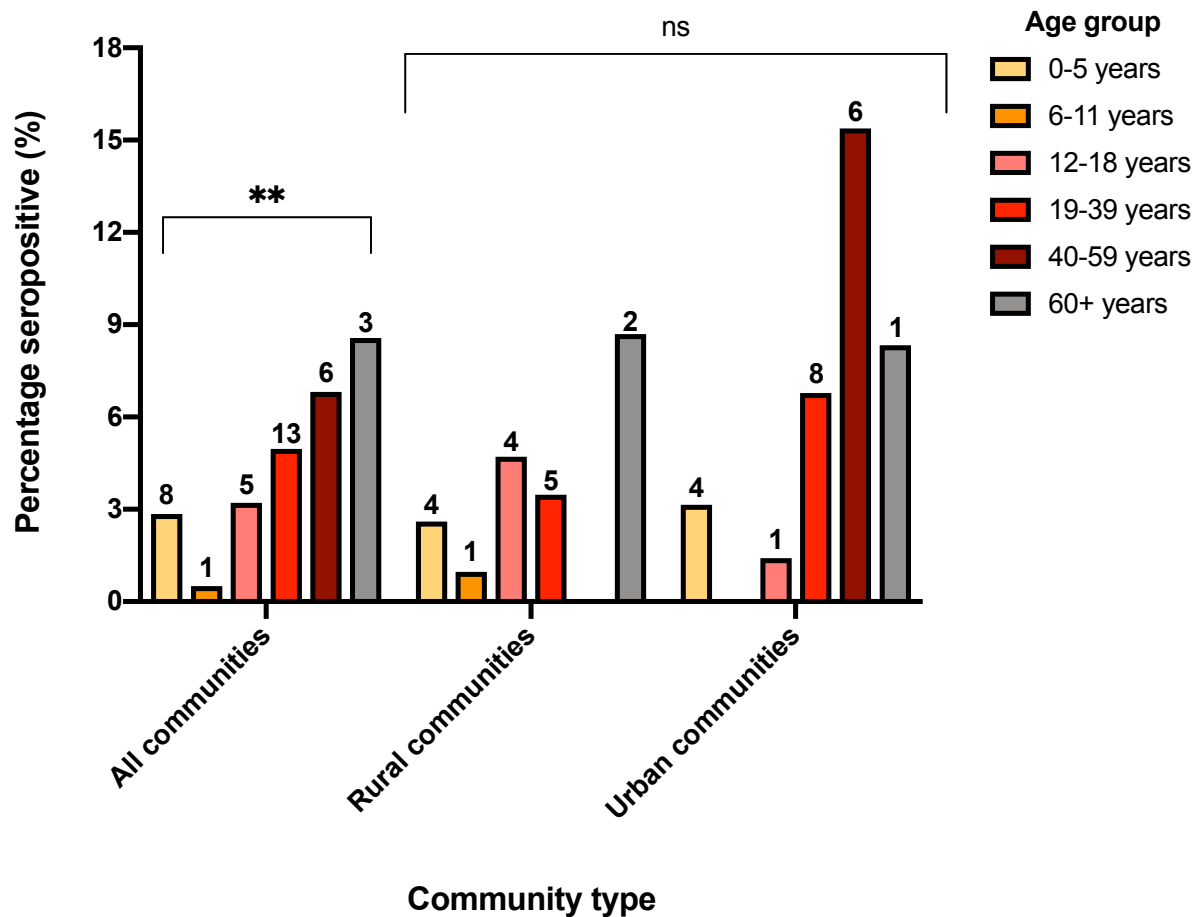
*Figure 4-9 Seropositivity rates by age group and community in household contacts. Numbers of seropositive participants are shown above the bars. **significant at p<0.01 for seropositivity increasing with age in rural and urban communities combined.*

4.3.3.3 Seropositivity and symptomatology

Approximately 4% of household contacts described symptoms during the time of the Ebola epidemic in their community, which met the case definition for Ebola. A further 0.9% described some symptoms,

but that either did not meet the case definition for Ebola or were not related to infection with Ebola. Over 95% of household contacts did not experience any symptoms during the time the Ebola epidemic affected their community.

Household contacts - symptoms experienced
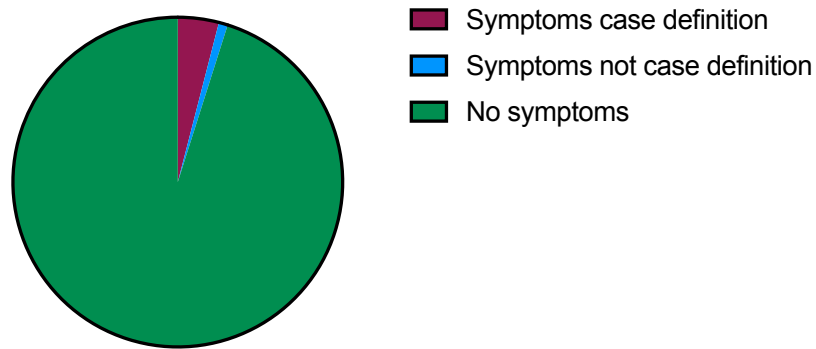
Total=1021

Figure 4-10 Symptoms experienced by household contacts during the time of the Ebola epidemic affecting their community.

While few household contacts experienced symptoms during this time, if they did experience symptoms which met the case definition for EVD then they were more likely to be positive for Ebola IgG, with nearly a third (31.7%) of symptomatic individuals being seropositive (OR=19.3, p<0.001, 95% CI=8.9-42.0). This increased significantly when symptoms were related to risk exposure level. Figure 4-11 below demonstrates that the household contacts who had contact with a dead body (risk exposure level 1) and experienced symptoms meeting the case definition for Ebola, had seropositivity rates of 53.9%. These risk level 1, symptomatic household contacts, when compared to all other household contacts, were 39.4 times more likely to be seropositive for anti-EBOV IgG (OR=39.4, p<0.001, 95% CI=12.5-124.5). Similarly, when looking at all household contacts who were seropositive, 36.1% experienced symptoms which met the case definition for EVD.

Rates of seropositivity in symptomatic household contacts by risk exposure level
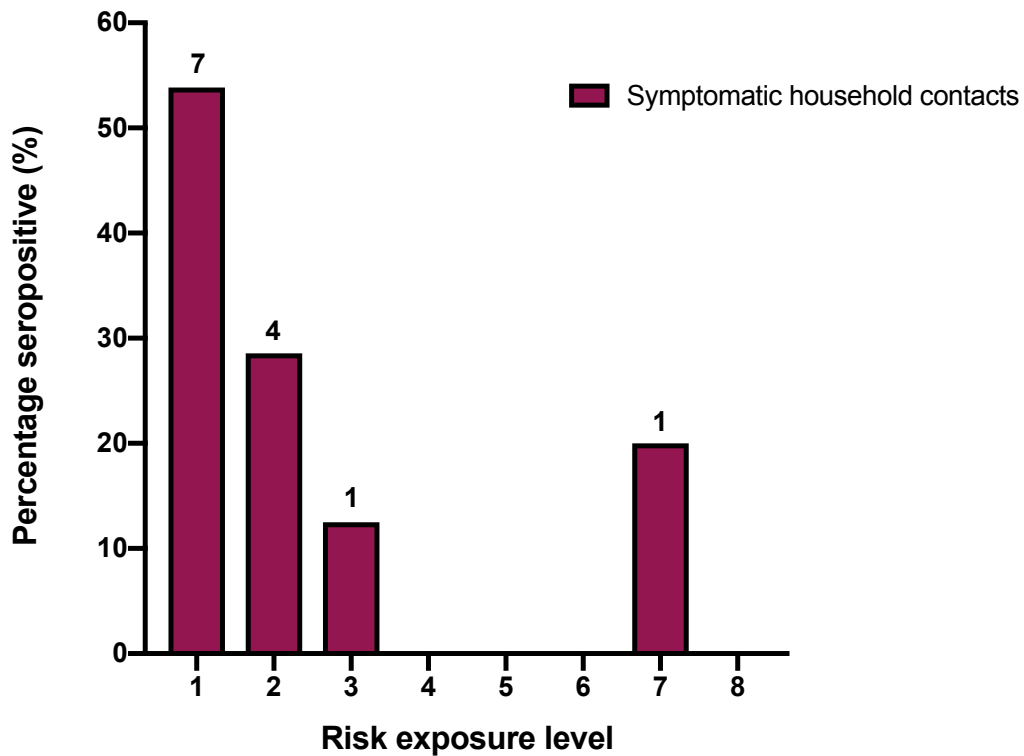
Figure 4-11 Seropositivity rates in household contacts with symptoms meeting the Ebola case definition by risk exposure level, where risk level one represents the highest level of exposure to Ebola (contact with a dead body) and risk level 8 the lowest level of exposure to Ebola (no known contact). Numbers of seropositive participants are shown above the bars.

Due to the interest in determining extreme phenotypes for the genetic analysis, the graph below (Figure 4-12) highlights the significant proportion of high exposure household contacts (level 1-3) who were both seronegative and experienced no symptoms. This will be discussed further in chapter 6.

Seronegativity and symptomaticity by risk level in household contacts
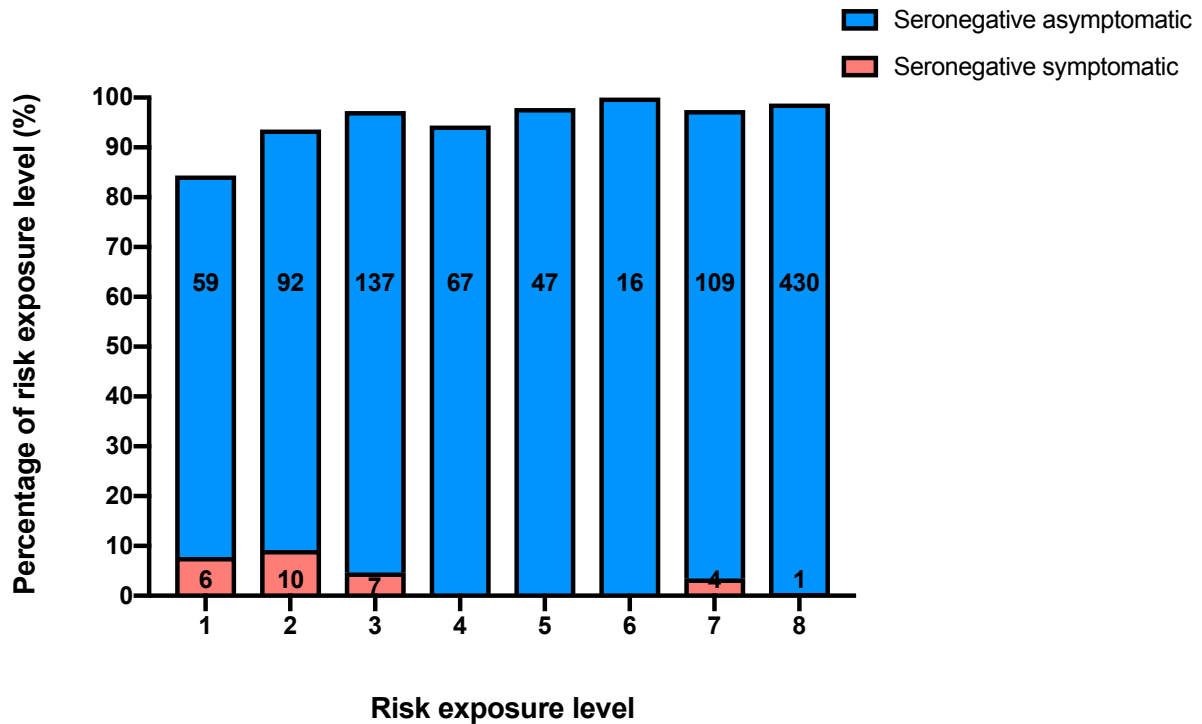
Figure 4-12 Seronegative household contacts by presence of symptoms and risk exposure level, where risk level one represents the highest level of exposure to Ebola (contact with a dead body) and risk level 8 the lowest level of exposure to Ebola (no known contact). Numbers of participants are shown within the bars.

4.3.4 Anti-EBOV IgG seropositivity in community controls

Overall seropositivity in the community controls was 1.8%. This varied with age and community type but was not significant. As over 99% of controls reported a risk exposure level of 8, risk exposure did not demonstrate a relationship with seropositivity for community controls. While eight community control participants reported a risk exposure other than level 8 (range 1-7), only one of these was seropositive (risk level 4). Similarly, due to small numbers of controls being seropositive there was no demonstrable relationship between seropositivity, age and community type either.

4.4 Discussion

4.4.1 Quantifying exposure to Ebola

As shown in Figure 4-5 and Figure 4-6, higher risk exposure levels were observed in urban communities and in adult age groups (p<0.001 for both). This observation could be explained by urban communities having a higher population density, where it was more likely that a person from an affected household

might come into direct contact with a sick individual in their household due to more cramped conditions. It is also known that Ebola is a 'disease of carers' and there are high rates of transmission of the disease to those individuals who directly care for an unwell individual, whether that is a healthcare worker who is not wearing appropriate personal protective equipment, or a family member. Young children are less likely to be utilised as carers of sick individuals and subsequently have a lower risk of exposure. This provides important insights into how to address disease transmission in communities, by targeting individuals who are likely to be identified as carers of the sick and improving awareness of the disease and its symptoms and signs, and providing information on how to isolate a patient until they can be transferred to an isolation facility.

The concept of Ebola as a 'disease of carers' is supported by the reported modes of contact from the prospectively recruited Ebola survivors, where over 70% reported caring for a family member and a further nearly 10% reported either caring for a friend or neighbour or being infected through their duties as a healthcare worker. As mentioned above this provides insight in who to prioritise for monitoring within communities and also who to target for community engagement and awareness measures. Just over 14% of survivors reported either sharing a bed with or having direct contact with a family member or friend. This emphasises another area that should be addressed in community engagement strategies. Families in affected locations often have very restricted incomes, it is not unusual for five people to share a bed. This is clearly a risk factor for becoming infected with Ebola. Interestingly, while funeral attendance is a known risk factor for transmission of Ebola, particularly in the context where attendees have direct contact with the corpse, this was not a commonly reported mode of transmission. While avoiding contact with a dead body and practicing safe and dignified burials is essential to the containment of an Ebola epidemic, it must be emphasised that transmission from attending a funeral, while a *high risk* form of transmission, is not a *common* mode of transmission.

4.4.2 Anti-EBOV IgG seropositivity in household contacts

The relationship of risk exposure level to seropositivity is consistent with the findings from the study conducted by *Glynn et al* which first described the risk exposure stratification[120]. The overall rates of seropositivity for the household contacts were similar to that found in the study conducted by *Glynn et al* and also the recently reported study by *Timothy et al*[121]. This demonstrates that increased contact with a person symptomatic of EVD placed an individual at higher risk of being seropositive, even if they did not experience any symptoms. Direct contact with a wet case of EVD, their bodily fluids or a dead body almost trebled the chances of being seropositive to EVD. The finding that 3.5% of household contacts had minimally symptomatic or asymptomatic Ebola virus disease and were never tested raises

the question of the true number of EVD cases from the West African epidemic. Furthermore, it raises the question of what the risk of disease transmission is from such individuals, either in the acute phase or in the recovery phase via sexual transmission. While it is likely people with minimally symptomatic disease have only a low level viraemia and so are less likely to express virus in bodily fluids, they may still present an unseen transmission risk [17, 44, 59, 138, 139].

The variation of seropositivity with age and high risk exposure is not unexpected. As described above, adults were more likely to have higher risk exposure levels than children, as were participants from urban communities. It is likely that it is the higher risk levels older participants were exposed to that accounts for the increased seropositivity in older age groups, although this does not fully explain the situation as when controlling for high risk exposure, age remains a significant risk factor for being seropositive. The odds of being seropositive as an adult in an urban area (where there were higher odds of high risk exposure, OR 4.38, p<0.001, CI 3.21-5.98) was 3.9 times higher than all other household contacts (OR 3.85, p<0.001, CI 1.94-7.64).

While the number of participants with symptoms meeting the EVD case definition was only a small proportion of the household contacts, they had an increased risk of being seropositive (OR=19.3, p<0.001). This difference was further enhanced when comparing risk exposure level and symptoms and emphasises the importance of good contact tracing in communities and regular assessment of symptoms in individuals with a history of high risk contact. A small number of household contacts had symptoms meeting the EVD case definition but were seronegative. These individuals were also pre-dominantly in the high risk exposure group. As the majority of these participants were highly exposed to EBOV it needs to be considered whether they are individuals who were infected but who did not mount a sustained antibody response. In the absence of a positive antibody test however, and given the similarities of mild symptoms of Ebola with symptoms of many other infectious diseases in the region, this is difficult to ascertain.

It was noted that during the epidemic, EVD contacts in some of the urban communities were subjected to Ebola PCR testing on the basis of their contact history despite an absence of symptoms. A total of 116 household contacts in this study were tested for EVD during the epidemic, 78% of these were in urban communities. It is recommended that PCR testing should not be conducted in the absence of symptoms as it is unlikely to yield a positive result due to very low levels of viraemia[140]. Of the 116 participants who were tested in the absence of symptoms, 72% tested negative and the remainder did not receive their test results. Twenty of these participants later went on to develop symptoms and 16

have now been identified as seropositive. Of the 16 who are seropositive 69% had symptoms meeting the EVD case definition. These individuals did not seek further testing despite developing symptoms of EVD. This emphasises the importance of not testing for Ebola in the absence of symptoms, as a negative result may be falsely reassuring and result in an individual who becomes symptomatic later on remaining in the community and potentially further transmitting the disease.

4.4.3 Anti-EBOV IgG seropositivity in community controls

The results described here indicate that there is still a risk of being sub-clinically infected with EVD when living in an affected community even if no one in a person's household is symptomatic and there is no known contact with anyone who is unwell. While this risk is low, it is important to consider the possibility of EBOV infection in individuals from affected communities, even if there is no history of direct contact with the disease.

Overall the results indicate that a person who is a household contact from an affected community is twice as likely as a community member from an unaffected household in the same community to be seropositive, 3.5% vs 1.8% respectively (OR 2.00, p=0.018, 95% CI 1.13-3.55). While this is expected, it puts further emphasis on the importance of history taking at health facilities in patients presenting with symptoms consistent with EVD, particularly those in the early stages of the disease where symptoms may represent any number of infectious diseases.

4.4.4 Relevance to genomic analysis

The antibody data described here shows that there were both highly exposed individuals who are antibody negative ('resistant to infection') and antibody positive individuals with subclinical disease ('resistant to disease'). This supports the hypothesis that such groups of individuals do exist and is suggestive that they may carry genetic variations that help them to resist or contain the virus.

Chapter 5 Ebola IgG antibody level in Ebola survivors and relationship to PES

5.1 Introduction

This chapter describes the results of Ebola IgG antibody studies conducted on the recruited Ebola survivors. Initially the possibility of sero-reversion in Ebola survivors is considered due to 11% of the recruited survivor participants being anti-EBOV IgG negative. Then variations in antibody level in relation to participant characteristics such as age and sex are examined. Following this, severity of acute disease and severity of post-Ebola symptoms and their relationship to antibody level are examined. Lastly, the use of the antibody data from chapters 4 and 5 to define the questions and hypotheses for the genetic analysis is discussed.

5.1.1 Sero-reversion of Ebola survivors

The data discussed initially in this chapter describes seropositivity within survivors in order to classify them as true Ebola survivors who have clear evidence of previous infection with EBOV. This preliminary discussion is focussed on the 11% of recruited Ebola survivors who were anti-EBOV IgG negative. Reasons for the 11% of Ebola survivors in this study testing anti-EBOV IgG negative are discussed and, using matched samples from a small group of Ebola survivors separated in time by 18 months, the hypothesis is explored as to whether over time some Ebola survivors may sero-revert and no longer have anti-EBOV IgG antibodies.

5.1.2 Seropositivity versus antibody level

The majority of this chapter principally analyses the antibody level (or index) within survivors who are anti-EBOV IgG positive and compares this across different parameters, particularly the prevalence and severity of post-Ebola symptoms. Seropositivity denotes the proportion of a group that are anti-EBOV IgG positive and is expressed as a percentage. The antibody level (or index) is a measure of the quantity of antibody within an individual and reflects the concentration of antibody in plasma. These two parameters are separate concepts and are not representative of one another, although in order to have an antibody level an individual must by definition be sero-positive.

5.1.3 Factors associated with antibody level

Anti-EBOV IgG analysis was initially conducted with the purpose of determining true survivors, with the exclusion from analysis of any survivor testing anti-EBOV IgG negative. However, as antibody studies were being conducted, they also served to determine whether there were variations in antibody level both with differences in survivor characteristics (such as age, sex and ethnicity), and also with different disease phenotypes within those survivors confirmed to be anti-EBOV IgG positive. While it was hypothesised there may be variations in antibody level in relation to the severity of acute symptoms experienced by an individual, other factors such as age, sex and ethnicity, were not deemed likely to result in significant variations in antibody level between Ebola survivors.

5.1.4 Post-Ebola symptoms (PES)

As described in chapter 3, there were different disease phenotypes in relation to whether an Ebola survivor experienced post-Ebola symptoms and the severity of those symptoms. This chapter therefore also explores whether there were variations in the amount of anti-EBOV IgG that survivors with and without post-Ebola symptoms produced. It was hypothesised that those survivors who had PES might have different antibody levels (either greater or less) reflecting a different immune response to the virus. The concept underlying this hypothesis relates to the understanding that post-Ebola syndrome might be related to the persistence of virus in immune privileged sites[12, 53, 56-58, 60-63, 66, 67, 141]. This persistence in virus might repeatedly trigger a humoral immune response resulting in production of antibody and subsequently a higher anti-EBOV IgG level in survivors with PES. Conversely, lower antibody levels in those with PES might indicate an inadequate immune response and failure to clear the infection. Thus, determining variations in antibody level in survivors may help to further phenotype the study groups for the genomic analysis.

5.2 Methods

The methods of participant recruitment of Ebola survivors are described briefly in chapter three and in more detail in Appendix A. The classification of post-Ebola syndrome is also described in chapter 3, along with the features of PES that were deemed to constitute 'severe PES'. The methods of antibody detection are described in chapter 4 above.

5.2.1 Assessing sero-reversion in Ebola survivors

As described in chapter 3, all recruited Ebola survivors were required to provide an Ebola survivor certificate at recruitment, but in order to verify Ebola survivors as 'true Ebola survivors' for accurate

disease phenotyping, it was necessary to determine that they had clear evidence of previous infection with EBOV. As the antibody samples were obtained approximately two years after their acute infection however, there was the possibility that some Ebola survivors may have sero-reverted and no longer have identifiable anti-EBOV IgG. As collaborators had recruited a cohort of survivors approximately 18 months prior to this study, fifteen survivors were identified who had been recruited to both studies. This meant it was possible to do a comparison of the survivor's antibody index from the sample taken during this study's recruitment and the sample taken during the collaborator's recruitment 18 months before to determine if antibody levels had decreased. This analysis utilised the same *Kalon Diagnostic Ltd* kits purchased for the current study and was conducted at the same time, and in the same environment for both sets of samples to control for assay variation in different environments.

5.2.2 Understanding variation in antibody level among Ebola survivors

The analysis described below was designed to determine if there was variation in antibody level along different parameters of participant features, such as age, sex, ethnicity and community type. Multiple regression analysis was performed to determine the impact of these features individually, but also whether any significant results observed were confounded by one of the other parameters. Similarly based on the definitions of mild and severe PES described in chapter 3 logistic regression was conducted to determine if there was any variation in antibody index between those with no PES, PES or severe PES, and was adjusted for potential confounding by age, sex, ethnicity and community type.

5.2.3 Statistical analysis

The statistics used for data analysis in this chapter include percentages for seropositivity and prevalence of PES, as well as geometric means for antibody level with 95% confidence intervals. The non-parametric Wilcoxon rank test was used to compare the geometric mean antibody level between two groups. Where geometric mean antibody level was compared across multiple groups a one way ANOVA (analysis of variance) was used. For analysis of trends in antibody level across groups linear and logistic regression was used and for correlations of prevalence of PES linear regression analysis was conducted. All statistics were calculated in Stata 14.2 (*StataCorp LP*), 'R Studio' 1.1 (*RStudio Inc*) and graphs were created using GraphPad Prism version 8.0.1 (*GraphPad Software Inc*).

5.3 Analysis

5.3.1 Assessment of sero-reversion in Ebola survivors

Out of the 504 Ebola survivors prospectively recruited, 448 were positive for anti-EBOV IgG (89%) and 56 survivors (11%) were IgG negative. To determine if these negative results were due to sero-reversion of survivors a comparison was conducted between 15 matched samples from a collaborator's recruitment (JG) undertaken 18 months prior to this study (NM). The resulting optical densities and antibody indexes are shown in Table 5-1 below, with a scatter plot comparing the antibody indexes (Figure 5-1).

Participant ID	JG Optical density	JG antibody index		NM Optical density	NM Antibody index
1	0.13	2.175		0.087	1.454
2	0.29	4.825		0.22	3.742
3	0.16	2.65		0.37	6.083
4	0.20	3.392		0.07	1.158
5	1.49	24.867		0.15	2.492
6	0.17	2.867		0.20	3.392
7	0.51	8.533		0.53	8.783
8	0.26	4.358		0.65	10.892
9	0.081	1.35		0.13	2.167
10	0.32	5.317		0.19	3.092
11	0.40	6.692		0.21	3.525
12	0.32	5.383		0.72	12.075
13	0.19	3.092		0.22	3.617
14	0.13	2.2		0.17	2.875
15	0.22	3.7		0.30	4.958

Table 5-1 Comparison of survivor anti-EBOV IgG results based on samples of the same Ebola survivors from 2015 (JG) and 2017 (NM). Red = IgG level decreased, black = IgG level similar, blue = IgG level increased.

Comparison of Ebola survivor antibody levels
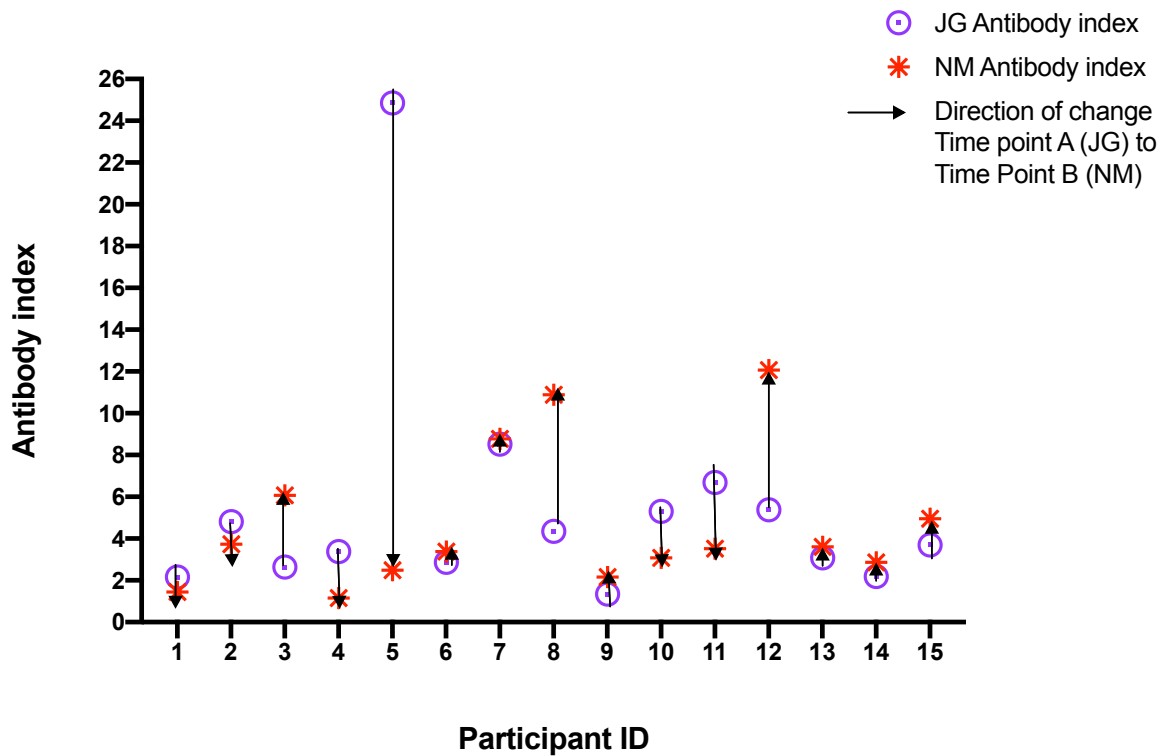
Figure 5-1 Scatter plot comparing the antibody index of Ebola survivor samples from Time point A (JG) and Time point B (NM). Arrow indicates direction of change from Time point A to Time point B.

5.3.2 Features of seronegative Ebola survivors compared with seropositive Ebola survivors

The number of seronegative Ebola survivors appeared to predominate from treatment facilities in one particular location (p<0.0001). These survivors were also noted to have much lower rates of PES (p<0.0001) and included a slightly larger proportion from rural areas (p=0.02) but did not vary significantly in age or sex distribution, as shown in Figure 5-2 below.

Seronegative vs seropositive survivor comparison

Legend:
- Sero-positive Survivors
- Seronegative survivors
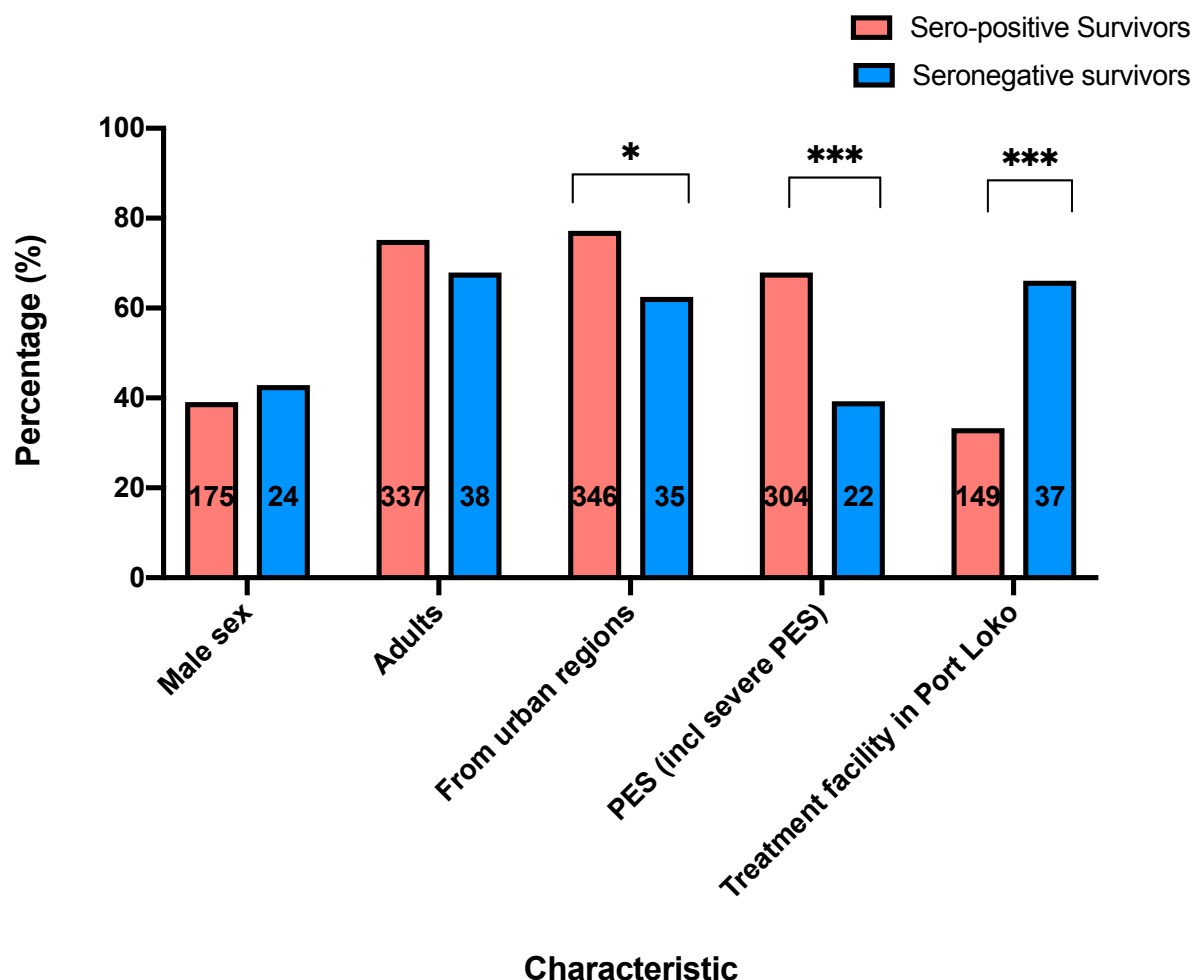
Figure 5-2 Comparison of various characteristics of the seronegative (n=56) and seropositive survivor populations (n=448). Participant numbers are shown within the bars. ***significant at p<0.001, *significant at p<0.05.

5.3.3 Variations in antibody index of Ebola survivors

As it was not possible to determine if seronegative survivors were true Ebola survivors, these participants were excluded from further analysis in relation to prevalence of PES, antibody levels and genetic analysis. The figures below demonstrate the parameters in which the antibody index, and thus the level of anti-EBOV IgG varied in those Ebola survivors who were anti-EBOV IgG positive. Figure 5-3 shows that the anti-EBOV IgG antibody level varied with age, with adults and older Ebola survivors having lower antibody levels, with the exception of the 6-11 year age group. Overall adults had a lower antibody index than children under 19 years of age, with a geometric mean AI of 3.18 (95% CI=2.98-3.39) in adults versus 3.82 (95% CI=3.39-4.31) in children (p=0.008). Linear regression was used to determine the correlation of log antibody index with age but was not significant with p=0.096. The

geometric mean antibody index and 95% confidence interval by age group are demonstrated in Table 5-2 and the linear regression analysis in Table 5-3.

Age Group	Geometric mean antibody index	95% confidence interval
0-5	4.87	3.22-7.38
6-11	3.25	2.66-3.97
12-18	3.97	3.42-4.60
19-39	3.16	2.94-3.40
40-59	3.34	2.86-3.90
60+	2.53	1.69-3.80

Table 5-2 Geometric mean antibody index and 95% confidence interval by age group in Ebola survivors.
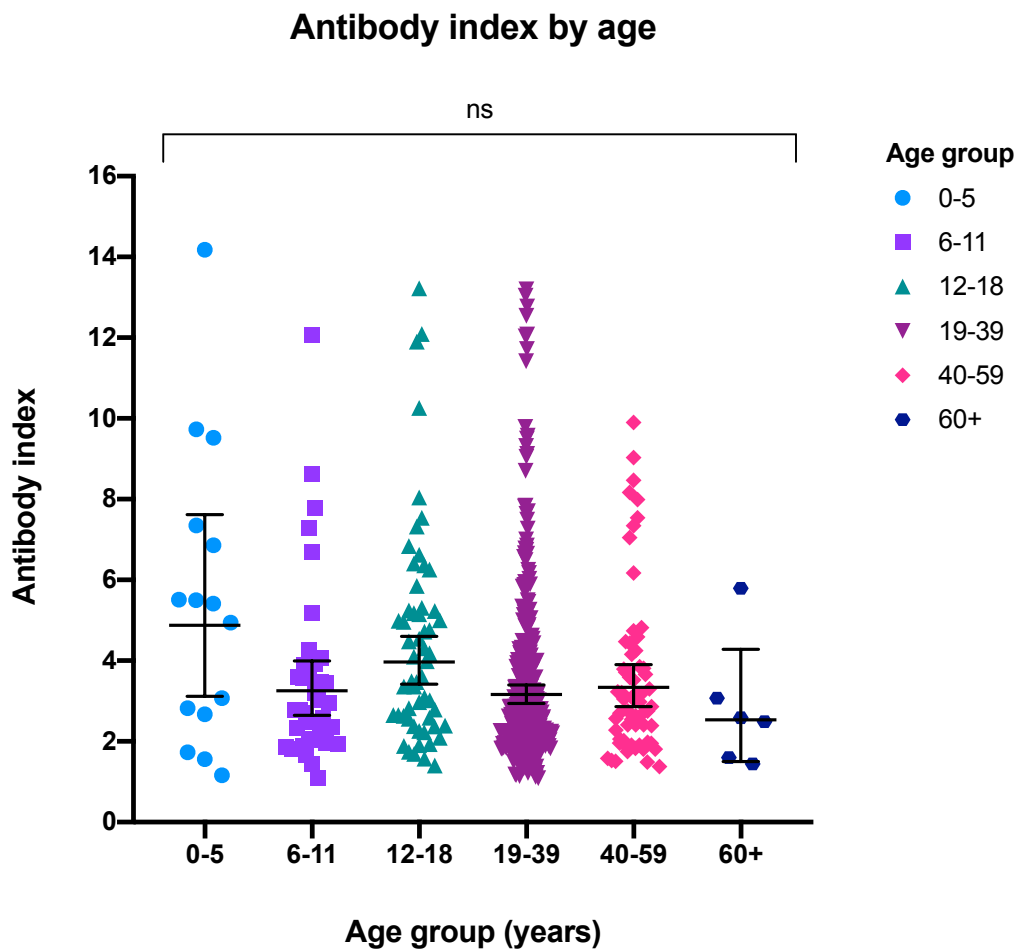
Figure 5-3 Variations in anti-EBOV IgG levels with survivor age group, with geometric mean and 95% CI, ns=not significant.

The three further figures below (Figure 5-4, Figure 5-5, Figure 5-6) demonstrate there was no difference in antibody index between male and female Ebola survivors (geometric mean AI 3.32 vs 3.34, p=0.91), but there was some variation between different ethnicities (p=0.023) and there was a difference between the geometric mean AI in survivors from rural and urban areas (2.84, 95% CI=2.55-3.17, vs 3.49, 95% CI=3.27-3.73, p=0.002).
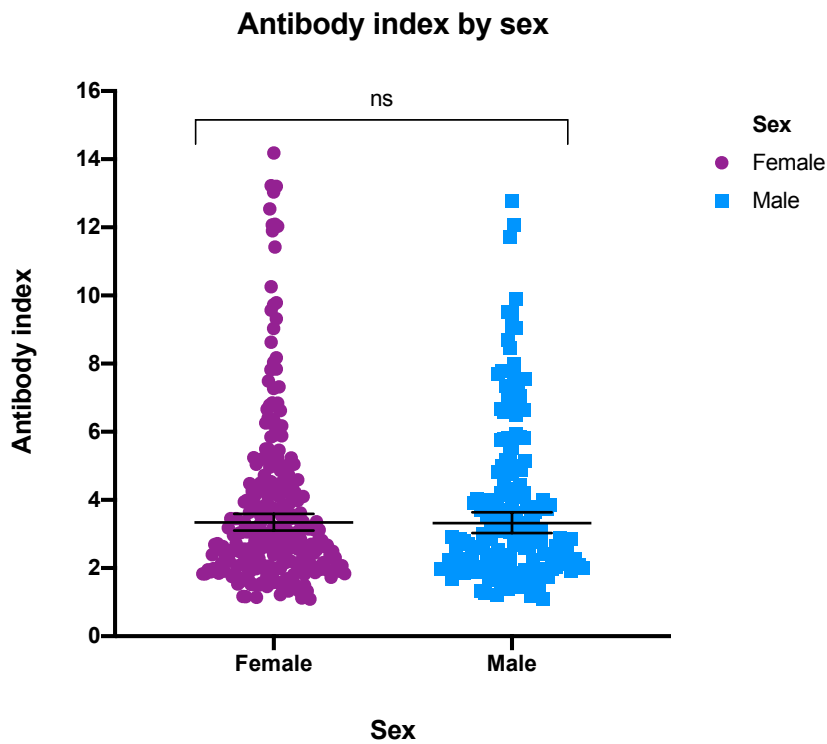
Figure 5-4 Anti-EBOV IgG antibody level by sex with geometric mean and 95% CI. ns =not significant.
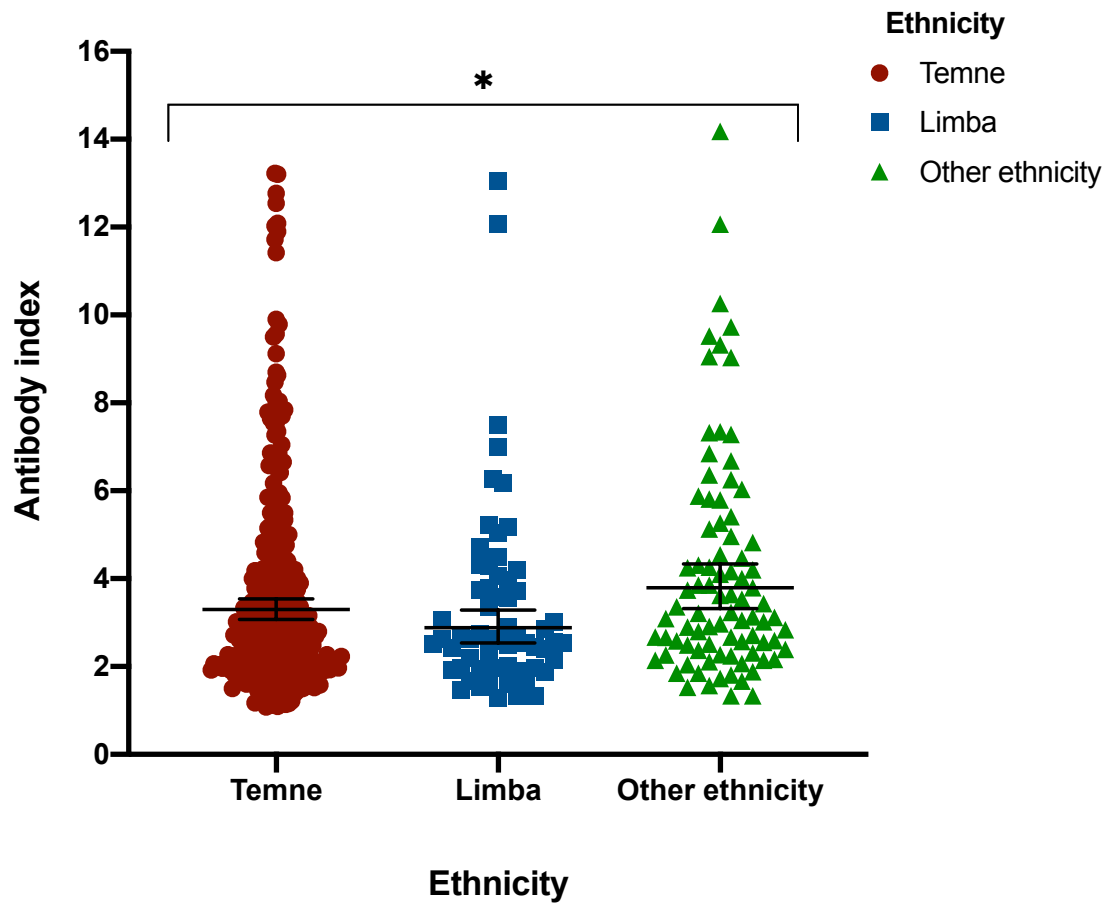
Figure 5-5 Variation in anti-EBOV IgG by the four main ethnicities of survivor participants, with geometric mean and 95% CI. *significant at p<0.05.

Antibody index by community type
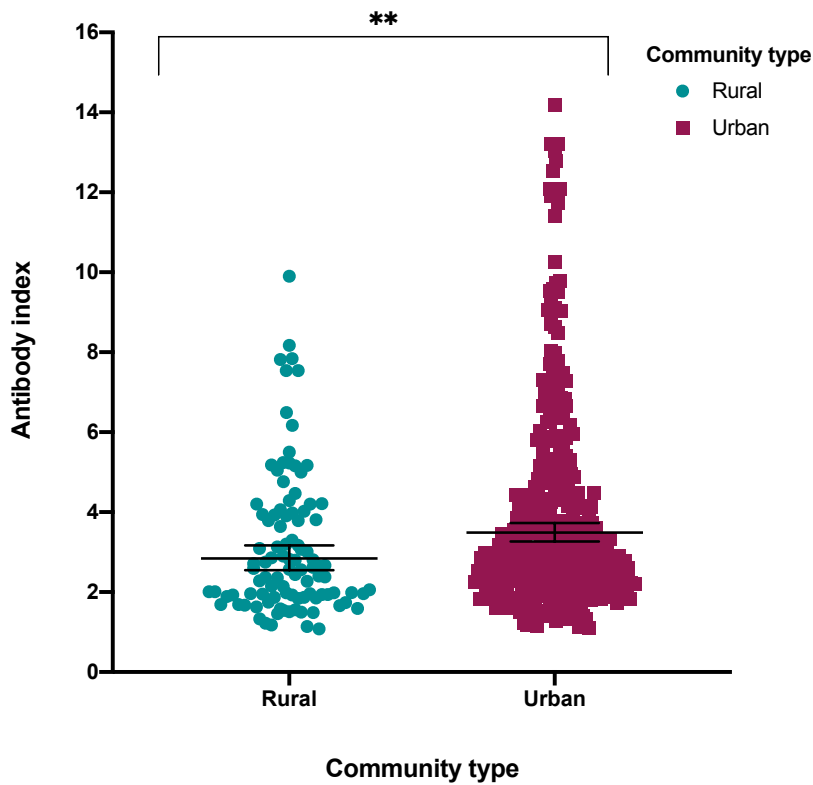
*Figure 5-6 Variations in anti-EBOV IgG by survivor community type, with geometric mean and 95% CI. **significant at p<0.01.*

To determine the robustness of these findings linear regression was performed to determine the odds ratio of the log antibody index geometric mean for these parameters. The unadjusted and adjusted results are demonstrated in Table 5-3 below. It is apparent from these figures that the suggestion of a lower mean antibody index for rural communities is maintained (p=0.003), although this is partially confounded by other parameters such as age, sex and ethnicity and so reduces somewhat in significance in the adjusted model (p=0.024). Similarly, the variation in antibody index with ethnicity is not maintained and is no longer significant.

	Variable		Unadjusted geometric mean odds ratio (95% CI)	p-value	Adjusted geometric mean odds ratio (95% CI)	p-value
Log antibody Index						
	Age (per 10 years)		0.96 (0.92-1.00)	0.096	0.97 (0.93-1.01)	0.118
	Male sex		0.99 (0.88-1.12)	0.911	0.99 (0.88-1.11)	0.800
	Rural community		0.81 (0.71-0.93)	0.003	0.85 (0.74-0.98)	0.024
	Ethnicity	*Limba*	0.88 (0.74-1.04)	0.119	0.90 (0.76-1.06)	0.206
		Other/mixed	1.15 (1.00-1.33)	0.055	1.12 (0.97-1.29)	0.127
	Profuse haemorrhagic acute disease		0.69 (0.52-0.92)	0.011	0.69 (0.52-0.92)	0.012

Table 5-3 Linear regression demonstrating which participant parameters vary with the log of the antibody index.

Of particular note is that there was a difference in geometric mean antibody index for survivors who suffered the most severe form of acute disease, versus survivors with a milder acute disease (Figure 5-7), although a comparison of all groups did not demonstrate a significant difference (p=0.06). The geometric mean AI was relatively similar for survivors with dry symptoms, wet symptoms or milder haemorrhagic disease during their acute disease process at 3.47 (95% CI=2.77-4.35), 3.31 (95%=3.07-3.57) and 3.51 (95% CI=3.15-3.90) respectively. However, survivors who experienced profuse haemorrhagic disease (defined as bleeding from three or more sites) had a significantly lower geometric mean antibody index at 2.33 (95% CI=1.94-2.80, p<0.001, when compared to those with milder haemorrhagic disease, p=0.001 when compared to all three groups). Linear regression of the log antibody index for those with profuse haemorrhagic disease confirmed this group were significantly more likely to have a lower antibody index than those without profuse haemorrhagic disease (p=0.011),

and this significance was maintained when accounting for potential confounding with age, sex, ethnicity and community type (p=0.012).

Antibody index by symptoms during acute EVD
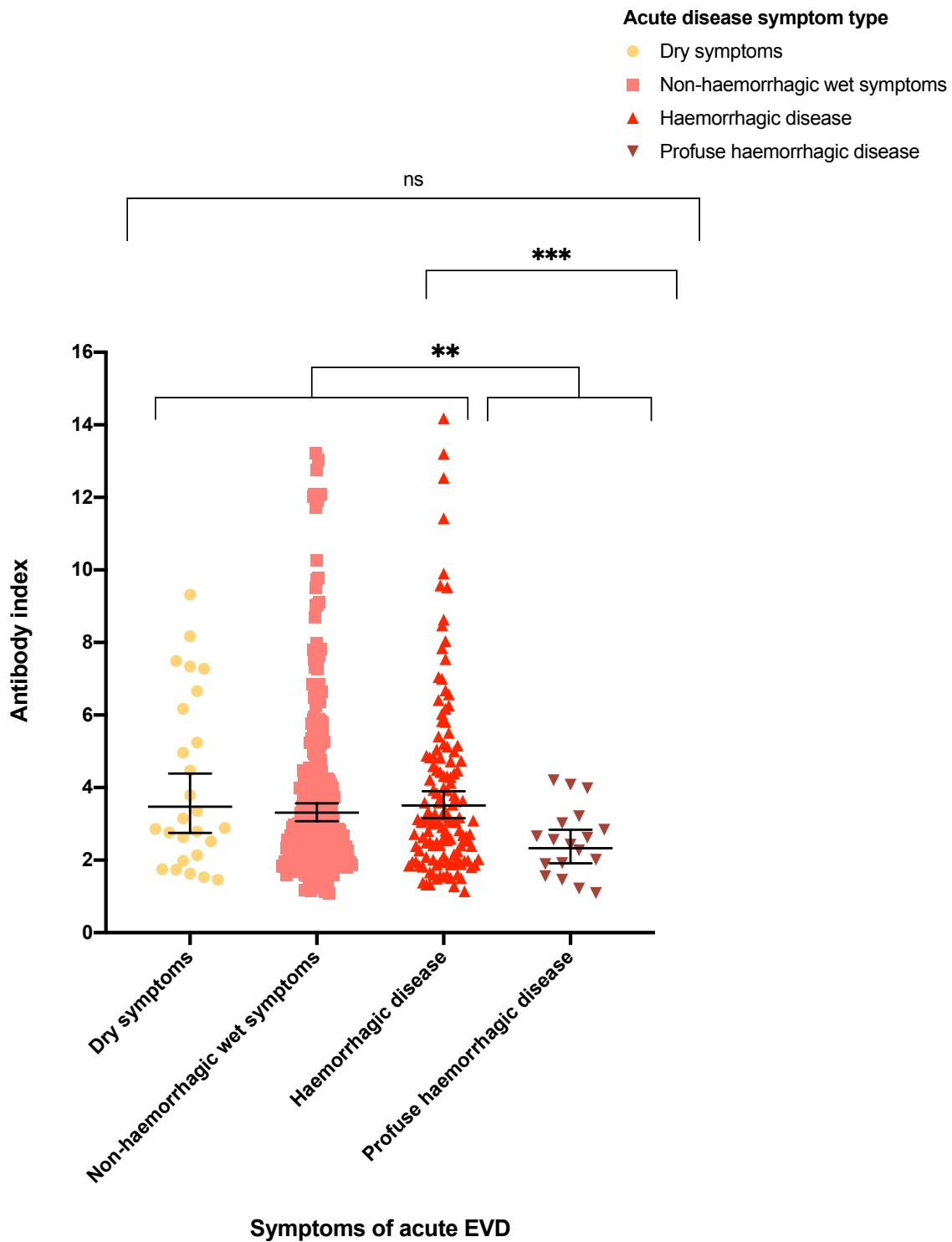
Figure 5-7 Variations in anti-EBOV IgG levels depending on severity of symptoms during acute EVD, with geometric mean and 95% CI. ns=not significant, **significant at p<0.01, ***significant at p<0.001.

5.3.4 Variations in antibody index with Post Ebola symptoms

In chapter 3 different risk factors were described for the development of post-Ebola symptoms and severe post-Ebola symptoms, these were increasing age (PES OR 1.69, CI 1.34-1.92, p<0.001; severe PES OR 1.26, CI 1.08-1.46, p=0.003), female sex (PES OR 1.80, CI 1.18-2.74, p=0.006; severe PES OR 1.53, CI 1.01-2.30, p=0.043), and ethnicities other than Temne (severe PES Limba OR 1.89, CI 1.07-3.33, p=0.028; other ethnicities OR 2.09, CI 1.27-3.44, p=0.004).

Variations were noted in anti-EBOV IgG antibody levels in Ebola survivors with different forms and levels of severity of PES (Figure 5-8), with a one way ANOVA demonstrating a significant difference with p=0.030. The mean AI of 3.13 (95% CI=2.84-3.26) in those with more severe forms of PES was significantly different to the mean AI of 3.70 (95% CI=3.32-4.13) in those with milder PES (p=0.017). Those with no PES however had a mean AI of 3.24 (95%=2.82-3.73), which was just significantly different from those with milder PES (p=0.049). Due to the apparent non-linear differences in mean AI across PES groups, logistic regression was performed which demonstrated that while antibody index was negatively correlated with the prevalence of any type of PES in Ebola survivors, this was not significant (p=0.296 for PES and p=0.129 for severe PES).

Antibody index by severity of PES
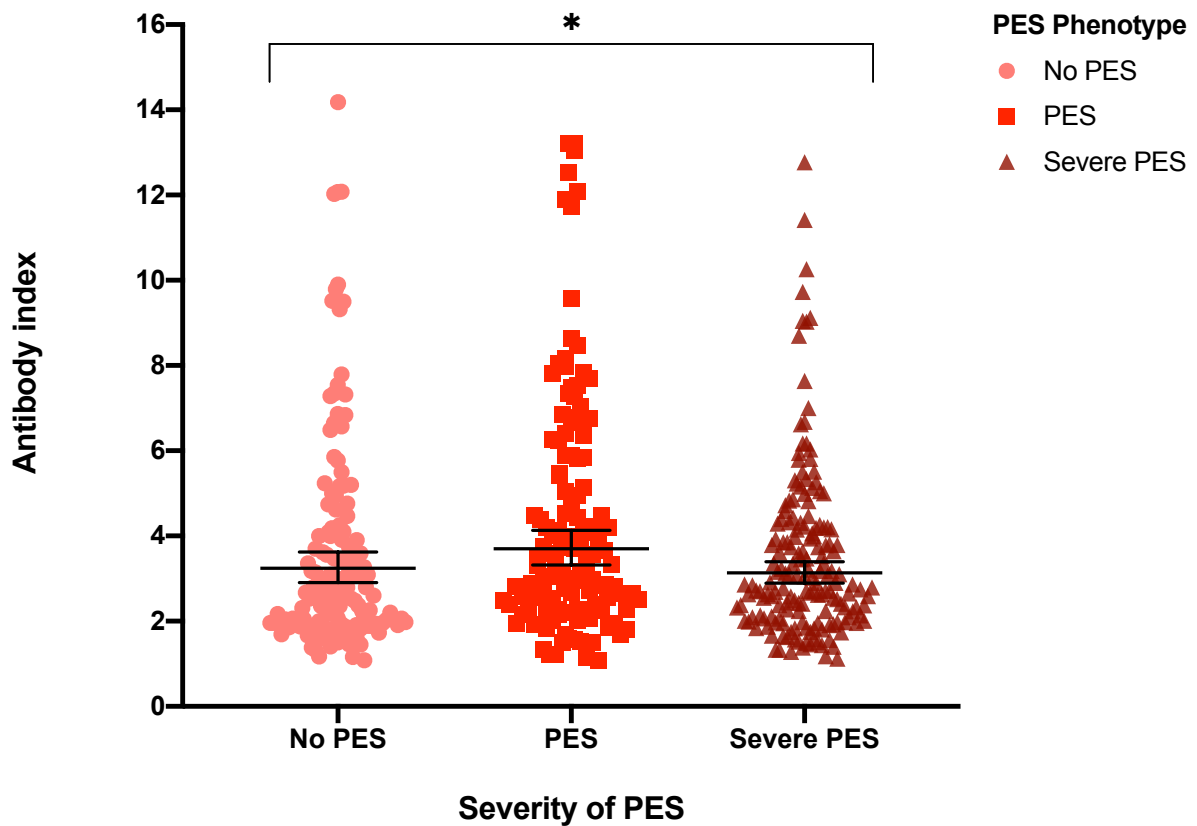
*Figure 5-8 Variation in anti-EBOV IgG by presence/severity of PES, with geometric mean and 95% CI. While there is significant variation between groups, logistic regression demonstrated that a negative correlation of antibody index with severity of PES was not significant. *significant at p<0.05.*

If this is broken down to determine if there are differences by the type of PES, differences were observed (Figure 5-9), particularly in relation to the more well defined types of severe PES, such as cataracts/blindness (Figure 5-10) where the geometric mean AI is 2.72 (95% CI=2.22-3.33). This is significant when compared with the mild PES and no PES categories with p=0.044, and also when compared with the no PES, mild PES and severe PES categories (p=0.040). The mean AI for other severe PES types (sleep disturbance, recurrent headaches, regular joint pain and fertility problems) ranged from 3.02-3.17, and for the severe uveitis and complex neuro forms of PES was 3.31 and 3.96 respectively.

Antibody index in types of severe PES
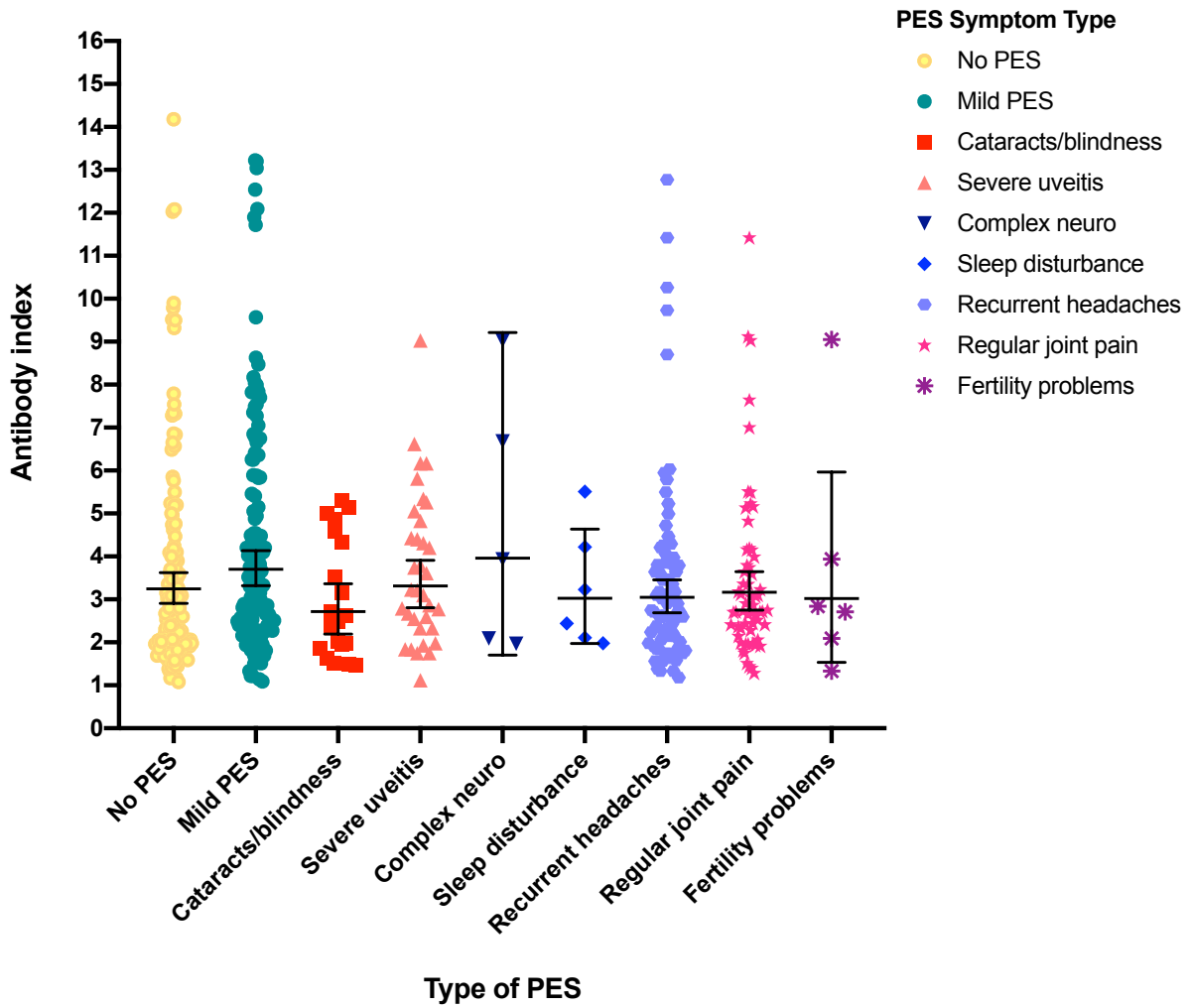
Figure 5-9 Variations in anti-EBOV IgG level in different forms of PES, with geometric mean and 95% CI.

Comparison of antibody index by severity of PES and severe eye disease
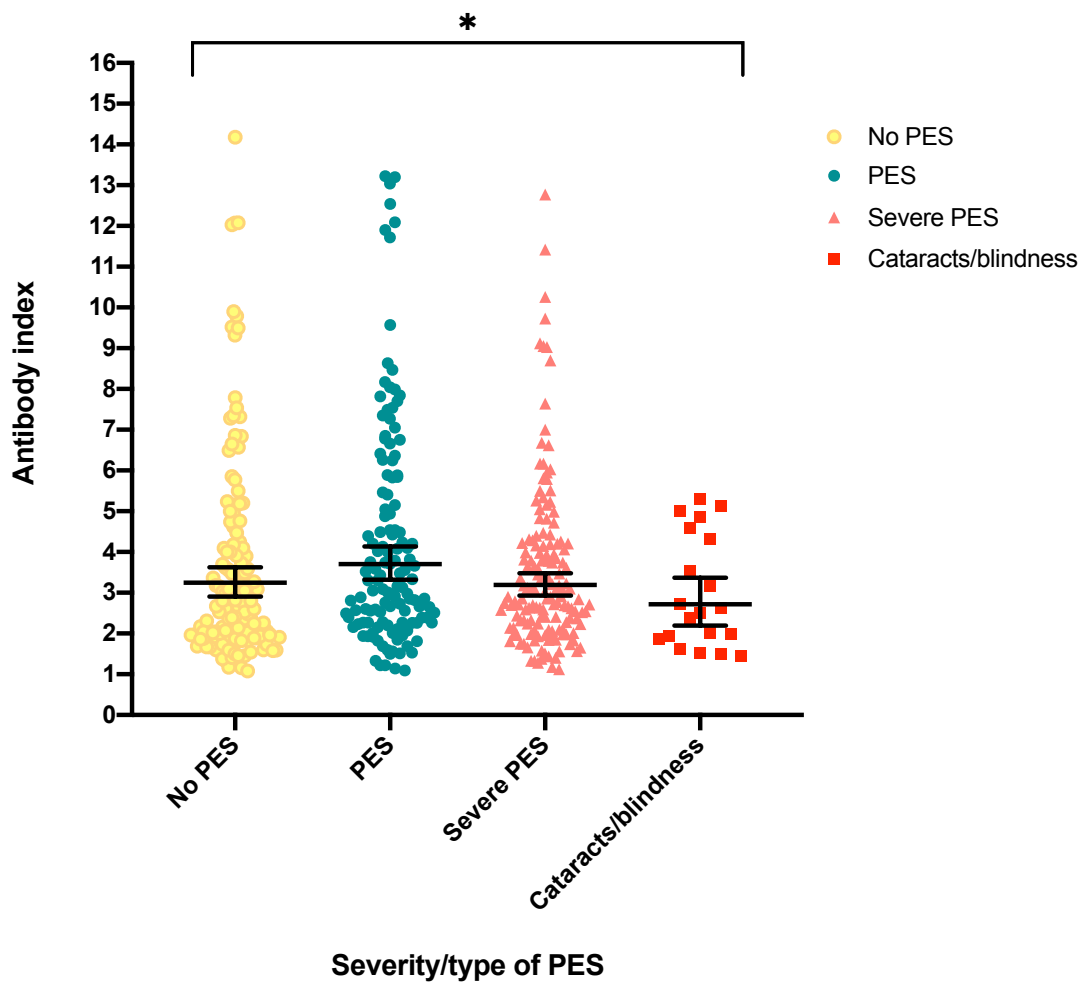
*Figure 5-10 Comparison of anti-EBOV IgG level in cataracts/blindness versus the different severities of PES, with geometric mean and 95% confidence interval. *significant at $p<0.05$.*

The suggestion of the antibody index being lower in more severe acute disease and in some severe forms of PES is reflected by the findings of the antibody index in those whose PES had yet to resolve or improve at the time of recruitment. While Table 5-4 and Figure 5-11 below show a downward trend in this regard, this correlation was not significant with $p=0.291$.

PES status	Mean (geometric) antibody index	95% Confidence interval
Resolved	3.67	3.10-4.16
Improving	3.35	3.02-3.53
Continuing the same/worsening	3.11	2.42-3.80

Table 5-4 Mean (geometric) antibody index and 95% CI by status of PES symptoms.

Antibody index by current PES status
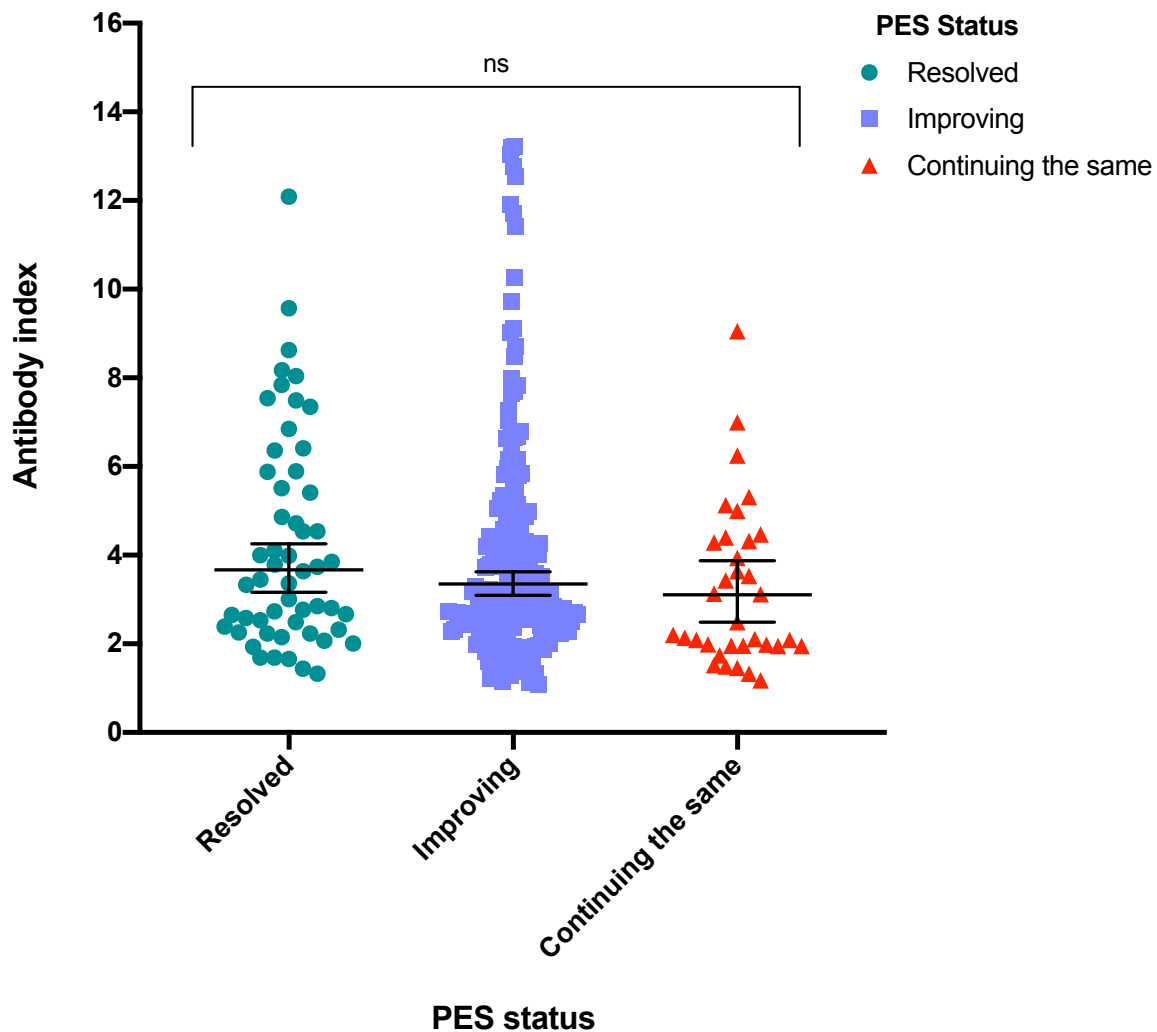
Figure 5-11 Variations in anti-EBOV IgG by PES status at the time of participant recruitment, with geometric mean and 95% CI. ns = not significant.

Given variations in antibody index across a variety of parameters, it seemed prudent to conduct a comparison of the geometric mean antibody index by age versus the prevalence of PES and severe PES by age (Figure 5-12) as they appeared to be negatively correlated, with antibody index decreasing with age and prevalence of PES increasing with age (as described in chapter 3). Both PES and severe PES were positively correlated with age with $p<0.001$ and $p=0.003$, respectively. The odds ratios described in chapter 3 (Table 3-10) correspond to an increase in the odds of developing PES by 60% for every 10 years of life gained and an increase in the odds of developing severe PES by 25% for every 10 years of life gained. While antibody index appeared negatively correlated with age, this was not significant ($p=0.096$), similarly while the antibody index appeared negatively correlated with overall rates of PES and severe PES, these results were not significant. Figure 5-12 shown below indicates that:

1. Mean anti-EBOV IgG antibody level appears to decrease with age (although this is not significant, p=0.096)

2. Post-Ebola syndrome prevalence increases with age (PES p=<0.001, severe PES p=0.003)

3. Anti-EBOV IgG antibody level decreases as prevalence of Post-Ebola syndrome increases, although this is not significant (p=0.296 for PES and p=0.129 for severe PES).

PES and Geometric Mean AI (with 95% CI) by age
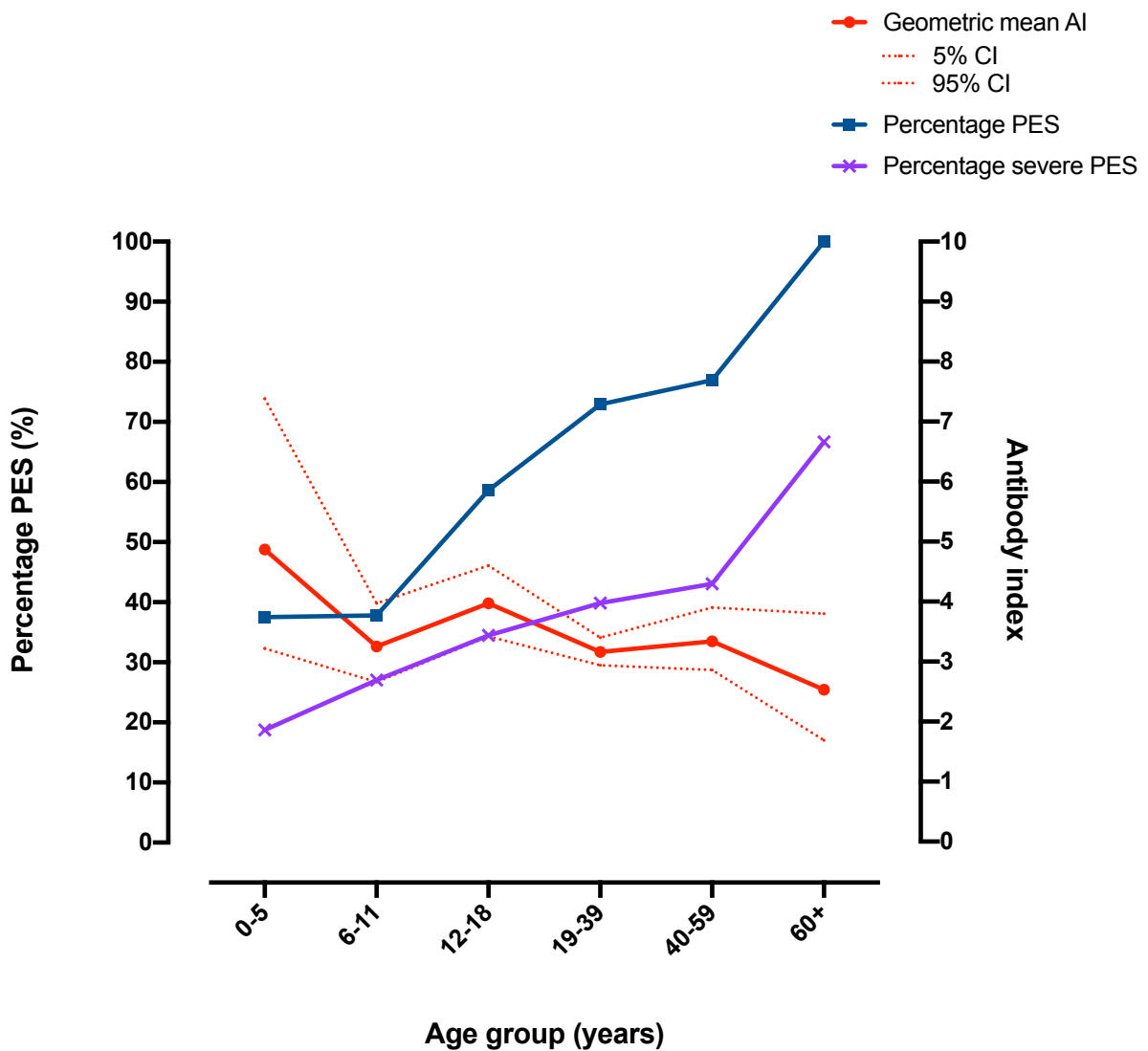
Figure 5-12 Prevalence of PES and severe PES increasing with age (p<0.001 and p<0.01 respectively) while anti-EBOV IgG antibody level decreases with increasing prevalence of PES/severe PES (not significant) and increasing age (not significant).

5.3 Discussion

5.3.1 Sero-reversion of Ebola survivors

From both Table 5-1 and Figure 5-1 it can be seen that there is variation in the anti-EBOV IgG result from Ebola survivors in 2015 when compared with the same survivors in 2017, however this variation is not consistently lower. The 2017 result was lower in six participants, almost the same in three participants and higher in six participants. Based on this very small sample set it cannot be deduced that Ebola survivors consistently sero-revert, although it is possible that some may. Further longitudinal studies are required to determine if sero-reversion is a feature in certain groups of survivors.

As sero-reversion was not confirmed other explanations were needed to account for the 11% of EBOV survivors identified to be sero-negative. As the EBOV assay was known to be highly sensitive and specific (>95% and >99%,[120, 132]) it was unlikely this was due to a lack of sensitivity in the assay. It is possible that a small proportion of the negative survivor results were as a result of intra and inter-assay variability, however this is unlikely to account for 11% of recruited Ebola survivors being anti-EBOV IgG negative and the intra and inter-assay variability was within acceptable limits for this type of assay. Despite the adjustment of the cut-off values for the assay reducing the sensitivity to 88.7%, 55 of the 56 Ebola survivors that tested negative for anti-EBOV IgG were definitively negative and did not show borderline results which may have resulted in the calling of false negatives. It was therefore assumed that these negative results were likely to be true negative results. All recruited Ebola survivors were required to provide an Ebola survivor certificate at recruitment, and if this had been forgotten or misplaced the chair of the local survivor's association was required to verify that at some stage, they had witnessed an Ebola survivor certificate for the participant. This method, while prone to potential abuse, was the optimum method achievable for confirming a survivor's status, as it was not possible to trace identifiable data from Ebola treatment facilities on patients who had survived the disease. It was noted during recruitment and while conducting the EIAs that there seemed to be a predominance of seronegative 'Ebola survivors' from communities in one particular district and who had been managed at treatment facilities in and around Port Loko town. This is demonstrated in Figure 5-2, which reveals that of the seronegative survivors, 66% were from treatment facilities in Port Loko, whereas for the seropositive survivors only 33% were from treatment facilities in Port Loko. Similarly, there were differences between these two groups in relation to the prevalence of PES, with seronegative survivors experiencing much lower rates than seropositive survivors (39.3% versus 67.9%). From this data it would suggest that not all of the seronegative 'Ebola survivors' are true Ebola survivors, despite the majority of them presenting survivor certificates.

There could be several reasons for this, firstly there may have been confusion over diagnostic tests within treatment facilities and the wrong patient attributed a confirmed diagnosis. Alternately it may have been that Ebola treatment facilities were issuing certificates to individuals who were not true Ebola survivors at discharge or that some individuals fraudulently produced Ebola survivor certificates for themselves. That some individuals were fraudulently claiming to be Ebola survivors seems the most plausible explanation as Ebola survivors are entitled to free medical care and handouts of money, clothing and food from the government and NGOs. In Port Loko district in particular one NGO was providing very generous financial donations to Ebola survivors.

5.3.2 Decline in antibody level with age and severity of acute disease

The variations in antibody index and therefore anti-EBOV IgG against certain parameters are interesting, particularly the variations in relation to age and disease severity in acute disease. While it may be expected that children may have higher antibody levels overall, as they have a primed and active immune system[142], the general downward trend of antibody level with age is interesting although not significant (p=0.096). This could be accounted for by the low numbers of participants in the two youngest age groups, which are representative of the fact that young children were less likely to be affected during the West African epidemic[22, 38, 122], and that children under the age of 5 had much higher mortality rates than older children and adults[22]. Subsequently there are far fewer child Ebola survivors, particularly in those under the age of 5 years. In relation to disease severity, older adults were more likely to have negative disease outcomes than young to middle age adults, both in respect of disease severity and mortality, but also in respect of prevalence of PES in Ebola survivors (as described in chapter 3). It might be assumed therefore that they would have higher rather than lower antibody levels than teenagers or younger adults as it might be expected that someone with a more severe acute disease would have had a higher viraemia, a longer disease course and as such a potentially higher antibody response. This finding is similar though to the significant finding of a lower level of antibody in those who suffered with profuse haemorrhagic symptoms during their acute disease (p=0.012). Both of these findings raise questions about the pathogenesis of EVD and the relationship of anti-EBOV IgG levels to disease severity.

5.3.3 Antibody level and Post-Ebola syndrome

The results of anti-EBOV IgG levels in Ebola survivors with severe PES reinforces the suggestion earlier in the chapter that individuals with a more severe acute disease process were more likely to have lower antibody levels. In line with this a relationship is also apparent between the prevalence of PES and the severity of acute disease. Those survivors who suffered from haemorrhagic symptoms or profuse

haemorrhagic disease experienced higher rates of PES/severe PES (73.4% and 83.3% respectively) than those who did not experience haemorrhagic disease (61.5% for those with dry symptoms and 64.5% for those with wet symptoms). While it appeared that Ebola survivors with the severe form of PES were more likely to have lower antibody levels, the trend across PES groups did not demonstrate declining antibody index with PES severity. However when breaking PES down into specific phenotypes, a significant result was apparent in relation to survivors with cataracts or blindness (p=0.04), the most concrete severe diagnosis it was possible to determine in the context of self-reporting of symptoms. While not significant, there is the suggestion of increasing antibody level in relation to the resolution of PES, where those with continuing or worsening symptoms had lower antibody levels than those with improving symptoms and those whose symptoms had resolved. It must be noted however that this result was not statistically significant.

Overall there is the suggestion that those with severe acute disease or severe ongoing problems had a tendency towards lower antibody levels than those with milder disease. This is underscored by the latter figure (Figure 5-12) demonstrating that lower antibody levels in older age groups correlate with prevalence of PES and severe PES in these age groups, although this correlation was not significant.

As the data is dependent on self-reporting of symptoms and disease severity it must be acknowledged that there could be several potential confounding factors, including the element of subjectivity in self-reporting and the possibility of overinflating symptoms. These factors may also account for discrepancies in the data such as the geometric mean antibody level for milder PES being higher than that of experiencing no PES.

In summary the data demonstrate the following:

1. Anti-EBOV IgG antibody level appears to decrease with age (Figure 5-3 and Figure 5-12), although this was not significant (p=0.096)
2. Post-Ebola syndrome prevalence increases with age (Figure 5-12), which was significant at p<0.001 for PES and p=0.003 for severe PES.
3. Anti-EBOV IgG levels are lower in those who experienced the most severe acute disease (Figure 5-7), p=0.012.
4. Anti-EBOV IgG antibody level appears to decrease as prevalence of Post-Ebola syndrome increases (Figure 5-12), although this is not significant.

The implication of this data is that those who experience a more severe disease phenotype have lower anti-EBOV IgG antibody levels. As mentioned above it was anticipated that Ebola survivors with severe disease or ongoing problems would be more likely to have higher antibody levels due to longer persistence of virus and thus longer stimulation of the immune response and antibody production. In contrast it appears to be the opposite. This could be explained by Ebola survivors with severe ongoing problems having a poorer initial immune response to the infection with more limited antibody production resulting in a higher viraemia and a more severe disease process with persistent inflammation or end organ damage. An alternative possibility is that the very severely affected patient's populations of T and B cells are destroyed by the virus resulting in poor long term antibody production.

While this data requires further validation to understand the relationship between lower antibody levels and severe disease, recently published findings from a large cohort of survivors in Liberia demonstrate decreasing antibody level with age[60]. However the authors of this paper did not identify differences in antibody level with individual symptoms and did not conduct an analysis of antibody level versus symptoms grouped as PES, nor stratify for symptom severity[60].

5.3.4 Implications of the antibody results and phenotypes identified for genetic analysis

While these findings require further validation, they do help to define different phenotypes for the genetic analysis and add further weight to the hypothesis that there is a genetic predisposition to susceptibility to, and severe outcome from, EVD.

The literature and data thus far suggest a likely genetic component to EVD susceptibility and severity by demonstrating that:

a. EVD is a severe disease process with high mortality rates, yet there are people who survive the disease despite limited access to supportive care, some with only mild symptoms.

b. There are different, but very characteristic, modes of death described for those who succumb to EVD.

c. There is asymptomatic/minimally symptomatic Ebola virus infection in undiagnosed people in affected communities.

d. There are people highly exposed to Ebola virus who do not become infected.

e. There are variations in antibody levels in Ebola survivors, with those with more severe acute disease demonstrating lower antibody levels.

The studies on seropositivity in the household contacts and community controls have provided clear evidence for a population of EBOV exposed individuals, some of whom show no evidence of infection, while others show evidence of infection but remain asymptomatic or minimally symptomatic. The differing antibody levels in Ebola survivors with the most severe forms of acute disease raise the possibility of genetic factors influencing each step in the evolution of the disease and progression from exposure to severe illness, recovery, persisting symptoms or death. Based on the antibody results several groups of patients can be identified in which to explore different types of hypothesised genetic effect:

1. The highly exposed, seronegative, asymptomatic group are naturally resistant to infection and may carry gene variants that prevent infection (such as absent or mutated receptors required for viral entry).
2. The seropositive but asymptomatic/minimally symptomatic group are resistant to disease but not resistant to infection and may carry gene variants that permit viral entry to the cell but limit replication.
3. The severely infected group who have low anti-EBOV IgG antibody levels may have gene variants that delay, or limit effective acquired immune responses.
4. These same low antibody responders may be more genetically predisposed to developing severe PES.

These groups will be compared, alongside a deceased group of patients, through exome sequencing and GWAS studies in chapters 6 and 7.

Chapter 6 Identifying rare variants associated with susceptibility to Ebola virus disease

6.1 Introduction

In chapter one the use of exome sequencing was described in identifying rare Mendelian variants affecting susceptibility to infectious diseases. Exome sequencing is a powerful tool for identifying rare and highly penetrant genetic defects underlying disease. Extreme presentations of EVD (such as complete resistance to infection following high level exposure) are particularly suited to this type of analysis, as they are most likely to be due to rare mutations directly affecting protein function in the participants concerned. As mentioned in chapter 1, no large scale genetic studies had previously been conducted of EVD in humans, therefore it was not possible in advance to identify specific candidate genes or gene pathways to look for. Given molecular findings in lipid pathways however, it was postulated that candidate genes or gene pathways might be highlighted through conducting a broad analysis of exome sequencing data for different disease phenotypes. In this chapter the exome sequencing conducted on 250 participants is described with a broad preliminary analysis conducted of different extreme phenotypes of Ebola susceptibility, resistance and disease severity, from the recruited participants and deceased patients. The findings from the preliminary analysis are highlighted demonstrating variants initially identified through family studies, which were then further analysed using gene burden and pathway analysis. These identified variants are then placed in a biological pathway which suggests that the cholesterol metabolism pathway plays a key role in the pathogenesis of EVD. The advantages and limitations of both candidate gene analysis and pathway analysis are described in chapter one, but both are dependent on an underlying knowledge of the disease process and the structure of potential pathways involved. They have the advantage though of identifying significant variants that may not have been identified in a more broader analysis due to lack of statistical power or small effect size[128, 143].

In earlier chapters the different disease phenotypes were described that would be investigated through the exome and genotyping analysis. These disease phenotypes include deceased and surviving participants, including those who survived with post-Ebola symptoms and those who survived with no ongoing problems, as well as antibody positive individuals with subclinical disease, highly exposed antibody negative participants and unexposed controls. A further refining of these disease phenotypes into specific classifications utilised for the exome analysis is described below along with the specific numbers of each disease phenotype selected for the exome sequencing.

6.2 Methods

6.2.1 DNA quantification

As stated in chapter 3 the DNA was extracted from all recruited samples using an automated DNA extractor, the *QIAsymphony*, by myself and my colleague Samuel Alberman. The extracted DNA was quantified on a *QIAxpert* instrument manufactured by *Qiagen*. The *QIAxpert* uses spectrophotometer-based technology to quantify DNA, similar to the *Nanodrop 2000* (*Thermo Scientific*), but with a more robust platform that utilises 2ul of sample filtered through a capillary to the point at which it is read. Each cartridge can take 16 samples, which makes quantification significantly quicker than on a *Nanodrop 2000*. The literature from *Qiagen* demonstrates it to be more accurate than the Nanodrop and as accurate as the *Qubit 3* (*Invitrogen*) (as per manufacturer's information, see Appendix D).

The majority of the extractions on the *QIAsymphony* were successful with adequate DNA yields achieved. On initial runs it became apparent that the concentration of DNA was too low. In order to reduce the need to concentrate samples at a later stage, the final elution volume was adjusted to 150ul instead of 200ul. This ensured concentrations over 30ng/ul, with the majority of samples achieving concentrations of >50ng/ul. The extraction protocol for the OG-575 samples, which were only 1.5ml in total, extracted from just 350ul of sample. The DNA yield for these samples was too low, subsequently these samples were run on the OG-500 protocol instead, which extracted from 1ml of sample. This ensured an adequate DNA yield but meant that in some cases there was not a lot of sample left if extraction failed.

If extraction failed and there was sufficient sample left, a repeat run was performed on the *QIAsymphony*. If extraction failed again, a manual extraction was performed using the *DNA Genotek PrepIt2* manual extraction protocol (as per manufacturer's instructions, see Appendix D), which is an ethanol precipitation based procedure. Following this process only one sample out of 1525 household contact and survivor samples failed to provide an adequate yield of DNA. Out of the 1004 Community control samples 71 did not provide an adequate yield of DNA, however a manual extraction procedure was not performed on these samples as there were sufficient numbers of control samples.

6.2.2 Sample concentrating and plating

As mentioned earlier many of the biobank samples yielded a low DNA concentration due to sample degradation. It was therefore necessary to concentrate the samples in order to have an adequate concentration for exome sequencing and genotyping. This concentration is ideally greater than 30ng/ul. Using a vacuum concentrator (*Eppendorf Concentrator Plus*) all the low yield samples were

concentrated down by myself and Samuel Alberman. After concentrating the samples, 317 of the 492 biobank samples had an adequate DNA concentration for sequencing. This consisted of 75 of 145 EVD survivors and 242 of 345 EVD deceased from the biobank samples.

Once the biobank samples had been concentrated and re-quantified, the samples were plated according to a plate template developed by myself for the exome sequencing, and then again separately for the genotyping. All plates had a mixture of samples from each of the phenotypes under investigation, to ensure that if a plate failed during sequencing no one specific sample type was adversely affected. Three plates were sent for exome sequencing, an example of the exome sequencing plate template is shown in Appendix C, along with the plate template for the SNP GWAS. At the request of the sequencing facility, the three plates for exome sequencing were also treated with RNase prior to sending to the facility to prevent any errors in sequencing caused by RNA contamination.

6.2.3 Classifying disease phenotypes

The underlying hypothesis for the work in this chapter was that individuals who carry rare genetic variants may explain differences in response to infection with Ebola virus. While the effect of common variants can be detected by the GWAS approach (chapter 7), rare variants (allele frequency less than 1%) might explain extreme phenotypes. Given the high mortality rate of EVD, the investigation of families with multiple members either resistant to infection or resistant to disease can help elucidate the genetic architecture of the host response to Ebola.

To optimise the possibility of identifying rare genetic variants it is important to clearly define the disease phenotypes of the samples to be sequenced and, where possible, identify the most well defined phenotypes of disease. As there was little data available on potential targets to be identified, nor allele frequency of such targets in the Sierra Leone population, an apparently unexposed, anti-EBOV IgG negative control group was included to determine the allele frequencies that might be expected within the background population. This was to ensure any variants identified were not identified erroneously due to increased allele frequencies of this type within the study population. The phenotypes identified for exome sequencing are shown in Table 6-1 and then further described below:

Sample type	Defining criteria
Ebola deceased	Haemorrhagic symptoms +/- CT value<18
	CT value<18
Ebola survivor	Severe, defined PES
	No PES
Household contact	Anti-EBOV IgG positive
	Anti-EBOV IgG negative, risk level 1 exposure
	Anti-EBOV IgG negative, risk level 2 or 3 exposure
Community control	Anti-EBOV IgG positive
	Anti EBOV IgG negative, risk level 8 exposure

Table 6-1 Disease phenotypes identified for exome sequencing

Deceased cases were by definition considered to have suffered from severe disease. However, within the deceased group samples were selected where there was sufficient clinical information to further characterise them as a severe disease phenotype. This was done initially by selecting biobank samples where data from treatment facilities confirmed the patient had haemorrhagic disease as part of their acute disease process. Following this, biobank deceased samples were identified as having severe disease based on the CT value of their diagnostic PCR sample, as it is well known that patients with a high viral load experience a more severe disease process[21, 39]. While it would have been optimal to compare groups with the modes of death described in the introduction[18], clinical data was not available to sufficiently categorise samples into these groups. Data on haemorrhagic symptoms however was relatively reliably reported, at least where such symptoms had been overt. Therefore, deceased biobank samples were selected on the basis of having one or both of the following parameters:

1. Haemorrhagic disease

2. A high viral load (a CT value<18).

Ebola survivor study participants were anti-EBOV IgG positive and categorised into two groups to determine possible genetic effect:

1. No ongoing problems

    a. This group may have an immune response that successfully contained and cleared the virus.

2. Well defined, severe ongoing problems such as cataracts, blindness or complex neurological problems (paralysis or new onset seizure disorders).

    a. This group potentially had a sub-optimal immune response that might permit persistent infection or result in inflammation.

Household contact study participants were categorised into two groups

1. Those who were anti-EBOV IgG positive

    a. This group may be innately resistant to symptomatic disease but not to infection.

2. High level exposure (risk level 1-3) but anti-EBOV IgG negative

    a. This group are presumed to be innately resistant to infection.

Lastly community control study participants were also categorised into two groups:

1. Those who were anti-EBOV IgG positive

    a. This group may also be innately resistant to symptomatic disease but not to infection.

2. Those who were Anti-EBOV IgG negative with an exposure risk level of 8 (no known exposure).

    a. This is the actual control group with no known exposure, to be used as a population control.

6.2.4 Exome sequencing

The exome sequencing was conducted at the Wellcome Centre for Human Genomics in Oxford, who were able to provide an automated service. All samples passed the quality control, although a few had low yields. Following discussion with the project manager it was agreed to proceed with the *NimbleGen SeqCap EZ Library* for exome library preparation and all samples were successfully sequenced on an *Illumina HiSeq*, with a read length of 75bp paired end and an average coverage depth of 40x on-target. The complete data set was provided on the 21st December 2018 and as such only a preliminary analysis has been included within this thesis. The samples sent for exome sequencing totalled 250, which consisted of:

Sample type	Total no. samples	No. samples selected as family members
Household contact (study participants)	86	31
Ebola survivor (study participants)	50	2
Ebola deceased (biobank samples)	50	0 (relationships unknown)
Community control (study participants)	64	0
Total	250	33

Table 6-2 Sample types sent for Exome sequencing.

The breakdown of samples sent for exome sequencing by disease phenotype are shown below in Table 6-3:

Sample type	Defining criteria	Number
Ebola deceased	Haemorrhagic symptoms +/- CT value<18	25
	CT value<18	25
Ebola survivor	Severe, defined PES	26
	No PES	24
Household contact	Anti-EBOV IgG positive	36
	Anti-EBOV IgG negative, risk level 1 exposure	32
	Anti-EBOV IgG negative, risk level 2 or 3 exposure	18
Community control	Anti-EBOV IgG positive	18
	Anti EBOV IgG negative, risk level 8 exposure	46

Table 6-3 Disease severity definitions and sample numbers sent for exome sequencing.

6.2.5 Identifying families

As mentioned above initial analysis would involve a broad preliminary analysis to determine whether any candidate genes or gene pathways were apparent for which a candidate gene or pathway analysis could then be undertaken. To refine this preliminary analysis and improve the chances of identifying specific genes of interest it was decided to conduct a family analysis, utilising related family members who had similar disease phenotypes or where there had been several different extreme phenotypes identified within a family. Where possible this was done by looking at trios, three members of a family that had been affected either with the same disease phenotype or different extreme phenotypes, and who were ideally non-immediate family members, therefore sharing a smaller proportion of each other's genetic material.

While it was not possible to identify related samples among the biobank samples in advance of the exome sequencing due to limitations in the data recorded during the epidemic, it was possible to identify related samples in the study participant categories based on the data collected during participant recruitment. Subsequently where there was a household contact who was anti-EBOV IgG positive and there was a non-immediate family member who shared a proportion of DNA (such as a nephew/niece, uncle/aunt, cousin or grandparent) but was either an Ebola survivor or an anti-EBOV IgG negative, high risk exposure (level 1-3), they were included in the exome sequencing in order to undertake an analysis of specific families or trios. This was also done where there were known Ebola survivor relationships or prospectively recruited Ebola survivors who could be linked to household contacts. Through this method thirteen families were identified. All potentially identified relationships were then confirmed following the exome sequencing. From the exome data it was determined that some individuals who reported being related were not biologically related and that some separately recruited families were biologically related. These confirmations from the exome data resulted in:

1. The exclusion of three of the family groups

2. The division of one family group into two groups

3. The merger of two family groups into one group

Following the exome sequencing it also became possible to identify previously unknown relationships, based on there being evidence of a shared proportion of DNA. Specifically, this was done by calculating the kinship coefficient[144] for all possible sample pairs, which ranges from 0 for unrelated individuals to 1 for monozygotic twins. To further unravel familial relationships, the proportion of genomic sites at which two individuals share no alleles was calculated. This ranges from 0 for parent-offspring pairs to 1 for unrelated individuals. Through this process a further four families of household contacts/survivor study participants were identified for analysis purposes, as well as two first cousins in the deceased biobank samples. Two survivor study participants were also identified as being related. This resulted in a total of 16 pedigrees being identified; fourteen relationships within household contact/survivor families, one relationship between two Ebola survivor study participants, and one relationship of first cousins in the deceased biobank samples. The pedigrees for these relationships are shown in Figure 6-1, with titles for each pedigree suggesting the protective or detrimental mechanism being displayed. The potential gene variants identified are also listed by pedigree number in Table 6-5 and explained in further detail.

6.2.6 Identification of rare variants

The genomic analysis was carried out by Dr Evan Bellos with whom I collaborated closely in defining the hypotheses and analytic approach to each family and phenotype group. The raw sequencing data was aligned to the human reference genome (version GRCh37/hg19) using the Burrows-Wheeler alignment (BWA) tool[145] and then joint genotyping of the samples was performed using the Genome Analysis Toolkit (GATK)[146]. The resulting genomic variant calls were annotated using Ensembl's Variant Effect Predictor (VEP)[147]. These annotations include population allele frequencies (AF) from the gnomAD project[148] as well as the calculated consequence of each variant on protein function. In all downstream analyses, the results were filtered to exclude variants with no apparent effect on the resulting protein (e.g. synonymous variants) and variants annotated to be common in the African population (AF>5%).

As mentioned earlier initial analysis was undertaken using a familial segregation analysis, identification of these families is described above with the families identified listed in the pedigrees in Figure 6-1. As rare, deleterious mutations were expected to have the largest effect on disease susceptibility and severity, the following criteria were used to filter genetic variants within each of the pedigrees:

1. Exclude variants with minor allele frequencies higher than 1% in Africans (based on gnomAD data).

2. Exclude variants that don't impact the resulting protein sequence (synonymous variants).

3. Exclude variants that are predicted to result in "benign" changes in the protein sequence (using SIFT & PolyPhen predictions and a CADD score cutoff of 10).

This process reduced the number of potentially protective variants by an order of magnitude, from a few thousand to a few hundred for each family. Furthermore, in families with highly exposed but seemingly protected individuals, only those variants that segregated with the phenotype (i.e. antibody status and description of symptoms/sequelae) were focussed on and variants that were present in any of the deceased patients were excluded. This allowed further narrowing of the scope of the search for protective variants and enabled manual curation of the few remaining candidate variants. Table 6-4 shows the initial number of variants detected per family pedigree and then the number identified following the filtering criteria described.

Pedigree no.	Initial number of detected variants	Filtered variants segregating with phenotype
1	4692	41
2	6472	13
3	6471	19
4	5737	68
5	4611	61
6	6623	11
7	5817	33
8	5113	45
9	4459	50
10	6317	13
11	5188	32
12	4939	28
13	5827	39
14	4445	71
15	NA	NA
16	4367	88

Table 6-4 Number of protective gene variants identified per family pedigree prior to filtering and after applying the filtering criteria described. NA = not applicable (this was a family of deceased participants, so a different filtering approach was used to identify detrimental gene variants).

6.2.7 Gene wise enrichment and pathway analysis

A gene-wise enrichment analysis of rare deleterious variants in antibody negative versus antibody positive unrelated individuals (38 vs 97 samples respectively) was also performed. Variants were filtered according to the criteria listed above and then aggregated per gene to estimate the mutational burden in resistant versus susceptible individuals. The statistical significance of the results was calculated using permutation testing.

Based on findings within the family analysis and gene-wise enrichment analysis, as well as the significance of the cholesterol transport receptor NPC1 for Ebola viral entry and replication (discussed in chapter 1), it was hypothesised that genetic variants in relation to disease susceptibility and severity may be identified in cholesterol metabolism pathways. On these grounds the mutational burden in selected lipid metabolism pathways obtained from Reactome was examined. Specifically, these were the plasma lipoprotein assembly pathway (similar to the Kegg pathway of cholesterol metabolism shown in Figure 6-2) and the LDL-clearance pathway (a pathway subset of the cholesterol metabolism pathway - Figure 6-3 and Table 6-7).

6.3 Analysis

6.3.1 Family pedigrees identified and their associated variants

1. **Household contacts resistant to infection**

   PCSK9 variant – both HC
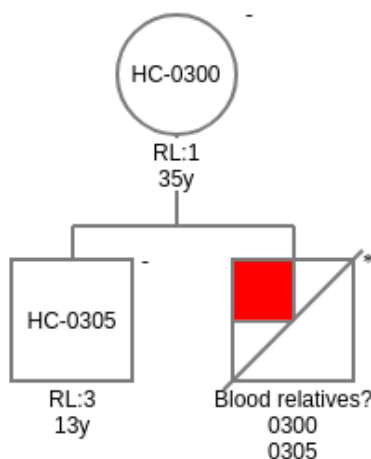
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

◯ Unknown Female -not recruited

◇ Unknown Individual – not recruited

■ Affected, deceased, recruited if PD code

■ Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

■ Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

− Anti-EBOV IgG negative

2. Household contacts resistant to disease, with possible resistance to infection

APOC1 variant – all HC
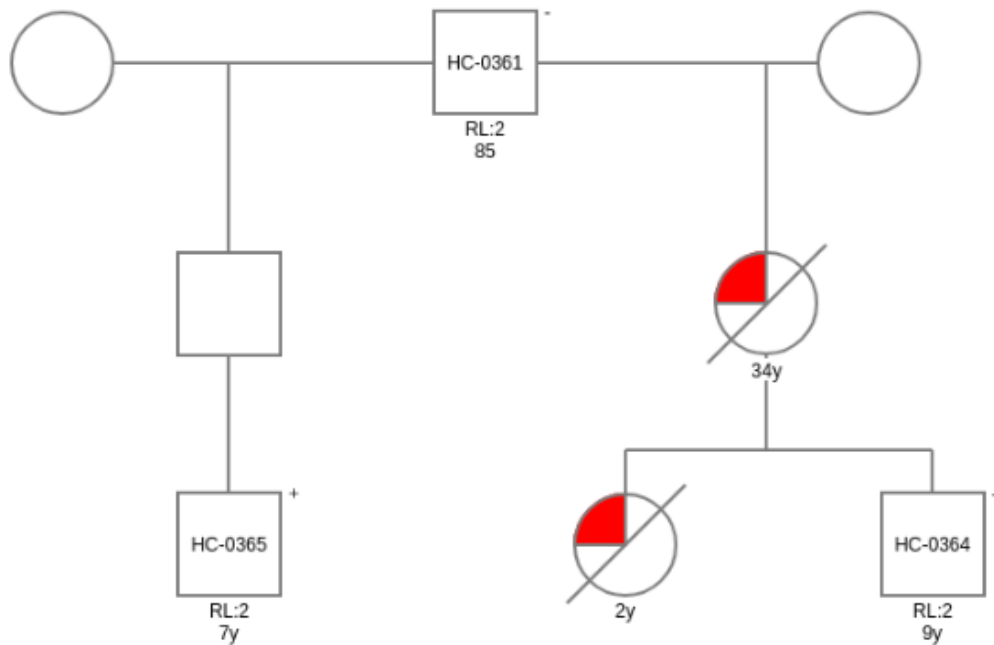
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

○ Unknown Female -not recruited

◇ Unknown Individual – not recruited

🟥 Affected, deceased, recruited if PD code

🟩 Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

🟨 Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

− Anti-EBOV IgG negative

3. Resistant to infection or resistant to disease

SQLE variant – all HC
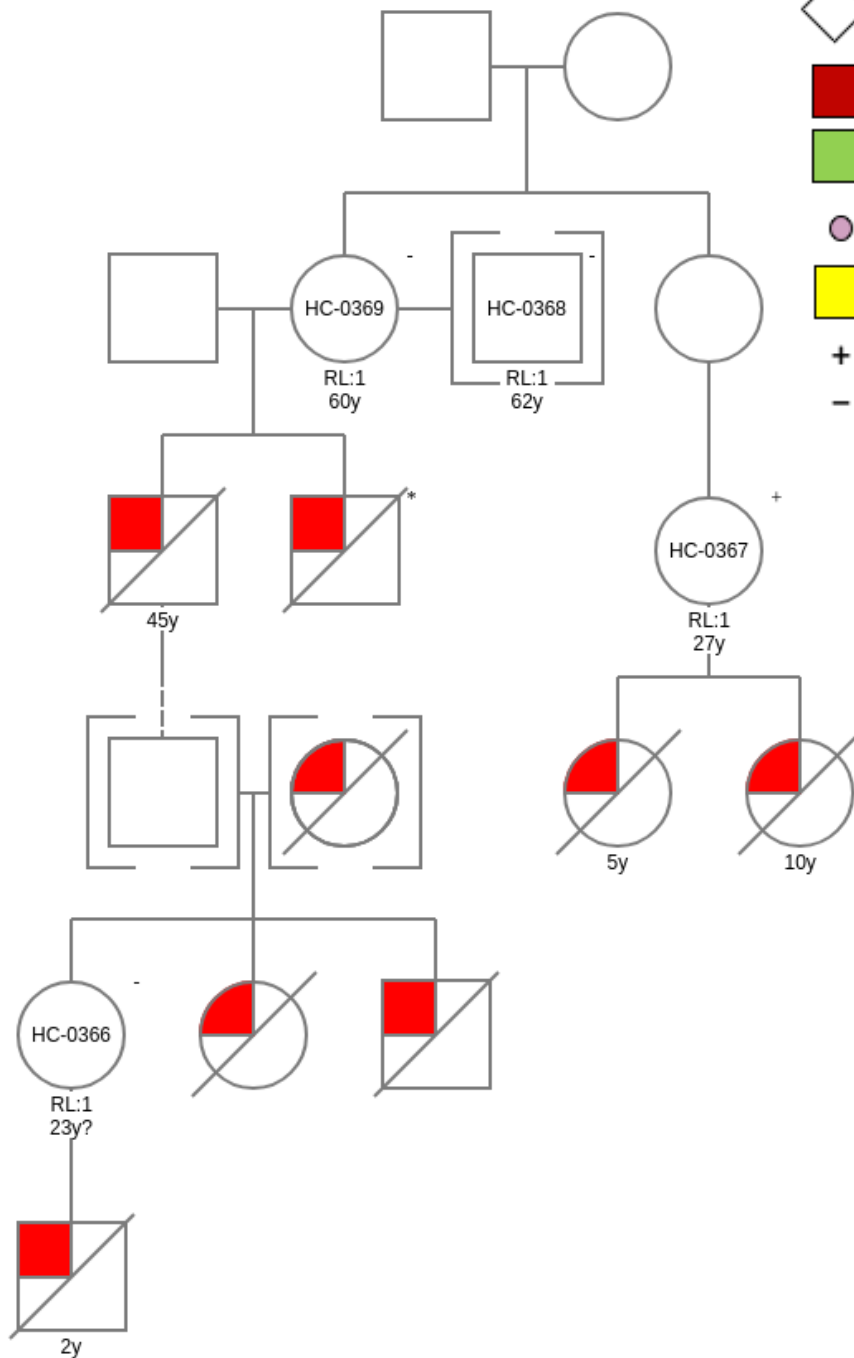
4. Household contacts resistant to disease, with possible resistance to infection

ABCA13 variant (rare frameshift mutation) – all HC

NPC1 and VPS11 variants – HC-0384
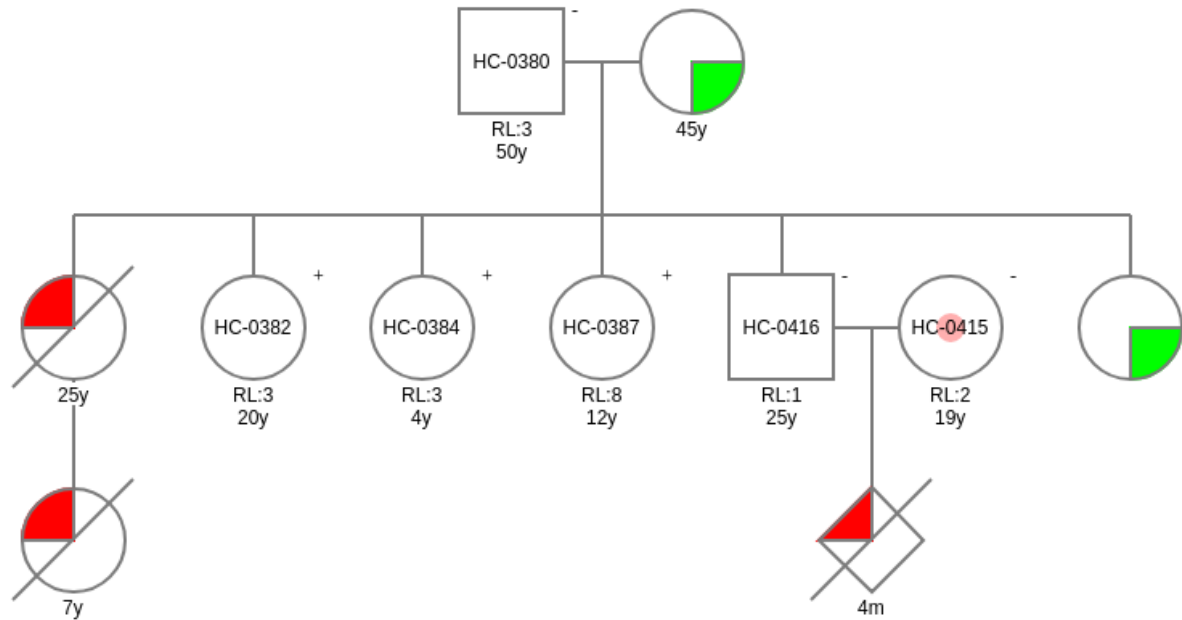
5. Household contacts resistant to infection

No variant identified
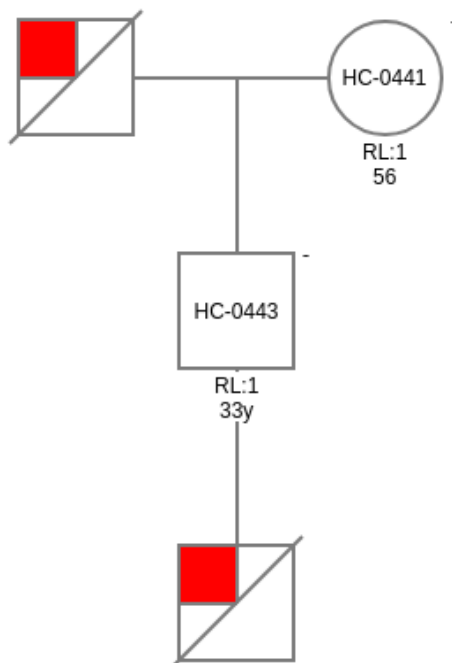
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

◯ Unknown Female -not recruited

◇ Unknown Individual – not recruited

■ Affected, deceased, recruited if PD code

■ Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

■ Conjunctivitis during acute disease

\+ Anti-EBOV IgG positive

\- Anti-EBOV IgG negative

6. Household contacts resistant to infection and disease

   ARSB variant – all HC
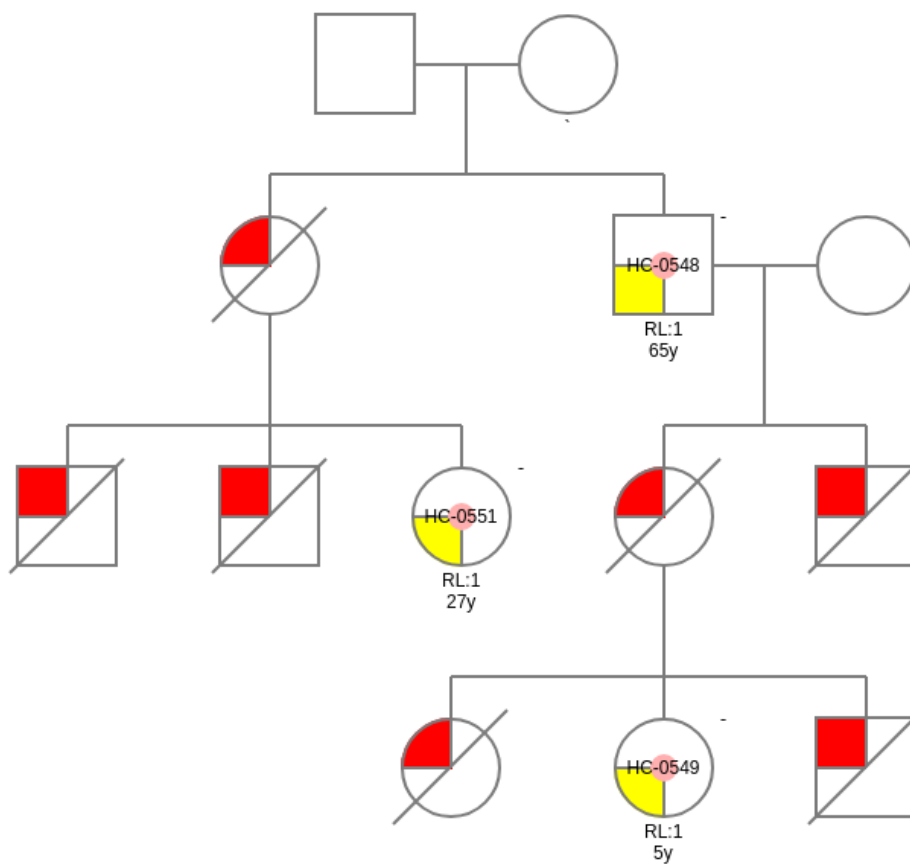
7. Household contacts resistant to disease or infection

   STARD9 and APOC1 variants – all HC
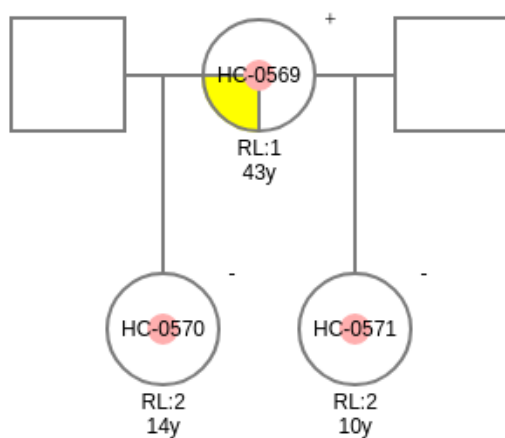
Three tenants living in same house died of Ebola

PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

◯ Unknown Female -not recruited

◇ Unknown Individual – not recruited

🟥 Affected, deceased, recruited if PD code

🟩 Affected, survived, recruited if SS code

🟣 Symptoms meeting EVD case definition

🟨 Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

− Anti-EBOV IgG negative

8. Household contacts resistant to disease or infection

ABCA8 variant – all HC
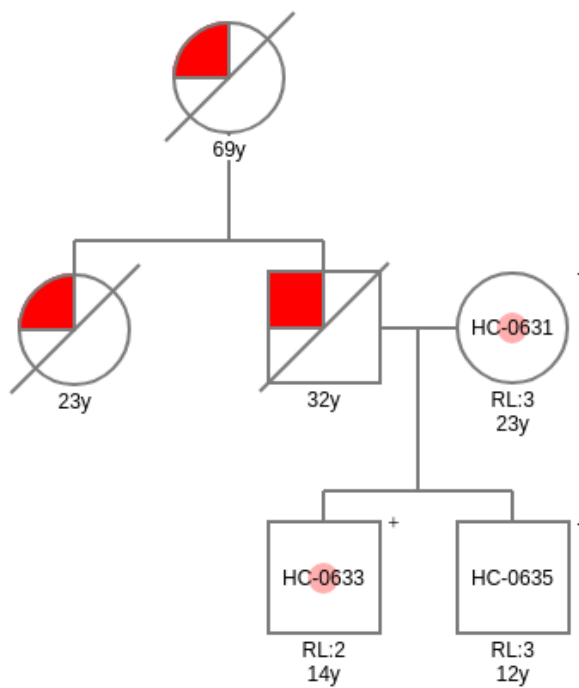
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

◯ Unknown Female -not recruited

◇ Unknown Individual – not recruited

■ Affected, deceased, recruited if PD code

■ Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

■ Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

– Anti-EBOV IgG negative

9. Household contacts resistant to disease or infection

No variant identified
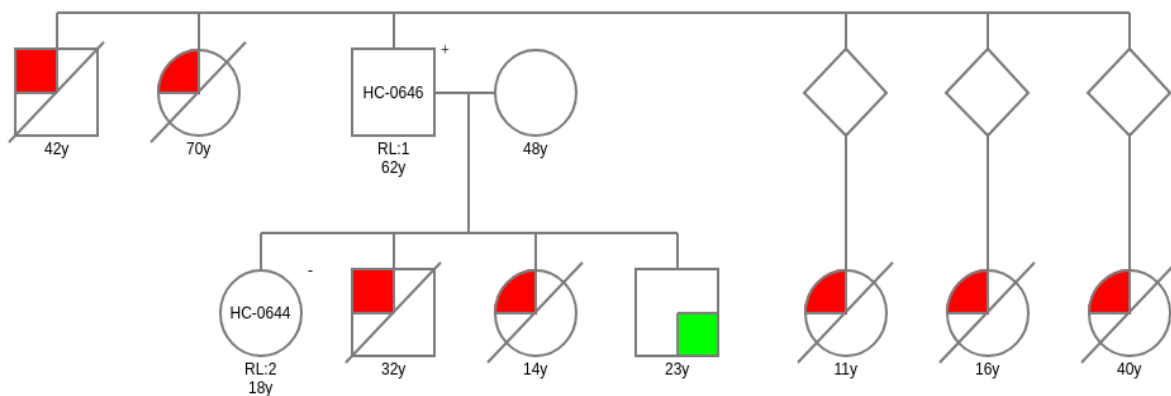
10. Household contacts resistant to disease or infection

LIPE variant – all HC
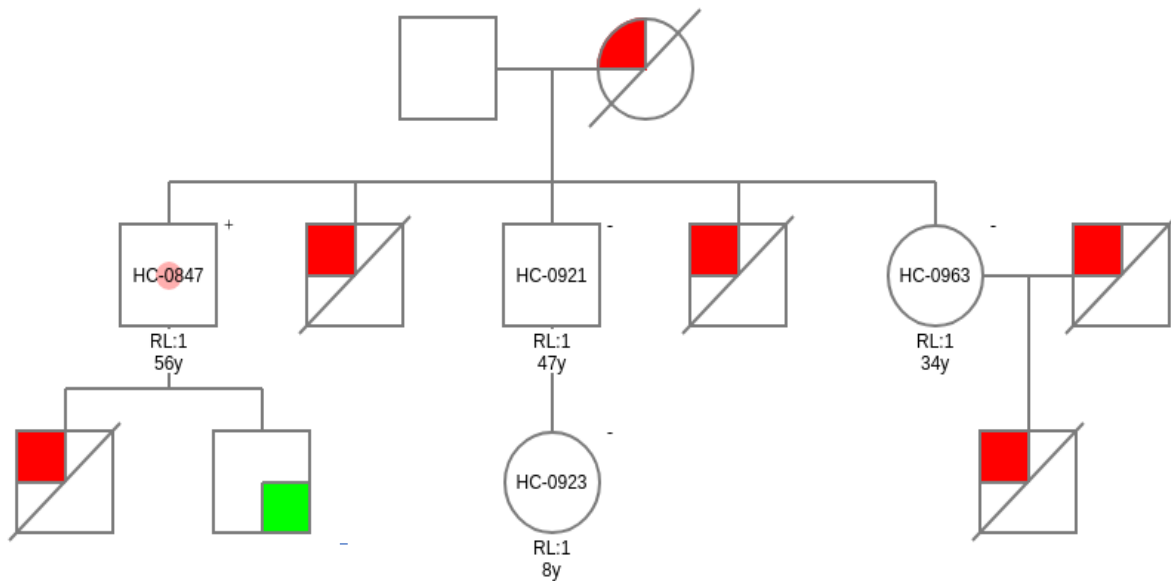
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

○ Unknown Female -not recruited

◇ Unknown Individual – not recruited

■ Affected, deceased, recruited if PD code

■ Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

☐ Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

− Anti-EBOV IgG negative

11. Household contacts and survivor resistant to disease/severe disease

STARD6 variant – all HC/SS
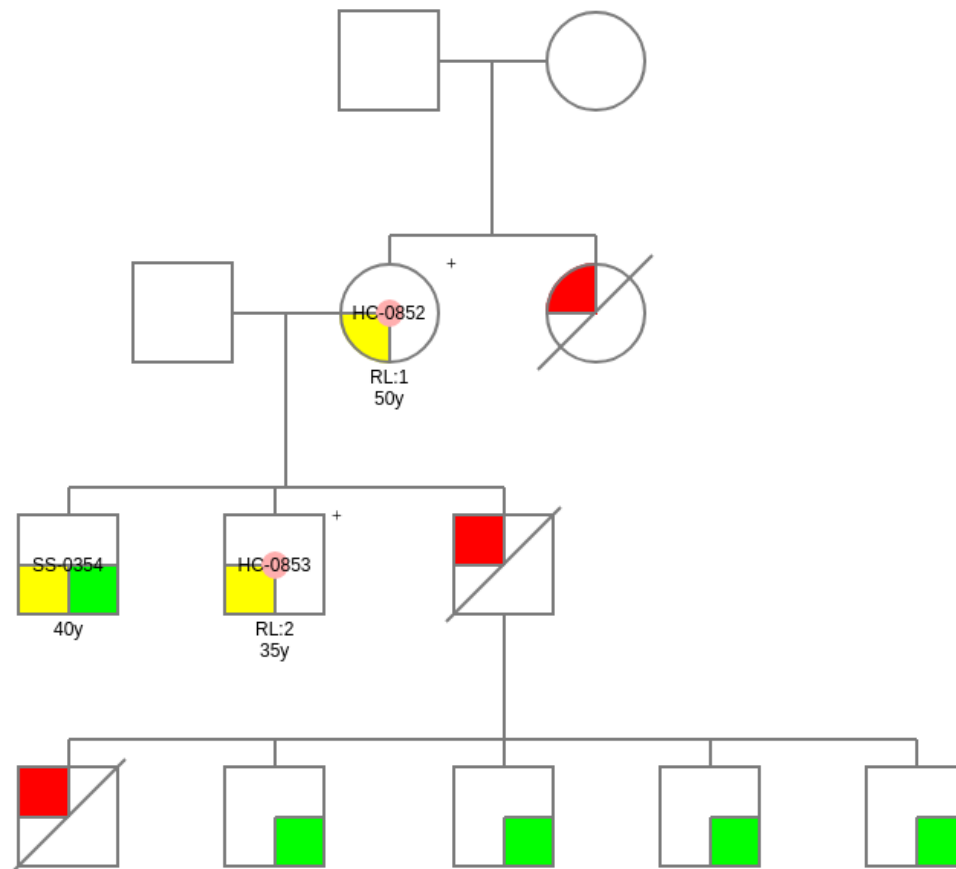
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

◯ Unknown Female -not recruited

◇ Unknown Individual – not recruited

🟥 Affected, deceased, recruited if PD code

🟩 Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

🟨 Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

– Anti-EBOV IgG negative

12. Household contact resistant to disease/Survivor resistant to severe disease

MX1 and LIPM variants – HC and SS

PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

☐ Unknown Male – not recruited

○ Unknown Female -not recruited

◇ Unknown Individual – not recruited

■ Affected, deceased, recruited if PD code

■ Affected, survived, recruited if SS code

● Symptoms meeting EVD case definition

■ Conjunctivitis during acute disease

+ Anti-EBOV IgG positive

– Anti-EBOV IgG negative

13. Household contacts resistant to disease or infection

LPIN2 variant – all HC
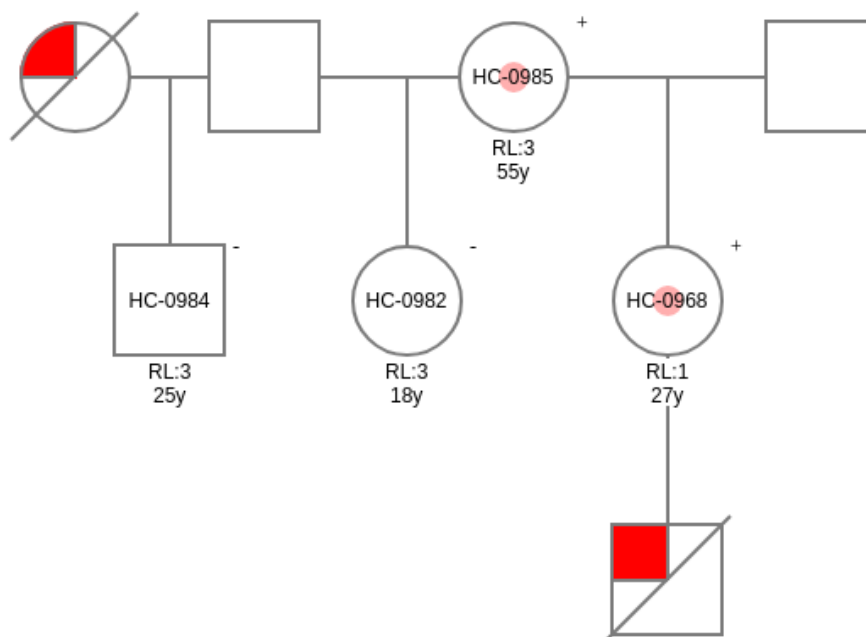
14. Household contacts resistant to disease

APOB variant – both HC

NPC1 variant – HC-1010
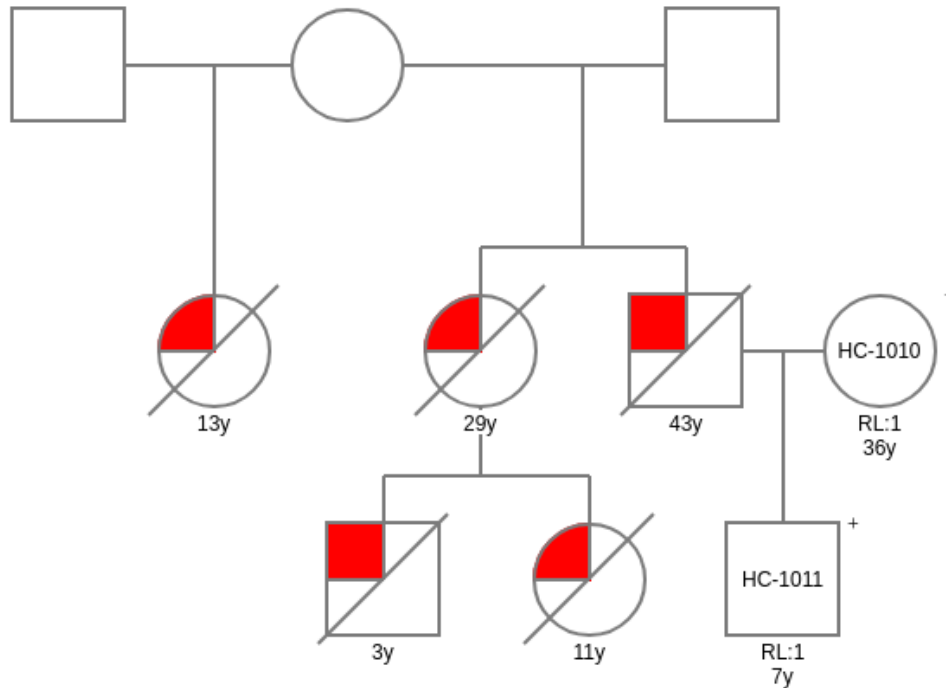
15. Susceptible to severe disease

SLC25A10 variant – both PD
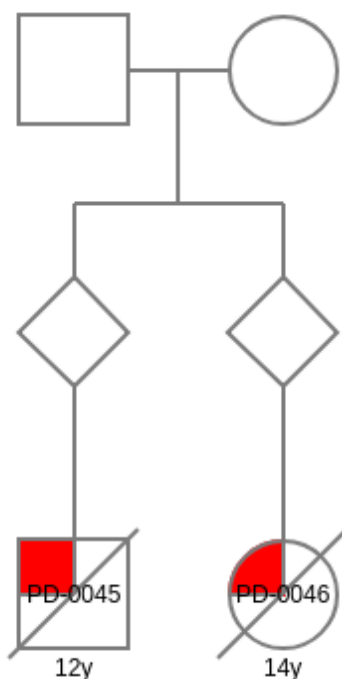
PEDIGREE LEGEND

HC = Recruited household contact

SS = Recruited EVD positive survivor

PD = Recruited EVD positive deceased

RL = Risk exposure level

Unknown Male – not recruited

Unknown Female -not recruited

Unknown Individual – not recruited

Affected, deceased, recruited if PD code

Affected, survived, recruited if SS code

Symptoms meeting EVD case definition

Conjunctivitis during acute disease

+   Anti-EBOV IgG positive

–   Anti-EBOV IgG negative

16. Survivors susceptible to disease, resistant to severe disease

STARD5 variant – both SS
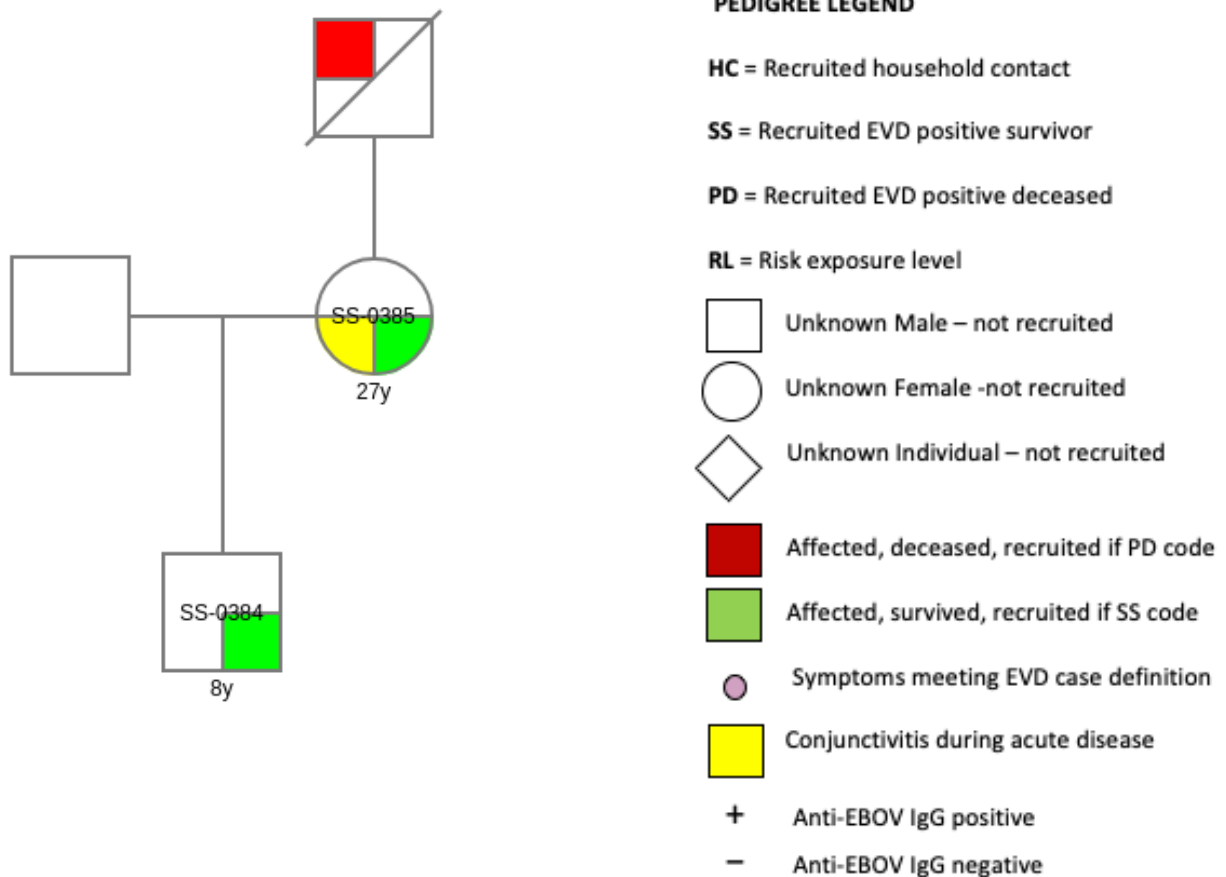
Figure 6-1 Pedigrees of families identified during the exome sequencing analysis.

6.3.2 Key rare variants identified through family analysis

Many of the key mutations identified through the family analysis highlighted variants within lipid metabolism pathways, which are described in Table 6-5, below, listed by the pedigree number the mutation was found in.

Pedigree number	Gene	Type	Who in pedigree?	Susceptibility or Severity	Function
1	PCSK9	Heterozygous Deleterious	All household contacts	Susceptibility Protective	Determines destination of endocytosed LDLR within the cell
2	APOC1	Heterozygous Deleterious	All household contacts	?Severity Protective	Inhibitor of lipoprotein binding to LDLR, LDLR related protein and VLDLR
3	SQLE	Heterozygous Deleterious	All household contacts	?Severity Protective	Catalyses first oxygenation step in

					sterol biosynthesis ?rate limiting enzyme
4	ABCA13	Compound heterozygous	All household contact siblings **Rare frameshift mutation**	?Severity Protective	Membrane transporter protein ?cholesterol transporter activity
	NPC1	Heterozygous	Sibling HC-0384	Severity Protective	Cholesterol receptor, primary receptor utilised by EBOV to enter cell cytoplasm from endosome.
	VPS11	Heterozygous	Sibling HC-0384	Severity Protective	Vacuolar protein sorting gene, forms part of the HOPS complex mediating endosome and lysosome fusion and endosome maturation.
5	None	NA	NA	NA	NA
6	ARSB	Heterozygous	All household contacts **Variant not previously reported**	?Susceptibility Protective	?lysosomal function Homozygous recessive results in mucopolysaccharidosis
7	STARD9	Heterozygous	All household contacts	?Severity Protective	intracellular sterol transport function Indirectly downregulates NPC1
	APOC1	Heterozygous Deleterious	All household contacts	?Severity Protective	Inhibitor of lipoprotein binding to LDLR, LDLR related protein and VLDLR
8	ABCA8	Heterozygous Deleterious	All household contacts	?Severity Protective	Membrane transporter protein ?regulates lipid metabolism
9	None	NA	NA	NA	NA
10	LIPE	Heterozygous	All household contacts	?Severity Protective	Converts cholesteryl esters to free cholesterol for steroid hormone production
11	STARD6	Heterozygous Deleterious	Household contacts & survivors	Severity Protective	Indirectly downregulates NPC1 intracellular transport of sterols/cholesterol and other lipids
12	MX1	Heterozygous Deleterious	Household contacts & survivors	Severity Protective	Antagonises replication process of RNA and DNA viruses
	LIPM	Heterozygous Deleterious	Household contacts & survivors	Severity Protective	?essential function in lipid metabolism

13	LPIN2	Heterozygous Deleterious	All related household contacts	?Severity Protective	Modulates lipid and triglyceride metabolism Controls metabolism of fatty acids
14	APOB	Heterozygous Homozygous	Mother Son	Severity Protective	Functions as a recognition signal for the cellular binding and internalisation of LDL Cholesterol receptor, primary receptor utilised by EBOV to enter cell cytoplasm from endosome
	NPC1	Heterozygous	Mother	Severity Protective	
15	SLC25A10	Homozygous Missense	Both deceased No other homozygotes in cohort	Severity Detrimental	Translocates substrates across mitochondrial membrane for use in Krebs cycle and fatty acid synthesis. Absence of this protein results in severe oxidative stress on cells.
16	STARD5	Homozygous Deleterious	Both survivors No other homozygotes in cohort	Severity Protective	Cholesterol/sterol transporter Indirectly down regulates NPC1

Table 6-5 A description of the gene variants identified within pedigrees in the preliminary analysis of the exome sequencing data. The gene variants identified are linked to the relevant pedigree number[82, 83, 113, 149-157].

6.3.3 Gene wise enrichment and pathway analysis

Through the preliminary analysis of rare variants and the familial analysis, one variant in particular stood out, in part due to its prevalence within those individuals who were highly exposed but antibody negative, but also because of its functional role in determining the fate of the endocytosed LDL receptor. This variant became strongly enhanced within the gene wise enrichment analysis. While only one family had the PCSK9 variant, when the entire group of 'resistant to infection' household contacts (highly exposed, antibody negative) were examined, five further individuals were identified as carrying rare deleterious PCSK9 variants. Thus, PCSK9 was the most enriched gene in the analysis, with seven antibody negative individuals carrying PCSK9 variants versus none of the antibody positive individuals (P=0.0018).

Due to the prevalence of variants within lipid metabolism pathways in the familial analysis, a further pathway analysis was undertaken to determine the mutational burden within lipid metabolism pathways between different disease phenotypes. The results of the mutational burden in lipid metabolism pathways are shown in Table 6-6, below. The mutational burden pathway analysis was conducted by comparing deceased Ebola patients (susceptible to severe disease) with highly exposed,

antibody negative participants (resistant to infection); and antibody positive participants (susceptible to infection and disease) with the highly exposed, antibody negative participants (resistant to infection).

Pathway	P-value Deceased vs IgG negative	P-value IgG positive vs IgG negative
Plasma Lipoprotein assembly pathway	Not significant	Not significant
LDL clearance pathway	0.018	0.025

Table 6-6 Significance values for the mutational burden in lipid metabolism pathways.

6.4 Discussion and future work

6.4.1 Relationship of mutations identified – Does Ebola utilise cholesterol metabolism pathways for intracellular replication?

A significant proportion of the mutations identified lie within lipid metabolism pathways. A good proportion of these are specific to cholesterol metabolism as shown in Figure 6-2, the Kegg pathway of cholesterol metabolism. The NPC1 receptor is established as essential for Ebola virus to exit the endosome and enter the cytoplasm of the cell where it can replicate[85, 87, 127]. The NPC1 receptor's normal function is as a cholesterol transport receptor[89], permitting free cholesterol (FC) to enter the hepatocyte cytoplasm and from there the endoplasmic reticulum, where it is converted into very low density lipoprotein (VLDL) and used in steroid and bile acid synthesis[82, 83, 152]. It is therefore of great interest that two families had individuals with NPC1 mutations who were resistant to symptomatic disease despite high levels of exposure. It is also of interest that three families had mutations in STARD genes, which are known to indirectly downregulate NPC1[151, 153].

The PCSK9 gene noted in pedigree 1 in Table 6-5 above, codes for a protein that is essential for targeting the endocytosed LDL receptor (LDLR) for degradation. Through the PCSK9 gene the endocytosed LDLR is prevented from being recycled and transitions to the late endosome/lysosome for degradation[113, 149, 150], where the NPC1 receptor resides. At various stages in the cholesterol metabolism process ApoB and ApoC play a role, both of which have been highlighted in pedigrees in Figure 6-1 and Table 6-5 above. ApoB appears to act as a ligand for LDL binding to the LDL receptor, facilitating LDL endocytosis into the cell, and ApoC plays a role in inhibiting ApoB-LDL binding to the LDL receptor[156, 157]. Similarly, several genes in the STARD family have been highlighted in the pedigrees above. While the exact role of some members of the STARD family is as yet unclear, what is apparent is that STARD1, STARD3 and STARD4 (not shown in the Kegg pathway below) appear to have a function within the endosome that involves cholesterol transport. Through this function they indirectly downregulate the

NPC1 receptor, which is regulated through a negative feedback loop[151, 153]. Lastly, a few individuals in the pedigrees appear to have variants in the NPC1 gene or in the VPS11 gene (part of the HOPS group mentioned in chapter 1). Variants within these genes are known to be directly related to EBOV's ability to exit the endosome into the cytoplasm of the cell where it can replicate[85, 127].

Our finding that a large proportion of individuals in the family based analysis who appear resistant to disease, carry variants that are functionally involved in processing cholesterol, suggests that Ebola viral entry into cells and replication in the cell cytoplasm may utilise the same pathway that cholesterol is metabolised through. Furthermore, the finding that PCSK9 mutations were not only identified in the family study, but also in five other non-familial individuals who were resistant to infection adds support to the concept that the cholesterol metabolism pathway is central to EBOV replication, as PCSK9 is a key regulator of the LDL-R. A gene burden analysis comparing highly exposed individuals who are antibody negative versus those who are anti-EBOV IgG positive (nominative p-value of 0.0018), suggests that PCSK9 is key to Ebola virus being successfully endocytosed into the cell and targeted to the late endosome/lysosome, where it can bind to NPC1 and enter the cell cytoplasm. As PCSK9 is specific to LDL receptors[113, 150], it is highly suggestive that Ebola virus binds to LDL receptors to be endocytosed into the cell.

PCSK9 is responsible for targeting bound LDL-R for degradation. It uses two methods to do this, an intracellular method whereby it targets endocytosed bound LDL-R to the late stage endosome/lysosome for degradation, and an extracellular method where it binds directly to the LDL-R complex on the cell surface targeting it for degradation. Previously reported PCSK9 gain of function mutations are associated with hypercholesterolaemia[113, 149, 150]. The suggested mechanism is excessive degradation of LDL-R through PCSK9 targeting, either during intracellular cycling of the LDL-R to the late endosome/lysosome or by secreted PCSK9 attaching to LDL-R on the cell surface. This increased extracellular degradation of LDL-R results in reduced expression of LDL-R on the cell surface and subsequent reduced uptake of cholesterol into the cell, resulting in increased levels of circulating cholesterol[113, 150]. PCSK9 loss of function mutations result in life long reduction in circulating LDL due to increased expression of LDL-R on cell surfaces and subsequent increased cellular uptake of circulating cholesterol [113]. It is not yet known if the mutations identified are gain of function or loss of function mutations, however it could be speculated that if their mechanism of action is similar to the reported gain of function mutations, they may result in reduction of the LDL receptor on cell surfaces. If EBOV mimics cholesterol in using the LDL-R to enter cells, this would explain resistance to viral entry into cells, resulting in absence of infection and thus absence of an antibody response.

Members of the STARD family of genes code for proteins that are involved in intracellular transport of cholesterol. NPC1 expression is regulated through a negative feedback loop, therefore if there is increased transport of cholesterol through STARD proteins from the endosome into the cytoplasm of the cell this will result in downregulation of the NPC1 receptor in the endosome[151]. On this basis it is hypothesised that a gain of function mutation in STARD genes would result in downregulation of the NPC1 receptor, reducing the frequency of EBOV entering the cell cytoplasm where it would replicate. As EBOV binding is specific to the NPC1 receptor[85] and not the STARD cholesterol transporter, a gain of function mutation in STARD genes would offer protection from a severe disease state due to reduced EBOV cytoplasmic entry and replication. This would explain why the STARD gene variants are found in family pedigrees with Ebola survivors and household contacts who were antibody positive but minimally symptomatic or asymptomatic. It is therefore hypothesised that a gain of function mutation in a STARD gene may make an individual resistant to disease or resistant to severe disease. In contrast a loss of function mutation in STARD genes may result in upregulation of NPC1 through the negative feedback loop, resulting in increased NPC1 expression in endosomes and increased EBOV entry into the cell cytoplasm. The role of APOB and APOC is less clear, but both are involved in the effective binding of cholesterol to the LDL-R, with APO-B acting as a ligand to the receptor and APO-C regulating the function of APO-B[156, 157]. If the above assumption is correct, then gain of function mutations in APO-B may be detrimental due to a resulting increase in EBOV binding to the LDL-R and being endocytosed into the cell, whereas the opposite would be true in a loss of function mutation. Conversely a gain of function mutation in APO-C may be protective due to increased inhibition of APO-B and reduced LDL binding of EBOV. Variants in NPC1 and the HOPS complex, as described in chapter 1, are known to prevent EBOV entry into the cell cytoplasm from the endosome[85] and therefore play a clear role in EBOV pathogenesis.

Lending further support to the suggestion that Ebola virus may bind to the LDL-R and that PCSK9 is important in ensuring the endocytosed virus reaches the late endosome/lysosome to bind to the NPC1 receptor, is the identification of the LDL clearance pathway being significant in the mutational burden of lipid metabolism pathways' analysis, shown in Table 6-7 and Figure 6-3 below. This pathway includes the PCSK9, LDL-R and LDLRAP1 genes. This latter gene has recently been identified to play a role in hypercholesterolaemia[158]. The LDL clearance pathway was significant in antibody negative, highly exposed participants when compared with both deceased participants and antibody positive participants (p=0.018 and p=0.025, respectively).

While the data described here is exciting and has achieved statistical significance in relation to a potential candidate gene and through a pathway analysis, it must be noted that it was limited by the number of families it was possible to identify in whom to conduct the familial analysis. It was also limited by the number of exomes it was possible to sequence within the financial constraints of the study and the variety of disease phenotypes that are of interest to compare. The data described here require validation either through sequencing the exomes of other participants within the larger study or through collaboration with other parties with cohorts from the West African epidemic.

6.4.2 Future work

This initial analysis has identified some potentially exciting data and variants on which to focus further work for validation. Such work could include initial cell based assays to identify the function of the PCSK9 protein in Ebola viral entry and replication and the role of the LDL and VLDL receptors. Similarly, animal studies could be conducted using knock out or knock down mice (using CRISPR-CAS9) to study the impact of Ebola infection in mice with and without LDL/VLDL receptors or without the PCSK9 gene. If such a study were undertaken it would be interesting to determine the correlation of circulating cholesterol phenotypes in these mice with their outcome when infected with EBOV. Should such a study yield significant results, the potential to return to Sierra Leone and obtain further samples from participants for measurement of cholesterol levels and to obtain serum samples in order to conduct cell based invasion assays should be considered. This would be subject to further ethical approval and the availability of sufficient funds to conduct such work, however.

As Ebola is a lipid envelope virus that is coated in glycoprotein, it is possible that other lipid envelope viruses and those coated in glycoprotein could well utilise a similar receptor or process[82, 83]. Therefore, it would be interesting to undertake similar cell based assays employing other viruses that are known to utilise similar systems to gain entry to cells, such as the vesicular stomatitis virus, HIV and hepatitis viruses.

This chapter has described the findings of rare variants and their potential role in EBOV pathogenesis. The following chapter will address methods to identify common variants in susceptibility to EVD through a genome wide association study.
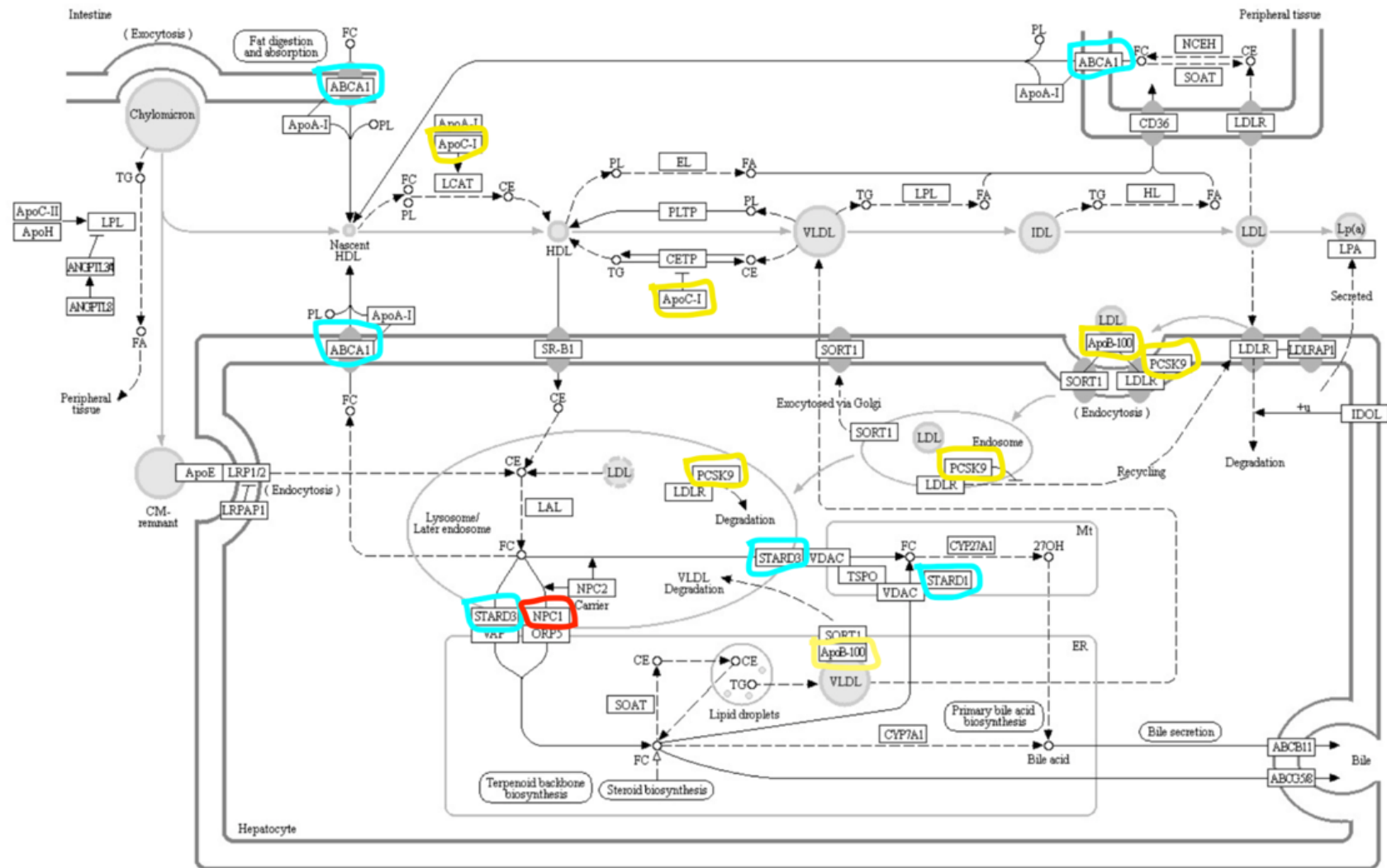
Figure 6-2 The Kegg pathway of cholesterol metabolism. Yellow = mutation identified, blue = mutation from same gene family identified, red = NPC1 receptor.

Gene	
AP2B1	ENSG00000006125
AP2S1	ENSG00000042753
SOAT1	ENSG00000057252
APOB	ENSG00000084674
LSR	ENSG00000105699
LIPA	ENSG00000107798
NPC2	ENSG00000119655
CLTA	ENSG00000122705
LDLR	ENSG00000130164
CLTC	ENSG00000141367
NPC1	ENSG00000141458
NCEH1	ENSG00000144959
LDLRAP1	ENSG00000157978
AP2M1	ENSG00000161203
SOAT2	ENSG00000167780
PCSK9	ENSG00000169174
CES3	ENSG00000172828
AP2A2	ENSG00000183020
AP2A1	ENSG00000196961

Table 6-7 List of genes in the LDL Clearance pathway (Reactome). Red indicates those genes of particular interest in relation to the expression and function of the LDL receptor (see Figure 6-3).
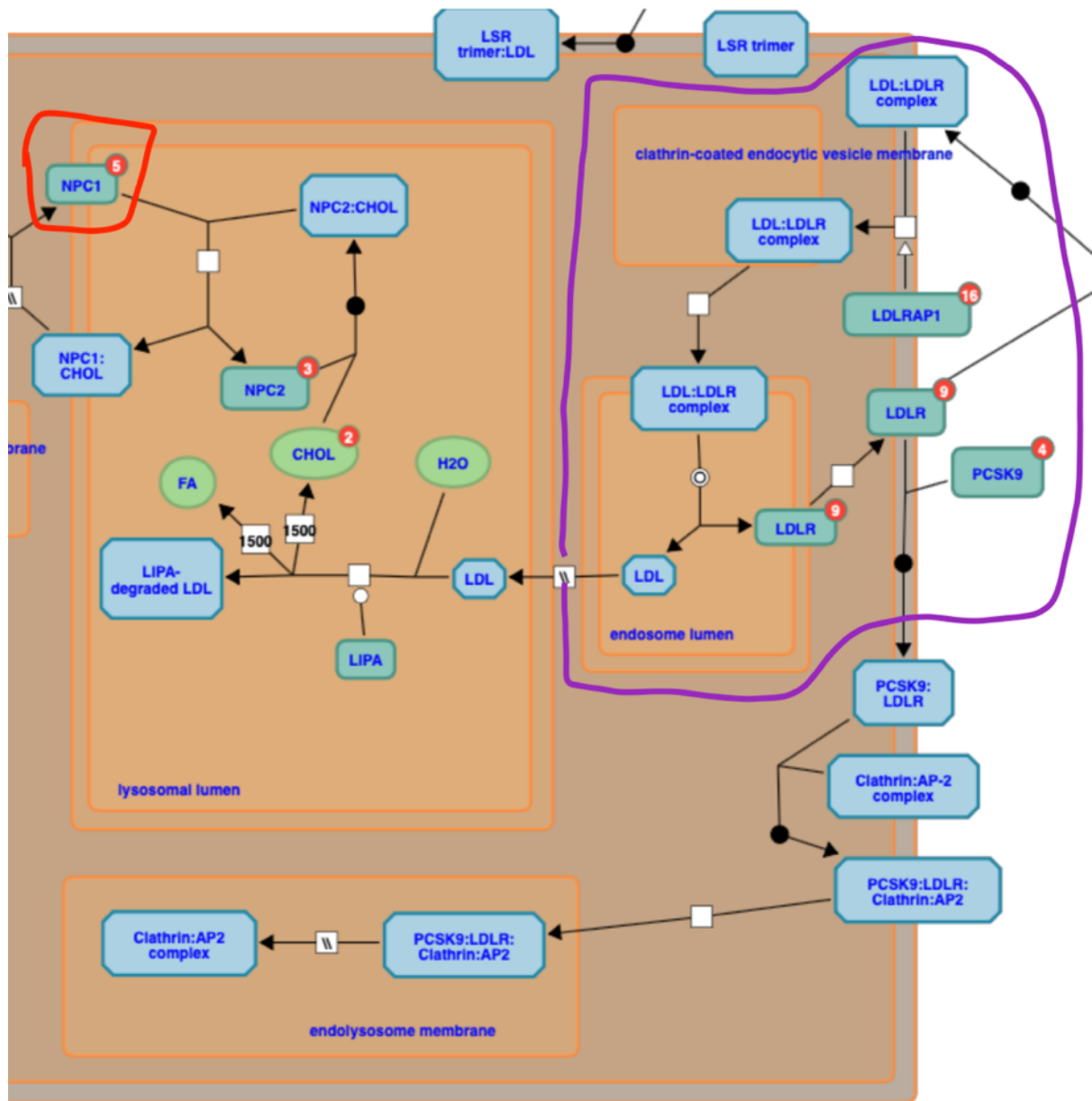
Figure 6-3 The section of the LDL clearance pathway that is of most interest, with key genes in turquoise (Reactome). The area circled in purple is of particular interest with LDLR, LDLRAP1 and PCSK9. NPC1 is circled in red.

Chapter 7 Identification of single nucleotide polymorphisms in susceptibility to and severity of Ebola virus disease

7.1 Introduction

In chapter one the use of genome wide association studies (GWAS) in identifying common variants that are associated with susceptibility or resistance to infectious diseases is described. GWAS are used to identify common variants associated with specific disease phenotypes within populations whereas exome sequencing, as mentioned above, is more suitable for identifying rare variants. GWAS has proved to be a powerful method for identifying reproducible genetic associations in a wide range of diseases. The improvements in genotyping methods, and the availability of population appropriate SNP arrays have enabled genotyping of large cohorts at much lower cost than whole exome or whole genome sequencing, facilitating interrogation of large datasets from different populations. Unlike exome sequencing the genotyping conducted in GWAS covers both coding and non-coding regions of the genome and utilises microarrays. These microarrays contain a large number of single nucleotide polymorphisms (SNP) that are known to be common among or within populations, depending on the array selected. Currently available SNP arrays include 500,000 SNPs to several million depending on the desired coverage of the genome and the population being studied. Advances in imputation enable the number of SNPs providing information to be greatly expanded by including imputed as well as directly genotyped SNPs.

The array utilised in this study was the H3 Africa array, which is a new array developed by The Human Heredity and Health in Africa Initiative and was first produced by Illumina in 2017. It includes SNPs from a wide range of different African regions[159]. To inform the SNP selection for the H3 Africa array, data was taken from African populations that constituted part of the 1000 Genomes Project and the African Genome Variation Project[159]. Data was also taken from further high and medium coverage sequenced samples belonging to the H3 consortium which covered a diversity of African populations, many of which are from West Africa, particularly Sierra Leone, Nigeria and The Gambia[159]. The array includes 2.4 million SNPs, made up of 1.7 million from existing Illumina bead pools applicable to the study of African populations and a further 700,000 selected from the H3 consortium sequenced samples, which include common but novel African variants and SNPs for specific diseases of interest[159].

African populations are considered the oldest populations from a genetic perspective and have had a longer period of time to develop variants adaptive to diverse climates, diets and infectious diseases[131, 159]. The migration of populations out of Africa resulted in a population bottle neck with significantly less diversity of the populations in the regions migrated to. Subsequently European and North American populations have significantly less diversity than African populations but have been some of the populations most studied from a genetic perspective. This has resulted in a discrepancy in the availability of SNP arrays applicable to the study of African populations[131, 159].

More genetic markers are required for the assessment and analysis of African samples due to the greater amounts of genetic diversity and presence of less linkage disequilibrium than non-African populations[131]. Linkage disequilibrium is the non-random association of alleles at different loci. Haplotype blocks are regions of the genome where SNPs are in strong linkage disequilibrium. African populations have smaller and fewer haplotype blocks, which means that a greater density SNP array is required as less SNPs can be inferred (or imputed) using linkage disequilibrium data[131]. On this basis it was important to select an appropriate array for this study. The H3 Africa array is currently the only array designed to contain SNPs selected to be specific predominantly for African populations. It also contains a large number of SNPs identified from West African populations, including populations from Sierra Leone. On these grounds it was deemed to be the most appropriate array for this study population.

In order to conduct a GWAS of genetic determinants of outcome from EVD, all participants exposed to Ebola that were recruited as part of the study or obtained through the Ebola biobank, and whose samples yielded sufficient DNA, were included. Additionally, a small number of community controls were included. The samples were from the same participants samples utilised in the exome sequencing and from the same disease phenotype categories. Below is a description of the number of samples which were included within each disease phenotype for the GWAS, with a further breakdown of the participant categories based on exposure level for household contacts and viral load/CT value for biobank samples.

7.2 Methods

7.2.1 Specification of samples

All viable samples from the household contact, Ebola survivor and Ebola deceased groups were sent for genotyping. A limited number of community control samples were sent for genotyping, which included

the eighteen antibody positive community controls mentioned above. In total 2153 samples were sent for genotyping, which consisted of:

Sample type	Number
Household contact (recruited participants)	1020
Ebola survivor (recruited participants)	448
Ebola survivor (biobank samples)	75
Ebola deceased (biobank samples)	242
Community control (recruited participants)	368
Total	2153

Table 7-1 Sample types sent for genotyping.

7.2.1.1 Sample subgroups sent for genotyping

As described in chapter 6 it was important to specifically define the samples sent for genotyping for analysis purposes. The breakdown of these groups is shown below (Table 7-2). As the antibody negative household contacts in risk levels 7 and 8 (very low exposure) essentially constituted a control group only a small number of the actual controls were genotyped. This was largely to provide an ethnically matched control group, as the unexposed, unaffected control group being included by collaborators in the US is predominantly from the Eastern part of Sierra Leone where the ethnicity is largely Mende. As over two thirds of the participants in this study were Temne and a further 15% were Limba, it was important to balance this by genotyping a small subset of controls from these ethnicities. Table 7-2 indicates the sample types, their defining criteria and the number of samples sent for genotyping.

Sample type	Defining criteria	Number
Ebola deceased (Biobank samples)	Haemorrhagic symptoms +/- CT value<18	34
	CT value<18	141
	Other/data unknown	67
Ebola survivor (Biobank samples)	PES status unknown	75
Ebola survivor (recruited participants)	Severe, defined PES	171
	Mild, defined PES	133
	No PES	144
Household contact	Anti-EBOV IgG positive	36
	Anti-EBOV IgG negative, risk level 1-3 exposure	310
	Anti-EBOV IgG negative, risk level 4-6 exposure	130
	Anti-EBOV IgG negative, risk level 7-8 exposure	544
Community control	Anti-EBOV IgG positive	18
	Anti EBOV IgG negative, risk level 8 exposure	350

Table 7-2 Disease severity definitions and sample numbers sent for genotyping.

7.2.2 DNA extraction and preparation

The methods of DNA extraction and then DNA concentration, purification and plating are described in chapters two and six respectively. Once the samples had been extracted, quantified and concentrated (where necessary), the samples were plated with a mixture of sample types to ensure that if a plate failed during sequencing no one specific sample type was adversely affected. Twenty-three plates were sent for genotyping, an example of the genotyping plate template is shown in Appendix C.

7.2.3 SNP array and genotyping

The samples for genotyping were sent to the *Illumina Fast Track Service* in California. In agreement with collaborators in the US and France, the *H3 Africa array* produced by *Illumina,* was selected for the genotyping. In order to avoid variation in genotyping between array batches and genotyping locations, samples from this study were sent to Illumina Fast Track services in California alongside those of our US collaborators. Due to delays in relation to the current Coronavirus pandemic our French collaborators have yet to complete their genotyping.

Of 2153 samples sent, 2052 were successfully genotyped, which is a genotyping success rate of 95.3%. The genotyping call rate was 99.7%. The high proportion of samples successfully genotyped, and the high call rate were particularly reassuring, as the DNA from deceased patients had been recovered from samples handled under far from ideal circumstances and obtained outside of a planned research study.

7.2.4 Analysis plan

Initial analysis was focussed on the deceased patient versus Ebola survivor phenotypes, as it was felt this would be most likely to yield a significant result given these were the most extreme phenotypes. Thereafter analysis will compare seropositive household contacts and seronegative high exposure contacts, as well as Ebola survivors with post-Ebola syndrome versus Ebola survivors with no ongoing problems. These latter two analyses will likely take the form of a meta-analysis with collaborators due to small sample numbers, particularly in the seropositive household contact group. Should significant SNPs be identified between any groups within this analysis, attempts to validate them with collaborators data will be made prior to further exploration of the impact of the causal allele on gene expression.

7.3 Analysis

The statistical analysis of the genotyping data was undertaken by Dr Clive Hoggart, with my guidance in relation to the patient groups to be analysed and specific details of their phenotypes. The initial

analysis conducted is a severity analysis to identify SNPs associated with fatal outcome and compares the differences in frequency of SNPs in deceased patients and Ebola survivors. Other planned comparisons will be performed later and are not included in this thesis. The analysis followed a standard approach to GWAS. Data was analysed using 'R for Statistical Computing' (*R Foundation*) and 'PLINK1.9 Whole Genome Association Analysis Toolset' (*Broad Institute, Harvard/MIT*).

From Illumina genotype data of 2052 individuals was received, genotyped at 2,213,564 SNPs. All samples had a call rate >0.01. Before imputation was conducted, stringent SNP quality control (QC) was performed in which SNPs with a call rate <0.01, minor allele frequency <0.01 or test for departure from Hardy-Weinberg equilibrium (p-value <0.0001) were removed leaving 1,790,722 SNPs.

Imputation was performed using the Michigan imputation server with the HRC reference panel. The HRC reference panel has 32,611 samples of predominately European ancestry but also includes the 1000 Genomes reference panel which has samples from West Africa and was used to inform some of the SNPs for the H3 Africa array. It has been demonstrated to perform well on African samples[160]. Imputation is limited by the accuracy of the reference panel being used to the population being studied. For this reason, as further analysis is conducted, additional reference panels more specific to African populations will be used for imputation, as access to these is anticipated through our US collaborators.

7.3.1 GWAS analysis of deceased patients versus Ebola survivors

The initial analysis aimed to identify genes associated with severity of the disease. As death was a clearly defined marker of severity, the distribution of SNPs in deceased patients and Ebola survivors were compared. Extremes of age were significant factors in outcome[6, 7, 22, 134], and therefore children five years and under and adults aged 60 and over were excluded from the analysis. Therefore, all analysed individuals were between the age of 6-59 years.

Therefore a GWAS on severity of Ebola infection comparing individuals who died to Ebola survivors aged between 6 and 59 years old, n=607 (480 Ebola survivors and 127 deceased samples), was performed. Before the GWAS, related individuals in this case/control grouping were removed excluding one member of each pair of samples with observed genomic relatedness greater than 0.025 using PLINK1.9. This removed 108 individuals leaving 499 participants (387 Ebola survivors and 112 deceased patients).

Multiple dimensional scaling (MDS), which is similar to principal component analysis (PCA), was then used to remove ethnic outliers. Samples with an MDS2 value greater than 0.02 were removed (shown in red in Figure 7-1). This removed a further 8 individuals (all survivors) leaving 112 Ebola deceased patients and 379 Ebola survivors.
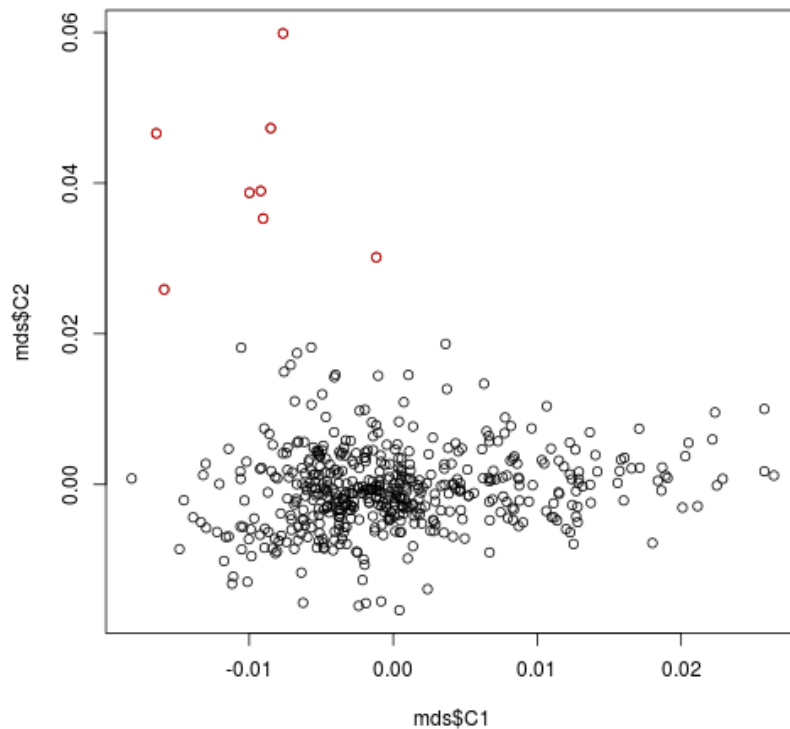
Figure 7-1 Plot of MDS1 vs MDS2 for the multiple dimensional scaling (similar to principal component analysis) of ethnicity. Eight ethnic outliers, shown in red, were removed.

The GWAS comparison of deceased patients and Ebola survivors was undertaken using logistic regression. Logistic regression was performed on all genotyped and imputed SNPs with minor allele frequency >0.05 and imputation quality score R^2 >0.8 (7,809,343 SNPs). The analysis controlled for MDS components associated with the outcome at p <0.1, these were 1, 4, 5, 9. The other MDS components were not included as they were not associated with outcome. Their inclusion would have had no effect on the analysis other than increasing noise. The association of components 1, 4, 5, 9 with disease outcome was corrected for within the logistic regression in order to prevent confounding of the results. Age between 6 and 59 years old was not associated with outcome and therefore was not included as a covariate. The variance inflation factor of the GWAS was 1.017. Population stratification was addressed by the removal of extreme outliers based on the MDS analysis. As the study participant categories were controlled for ethnicity and matched to similar locations to the origins of the biobank samples, it is reasonable to assume that there was a representative proportion of each ethnicity within each disease

phenotype, as demonstrated in chapter 3. On these grounds population stratification is unlikely to result in significant confounding of the results.

7.3.2 Identification of SNPs associated with fatal outcome

After imputation there were 7,809,343 SNPs available for analysis. Comparison of the deceased patients and Ebola survivors using logistical regression identified 20 SNPs showing nominally significant differences between the deceased patients and Ebola survivors. Table 7-3 lists all SNPs identified with a $p<5\times10^{-06}$ and Figure 7-2 shows a Manhattan plot of all the SNPs, displayed by chromosome on the X axis and by p-value for differences between deceased patients and Ebola survivors on the y axis. There were a number of closely linked SNPs that appeared significant. However, there was only one region where the SNP met the genome wide significance of $p<5\times10^{-08}$, which is shown in Figure 7-3 and listed at the top of Table 7-3.

Chromosome & *gene*	Position	SNP ID	Reference allele	Alternate allele	Odds Ratio	P-value	Typed/ imputed
16 – *CA5A*	87946292	rs114671281:C	G	C	5.41	2.9x10^{-08}	Imputed
10 – *LRRC20*	72073856	rs7909599:A	G	A	2.25	5.9x10^{-07}	Imputed
10 – *LRRC20*	72072894	rs3812700:T	C	T	2.25	5.9x10^{-07}	Imputed
10 – *LRRC20*	72074693	rs7086666:T	G	T	2.25	6.2x10^{-07}	Typed
16 – *CA5A*	87946133	rs116591916:T	C	T	4.13	6.6x10^{-07}	Imputed
12 – ?	14325852	rs7955915:G	A	G	2.67	7.7x10^{-07}	Typed
16 – *CA5A*	87945291	rs115144501:T	A	T	4.39	8.6x10^{-07}	Imputed
12 - ?	14332856	rs12580588:G	T	G	0.39	1.4x10^{-06}	Imputed
2 – *RTN4*	55286943	rs139391914:A	C	A	4.10	1.6x10^{-06}	Imputed
6 – *BEND6*	56852159	rs73455810:C	T	C	4.37	2.1x10^{-06}	Imputed
18 - ?	2458870	rs11080949:C	G	C	2.19	2.4x10^{-06}	Imputed
14 - ?	46068807	rs74862831:T	C	T	2.61	2.5x10^{-06}	Imputed
16 – *CA5A*	87946465	rs113329332:T	C	T	4.17	2.9x10^{-06}	Typed
1 - ?	242717608	rs1938334:T	C	T	0.39	3.4x10^{-06}	Typed
1 - ?	242734185	rs4658826:T	C	T	0.39	4.2x10^{-06}	Imputed
14 - ?	46040543	rs78134507:G	T	G	2.52	4.3x10^{-06}	Imputed
1 - ?	242722077	rs6700475:G	A	G	0.39	4.3x10^{-06}	Imputed
1 - ?	242749969	rs2096433:C	G	C	0.39	4.5x10^{-06}	Imputed
8 - ?	1971214	rs4875932:A	G	A	0.42	4.9x10^{-06}	Imputed
14 - ?	46042498	rs58163582:T	G	T	2.20	5.0x10^{-06}	Imputed

Table 7-3 Single nucleotide polymorphisms with p<5x10^{-06} identified in the GWAS comparing Ebola deceased patients versus Ebola survivors. ? = unknown/unreported gene in relation to SNP ID.

The SNP meeting genome wide significance is on chromosome 16 at position 87946292 and was significant at p=2.9x10^{-08}. In addition to this SNP three others in close proximity showed small p-values but did not meet genome wide significance. These SNPs are all located in the Carbonic Anhydrase 5a gene (CA5A) in intronic regions of the genome. Using linkage disequilibrium several SNPs have been inferred and imputed into the dataset, all of the SNPs in linkage with the lead SNP reside within intronic regions of the CA5A gene (Figure 7-3), helping to delineate the region within which the functional variant potentially resides.

7.4 Discussion of the GWAS analysis

7.4.1 The CA5A gene

The SNP meeting genome wide significance on chromosome 16 and the three other SNPs with small p-values that did not quite meet genome wide significance are all located within intronic regions of the Carbonic Anhydrase 5a gene (CA5A). There are multiple different carbonic anhydrase genes in the human body, and as a primary function they catalyse the reversible hydration of carbon dioxide within cells. They are very diverse in their tissue distribution and location within cells. In the context of carbonic anhydrase 5a, it is located within the mitochondria of cells and is primarily expressed within hepatocytes[161].

All SNPs identified to be in linkage disequilibrium with the lead SNP are within intronic regions of the CA5A gene (Figure 7-3). As the lead CA5A SNP identified is intronic and in the non-coding region of the genome it is possible that it is directly affecting the function of the CA5A gene itself, however it is also possible it could be regulating the expression of a nearby gene. While the molecular mechanism by which the variant may confer risk of disease is not yet entirely clear, as it is an intronic variant it is possible it affects CA5A splicing, although it could also be altering the expression of the CA5A gene or a neighbouring gene.

Loss of function mutations in the CA5A gene have been associated with early onset hyperammonaemia due to carbonic anhydrase 5a deficiency (CA5AD; 615751[162]). Three different homozygous mutations in the CA5A gene (114761.0001-114761.0003[162]) have been identified, all of which result in a loss of enzyme function[162]. CA5A deficiency is characterised clinically by acute onset of encephalopathy, hyperammonaemia and hypoglycaemia in infancy or early childhood, usually during times of stress, such as intercurrent illness. Biochemical abnormalities also include metabolic acidosis and raised serum lactate. Impaired provision of bicarbonate to essential mitochondrial enzymes results in impairment of mitochondrial functioning, particularly impacting on the urea cycle[162]. While such a disorder, if managed appropriately in the context of a robust healthcare system, may have a benign course, it is unlikely that individuals with this genetic disorder would survive in countries such as Sierra Leone due to the high burden of infections (including malaria) in early life and the limited health care available. Subsequently a significant deficiency of this enzyme is unlikely to be common and if present is probably fatal during early childhood. A heterozygous form with reduced enzymatic levels or partial loss of function, however, may well be consistent with the 'Encephalopathy' mode of death from EVD described in chapter one. This encephalopathy mode of death has features consistent with

encephalopathy, acidosis and hyperthermia[18], therefore the association of this gene with a fatal outcome presents a biologically plausible gene association.

Interestingly the gene also appears to have a key role in lipogenesis, although it is not entirely clear from the current literature how it is key in this process[161]. This finding may complement the findings within the exome data analysis that highlights several mutations in cholesterol metabolism pathways.

7.4.2 Future work

This chapter has presented only an initial analysis of the genome wide SNP data. Full analysis will be undertaken to compare each of the patient phenotypes with each other and the control group in due course. Provisional analysis by US collaborators of genotype data from Eastern Sierra Leone has not validated the CA5A SNP, however the sample size for that analysis was very small and achieving genome wide significance was unlikely. For this reason, further attempts to investigate whether the SNP found in the CA5A intronic region might influence gene expression have not been explored. This will be reconsidered following a meta-analysis of the data from this study and collaborators in the US and France, which will include data from Eastern Sierra Leone, Liberia and Guinea.

The discovery of a genome wide significant finding in the CA5A gene from this relatively small cohort of deceased patients and Ebola survivors remains of interest. Additionally, other associations with p-values short of genome wide significance require further exploration in combined analysis with the data from the Eastern Sierra Leone, Liberian and Guinean cohorts. This carefully phenotyped cohort with genome wide SNP data is likely to be a valuable resource for exploring the genetic basis of outcome from Ebola virus disease for years to come.
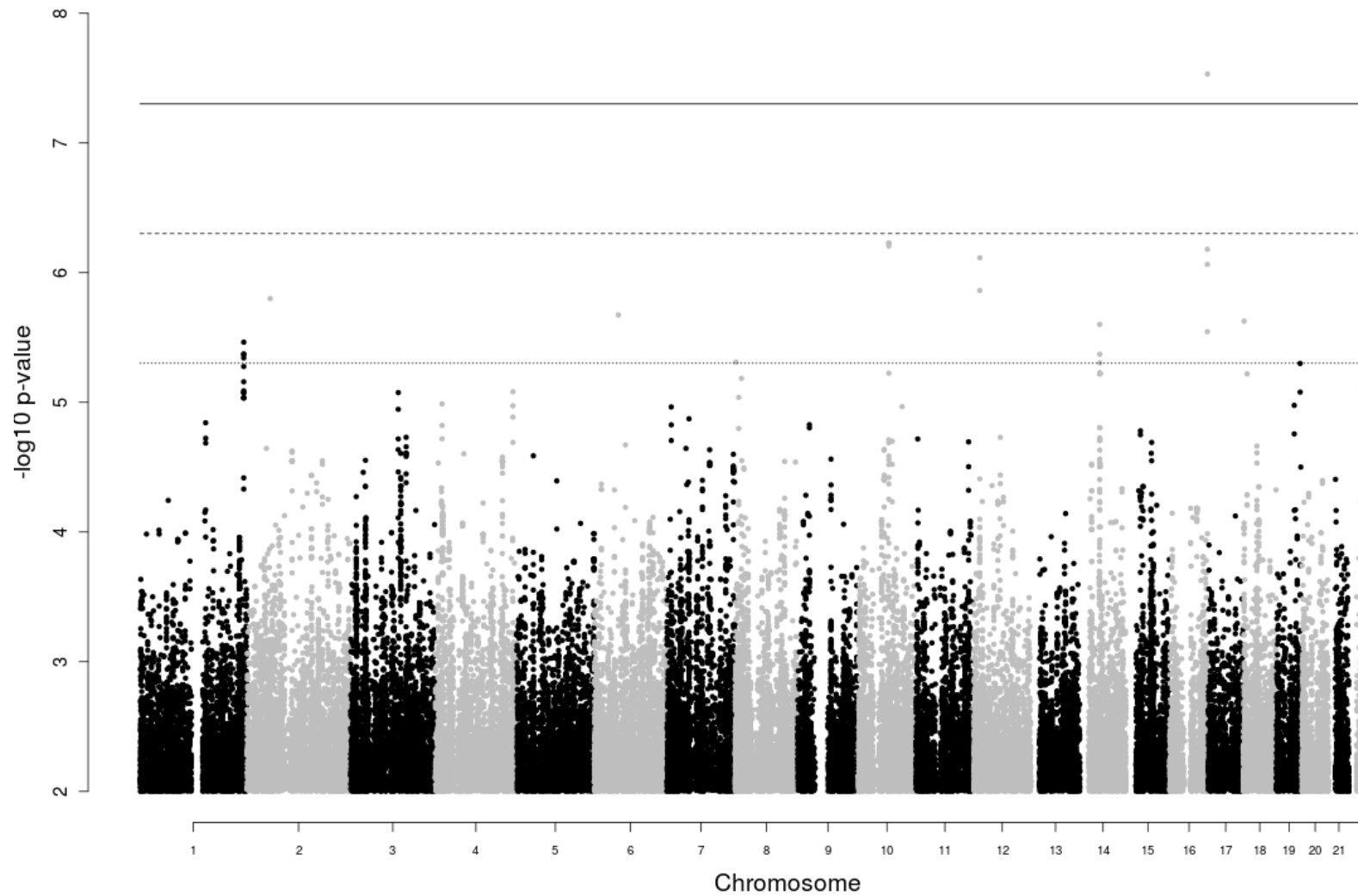
Figure 7-2 Manhattan plot of whole genome in the deceased patient versus Ebola survivor GWAS analysis.
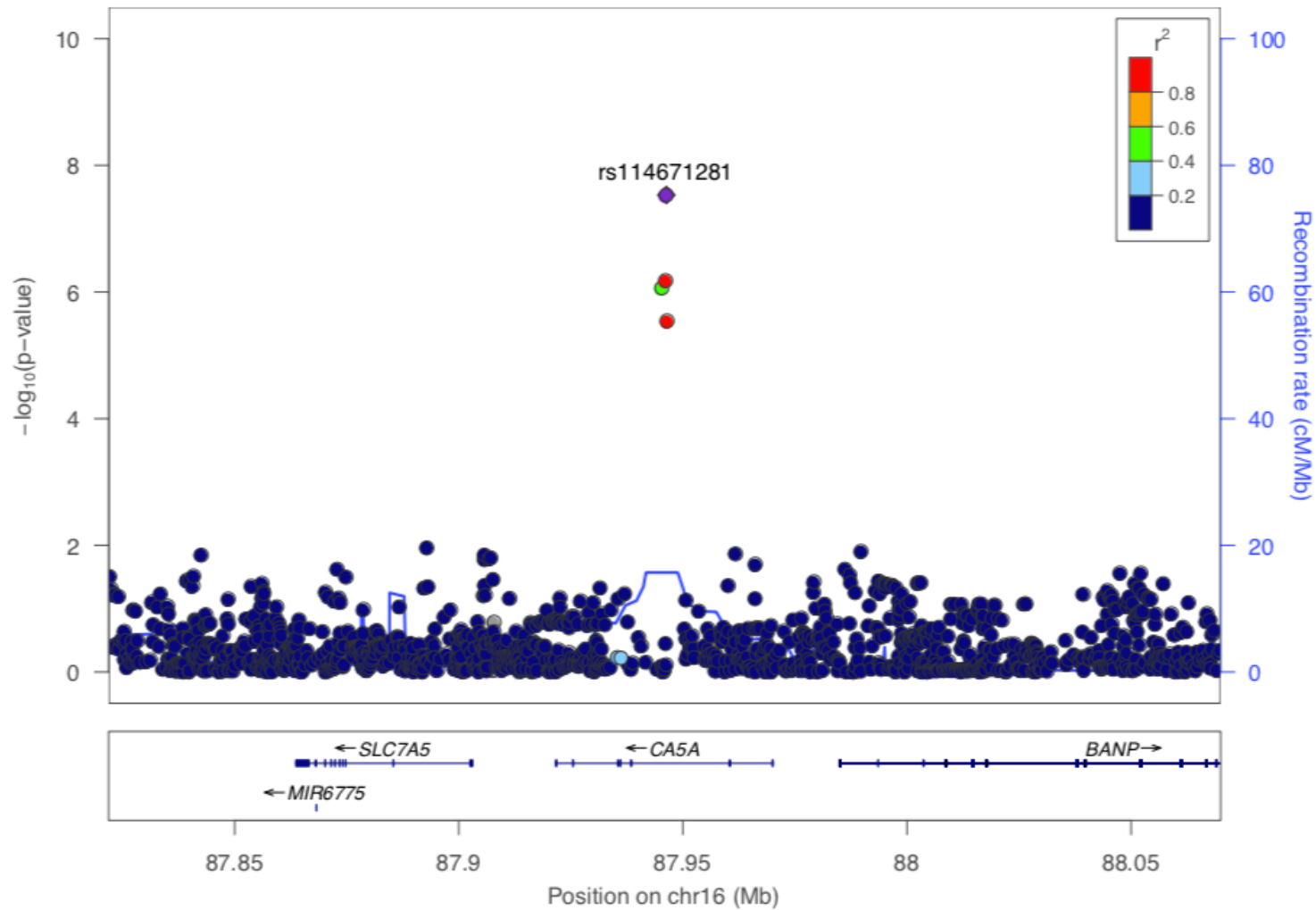
Figure 7-3 Regional plot of the significant SNP identified in the preliminary GWAS analysis, located on chromosome 16 and affecting the Carbonic Anhydrase 5a gene. X-axis shows position of gene on chromosome. Left Y-axis shows the -log$_{10}$ p-value, right Y-axis shows the recombination rate in centiMorgans/Megabase. r^2 = correlation with the top SNP, top SNP shown in purple.

Chapter 8 Conclusions

In undertaking the work described in this thesis, I aimed to explore the genetic basis of both susceptibility to and outcome from Ebola virus infection. The idea of examining the genetic basis of EBOV infection arose from my personal observation of the clinical variability of the disease in patients I cared for in Liberia. This ranged from asymptomatic or minimally symptomatic illness, to fulminant and rapidly fatal illness, the clinical data supporting this has since been published in the American Journal of Tropical Medicine and Hygiene[18]. In addition to the marked variability in outcome described in chapter 1, I observed that throughout the epidemic some family members and household contacts showed no sign of infection despite high levels of exposure, while other families and households experienced high rates of often fatal disease. The relatively simple question of whether genetic factors explain these observed differences has led to the research described in this thesis. The initial focus of the project was to establish well phenotyped cohorts of participants, with the goal of reaching the point of genetic sequencing. Given the scale of the project and the recruitment targets, it was an ambitious goal to aim for completion of the genetic sequencing and analysis. This latter aspect, while the most technically difficult part of the study, only took part in the last few months of the time allowed for my PhD. As such only the initial analysis of the GWAS and exome sequence data has been included in this thesis. The complete analysis of the genetic data has been further delayed by the current Coronavirus pandemic and requires many months more work and collaboration with colleagues both within Imperial College and at the Broad Institute (Harvard) and INSERM/Institut Pasteur in Paris. Through this collaboration we will meta-analyse the combined GWAS data from Sierra Leone, Liberia and Guinea.

Most of the work described in the thesis was focussed on identifying, recruiting and phenotyping the groups of participants who represented the different types of response to EBOV exposure, from resistance to infection, asymptomatic infection, symptomatic disease, PES and fatal outcome. In attempting to identify and categorise patients into groups with different responses to EBOV exposure, it was important to identify methods to describe the level of exposure, severity of clinical illness and to establish if infection had occurred, whether asymptomatically or symptomatically. A strategy was also needed to identify and score symptoms of post-Ebola syndrome. These methods were questionnaires and scoring systems to rank the level of exposure, determine family relatedness and identify post-Ebola symptoms; recording of standardised symptoms during acute disease; use of CT value to determine outcome of EVD when outcome data was untraceable; and measurement of anti-EBOV IgG in household contacts, community controls and prospectively recruited Ebola survivors.

This thesis has described my application of these methods in a prolonged period of fieldwork in Sierra Leone in which I established the groups shown in Table 8-1.

Clinical phenotype	Ebola antibody/PCR status	Genetic hypothesis
Highly exposed	Anti-EBOV IgG negative	Resistant to infection
Asymptomatic	Anti-EBOV IgG positive	Resistant to disease
Minimally symptomatic	Anti-EBOV IgG positive	Resistant to disease
Minimally symptomatic meets EVD case definition	Anti-EBOV IgG positive	Susceptible to disease
Symptomatic Ebola survivor no PES	Anti-EBOV IgG positive or Ebola PCR positive	Successful immune response
Symptomatic Ebola survivor PES	Anti-EBOV IgG positive or Ebola PCR positive	Partially successful immune response
Symptomatic Ebola survivor severe PES	Anti-EBOV IgG positive or Ebola PCR positive	Inadequate immune response
Symptomatic Ebola survivor with virus persistence +/- PES	Anti-EBOV IgG positive or Ebola PCR positive	Inadequate immune response
Symptomatic Ebola deceased	Ebola PCR positive	Failed immune response, inadequate control of virus

Table 8-1 Clinical phenotypes of Ebola virus disease.

This classification formed the basis of the initial genetic analyses described in chapters 6 and 7 and will be the basis of ongoing analyses in the future. In this concluding chapter I will discuss the major findings from the work described in the preceding chapters.

8.1 Interpretation of key findings

8.1.1 Anti-EBOV IgG Seroprevalence in communities

This study has shown that IgG positive individuals are present within apparently healthy populations of affected communities, as well as defined seroprevalence rates within different communities affected by Ebola virus disease in Sierra Leone. It has demonstrated that seroprevalence is higher within families where at least one member was diagnosed with EVD. However, it also indicated that there is a risk of being infected with EVD simply by living in a community affected by EVD. This risk is low, but it remains

a factor that must be taken into consideration when assessing patients in an epidemic or outbreak setting. Similarly, it identified that adults from urban communities were much more likely to be infected if a household member had EVD than children from urban communities. It is understandable that adult household contacts may have been more likely to be seropositive due to higher exposure levels to the disease while caring for unwell family members. This was also more likely in an urban setting due to higher population density.

The primary factor associated with seropositivity within communities was the level of contact a person had with an EVD case. Contact with a dead body presented the highest risk, while this may be expected, there is limited published data from previous viral haemorrhagic fever epidemics clearly demonstrating this. Study of level of exposure in household contacts also identified highly exposed individuals who remained antibody negative. These individuals are apparently resistant to infection and form one of the groups to be investigated genetically, and for which there is some initial genetic evidence to indicate how resistance might be mediated.

8.1.2 Disease severity

Table 8-1 highlights a variety of different phenotypes of response to EBOV exposure. These groups hint at a possible genetic influence, and as such a genetic hypothesis to be investigated for each of these groups in chapters 6 and 7 and in future work, is listed.

Within the symptomatic Ebola survivor and deceased groups there is variation in the acute symptomatology suffered, with a few patients only experiencing dry symptoms, a large proportion experiencing some form of wet symptom such as vomiting, diarrhoea or haemorrhage and just over a third of patients experiencing haemorrhagic symptoms. A small proportion of these experienced a more profuse form of haemorrhagic disease with bleeding from three or more locations. Interestingly those survivors who reported a profuse haemorrhagic disease appeared to have lower antibody levels than the other groups. This variation in clinical phenotype is strongly suggestive of genetic predisposition in terms of both disease susceptibility and disease severity, particularly in the context of a virus which did not demonstrate significant genetic variation within the geographical area that was studied.

8.1.3 Ebola survivors and PES

The study reveals some insight into the plight of Ebola survivors. As mentioned in the introduction, there was little information on the incidence and severity of ongoing problems in Ebola survivors prior to conducting this study. However, since completion of this work an article has been published about

the incidence of PES in Ebola survivors in Liberia[60]. The Liberian study described symptoms such as headache, fatigue, muscle pain, memory loss, joint pain and uveitis occurring in between 18.4% to 47.6% of Ebola survivors. The authors do not provide an overall figure for the rates of PES among their survivor population[60]. Approximately two-thirds of survivors in this study reported experiencing some form of PES, with 38% experiencing a more severe, life limiting form of PES. This study also found that PES and severe PES prevalence appears to increase with age. The manner in which the study in Liberia was conducted was very different to the study described here, which may account for the differences in findings. Similar to the study in Liberia, which followed survivors for up to 12 months after their acute disease[60], the majority of survivors reported their PES symptoms as resolved or improving at the point of recruitment, approximately two years after their acute disease. Less than 12% reported their symptoms as continuing the same and only one participant described their symptoms as worsening.

Symptoms fell within the categories of ocular symptoms such as uveitis, cataracts and blindness; joint pain; neurological problems including sleep disturbance, persistent headaches, new onset seizure disorders and specific paralysis syndromes; and fertility problems including impotence. Interestingly it appears that survivors with more severe forms of PES tended to have lower antibody levels than those who had milder forms of PES. This was particularly notable in survivors who suffered from cataracts or blindness, despite this being a relatively small group of survivors.

This study also found that while incidence of PES increased with age, antibody levels declined with age, although this did not achieve statistical significance. As lower antibody levels were found in patients with severe PES, who also had a more severe clinical presentation of acute EVD, impaired development of the acquired immune response may underlie PES. Future analysis to be undertaken on the GWAS and exome sequencing data will compare Ebola survivors with PES and those without to identify a possible genetic basis for PES. Furthermore, the genetic basis of the antibody response could be explored by undertaking a quantitative trait association analysis of antibody levels against genotype.

8.1.4 Exome sequencing and GWAS results

Despite concern over the power of the study the genomic data obtained has revealed novel findings that require further investigation. The exome sequencing has identified multiple mutations within cholesterol and lipid metabolism pathways that appear protective from infection with EBOV. This is consistent with previous understanding of Ebola virus requiring the Nieman Pick C1 cholesterol receptor to enter the cell cytoplasm where it can replicate[85, 87, 127]. Interestingly many of the

mutations identified are related to the LDL receptor and its endocytosis into the cell, as well as the targeting of the endocytosed receptor to the late endosome/lysosome where the NPC1 receptor resides (by PCSK9[113, 149, 150]). Other mutations identified are involved in cholesterol transport both external to the cell (APO B and APO C) and internal to the cell (STARD family genes)[153], some of which appear to have an indirect role in the downregulation of the NPC1 receptor[151]. The finding that individuals who appear to be resistant to infection or disease carry rare, deleterious variants in multiple genes in pathways related to cholesterol metabolism and transport, suggests that EBOV might be using similar mechanisms as cholesterol uses to enter cells and replicate. Due to the findings related to the LDL receptor it is also possible that EBOV binds with the LDL receptor to enter the cell, although this will require confirmation through functional studies, as will the impact of variants in cholesterol and lipid metabolic pathways on EBOV infectivity of cells. There may be similarities in relation to how the Hepatitis B and C viruses enter cells, as the role of APO B and APO C is important in relation to both these viruses[82, 84], but this will likewise require further work through functional studies. These initial findings however, together with the established role of NPC1 in EBOV replication may offer new approaches to treatment and prevention.

A number of drugs are now licensed and available that moderate cholesterol and lipid metabolism (such as PCSK9 inhibitors), which could be explored in the context of their use in controlling EBOV host cell entry and viral replication. As many of the variants identified might predispose an individual to high circulating LDL levels due to failure of uptake of LDL into the cell (APO-B/APO-C variants) or alteration of cholesterol metabolic pathways in the cell (PCSK9/NPC-1/STAR-D variants), it is possible that cholesterol lowering drugs that promote upregulation of LDL receptors on the cell surface could be detrimental by promoting EBOV cell entry. In turn those that alter cholesterol metabolism pathways that EBOV utilises to enter the cell cytoplasm could have a beneficial impact. Statins have an indirect effect of upregulation of LDL receptors on the cell surface[163] and while they do alter cholesterol metabolism within the cell they do not do this by affecting any of the variants identified as protective of EBOV infection, therefore they would likely have a detrimental effect by improving EBOV uptake by cells but not altering the pathway it utilises to enter the cell cytoplasm where it may replicate.

Two other variants have been identified which are not related to lipid metabolism but would be consistent in regard to disease severity. One of these is a homozygous recessive variant with a detrimental impact that affects the SLC25A10 gene, interfering with substrate transfer across the mitochondrial membrane and causing severe oxidative stress on cells[155]. Another is in the MX1 gene which is a gene known to interfere with viral replication in cells through many diverse methods[154].

The initial analysis of the GWAS SNP data has identified a genome wide significant polymorphism in the carbonic anhydrase 5a gene that is associated with poor outcome. Rare loss of function variants in this gene are associated with encephalopathy, acidosis and liver dysfunction[162], all of which may be features of severe EVD. While further work is required to functionally characterise this variant and to validate the findings, this initial finding gives support to the idea that genetic factors play a role in the outcome of EBOV infection. Interestingly the CA5A gene may also influence lipid metabolism[161], although the mechanism of this is not clearly described and further work would be required to understand this process.

8.2 Research impact: How this alters our understanding of EVD

8.2.1 Asymptomatic infection

Prior to this work it was unclear if asymptomatic Ebola infection occurred and the incidence of this within communities. Early studies were suggestive of this[120, 121] and the results of this study not only confirm this but are similar in terms of the prevalence demonstrated within affected communities. That asymptomatic infection occurs is important in the understanding of disease transmission within these communities and raises the question of whether asymptomatic individuals could contribute to disease spread. While this is unlikely it remains a possibility, particularly in relation to asymptomatic males and sexual transmission[62].

That asymptomatic infection exists suggests that rapid diagnostic tests for Ebola antigen and antibodies may well be useful in studying disease transmission in communities during an epidemic, but also immediately following an epidemic to determine if asymptomatic individuals contribute to disease spread, or in identifying males who may benefit from screening of semen for EBOV. It would also be interesting to determine if asymptomatic individuals experience PES. While this seems less likely as PES seems to occur more frequently in those with longer hospital admissions and more severe acute disease, it would be interesting to compare the prevalence of PES symptoms in this group, as well as their antibody levels with those of known Ebola survivors.

Asymptomatic infection and highly exposed anti-EBOV IgG negative participants also help us to better understand the potential pathogenesis of EVD, particularly in terms of genetic analysis. They may well

lend themselves better to specific study for vaccine or therapeutic targets due to their ability to resist or contain the virus, as has been suggested through the exome sequencing results.

8.2.2 Disease susceptibility and severity

The identification of different groups of patients that respond differently to EBOV exposure (Table 8-1) and the identification of genetic variants in these groups, which appear to moderate both disease severity and susceptibility, are findings which have not been previously identified. These groups of patients and the findings reported on genetic associations of different responses, provide the first evidence for a genetic effect on infection and outcome following EBOV exposure. Furthermore, the variants implicated may help us to better understand the pathogenesis of EVD and in due course could shed light on patients who are more likely to succumb to the disease and may require specific therapies or more intensive clinical care. Similarly, the variants identified might be implicated in susceptibility and severity of other infections which utilise the same pathways for viral cell entry (such as hepatitis C), and so may also provide insights into these diseases and therapies that may be applicable to them also.

8.2.3 Ebola entry receptors and drug/vaccine targets

Our data further support previous studies that identified the NPC1 receptor as essential for EBOV entry into the cell cytoplasm[85-87, 127], with evidence of several protective variants in the STARD family of cholesterol transporters, which indirectly downregulate NPC1. This suggests the STARD family may be a key area to target with novel therapeutics. Most importantly however is the finding of the PCSK9 mutation and significant result in the LDL clearance pathway analysis, this suggests Ebola utilises the LDL receptor and cholesterol metabolic pathways to be endocytosed into the cell. As LDL receptors are present on the membranes of all cell types, a treatment which potentially alters Ebola's ability to bind to the LDL receptor could convey a protective effect following exposure or help reduce the viral load following infection. As altering the function of PCSK9 is a focus of research into managing hypercholesterolaemia, there are already monoclonal antibody products available that target PCSK9[163]. These products are licensed and so have already demonstrated an adequate safety profile; while the use of monoclonal antibody products can be challenging in low income settings, the widespread use of products such as ZMAPP and ZMAB for management of EVD in recent African epidemics demonstrates that such products can be safely stored and administered. Future studies should explore the role of manipulation of cholesterol and lipid pathways as a means of reducing viral replication. PCSK9 inhibitors may have a role to play, as inhibition of intracellular PCSK9 may prevent Ebola bound to the LDL-R reaching the endosome. Further functional work on the effect of PCSK9

inhibitors both intracellularly and extracellularly is required however to ensure inhibition does not indirectly result in increased LDL-R expression on cell surfaces.

This information is also important as there are advocates of trialling statins in patients with Ebola due to their putative role in preventing endothelial damage. It appears these were 'trialled' at one stage during the epidemic in Sierra Leone, although not under the WHO's compassionate care approval and with no formal data on outcomes published[164]. Statins target HMG-CoA reductase to prevent the formation of free cholesterol in hepatocytes, but they have an additional secondary effect, which is the upregulation of LDL receptors on the cell surface, resulting in increased cholesterol removal from the circulation and into cells[163]. If Ebola does use the LDL receptor to enter cells, then potentially the administration of statins could have significantly detrimental effects and result in an increased viral load and more severe disease process.

8.3 Research Impact: How this alters our understanding of severe viral infections/infectious diseases

8.3.1 Genetic component of disease susceptibility/severity

This study lends further support to the existing evidence, described in chapter 1, for genetic susceptibility to infectious diseases. It also furthers our understanding of the potential mechanisms that Ebola uses to infect cells. As many viruses have a cholesterol rich lipid membrane these findings may well be applicable to other viruses.

8.3.2 Potential vaccine/drug targets for other VHFs/pandemic disease

While far more work is required to validate the genetic variants identified here, as well as functional work to determine how these variants may impact on cellular pathogenesis, it is possible that the genetic variants identified through the data from this project may provide potential drug and vaccine targets for other lipid envelope viruses and those coated with glycoprotein. Many of these viruses may also utilise lipid, and specifically cholesterol, metabolism pathways. As a large proportion of RNA viruses are lipid envelope viruses with glycoprotein coated membranes, it is possible this data could in time lead to the development of potential therapies or vaccines to treat a large number of other viruses that utilise lipid metabolism pathways to enter cells or the cell cytoplasm and replicate, such as Marburg virus and the hepatitis viruses.

8.4 Study limitations

8.4.1 Data collection

The prospective data collection relied on participant recall of events and the participant choosing to provide honest and accurate data. It is apparent that this was not always the case, as was noted above when discussing risk exposure levels and 'true' Ebola survivors. Where possible honesty was advocated for, explaining the data was confidential and there would be no repercussions from the information shared. However, when it was apparent that individuals were not being honest with the information provided, this was noted on the recruitment form and additional information sought from the chief or community leader to minimise the inaccuracy of the data. Verbal reports of Ebola survivors' acute disease and any ongoing problems they now have were also relied upon. For this reason, the symptoms were deliberately stratified by severity prior to recruitment and in advance of any analysis being conducted on the data. This helped to differentiate Ebola survivors who have life limiting problems, from those who may only have very mild problems, or who may be exaggerating their symptoms.

The clinical data obtained from treatment facilities varied in terms of the data collected and, in many instances, it was not possible to access data on the patients. Much of the clinical data, even just on outcome, had yet to be collated by the end of 2017. Therefore determining outcomes and clinical syndromes was based on data shared by NGOs who operated treatment facilities. Subsequently it was possible to trace clinical data for a small proportion of patients, but outcome data for a lot more. To enhance the numbers for whom outcome data was available, a cut off CT value was established for determining the likelihood that a patient died, this had a sensitivity of >98%. Due to the lack of clinical data it was not possible to define the groups as accurately for the analysis as would have been desired. However, for the crude and less 'symptom specific' analysis the data is robust, with groups large enough to compare in relation to symptom incidence and genetic variation.

8.4.2 Serology

Different studies have been conducted on sero-prevalence within communities following the West African epidemic. Many of these have used different antibody assays with variation in the sensitivity and specificity of the assay. The assay used in this study was selected based on its accessibility in respect of the communities, but also due to it having a high sensitivity (>95%) and specificity (>99%) demonstrated in a study conducted in Sierra Leone[120, 132]. IgG capture assays are generally very sensitive, and as the OD cut-off was corrected for based on the results of the controls and survivors from this study the sensitivity was optimised. The results reported here are also similar to the results

in two studies using this assay[120, 121]. There is a small risk of false negative results with such assays if a participant had a low total IgG level, but this is relatively rare.

8.4.3 Exome data

Due to the costs of exome sequencing, only a limited number of samples could be sent for exome sequencing, which will affect the power of the study to detect significant mutations. However, given the mutations that have already been identified, the suggestion is there is sufficient power to detect some pathway mutations. Some of the DNA samples which were obtained from Ebola patients during the epidemic contained low levels of DNA, some of which was quite fragmented and could impact on the exome analysis. While this study may have been underpowered, there is sufficient DNA left from most samples to conduct additional exome sequencing if funds are available.

8.4.4 Genotyping/GWAS

Due to the low DNA yields and possible fragmentation of DNA mentioned above, it was not possible to obtain genotyping data from all the clinical samples obtained during the epidemic. This will have impacted the power of the GWAS as only a relatively small number of deceased samples (n=112) were included in the analysis. Despite this however, a significant SNP has been identified between deceased and surviving patients. Hopefully also, the collaboration that has been established will help to overcome the low power of the study.

Traditionally there has not been a lot of genotyping data available for African populations. Although recently this has increased, and it was possible to access the H3 Africa array, which is an array specific for African populations and includes data on SNPs from a large number of people from West Africa. Hopefully this has helped to overcome potential inaccuracies in SNP calls that may occur using arrays from significantly ethnically diverse populations. There are a wide range of ethnicities within Sierra Leone, but many of these ethnicities are not too diverse and there has often been inter-mixing of ethnicities, further reducing the diversity. To combat potential discrepancies between participant categories recruitment intentionally occurred from similar, if not identical, communities, or similar regions of the country. Multiple dimensional scaling was also used to help exclude any ethnicities that appeared to significantly diverge from the group as a whole.

To achieve the recruitment targets and understand transmission dynamics within communities, the method of recruiting utilised entire families for the household contacts and community controls. This was necessary to obtain optimum numbers of recruited participants and was difficult to avoid due to

significant amounts of inter-relatedness within rural communities. While ideal for the exome sequencing analysis as it permitted the identification of multiplex families which could then be linked genetically to recruited survivor participants and biobank samples, it created the problem of relatedness of many of the samples for the GWAS analysis. This meant some samples had to be excluded if they were too closely related within the same participant sub-category, further reducing numbers and the power of the study. There is much less relatedness within the survivor cohort as these participants were recruited individually rather than in families, although there remains some relatedness due to recruitment from similar regions of Sierra Leone. Relatedness has also been identified within the biobank samples, these samples were anonymised by the biobank, therefore it was not possible to trace relatedness of samples in relation to other biobank samples or to recruited household contact, control or survivor participants in advance of receiving the samples. It was only following sequencing and genotyping that it was possible to establish the relatedness of biobank samples to each other and to the recruited participants.

8.5 Future work

8.5.1 Further exome/GWAS data interrogation

The data reported in this thesis is an early analysis of the exome sequencing and genotyping data of the cohort I have established. A further in depth analysis of other pathways and potential mutations identified in the exome sequencing is now required, particularly a look at the involvement of the VLDL metabolism and clearance pathways and identification of variants when solely comparing deceased patients versus survivors. Similarly, the GWAS analysis has only examined the deceased patients and Ebola survivors. Further analysis will compare each of the other groups (Table 8-1) with each other and the controls. Additionally, combination of my Sierra Leone cohort with the cohorts from Eastern Sierra Leone, Liberia and Guinea may provide sufficiently large groups to increase the power of the study to detect further genome wide significant associations. Likewise, the rare variants identified in the exome analysis might provide a clue to which genes and pathways should be focussed on in the GWAS analysis, specifically investigating the possibility of there being genotypic variants within lipid metabolism pathways. Regrettably due to the Coronavirus pandemic this work is yet to be completed.

8.5.2 Data validation and GWAS meta-analysis with collaborators

While we are currently analysing our genotyping data separately from our collaborators, we are planning a meta-analysis of the Sierra Leone, Liberia and Guinea cohorts with our collaborators in the US and France. We will also use each other's datasets to cross validate one another's findings. This meta-analysis will include DNA from approximately 10,000 individuals defined according to the same

phenotypes and following similar study protocols. This means it is likely it will be well powered to identify possible significant SNPs, which our individual studies may not have been adequately powered to identify.

8.5.3 Functional studies

Essential to any genetic study is to undertake functional studies of any potential gene variants identified to determine if these mutations have a true in vitro, and later in vivo, biological basis to them. Based on my current data analysis I plan to undertake functional studies to confirm if Ebola utilises the LDL and VLDL receptors to enter cells, and also if there is a role for LDLRAP1. I would also like to further investigate the role of gain and loss of function mutations in PCSK9 and EBOV cell entry, as well as the role of STARD receptors and APOB and APOC from a functional perspective. Lastly the SLC25A10 homozygous mutation appears to have a highly detrimental effect and understanding this from a functional perspective is also important. I have already commenced discussions with the molecular virology laboratory at Imperial College about possible models to use to undertake the in vitro cell based assay work, but this has been delayed by the coronavirus pandemic. It is also likely that other gene variants may be identified for additional functional studies from the further data interrogation and meta-analysis I will conduct in due course.

8.5.4 Drug/vaccine design/development

Lastly, based on the findings of the functional studies, it is possible that these findings may lend themselves to development of novel therapeutics or vaccines. For this to occur it would require developing a collaboration with the pharmaceutical industry to take forward drug or vaccine development and then initiate the various stages of clinical trials. If a product, such as the monoclonal antibody product to PCSK9, is already available on the market and targets variants identified in this data set, then it is possible that this could be fast tracked to animal studies and WHO compassionate care licensing. Provided adequate safety data is available such a product could be trialled directly in a clinical trial of human patients with EVD.

8.6 Data sharing and open access

All genomic data from this study will be placed in open access repositories following publication of the data to enable access by other parties who may wish to conduct further analyses. All papers will be published open access to ensure the data is available to all who may wish to use it. Any publications will be shared with the relevant individuals and authorities in Sierra Leone who may benefit from them.

8.6.1 Informing communities

As a significant proportion of the participants in this study do not have a secondary level education and many are also illiterate, I committed to them that I would return to explain the results of the study. Provided funds are available I will travel to Sierra Leone once the analysis is completed and visit each individual community, along with a member of my field team, to communicate the results of the study to the chiefs and the community members. In the interests of conducting high integrity scientific research and encouraging these communities to continue to be involved in research, I believe it is important they are informed of the outcome of the research they were involved in.

8.7 Closing remarks

This thesis represents a project that started during a horrific epidemic, where thousands of people lost their lives in part due to a lack of understanding about the disease and its management, and its ability to spread in urban settings. Ironically this project, save for the additional analyses to be conducted, now comes to an end in the middle of a worldwide pandemic that has taken millions of lives.

The world learnt many lessons from the West African Ebola epidemic, as did I. I hope the data contained in this thesis will go on to not only reduce disease transmission within communities in future epidemics, but also to help reduce mortality from EVD through the development of targeted therapeutics and vaccines. Hopefully it will also provide insights into other diseases with pandemic potential and facilitate development of vaccines and therapeutics that may be useful for such epidemics. It may also help in preparations for an epidemic of 'disease X' and will be interesting in light of the current SARS-COV2 pandemic to determine if there are any similarities in the susceptibility and severity genes identified, and whether the data presented here may have been of benefit early on in the pandemic or for future interventions.

The data in this thesis belongs to the people of Sierra Leone, and I am grateful to them for their cooperation and willingness to permit the study and participate in it. While the communities which participated in this study may seem far removed from us, let us not forget them and the impact this disease has wrought on their lives, families, communities and country. Given the current worldwide pandemic affecting our own nations, we are now well able to understand the toll an epidemic takes and the implications of epidemics and pandemics for affected nations.

References

1. Baseler, L., et al., *The Pathogenesis of Ebola Virus Disease.* Annual Review of Pathology: Mechanisms of Disease, Vol 12, 2017. **12**: p. 387-418.
2. Rougeron, V., et al., *Ebola and Marburg haemorrhagic fever.* Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology, 2015. **64**: p. 111-119.
3. Paessler, S. and D.H. Walker, *Pathogenesis of the viral hemorrhagic fevers.* Annual review of pathology, 2013. **8**: p. 411-440.
4. Bausch, D.G., et al., *Treatment of Marburg and Ebola hemorrhagic fevers: a strategy for testing new drugs and vaccines under outbreak conditions.* Antiviral research, 2008. **78**(1): p. 150-161.
5. Schieffelin, J.S., et al., *Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone.* The New England journal of medicine, 2014. **371**: p. 2092-2100.
6. Helleringer, S., K.A. Grepin, and A. Noymer, *Ebola virus disease in West Africa--the first 9 months.* N Engl J Med, 2015. **372**(2): p. 188-9.
7. Team, W.H.O.E.R., et al., *West African Ebola epidemic after one year--slowing but not yet under control.* N Engl J Med, 2015. **372**(6): p. 584-7.
8. Falasca, L., et al., *Molecular mechanisms of Ebola virus pathogenesis: focus on cell death.* Cell Death and Differentiation, 2015. **22**(8): p. 1250-1259.
9. Sadek, R.F., et al., *Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995: determinants of survival.* Journal of Infectious Diseases, 1999. **179 Suppl 1**: p. S24-7.
10. Bwaka, M.A., et al., *Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: clinical observations in 103 patients.* Journal of Infectious Diseases, 1999. **179 Suppl 1**: p. S1-7.
11. Matua, G.A., D.M. Van der Wal, and R.C. Locsin, *Ebola hemorrhagic fever outbreaks: strategies for effective epidemic management, containment and control.* Braz J Infect Dis, 2015. **19**(3): p. 308-13.
12. Clark, D.V., et al., *Long-term sequelae after Ebola virus disease in Bundibugyo, Uganda: a retrospective cohort study.* The Lancet infectious diseases, 2015. **15**(8): p. 905-912.
13. Goldstein, T., et al., *The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses (vol 3, pg 1084, 2018).* Nature Microbiology, 2018. **3**(12): p. 1486-1486.
14. Han, H.J., et al., *Bats as reservoirs of severe emerging infectious diseases.* Virus Research, 2015. **205**: p. 1-6.
15. Brierley, L., et al., *Quantifying Global Drivers of Zoonotic Bat Viruses: A Process-Based Perspective.* The American naturalist, 2016. **187**(2): p. E53-64.
16. Bausch, D.G., et al., *Assessment of the risk of Ebola virus transmission from bodily fluids and fomites.* Journal of Infectious Diseases, 2007. **196 Suppl 2**: p. S142-7.
17. Vetter, P., et al., *Ebola Virus Shedding and Transmission: Review of Current Evidence.* J Infect Dis, 2016. **214**(suppl 3): p. S177-S184.
18. Mobula, L.M., et al., *Clinical Manifestations and Modes of Death among Patients with Ebola Virus Disease, Monrovia, Liberia, 2014.* American Journal of Tropical Medicine and Hygiene, 2018. **98**(4): p. 1186-1193.
19. Bah, E.I., et al., *Clinical presentation of patients with Ebola virus disease in Conakry, Guinea.* The New England journal of medicine, 2015. **372**(1): p. 40-47.

20. Dallatomasina, S., et al., *Ebola outbreak in rural West Africa: epidemiology, clinical features and outcomes.* Trop Med Int Health, 2015. **20**(4): p. 448-54.

21. Haaskjold, Y.L., et al., *Clinical Features of and Risk Factors for Fatal Ebola Virus Disease, Moyamba District, Sierra Leone, December 2014-February 2015.* Emerg Infect Dis, 2016. **22**(9): p. 1537-44.

22. Team, W.H.O.E.R., et al., *Ebola virus disease among children in West Africa.* N Engl J Med, 2015. **372**(13): p. 1274-7.

23. Malvy, D., et al., *Ebola virus disease.* Lancet, 2019. **393**(10174): p. 936-948.

24. Team, W.H.O.E.R., et al., *After Ebola in West Africa--Unpredictable Risks, Preventable Epidemics.* N Engl J Med, 2016. **375**(6): p. 587-96.

25. Secretariat, C. *Sierra Leone: Overview*. 2018 [cited 2019 29/04/19]; Available from: http://thecommonwealth.org/our-member-countries/sierra-leone.

26. Secretariat, C. *Sierra Leone: Constitution and politics*. 2018 [cited 2019 29/04/19]; Available from: http://thecommonwealth.org/our-member-countries/sierra-leone/constitution-politics.

27. Secretariat, C. *Sierra Leone: History*. 2018 [cited 2019 29/04/19]; Available from: http://thecommonwealth.org/our-member-countries/sierra-leone/history.

28. Dean, N.E., et al., *Transmissibility and Pathogenicity of Ebola Virus: A Systematic Review and Meta-analysis of Household Secondary Attack Rate and Asymptomatic Infection.* Clinical Infectious Diseases, 2016. **62**(10): p. 1277-1286.

29. Lindblade, K.A., et al., *Secondary Infections with Ebola Virus in Rural Communities, Liberia and Guinea, 2014-2015.* Emerg Infect Dis, 2016. **22**(9): p. 1653-5.

30. Plowright, R.K., et al., *Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir-Host Populations.* PLoS Negl Trop Dis, 2016. **10**(8): p. e0004796.

31. Simon-Loriere, E., et al., *Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic.* Nature, 2015. **524**(7563): p. 102-4.

32. Rugarabamu, S., et al., *Forty-two years of responding to Ebola virus outbreaks in Sub-Saharan Africa: a review.* BMJ Glob Health, 2020. **5**(3): p. e001955.

33. Muzembo, B.A., et al., *Local perspectives on Ebola during its tenth outbreak in DR Congo: A nationwide qualitative study.* PLoS One, 2020. **15**(10): p. e0241120.

34. Masumbuko Claude, K., J. Underschultz, and M.T. Hawkes, *Social resistance drives persistent transmission of Ebola virus disease in Eastern Democratic Republic of Congo: A mixed-methods study.* PLoS One, 2019. **14**(9): p. e0223104.

35. Secretariat, C. *Sierra Leone: Society*. 2018 [cited 2019 29/04/19]; Available from: http://thecommonwealth.org/our-member-countries/sierra-leone/society.

36. Secretariat, C. *Sierra Leone: Economy*. 2018 [cited 2019 29/04/19]; Available from: http://thecommonwealth.org/our-member-countries/sierra-leone/economy.

37. Barry, M., et al., *Clinical predictors of mortality in patients with Ebola virus disease.* Clinical Infectious Diseases, 2015. **60**(12): p. 1821-1824.

38. Chérif, M.S., et al., *Ebola virus disease in children during the 2014-2015 epidemic in Guinea: a nationwide cohort study.* European journal of pediatrics, 2017. **176**(6): p. 791-796.

39. Hunt, L., et al., *Clinical presentation, biochemical, and haematological parameters and their association with outcome in patients with Ebola virus disease: an observational cohort study.* Lancet Infect Dis, 2015. **15**(11): p. 1292-9.

40. Leligdowicz, A., et al., *Ebola virus disease and critical illness.* Critical care (London, England), 2016. **20**(1): p. 217.

41. Uyeki, T.M., et al., *Clinical Management of Ebola Virus Disease in the United States and Europe.* The New England journal of medicine, 2016. **374**(7): p. 636-646.

42. Team, W.H.O.E.R., et al., *Ebola Virus Disease among Male and Female Persons in West Africa.* N Engl J Med, 2016. **374**(1): p. 96-8.

43. Mobula, L.M., et al., *Clinical Manifestations and Modes of Death among Patients with Ebola Virus Disease, Monrovia, Liberia, 2014.* Am J Trop Med Hyg, 2018. **98**(4): p. 1186-1193.

44. MacDermott, N.E., S. De, and J.A. Herberg, *Viral haemorrhagic fever in children.* Arch Dis Child, 2016. **101**(5): p. 461-8.

45. Fletcher, T.E., T.J.G. Brooks, and N.J. Beeching, *Ebola and other viral haemorrhagic fevers.* BMJ (Clinical research ed.), 2014. **349**: p. g5079.

46. Shah, T., et al., *Inpatient signs and symptoms and factors associated with death in children aged 5 years and younger admitted to two Ebola management centres in Sierra Leone, 2014: a retrospective cohort study.* The Lancet. Global health, 2016. **4**(7): p. e495-501.

47. Rosenke, K., et al., *Plasmodium Parasitemia Associated With Increased Survival in Ebola Virus-Infected Patients.* Clin Infect Dis, 2016. **63**(8): p. 1026-33.

48. Colebunders, R.L., *Is Plasmodium Species Parasitemia Really Associated With Increased Survival in Ebola Virus-Infected Patients?* Clinical Infectious Diseases, 2017. **64**(2): p. 231-232.

49. Waxman, M., et al., *Characteristics and survival of patients with Ebola virus infection, malaria, or both in Sierra Leone: a retrospective cohort study.* The Lancet infectious diseases, 2017. **17**(6): p. 654-660.

50. Kibadi, K., et al., *Late ophthalmologic manifestations in survivors of the 1995 Ebola virus epidemic in Kikwit, Democratic Republic of the Congo.* The Journal of Infectious Diseases, 1999. **179 Suppl 1**: p. S13-4.

51. Rowe, A.K., et al., *Clinical, virologic, and immunologic follow-up of convalescent Ebola hemorrhagic fever patients and their household contacts, Kikwit, Democratic Republic of the Congo. Commission de Lutte contre les Epidémies à Kikwit.* Journal of Infectious Diseases, 1999. **179 Suppl 1**: p. S28-35.

52. Bausch, D.G., *Sequelae after Ebola virus disease: even when it's over it's not over.* The Lancet infectious diseases, 2015. **15**(8): p. 865-866.

53. Mattia, J.G., et al., *Early clinical sequelae of Ebola virus disease in Sierra Leone: a cross-sectional study.* The Lancet infectious diseases, 2016. **16**(3): p. 331-338.

54. Vetter, P., et al., *Sequelae of Ebola virus disease: the emergency within the emergency.* Lancet Infect Dis, 2016. **16**(6): p. e82-e91.

55. Fowler, R., S. Mishra, and A.K. Chan, *The crucial importance of long-term follow-up for Ebola virus survivors.* Lancet Infect Dis, 2016. **16**(9): p. 987-989.

56. Jacobs, M., et al., *Late Ebola virus relapse causing meningoencephalitis: a case report.* Lancet, 2016. **388**(10043): p. 498-503.

57. Steptoe, P.J., et al., *Novel Retinal Lesion in Ebola Survivors, Sierra Leone, 2016.* Emerging infectious diseases, 2017. **23**(7): p. 1102-1109.

58. Varkey, J.B., et al., *Persistence of Ebola Virus in Ocular Fluid during Convalescence.* The New England journal of medicine, 2015. **372**(25): p. 2423-2427.

59. MacDermott, N.E. and D.G. Bausch, *Virus persistence and recrudescence after Ebola virus disease: what are the risks to healthcare workers?* J Hosp Infect, 2016. **94**(2): p. 113-5.

60. Sneller, M.C., et al., *A Longitudinal Study of Ebola Sequelae in Liberia.* New England Journal of Medicine, 2019. **380**(10): p. 924-934.

61. Deen, G.F., et al., *Ebola RNA Persistence in Semen of Ebola Virus Disease Survivors - Final Report.* N Engl J Med, 2017. **377**(15): p. 1428-1437.

62. Mate, S.E., et al., *Molecular Evidence of Sexual Transmission of Ebola Virus.* The New England journal of medicine, 2015. **373**: p. 2448-2454.

63. Thorson, A., et al., *Systematic review of the literature on viral persistence and sexual transmission from recovered Ebola survivors: evidence and recommendations.* BMJ open, 2016. **6**(1): p. e008859.

64. Deen, G.F., et al., *Implementation of a study to examine the persistence of Ebola virus in the body fluids of Ebola virus disease survivors in Sierra Leone: Methodology and lessons learned.* PLoS Negl Trop Dis, 2017. **11**(9): p. e0005723.

65. Blackley, D.J., et al., *Reduced evolutionary rate in reemerged Ebola virus transmission chains.* Science Advances, 2016. **2**(4): p. e1600378-e1600378.

66. Den Boon, S., et al., *Ebola Virus Infection Associated with Transmission from Survivors.* Emerging Infectious Diseases, 2019. **25**(2): p. 240-246.

67. Diallo, B., et al., *Resurgence of Ebola Virus Disease in Guinea Linked to a Survivor With Virus Persistence in Seminal Fluid for More Than 500 Days.* Clin Infect Dis, 2016. **63**(10): p. 1353-1356.

68. Gregory, S.M., et al., *Ebolavirus entry requires a compact hydrophobic fist at the tip of the fusion loop.* Journal of virology, 2014. **88**(12): p. 6636-6649.

69. Diehl, W.E., et al., *Ebola Virus Glycoprotein with Increased Infectivity Dominated the 2013-2016 Epidemic.* Cell, 2016. **167**(4): p. 1088-1098.e6.

70. Martinez, O., et al., *A mutation in the Ebola virus envelope glycoprotein restricts viral entry in a host species- and cell-type-specific manner.* Journal of virology, 2013. **87**(6): p. 3324-3334.

71. Ning, Y.J., et al., *The roles of ebolavirus glycoproteins in viral pathogenesis.* Virologica Sinica, 2017. **32**(1): p. 3-15.

72. Takada, A. and Y. Kawaoka, *The pathogenesis of Ebola hemorrhagic fever.* Trends in Microbiology, 2001. **9**(10): p. 506-511.

73. Urbanowicz, R.A., et al., *Human Adaptation of Ebola Virus during the West African Outbreak.* Cell, 2016. **167**(4): p. 1079-1087.e5.

74. Kawaoka, Y., *Clinical implications of basic research: How ebola virus infects cells.* New England Journal of Medicine, 2005. **352**(25): p. 2645-2646.

75. Kirchdoerfer, R.N., et al., *Assembly of the Ebola Virus Nucleoprotein from a Chaperoned VP35 Complex.* Cell Rep, 2015. **12**(1): p. 140-149.

76. Leung, D.W., et al., *An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions.* Cell reports, 2015. **11**(3): p. 376-389.

77. Wang, H., et al., *Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1.* Cell, 2016. **164**(1-2): p. 258-268.

78. Garcia, M., et al., *Productive replication of Ebola virus is regulated by the c-Abl1 tyrosine kinase.* Sci Transl Med, 2012. **4**(123): p. 123ra24.

79. Education, E.U.D.o.M.V., *Ebola virus: Exploring the structure.* 2014, Emory University VME Gallery.

80. Basler, C.F., *Innate immune evasion by filoviruses.* Virology, 2015. **479-480**: p. 122-30.

81. Bray, M. and T.W. Geisbert, *Ebola virus: The role of macrophages and dendritic cells in the pathogenesis of Ebola hemorrhagic fever.* International Journal of Biochemistry & Cell Biology, 2005. **37**(8): p. 1560-1566.

82. Heaton, N.S. and G. Randall, *Multifaceted roles for lipids in viral infection.* Trends in Microbiology, 2011. **19**(7): p. 368-375.

83. Sviridov, D. and M. Bukrinsky, *Interaction of pathogens with host cholesterol metabolism.* Current Opinion in Lipidology, 2014. **25**(5): p. 333-338.

84. Lindenbach, B.D. and C.M. Rice, *The ins and outs of hepatitis C virus entry and assembly.* Nature Reviews Microbiology, 2013. **11**(10): p. 688-700.

85. Carette, J.E., et al., *Ebola virus entry requires the cholesterol transporter Niemann-Pick C1.* Nature, 2011. **477**(7364): p. 340-343.

86. Côté, M., et al., *Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection.* Nature, 2011. **477**(7364): p. 344-348.

87. Herbert, A.S., et al., *Niemann-pick C1 is essential for ebolavirus replication and pathogenesis in vivo.* mBio, 2015. **6**(3): p. e00565-15.

88. Coleman, E.M., T.N. Walker, and J.E. Hildreth, *Loss of Niemann Pick type C proteins 1 and 2 greatly enhances HIV infectivity and is associated with accumulation of HIV Gag and cholesterol in late endosomes/lysosomes.* Virol J, 2012. **9**: p. 31.

89. Pfeffer, S.R., *Clues to NPC1-mediated cholesterol export from lysosomes.* Proc Natl Acad Sci U S A, 2016. **113**(29): p. 7941-3.

90. Dahlmann, F., et al., *Analysis of Ebola Virus Entry Into Macrophages.* J Infect Dis, 2015. **212 Suppl 2**: p. S247-57.

91. Kwok, A.J., A. Mentzer, and J.C. Knight, *Host genetics and infectious disease: new tools, insights and translational opportunities.* Nat Rev Genet, 2021. **22**(3): p. 137-153.

92. Casanova, J.L. and L. Abel, *The human genetic determinism of life-threatening infectious diseases: genetic heterogeneity and physiological homogeneity?* Hum Genet, 2020. **139**(6-7): p. 681-694.

93. Casanova, J.L., *Human genetic basis of interindividual variability in the course of infection.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(51): p. E7118-E7127.

94. Schwarze, K., et al., *Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature.* Genet Med, 2018. **20**(10): p. 1122-1130.

95. Lacey, S., J.Y. Chung, and H. Lin, *A comparison of whole genome sequencing with exome sequencing for family-based association studies.* BMC Proc, 2014. **8**(Suppl 1 Genetic Analysis Workshop 18Vanessa Olmo): p. S38.

96. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-753.

97. Manolio, T.A., *Genomewide association studies and assessment of the risk of disease.* The New England journal of medicine, 2010. **363**(2): p. 166-176.

98. Casanova, J.L., *Severe infectious diseases of childhood as monogenic inborn errors of immunity.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(51): p. E7128-E7137.

99. Casanova, J.L. and L. Abel, *Human genetics of infectious diseases: Unique insights into immunological redundancy.* Seminars in Immunology, 2018. **36**(C): p. 1-12.

100. Bartha, I., et al., *A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control.* eLife, 2013. **2**: p. e01123.

101. Carrington, M., et al., *Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences.* Human molecular genetics, 1999. **8**(10): p. 1939-1945.

102. Carrington, M., A.A. Bashirova, and P.J. McLaren, *On stand by: host genetics of HIV control.* AIDS (London, England), 2013. **27**(18): p. 2831-2839.

103. Shea, P.R., et al., *Host genetics of HIV acquisition and viral control.* Annual review of medicine, 2013. **64**: p. 203-217.

104. Jiang, D.-K., et al., *Genetic variations in STAT4,C2,HLA-DRB1 and HLA-DQ associated with risk of hepatitis B virus-related liver cirrhosis.* Scientific reports, 2015. **5**: p. 16278.

105. Matsuura, K., M. Isogawa, and Y. Tanaka, *Host genetic variants influencing the clinical course of Hepatitis B virus infection.* Journal of medical virology, 2015.

106. Stättermayer, A.F. and P. Ferenci, *Effect of IL28B genotype on hepatitis B and C virus infection.* Current opinion in virology, 2015. **14**: p. 50-55.

107. Davila, S., et al., *Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease.* Nature genetics, 2010. **42**(9): p. 772-776.

108. Geishofer, G., et al., *4G/5G promoter polymorphism in the plasminogen-activator-inhibitor-1 gene in children with systemic meningococcaemia.* European journal of pediatrics, 2005. **164**(8): p. 486-490.

109. Haralambous, E., et al., *Role of functional plasminogen-activator-inhibitor-1 4G/5G promoter polymorphism in susceptibility, severity, and outcome of meningococcal disease in Caucasian children.* Critical Care Medicine, 2003. **31**(12): p. 2788-2793.

110. Hibberd, M.L., et al., *Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. Meningococcal Research Group.* The Lancet, 1999. **353**(9158): p. 1049-1053.

111. Kondaveeti, S., et al., *Effect of the Factor V Leiden mutation on the severity of meningococcal disease.* The Pediatric Infectious Disease Journal, 1999. **18**(10): p. 893-896.

112. Smirnova, I., et al., *Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(10): p. 6075-6080.

113. Poirier, S., et al., *Trafficking Dynamics of PCSK9-Induced LDLR Degradation: Focus on Human PCSK9 Mutations and C-Terminal Domain.* PLoS One, 2016. **11**(6): p. e0157230.

114. Boyd, A.R. and E.M. Mortensen, *Are statins beneficial for viral pneumonia?* Eur Respir J, 2013. **41**(5): p. 1010-1.

115. Grudzinska, F.S., et al., *Statin therapy in patients with community-acquired pneumonia.* Clin Med (Lond), 2017. **17**(5): p. 403-407.

116. Fitzpatrick, G., et al., *The Contribution of Ebola Viral Load at Admission and Other Patient Characteristics to Mortality in a Médecins Sans Frontières Ebola Case Management Centre, Kailahun, Sierra Leone, June-October 2014.* The Journal of Infectious Diseases, 2015. **212**(11): p. 1752-1758.

117. Van der Waals, F.W., et al., *Hemorrhagic fever virus infections in an isolated rainforest area of central Liberia. Limitations of the indirect immunofluorescence slide test for antibody screening in Africa.* Tropical and geographical medicine, 1986. **38**(3): p. 209-214.

118. Tomori, O., et al., *Viral hemorrhagic fever antibodies in Nigerian populations.* The American journal of tropical medicine and hygiene, 1988. **38**(2): p. 407-410.

119. Knobloch, J., E.J. Albiez, and H. Schmitz, *A serological survey on viral haemorrhagic fevers in liberia.* Annales de l'Institut Pasteur / Virologie, 1982. **133**(2): p. 125-128.

120. Glynn, J.R., et al., *Asymptomatic infection and unrecognised Ebola virus disease in Ebola-affected households in Sierra Leone: a cross-sectional study using a new non-invasive assay for antibodies to Ebola virus.* Lancet Infect Dis, 2017. **17**(6): p. 645-653.

121. Timothy, J.W.S., et al., *Early transmission and case fatality of Ebola virus at the index site of the 2013-16 west African Ebola outbreak: a cross-sectional seroprevalence survey.* Lancet Infectious Diseases, 2019. **19**(4): p. 429-438.

122. Bower, H., et al., *Exposure-Specific and Age-Specific Attack Rates for Ebola Virus Disease in Ebola-Affected Households, Sierra Leone.* Emerg Infect Dis, 2016. **22**(8): p. 1403-11.

123. Sanchez, A., K.E. Wagoner, and P.E. Rollin, *Sequence-based human leukocyte antigen-B typing of patients infected with Ebola virus in Uganda in 2000: identification of alleles associated with fatal and nonfatal disease outcomes.* Journal of Infectious Diseases, 2007. **196 Suppl 2**: p. S329-36.

124. Wauquier, N., et al., *Association of KIR2DS1 and KIR2DS3 with fatal outcome in Ebola virus infection.* Immunogenetics, 2010. **62**(11-12): p. 767-771.

125. Liu, X., et al., *Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus.* Genome biology, 2017. **18**(1): p. 4.

126. Kuroda, M., et al., *Interaction between TIM-1 and NPC1 Is Important for Cellular Entry of Ebola Virus.* J Virol, 2015. **89**(12): p. 6481-93.

127. Kondoh, T., et al., *Single-Nucleotide Polymorphisms in Human NPC1 Influence Filovirus Entry Into Cells.* J Infect Dis, 2018. **218**(suppl_5): p. S397-S402.

128. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.

129. Purcell, S., S.S. Cherny, and P.C. Sham, *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.* Bioinformatics, 2003. **19**(1): p. 149-50.

130. McEvoy, B.P., et al., *Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs.* Genome Res, 2011. **21**(6): p. 821-9.

131. Campbell, M.C. and S.A. Tishkoff, *African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 403-33.

132. Tedder, R.S., et al., *Detection, characterization, and enrollment of donors of Ebola convalescent plasma in Sierra Leone.* Transfusion, 2018. **58**(5): p. 1289-1298.

133. Diallo, M.S.K., et al., *Prevalence of infection among asymptomatic and paucisymptomatic contact persons exposed to Ebola virus in Guinea: a retrospective, cross-sectional observational study.* Lancet Infect Dis, 2019. **19**(3): p. 308-316.

134. Glynn, J.R., *Age-specific incidence of Ebola virus disease.* Lancet, 2015. **386**(9992): p. 432.

135. Leung, C.H., et al., *Clinical characteristics of children and adults hospitalized for influenza virus infection.* J Microbiol Immunol Infect, 2014. **47**(6): p. 518-25.

136. Ukwuani, F.A., A.O. Tsui, and C.M. Suchindran, *Condom use for preventing HIV infection/AIDS in sub-Saharan Africa: a comparative multilevel analysis of Uganda and Tanzania.* J Acquir Immune Defic Syndr, 2003. **34**(2): p. 203-13.

137. Ayiga, N., *Rates and predictors of consistent condom-use by people living with HIV/AIDS on antiretroviral treatment in Uganda.* J Health Popul Nutr, 2012. **30**(3): p. 270-80.

138. Nelson, J.M., et al., *Live neonates born to mothers with Ebola virus disease: a review of the literature.* Journal of perinatology : official journal of the California Perinatal Association, 2016. **36**(6): p. 411-414.

139. Moreau, M., et al., *Lactating mothers infected with Ebola virus: EBOV RT-PCR of blood only may be insufficient.* Euro Surveill, 2015. **20**(3).

140. Broadhurst, M.J., T.J. Brooks, and N.R. Pollock, *Diagnosis of Ebola Virus Disease: Past, Present, and Future.* Clin Microbiol Rev, 2016. **29**(4): p. 773-93.

141. Green, E., et al., *Viraemia and Ebola virus secretion in survivors of Ebola virus disease in Sierra Leone: a cross-sectional cohort study.* Lancet Infect Dis, 2016. **16**(9): p. 1052-1056.

142. He, X.S., et al., *Cellular immune responses in children and adults receiving inactivated or live attenuated influenza vaccines.* J Virol, 2006. **80**(23): p. 11756-66.

143. Wayne, M.L. and L.M. McIntyre, *Combining mapping and arraying: An approach to candidate gene identification.* Proc Natl Acad Sci U S A, 2002. **99**(23): p. 14903-6.

144. Manichaikul, A., et al., *Robust relationship inference in genome-wide association studies.* Bioinformatics, 2010. **26**(22): p. 2867-73.

145. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

146. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

147. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

148. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

149. DeVay, R.M., D.L. Shelton, and H. Liang, *Characterization of Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Trafficking Reveals a Novel Lysosomal Targeting Mechanism via Amyloid Precursor-like Protein 2 (APLP2).* Journal of Biological Chemistry, 2013. **288**(15): p. 10805-10818.

150. Horton, J.D., J.C. Cohen, and H.H. Hobbs, *Molecular biology of PCSK9: its role in LDL metabolism.* Trends Biochem Sci, 2007. **32**(2): p. 71-7.

151. Jefcoate, C.R. and J. Lee, *Cholesterol signaling in single cells: lessons from STAR and sm-FISH.* J Mol Endocrinol, 2018. **60**(4): p. R213-R235.

152. Myant, N.B., *The metabolism of plasma cholesterol.* Biochem J, 1971. **122**(1): p. 2P-3P.

153. Soccio, R.E., et al., *The cholesterol-regulated StarD4 gene encodes a StAR-related lipid transfer protein with two closely related homologues, StarD5 and StarD6.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6943-8.

154. Verhelst, J., P. Hulpiau, and X. Saelens, *Mx Proteins: Antiviral Gatekeepers That Restrain the Uninvited.* Microbiology and Molecular Biology Reviews, 2013. **77**(4): p. 551-566.

155. Zhou, X., et al., *The mitochondrial carrier SLC25A10 regulates cancer cell growth.* Oncotarget, 2015. **6**(11): p. 9271-83.

156. Luo, M. and D. Peng, *The emerging role of apolipoprotein C-III: beyond effects on triglyceride metabolism.* Lipids Health Dis, 2016. **15**(1): p. 184.

157. Shapiro, M.D. and S. Fazio, *Apolipoprotein B-containing lipoproteins and atherosclerotic cardiovascular disease.* F1000Res, 2017. **6**: p. 134.

158.   Wang, S., et al., *Deficiency of receptor-associated protein attenuates angiotensin II-induced atherosclerosis in hypercholesterolemic mice without influencing abdominal aortic aneurysms.* Atherosclerosis, 2012. **220**(2): p. 375-80.

159.   Mulder, N., et al., *H3Africa: current perspectives.* Pharmacogenomics & Personalized Medicine, 2018. **11**: p. 59-66.

160.   Vergara, C., et al., *Genotype imputation performance of three reference panels using African ancestry individuals.* Human Genetics, 2018. **137**(4): p. 281-292.

161.   Supuran, C.T., *Carbonic anhydrase inhibitors and activators for novel therapeutic applications.* Future Medicinal Chemistry, 2011. **3**(9): p. 1165-1180.

162.   van Karnebeek, C.D., et al., *Mitochondrial Carbonic Anhydrase VA Deficiency Resulting from CA5A Alterations Presents with Hyperammonemia in Early Childhood.* American Journal of Human Genetics, 2014. **94**(3): p. 453-461.

163.   Young, S.G. and L.G. Fong, *Lowering plasma cholesterol by raising LDL receptors--revisited.* N Engl J Med, 2012. **366**(12): p. 1154-5.

164.   Fedson, D.S., et al., *Treating the Host Response to Ebola Virus Disease with Generic Statins and Angiotensin Receptor Blockers.* mBio, 2015. **6**(3): p. e00716.

Appendix

A. Fieldwork undertaken to establish cohorts for genetic analysis

A.1 Early stages

In early 2016 I developed a relationship with the University of Makeni. This led to the utilisation of their infectious diseases research laboratory as a base for the fieldwork, and a location to conduct the initial laboratory work including heat treatment of saliva samples for shipment and the EIA on the oral fluid samples. Sierra Leone had no formal shipping channels for laboratory equipment and consumables, and it was not possible to purchase any items in country. There was also no infrastructure for onward shipment by freight companies from the airport in Freetown to Makeni where the project would be based (approximately a three hour drive from Freetown). I collaborated with Public Health England (PHE) and my supervisor Dr Tim Brooks based at PHE Porton Down and was able to ship our equipment alongside one of PHE's shipments. This was collected by the PHE team in Freetown and stored until I arrived in country and transported the items to Makeni. In the end I shipped 63 boxes of equipment worth a total value of just over £60,000.

PHE were operating in Sierra Leone under DfID, who held a duty waiver for aid and scientific research supplies under which the shipment was due to progress. A week before the shipment was due to leave the UK in July of 2016, the government of Sierra Leone rescinded the duty waiver as they deemed the epidemic emergency to be over. This resulted in an additional charge of approximately £17,000 for the shipment to clear customs at the airport (15% VAT, 20% duty tax), which had not been budgeted for.

I arrived in Freetown in the middle of November 2016 and was met by the two drivers and vehicles I had arranged. We collected the equipment that I had shipped in the summer of 2016 from the PHE compound in Freetown and drove the items to Makeni, this involved a second trip to Freetown to collect some remaining items which did not fit in the vehicles for the initial trip. On the drive to Makeni we were stopped at a police check point where there was a clear intention of trying to extort a bribe. I navigated this without paying a bribe and there were no further incidents during the journey to Makeni.

A.2 Employing and training a field team

Through collaborators at the London School of Hygiene and Tropical Medicine and Cambridge University, I identified six field team members. I deliberately ensured the field team members had previous understanding of what Ebola patients and their families had endured, knowledge of infection

prevention and control practices and for half of the team, experience of having undertaken a similar form of recruitment using the risk stratification. I included two Ebola survivors from two different regions that had been affected during the epidemic, as this provided a safe means by which to collect samples and helped us to engage better with communities and Ebola survivors. The field team were also identified for the languages they were able to speak to ensure all the main local languages of the participants we were targeting for recruitment were covered. The University of Makeni provided the field team with formal employment contracts and arranged their payment through money transferred to the University from Imperial College.

The training for the field team included the nature of the study, how to obtain consent and how to recruit to the study. This was largely conducted through repeated scenarios of obtaining consent and conducting recruitment and completing the fieldwork forms. There was also the opportunity for all of them to provide a saliva and oral fluid sample themselves, so they would understand how to encourage the study participants when they were providing samples. During this time, we also discussed how we would approach the communities and what incentives the team felt were appropriate. After ten days of training the team were ready to commence recruitment and so contact was initiated with the community leaders of the first recruitment location.

A.3 Purchasing equipment

I purchased a -20°C freezer for the lab at the market in Makeni, as well as all the community incentives/gifts. Following discussion with the field team it had been decided that each household of four people would be provided with a bucket with spigot and a bar of soap to reiterate the importance of hand washing. Each participant would also be provided with a bag of water, small packet of biscuits and a bottle of soft drink. Children would also receive a balloon and a lollipop to encourage the younger children to be willing to have a mouth swab put in their mouths in exchange for a lollipop later. All the items were then safely stored at the laboratory at the University of Makeni.

I used the University printing service to print all of the recruitment and consent forms at regular intervals throughout the recruitment. All recruitment equipment, except for sample collection kits, incentives and plastic tables, was transported to the field in one of the large Peli Storm cases I had brought with me, to keep it dry and dust free. This included recruitment and consent forms, barcoded sample labels, stationary, personal protective equipment, Safety Gel and Virkon for managing spills, high level disinfectant wipes for surface/sample tube cleaning and a first aid kit.

A.4 Identifying communities for recruitment and selecting sample kits

The population of Sierra Leone is made up of several different ethnicities. There are more dominant ethnicities such as Temne, Mende and Limba and other much smaller ethnicities such as Susu, Sherbro and Yalunka. Data on ethnicities of patients was not collected during the epidemic and may not have been reliable as some people may have been of mixed ethnicity but only reported their maternal or paternal ethnicity. In some cases, people also adopt a different ethnicity if they are from a more minor tribe but are living in an area dominated by a different ethnicity. Many urban communities in Sierra Leone are also of mixed ethnicity. In order to control for ethnicity, we attempted to trace the communities linked to the patient samples we had access to from the MOHSL-PHE Ebola biobank. This enabled us to recruit our household contacts, community controls and survivor populations from the same communities. The recruitment distribution maps below (Figure A-1) demonstrate the locations of the samples from the MOHS-PHE Ebola biobank (red) and the samples recruited through fieldwork (blue, purple and green). While not exactly matched, this helped prevent variations in group genomic data secondary to dramatic differences in ethnicity.
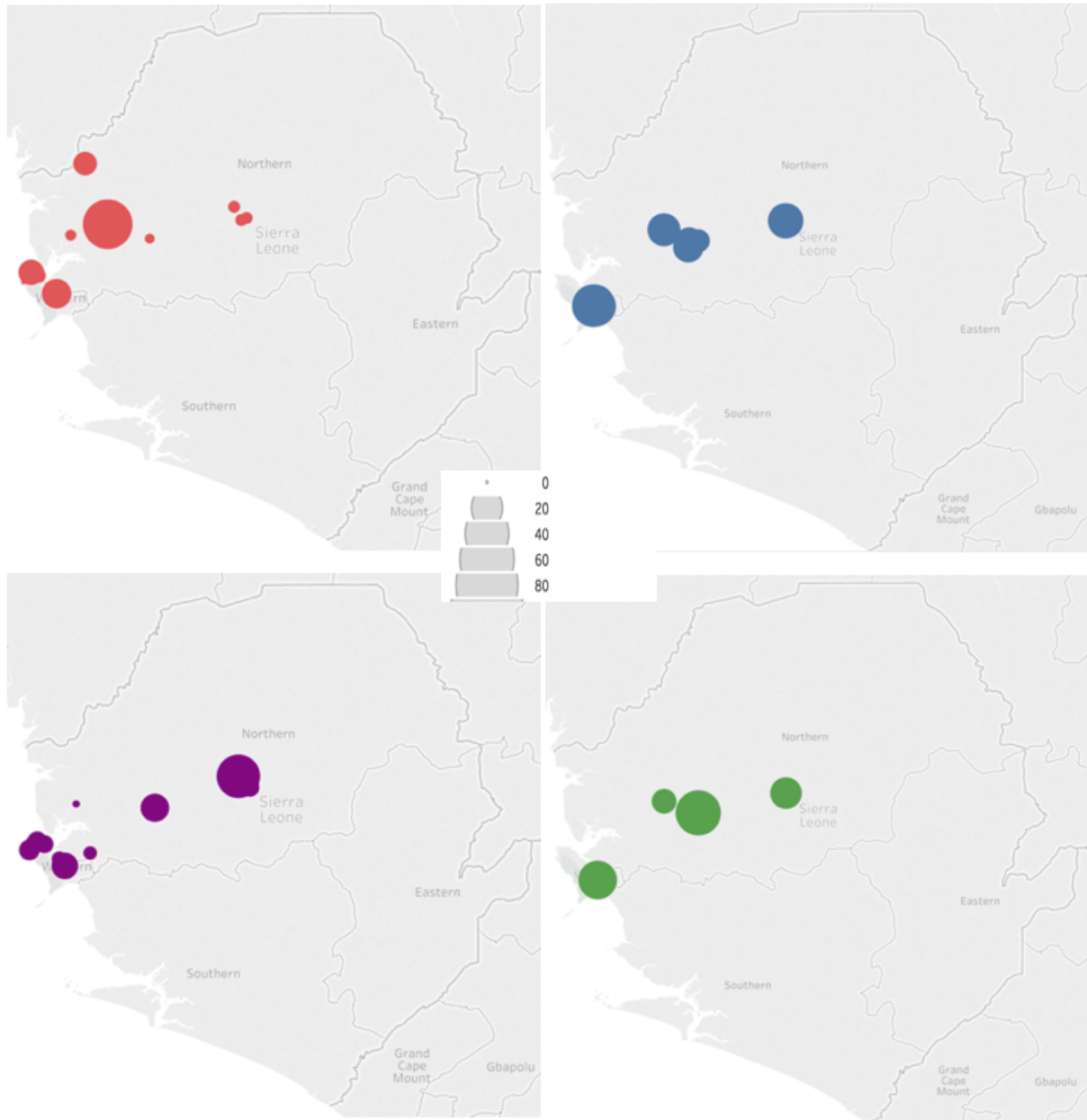
Figure A-1. Maps of Sierra Leone demonstrating participant recruitment location and distribution. MOHSL-PHE Ebola biobank samples = Red, Household contacts = Blue, Prospective Ebola survivors = Purple, and Community Controls = Green.

Communities were therefore identified for recruitment based on access to biobank samples. Through a process described below, communities were approached and sensitised about the study. To reduce community concerns and fear of taking part in the study I tried to minimise the need for invasive sampling and selected a saliva collection kit manufactured by *DNA Genotek*, the *Oragene OG-500* or *OG-575* (for young children) kits, and an oral fluid Ebola IgG capture assay designed by Richard Tedder and his team at PHE, but now in manufacture by *Kalon Diagnostics Ltd*.

The communities were relatively receptive to these methods of sampling, although had their apprehensions about putting the mouth swab in their mouths in the event it was contaminated. We attempted to allay fears when sensitising communities by requesting a community member select a sealed swab from a large box of swabs. One of the field team would then open the swab and rub it around their mouth themselves, to reassure the community the procedure was safe, and no harm was intended.

A.5 Community engagement/sensitisation strategy

The communities targeted for recruitment had all been significantly affected by Ebola. As a consequence, they remained extremely afraid of any discussions on the topic and continued to harbour many superstitions about the epidemic. This resulted in reticence to engage with a study on Ebola. In an attempt to overcome this situation, I not only ensured I had the appropriate official governmental approvals for the research to be conducted, I also approached the communities through the traditional hierarchical structure.

Due to years of civil war in Sierra Leone, which had resolved less than 10 years before the epidemic, communities are generally untrusting of political leaders and entities. Communities are much more trusting of their immediate chief, this is in part because this chief is known to them and either has been elected by the community or has inherited his chiefdom through a family line that are well known to the community. There is significantly more onus and responsibility on the local chief to protect his people and do good for them, than is perceived of a political leader.

The traditional structure in Sierra Leone varies in different parts of Sierra Leone and from rural to urban areas. However, the predominant structure in the regions I recruited in was that commencing with a Paramount Chief, who covers a large region. This is followed by a Section Chief who oversees several local communities. Under each Paramount Chief there are multiple Section Chiefs. For example, in Marampa Kingdom there were approximately 30 Section Chiefs under the Paramount Chief. After the Section Chief is the Local Chief who lives in the immediate community in which he is the Chief. Under each Section Chief there are several communities (approximately 5-10), but the Local Chief is Chief only to the community in which he resides. While many challenges were faced in recruitment, the communities where the greatest resistance was experienced, or where it was not possible to recruit at all, were those where the local chief had a poor relationship with the section chief, or where there was division in the community with some community members disrespecting the local chief.

In order to commence recruitment, we first arranged to meet the Paramount Chief and explain the study. With his approval in place, he would gather the Section Chiefs for his region, and we would meet with them and explain the study. The Section Chiefs who oversaw communities that had been affected by Ebola would then travel with us and introduce us to the Local Chief of each affected community, where we would again explain the study. With the approval of all chiefs in place we would then arrange with the Local Chief to return to their community at a later date to sensitise the entire community to the study. The Local Chief would introduce us to the community and would provide his backing and support to the study to encourage community members to be involved. Once recruitment began, we would always commence by recruiting the chief's household first to demonstrate his support to the study and that it was safe to be involved.
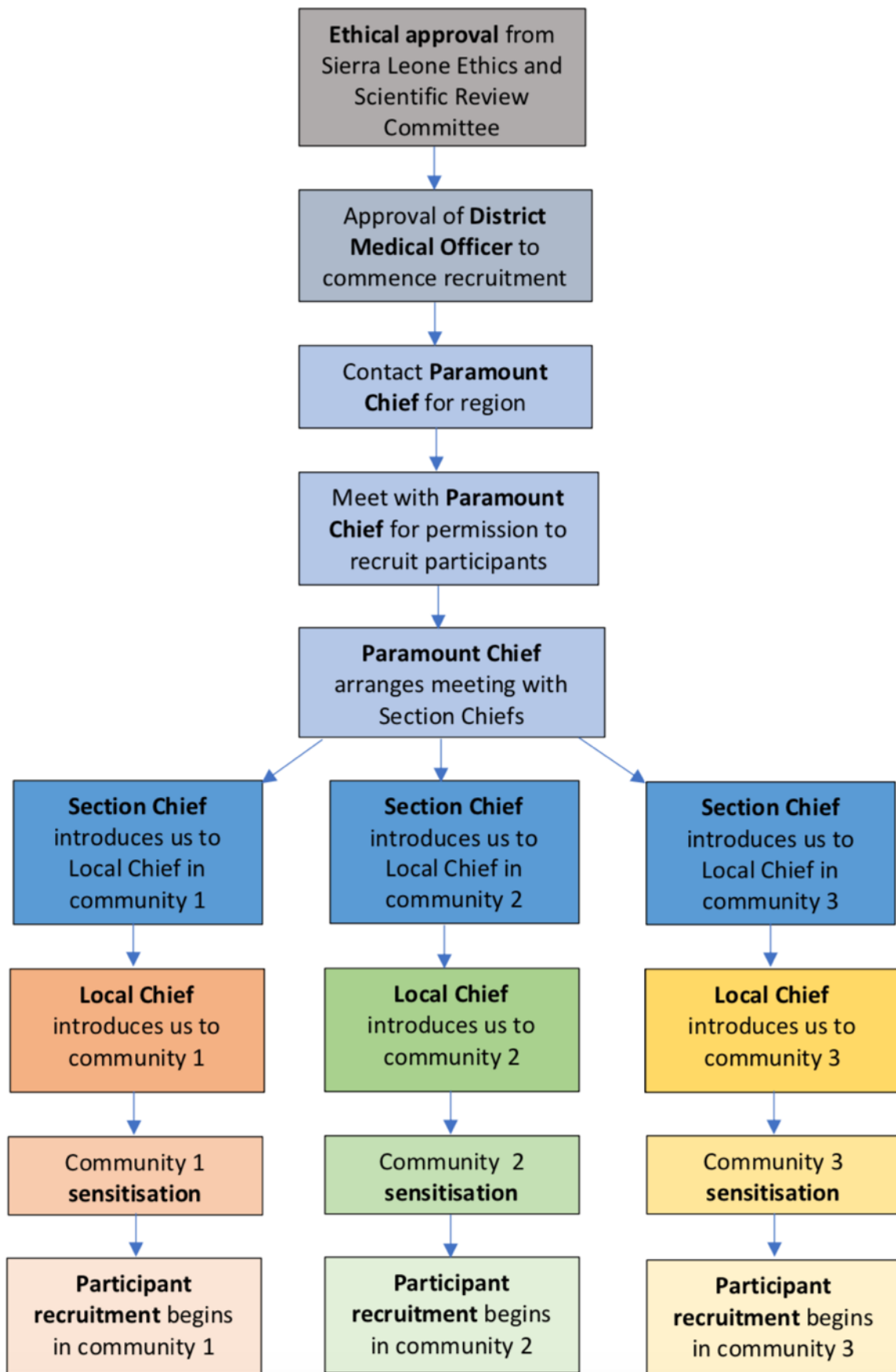
Figure A-2 Flow chart of the community engagement and sensitisation process.

A.6 Undertaking the fieldwork

Participants were recruited in their communities in a neutral location, either the community meeting place, outside the chief's house or under a centrally located mango tree. At the first community the chief was keen for us to conduct recruitment outside people's homes, but it became clear within the first few days that this was perceived as threatening by community members and a neutral, central location was preferable. In each community two recruitment stations were set up, dividing the field team into two groups of three. A survivor was present in each group of three to assist with obtaining the samples. Below is a list of the equipment taken to the field for recruitment of 100 participants:

- Satellite phone & navigation system (Iridium 9575 Extreme)
- 2x Plastic table
- Study questionnaire – 100 copies
- Pens – 4 items
- Fine tipped indelible markers – 4 items
- Adhesive barcode study ID labels – 1 roll community controls & 1 roll household contacts
- Oragene DNA OG-500 – 75 items
- Oragene DNA OG-575 – 25 items
- Oral swabs – 100 items
- Oral swab storage receptacle – 100 items
- Chlorhexidine skin wipes – 150 items
- Gauze – 1 pack
- Micropore – 4 rolls
- Leak proof, sealable plastic biohazard bags & biohazard sticker – 300 items
- Medipal high level disinfectant wipes 225/canister – 2 canisters
- iM3075 Storm case – 1 item
- 80QT wheeled cooler (containing ice) – 1 item
- Heavy duty clinical waste bags – 1 roll (200/roll)
- Spray can containing Virkon tablets (to be filled with water if necessary)
- Virkon granules in Squeeze-puff container
- Safetygel shakerpot x 2
- Shovel
- Alcohol gel – 4 bottles
- Bottled water – 21 litres (3 litres/person for drinking)
- Face visor & mask – 1 pack (25/pack)

- FFP3 facemasks plus valve – 1 box

- Nitrile gloves M/L – 1 box each

- Thumbs up gowns – 1 pack

- Rubber boots – 1 pair/team member

- Multicoloured balloons

- Community incentives – bucket with spigot, bar of soap, water, soft drinks, biscuits, lollipops

Recruitment was conducted using handwritten paper forms (see Appendix G, H and I), while this was not as efficient for data entry as an electronic method, it was the most straight forward method given the location of the communities, absence of network signals, and internet data charges. Recruitment was conducted with all the members of the family present who were willing to participate. The majority of study participants did not speak English or Krio, and many were not able to read or write. For this reason, the purpose of the study and what would be expected of participants was clearly explained and demonstrated by a member of the field team during the community sensitisation process described above. The field team were selected to cover all the major languages of the peoples of Sierra Leone including, English, Krio, Temne, Limba, Fula, Mende and Loko. Therefore, depending on the community being addressed the field team member(s) speaking the dominant language of the population conducted the explanation of the study.

At the time of recruitment, the purpose of the study, requirements of the participants and consent form were explained again to the individual participating families in the relevant language, and participation was agreed by the head of the household and all household members old enough to consent or assent. All participants signed, or thumb printed their consent to take part, parents/guardians consented for children under the age of 18 years. For the household contacts, a household inventory was completed first to determine how members of each family were related and who was living together in the same house (see Appendix G). This included Ebola deceased and surviving members of the family. Individual questionnaires were then completed for each participant by a member of the field team reading the questions to the participant and marking their answers on the questionnaire. Samples were then obtained, labelling both the samples and the inventory and questionnaire with the relevant participant ID sticker and barcode. A socio-economic status questionnaire was also completed for each household to determine access to basic hygiene and sanitation facilities (see Appendix G and H). Following collection of all samples, discussion took place with the whole family to determine levels of contact within the household or community. This was then recorded for each household member on the risk assessment at the end of each individual

questionnaire. For recruitment of community controls a similar process occurred, but a household inventory was not completed. Once recruitment of a household had been concluded, each household was provided with the gifts described above to acknowledge appreciation of their participation.

Following recruitment of each household I checked the recruitment forms to ensure no data was missing and to clarify any apparent errors. I then collected the samples and placed the oral fluid samples in a cooler on ice for transport back to the lab. Saliva samples were placed in a second Storm case, manufactured by *Peli Products UK*, a dust and weather resistant, pressure equalising case for transport back to the lab. On return to the lab samples were offloaded, the oral fluid swabs placed in a -20 °C freezer and the saliva samples heat treated at 60 °C for 60 minutes and then stored in specifically designed *Oragene* tube storage boxes. These were then placed in a sealed outer plastic container.

At times temperatures reached 40°C in the middle of the day during recruitment. While one might imagine that it was difficult to produce enough saliva and oral fluid for the samples in such temperatures, the majority of participants did not struggle. Those most likely to struggle were the very elderly, but we permitted them all the time they needed, and the field team actively encouraged them. The majority of children were very cooperative, particularly when they realised they would receive a lollipop at the end, and did not struggle to produce saliva or spit in the tubes. Children under the age of 4 tended to struggle as they were not yet able to spit. In these cases the OG-575 tubes came with absorbent sponges that could be placed under the tongue and then wrung out into the collection device. In most children this did not take longer than a few minutes, but in the youngest children aged 2-3 years old this was more of a struggle and could take a bit longer. Where possible the mother would hold the younger children on her lap and place the sponge under their tongue returning it to the field team member to wring into the collection device. Very few children did not cooperate with this process when they realised they would receive a lollipop afterwards. Of the handful who did not wish to cooperate, entertainment with balloons seemed to solve the problem and all samples were successfully obtained with minimum distress. On average approximately 70 participants were recruited each day, with a maximum of 105 recruited on one day. In the final full week of recruitment in March 2017 445 participants were recruited.

A.7 Engaging Ebola survivors

To recruit Ebola survivors a slightly different approach was used, as discussion and approval was additionally required from the local Ebola survivors committee. Provided the committee were supportive of the recruitment they would disseminate information about the study to the Ebola

survivors within their locality and invite them to attend for recruitment at either the local Ebola survivor clinic or the offices/meeting hall of the local Ebola survivors committee. The field team was scaled down to three members, with two members conducting recruitment and one member (an Ebola survivor) collecting the samples from participants.

While we attempted to obtain clinical records from the Ebola survivor clinics in relation to the ongoing problems of the survivors we recruited, it became apparent early on that this was going to be neither a robust nor consistent source of information. The survivor clinics had been largely operated by NGOs in the first six to twelve months following the epidemic, but as soon as the government of Sierra Leone declared the Ebola emergency over, the government coordinated this service instead. In many cases this resulted in the cessation of any formal clinical services provided to Ebola survivors, with no patient records available. For this reason, the participant questionnaire that was utilised for the Ebola survivors included a self-reporting of clinical diagnoses/symptoms (Appendix I). In order to account for the possibility of false reporting, there was the option to select certain milder options to help differentiate from those suffering from life limiting symptoms. This was done in order to help stratify Ebola survivors for severity of ongoing symptoms when conducting the analysis of the antibody and genomic data.

Ebola survivors were recruited individually, rather than in families, however documentation was made if any of the survivors were related to one another. Each survivor completed a questionnaire and provided an *Oragene* saliva sample and an oral fluid swab. These were handled in a similar manner to the samples from the household contact and community control recruitment and stored at the laboratory at the University of Makeni. Each survivor was also provided with a gift of appreciation for their participation, which included a blanket, a bag of water to rinse their mouth with and a soft drink. Children were also given a lollipop and a balloon. As survivor recruitment took place in a central location for the region, survivors travelled to the recruitment location and so were reimbursed for their transport costs (Le 10,000-20,000, equivalent to £1.25-2.50).

A.8 Managing fieldwork data

As all study participants were recruited using paper forms, this posed challenges in terms of storage, backup and transport of data forms to the UK for data entry. Upon completion of recruitment, the participant recruitment questionnaires filled 20 large ring binders. As the forms contained participant identifiable information, they were stored in a locked cabinet in the laboratory office. For the purposes of having an electronic record of all the forms and for transporting the data back to the UK, a small portable scanner was purchased in the UK and transported to Sierra Leone. All forms were then scanned to ensure an electronic copy of all participant recruitment and consent forms. These copies

are stored on password protected computers, and through the employment of a data entry clerk, all data was entered in anonymised format into an online database that is backed up to the Imperial College server on a daily basis. This database is also password protected.

A.9 Fieldwork challenges

In November 2016 the government of Sierra Leone removed the fuel tax subsidy they normally applied, in order to disincentivise a black market trading ring in Guinea. This resulted in a doubling of the fuel cost in Sierra Leone from Leones (Le) 3000 per litre to Le 6000 per litre. Such an increase had not been budgeted for in the grant costings and at times £60-80 was spent per day just on fuel for the two vehicles used.

Accessing funds in Sierra Leone was difficult. Credit and debit cards are not widely used and there are few functioning ATMs outside Freetown. For this reason, it was arranged that Imperial College would transfer the money to pay the field team and for general fieldwork costs to the University of Makeni who would provide me with cash in Leones. Unfortunately, there were frequent delays and obstacles in the money being transferred resulting in the field team not being paid for prolonged periods and there being no money for me to access for fuel and community incentives costs. This meant attempting to withdraw the fuel costs from my personal bank account via an ATM. The ATM's in Makeni were often out of order or would frequently run out of money.

Despite being promised nearly new vehicles by the vehicle hire company we were provided with vehicles which were approximately 20 years old, subsequently there were regular breakdowns which often impacted on our ability to take the full team to the field to recruit participants. Vehicles were also often swapped to provide them with 'maintenance' and sub-standard vehicles provided in their place. The drivers were also often not paid, resulting in them asking to borrow money or needing to return the vehicle to Freetown as they could not afford to remain in Makeni. On one occasion we were involved in a road traffic accident due to the poor quality of the vehicle and one of the nearly bare tires exploding while we were driving on the highway.

Several factors compounded the speed with which we were able to recruit study participants, fortunately despite these issues we were still able to reach our target recruitment numbers. While many of the communities were very receptive to the team and the study, some communities were very resistant towards us at the time of recruitment. At times this put the team at risk, but fortunately we were sensitive to the circumstances and left the community before any resistance turned to violence.

Similarly, at times we faced bribery and corruption, usually within local government departments, fortunately we were able to navigate this without paying any bribes, however it did delay the progress of the fieldwork at times. There were also struggles with attendance of some members of the field team, and at times we operated with just four team members instead of six.

B. Laboratory work in Sierra Leone

The laboratory work in Sierra Leone was conducted at the Infectious disease research laboratory at the University of Makeni, which had been set up by Prof Goodfellow at the University of Cambridge immediately following the Ebola epidemic in Sierra Leone.

I have described the sample handling and processing in chapter 2 above, along with the process for conducting the EIA. In September 2017 I coordinated the shipping of the saliva samples back to the UK for DNA extraction and sequencing. This was a protracted process of obtaining permissions and signatures from multiple different government offices, none of whom coordinated with each other. It also was not possible to obtain electronic signatures or arrange for documents to be emailed from one office to another. The original signed document had to be presented at each office in order to receive the next signed document, which often took several days to obtain. All of these offices were in Freetown, subsequently I travelled to Freetown approximately twice a week for several weeks. Despite a last minute error in a signed letter, all permissions were obtained and the samples collected from the lab by the courier and transported successfully to the UK.

B.1 Sierra Leone Laboratory challenges

While the laboratory work conducted in Sierra Leone was based at the laboratory at the University of Makeni, there was some laboratory equipment I required based in a lab at the Holy Spirit Hospital in Makeni. This lab had been developed by Prof Vittorio Colizzi from the University of Rome. At the beginning of November 2016, the lab was struck by lightning, which triggered an electrical fire resulting in the entire laboratory and all equipment being destroyed (Figure B-1). This meant an additional purchase of a plate reader to conduct the EIAs, which had to be transported out to Sierra Leone as hand luggage to protect it.

Figure B-1 The lab at the Holy Spirit Hospital following a lightning strike and lab fire in November 2016. Makeni, Bombali District, Sierra Leone. July 2017.

Power surges and lightning strikes were an ongoing problem in Sierra Leone and upon my arrival in November 2016 there was no mains power due to a lightning strike which damaged the main hydroelectric dam, this remained the case until February 2017. During this time conducting laboratory work was a challenge. The lab generator was also old and underpowered for the needs of the lab. If used for prolonged periods it would overheat and break down. There was no backup generator, so if the generator broke down it would mean several days to a week with no power, during which time the freezers would start to defrost. A new generator was purchased in February 2017, but unfortunately this was damaged by a lightning strike while my colleagues and I were in the laboratory in August 2017, causing problems conducting the EIA.

There were delays in accessing the EIA kits from PHE, due to the scientist who developed the EIA being diagnosed with terminal cancer and succumbing to the disease in early 2016. Subsequently the EIA was passed to a commercial company to manufacture, *Kalon Diagnostics Ltd*. The commercialised assay was not available until late spring 2017, meaning there was not the time to test run the assay in the UK prior to using it in Sierra Leone. Fortunately, the assay performed very consistently. Once I had accessed the EIA kits I required, I transported them out to Sierra Leone from the UK in my luggage. Unfortunately, there was an impromptu strike of all baggage handlers in Brussels resulting in £8000 worth of EIA kits sitting on the tarmac in Brussels in 30 °C heat for 36 hours before arriving in Freetown. Fortunately, the manufacturer reassured me the kits were stable at room temperature for up to one week.

C. Plate templates

Plate templates for the anti-EBOV IgG EIA

A. EIA plate template – Contact & control plate

	1	2	3	4	5	6	7	8	9	10	11	12
A	PC	Contact sample 3	Contact sample 11	Contact sample 19	Contact sample 27	Contact sample 35	Contact sample 43	Control sample 6	Control sample 14	Control sample 22	Control sample 30	Control sample 38
B	PC	Contact sample 4	Contact sample 12	Contact sample 20	Contact sample 28	Contact sample 36	Contact sample 44	Control sample 7	Control sample 15	Control sample 23	Control sample 31	Control sample 39
C	CO	Contact sample 5	Contact sample 13	Contact sample 21	Contact sample 29	Contact sample 37	Contact sample 45	Control sample 8	Control sample 16	Control sample 24	Control sample 32	Control sample 40
D	CO	Contact sample 6	Contact sample 14	Contact sample 22	Contact sample 30	Contact sample 38	Control sample 1	Control sample 9	Control sample 17	Control sample 25	Control sample 33	Control sample 41
E	NC	Contact sample 7	Contact sample 15	Contact sample 23	Contact sample 31	Contact sample 39	Control sample 2	Control sample 10	Control sample 18	Control sample 26	Control sample 34	Control sample 42
F	NC	Contact sample 8	Contact sample 16	Contact sample 24	Contact sample 32	Contact sample 40	Control sample 3	Control sample 11	Control sample 19	Control sample 27	Control sample 35	Control sample 43
G	Contact sample 1	Contact sample 9	Contact sample 17	Contact sample 25	Contact sample 33	Contact sample 41	Control sample 4	Control sample 12	Control sample 20	Control sample 28	Control sample 36	Control sample 44
H	Contact sample 2	Contact sample 10	Contact sample 18	Contact sample 26	Contact sample 34	Contact sample 42	Control sample 5	Control sample 13	Control sample 21	Control sample 29	Control sample 37	Control sample 45

B. EIA plate template – Survivor plate

	1	2	3	4	5	6	7	8	9	10	11	12
A	PC	Survivor sample 3	Survivor sample 11	Survivor sample 19	Survivor sample 27	Survivor sample 35	Survivor sample 43	Survivor sample 51	Survivor sample 59	Survivor sample 67	Survivor sample 75	Survivor sample 83
B	PC	Survivor sample 4	Survivor sample 12	Survivor sample 20	Survivor sample 28	Survivor sample 36	Survivor sample 44	Survivor sample 52	Survivor sample 60	Survivor sample 68	Survivor sample 76	Survivor sample 84
C	CO	Survivor sample 5	Survivor sample 13	Survivor sample 21	Survivor sample 29	Survivor sample 37	Survivor sample 45	Survivor sample 53	Survivor sample 61	Survivor sample 69	Survivor sample 77	Survivor sample 85
D	CO	Survivor sample 6	Survivor sample 14	Survivor sample 22	Survivor sample 30	Survivor sample 38	Survivor sample 46	Survivor sample 54	Survivor sample 62	Survivor sample 70	Survivor sample 78	Survivor sample 86
E	NC	Survivor sample 7	Survivor sample 15	Survivor sample 23	Survivor sample 31	Survivor sample 39	Survivor sample 47	Survivor sample 55	Survivor sample 63	Survivor sample 71	Survivor sample 79	Survivor sample 87
F	NC	Survivor sample 8	Survivor sample 16	Survivor sample 24	Survivor sample 32	Survivor sample 40	Survivor sample 48	Survivor sample 56	Survivor sample 64	Survivor sample 72	Survivor sample 80	Survivor sample 88
G	Survivor sample 1	Survivor sample 9	Survivor sample 17	Survivor sample 25	Survivor sample 33	Survivor sample 41	Survivor sample 49	Survivor sample 57	Survivor sample 65	Survivor sample 73	Survivor sample 81	Survivor sample 89
H	Survivor sample 2	Survivor sample 10	Survivor sample 18	Survivor sample 26	Survivor sample 34	Survivor sample 42	Survivor sample 50	Survivor sample 58	Survivor sample 66	Survivor sample 74	Survivor sample 82	Survivor sample 90

Plate template for exome sequencing

Exome sequencing plate template

	1	2	3	4	5	6	7	8	9	10	11	12
A	Deceased sample 1	Deceased sample 9	Deceased sample 17	Survivor sample 5	Survivor sample 13	AB+contact sample 1	AB+contact sample 9	AB+contact sample 17	Ab-contact sample 5	Ab-contact sample 13	Control sample 1	Control sample 9
B	Deceased sample 2	Deceased sample 10	Deceased sample 18	Survivor sample 6	Survivor sample 14	AB+contact sample 2	AB+contact sample 10	AB+contact sample 18	Ab-contact sample 6	Ab-contact sample 14	Control sample 2	Control sample 10
C	Deceased sample 3	Deceased sample 11	Deceased sample 19	Survivor sample 7	Survivor sample 15	AB+contact sample 3	AB+contact sample 11	AB+contact sample 19	Ab-contact sample 7	Ab-contact sample 15	Control sample 3	Control sample 11
D	Deceased sample 4	Deceased sample 12	Deceased sample 20	Survivor sample 8	Survivor sample 16	AB+contact sample 4	AB+contact sample 12	AB+contact sample 20	Ab-contact sample 8	Ab-contact sample 16	Control sample 4	Control sample 12
E	Deceased sample 5	Deceased sample 13	Survivor sample 1	Survivor sample 9	Survivor sample 17	AB+contact sample 5	AB+contact sample 13	Ab-contact sample 1	Ab-contact sample 9	Ab-contact sample 17	Control sample 5	Control sample 13
F	Deceased sample 6	Deceased sample 14	Survivor sample 2	Survivor sample 10	Survivor sample 18	AB+contact sample 6	AB+contact sample 14	Ab-contact sample 2	Ab-contact sample 10	Ab-contact sample 18	Control sample 6	Control sample 14
G	Deceased sample 7	Deceased sample 15	Survivor sample 3	Survivor sample 11	Survivor sample 19	AB+contact sample 7	AB+contact sample 15	Ab-contact sample 3	Ab-contact sample 11	Ab-contact sample 19	Control sample 7	Control sample 15
H	Deceased sample 8	Deceased sample 16	Survivor sample 4	Survivor sample 12	Survivor sample 20	AB+contact sample 8	AB+contact sample 16	Ab-contact sample 4	Ab-contact sample 12	Ab-contact sample 20	Control sample 8	Control sample 16

Plate template for genotyping

Genotyping plate template

	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	Contact sample 8	Contact sample 16	Contact sample 24	Contact sample 32	Contact sample 40	Survivor sample 4	Survivor sample 12	Survivor sample 20	BB-Deceased sample 4	Control sample 2	BLANK
B	Contact sample 1	Contact sample 9	Contact sample 17	Contact sample 25	Contact sample 33	Contact sample 41	Survivor sample 5	Survivor sample 13	BB-Survived sample 1	BB-Deceased sample 5	Control sample 3	Control sample 10
C	Contact sample 2	Contact sample 10	Contact sample 18	Contact sample 26	Contact sample 34	Contact sample 42	Survivor sample 6	Survivor sample 14	BB-Survived sample 2	BB-Deceased sample 6	Control sample 4	Control sample 11
D	Contact sample 3	Contact sample 11	Contact sample 19	Contact sample 27	Contact sample 35	Contact sample 43	Survivor sample 7	Survivor sample 15	BB-Survived sample 3	BB-Deceased sample 7	Control sample 5	Control sample 12
E	Contact sample 4	Contact sample 12	Contact sample 20	Contact sample 28	Contact sample 36	Contact sample 44	Survivor sample 8	Survivor sample 16	BB-Survived sample 4	BB-Deceased sample 8	Control sample 6	Control sample 13
F	Contact sample 5	Contact sample 13	Contact sample 21	Contact sample 29	Contact sample 37	Survivor sample 1	Survivor sample 9	Survivor sample 17	BB-Deceased sample 1	BB-Deceased sample 9	Control sample 7	Control sample 14
G	Contact sample 6	Contact sample 14	Contact sample 22	Contact sample 30	Contact sample 38	Survivor sample 2	Survivor sample 10	Survivor sample 18	BB-Deceased sample 2	BB-Deceased sample 10	Control sample 8	Control sample 15
H	Contact sample 7	Contact sample 15	Contact sample 23	Contact sample 31	Contact sample 39	Survivor sample 3	Survivor sample 11	Survivor sample 19	BB-Deceased sample 3	Control sample 1	Control sample 9	Control sample 16

D. List of manufacturers

Manufacturer	Instrument/Kit
BioTek	ELx50 TS Washer
DNA Genotek	Oragene OG-500/OG-575 kit
	PrepIt2 Manual extraction kit
Eppendorf	Concentrator Plus
Illumina	HiSeq
	H3 Africa Array
Invitrogen	Qubit 3
Kalon Diagnostics Ltd	EBOV IgG Capture EIA
Pelican	80QT Cooler with wheels
Peli Products	IM3075 Storm Case
Qiagen	QIAmp DNA Mini kit
	QIAxpert
	QIAsymphony SP
	QIAsymphony DNA Midi kit
Roche	NimbleGen SeqCap EZ Library
Sarstedt	10ml 79x16mm flat base, sterile tube
Thermo Scientific	Multiskan FC Microplate Photometer
	Nanodrop 2000

Table D-1 List of Manufacturers and associated instruments, kits or products.

E. Participant Information Leaflet and Consent Form – Household Contacts &

Community Controls

Participant information leaflet

Host susceptibility to Ebola virus infection

The study of host susceptibility to Ebola virus infection looks at whether there is a human reason that some people develop infection to Ebola virus and others do not. It also looks at if there is a reason in a person's body that some people survive Ebola infection and others do not. It will particularly look at whether there is a pathway in the immune system, the part of the human body that prevents and fights infection with viruses, that is different in those people who get infected with Ebola virus and survive compared to those people who do not survive. We find these immune pathways by looking at the genes that make up our bodies. Genes are what tell our bodies which proteins and cells to make. We compare different groups such as those people who were ill with Ebola virus and those who weren't and those people who survived Ebola virus and those who didn't. If we see a difference in the genes between these groups we look at what role the gene has in making proteins and cells for our immune system and if this plays a role in fighting the Ebola virus.

This study is a descriptive study as it does not involve the use of any trial medicines or tests. It will describe the results found from the laboratory tests performed on the samples given.

The study will help us to understand Ebola virus more and how it interacts with the human body and our immune systems. When we understand this better it helps us to make new and better medicines to treat the infection or vaccines to prevent infection. This hopefully will mean we see less Ebola virus infection in the future or if someone does become unwell with Ebola virus infection, we will have an effective treatment we can use to cure the disease. It will benefit your community in the respect that if Ebola virus made people in your community ill again we would understand the disease better, and hopefully have a vaccine that could prevent the disease or medicines that will be able to treat the disease, improving the chances of survival.

The samples we need from you

For the study we need both people who were infected with Ebola virus and people who never showed any signs of infection with the virus. The Ministry of Health and Sanitation of Sierra Leone have already granted us access to stored samples from people who were infected with Ebola, but we need to obtain samples from those who were not known to be infected. The samples we need involve a saliva sample and a mouth swab.

How we will collect the samples

Before we take any samples, we need you to not eat or drink anything, including water, for 1 hour before the samples are collected. If your mouth is very dry and you do not have enough saliva you can take a small sip of water to help, but not a large drink.

To take the saliva sample we need you to spit into a tube that we will provide. The tube is empty, and you will need to spit into the tube up until the line marked on the tube, we will then seal the tube. In small children who are unable to spit we will use mouth sponges to collect the saliva and then we will put these into the tube instead.

For the mouth swab this is a sterile, dry, soft cotton swab on a stick that will be pressed up against the inside of your cheek or gum for approximately 90 seconds. This swab will then be stored in a separate sealed tube.

All samples and information you give us will be completely anonymous and it will not be possible to trace the information back to you. Samples will not be stored with your name, they will be allocated a study number and all data obtained will be stored under this number. Only the fieldwork team will be aware of your name and which samples your name is linked to. All information will be stored on password protected computers in password protected files.

Risks to participants

There are no foreseeable risks to participants of the study from sample taking.

This study is a genetic study looking at human genes. We are only looking at differences in genes between groups of people, not individuals. This means we will not be able to look at samples and know if you are at risk of any other disease. We will not have any genetic information on your sample specifically and you do not need to worry that we would know something about you independent of the rest of the group. All samples and information you give us will be completely anonymous and it will not be possible to trace the information back to you. Any information we obtain from this study will be shared with you when the study is completed, if you would like to know about it.

One of the samples we will take will look at if you have immune cells that already recognise the Ebola virus; these cells are called antibodies. We do not yet understand what it means if you have antibodies to Ebola virus. We won't be able to give you these results directly as we will only be interpreting them by the study group you are in. This test is only being done to help us understand the genetic test results better.

Participation in the study

Participation in the study is voluntary. If at any time you change your mind about being part of the study, you can request to be removed from the study by contacting the study fieldwork team at the number below. If you have further questions regarding the study, please contact the fieldwork team on the number below and they would be happy to explain further.

There is no penalty for choosing not to be involved in the study.

If we are unable to obtain the information needed for the study from the sample provided you will be excluded from the study analysis. This will have no direct impact on you and will not affect you retaining your thank you gift.

The only people who will have contact with you for the purposes of the study will be Dr Nathalie MacDermott, Mr David Kargbo and the fieldwork team listed below, who you will meet when samples are collected. No one else will contact you regarding this study.

Incentives/thank you gifts

Those participants who choose to take part in the study will be provided with a bag of water following obtaining samples. Each family involved (even if only one member of the family participates) will receive a hand wash bucket with soap with accompanying explanation of the importance of hand washing in prevention of diarrhoeal illnesses. Should a participant choose to withdraw from the study at a later date they will still retain their thank you gift.

All gifts will be distributed to participants on the same day following sample collection, where possible.

Contact details

If you have any questions related to the study, you can contact the fieldwork team on the mobile phone numbers below:

Mr Daniel Sesay
077-373-828

Mr Santigie Dobson Kanu
076-936-352

Mr Osman Kamara
076-799-162
077-260-070

Mr Collier
078-591-817
030-809-318

Mr Alim Billo Bah
077-385-011

Ms Cecilia Touray
077-595-270

Dr Nathalie MacDermott
099-368-296

If you have concerns regarding the study or feel you have suffered an adverse consequence of taking part, you may contact the Ethics and Scientific Review Committee at any time at the address below:

Chair: Prof H Morgan
Sierra Leone Ethics and Scientific Review Committee
PCMH
Fourah Bay Road, Freetown
076-629-251

Consent form

Study of host susceptibility to Ebola virus infection

I consent to providing a saliva sample and oral swab for use in the study of host susceptibility to Ebola virus infection. I acknowledge that I have been provided with information regarding this study and I understand the information provided to me. Any questions have been answered in a manner I understand, and I am satisfied with the answer.

I am aware that my participation in the study is voluntary and I can withdraw from the study at any time by contacting the fieldwork team. I understand there will be no penalty for withdrawing from the study.

I acknowledge that I have not been coerced into taking part in this study and I do so of my own free will.

I acknowledge that if I sustain an injury or adverse consequence as a result of this study I may notify the Sierra Leone ethics and scientific review committee to report this.

Participant name:

Signature or thumb print of participant   Date
(In children <18 years of age signature and name of guardian)

Signature and name of fieldworker obtaining consent   Date

I confirm that the information herein is written in good faith and all procedures described will be complied with, to the best of our ability.

Professor Michael Levin
Imperial College London

F. Participant Information Leaflet and Consent Form – Survivors

Participant information leaflet

Host susceptibility to Ebola virus infection

The study of host susceptibility to Ebola virus infection looks at whether there is a human reason that some people develop infection to Ebola virus and others do not. It also looks at if there is a reason in a person's body that some people survive Ebola infection and others do not, and if they do survive whether they will have ongoing problems afterwards. It will particularly look at whether there is a pathway in the immune system, the part of the human body that prevents and fights infection with viruses, that is different in those people who get infected with Ebola virus and survive compared to those people who do not survive. We find these immune pathways by looking at the genes that make up our bodies. Genes are what tell our bodies which proteins and cells to make. We compare different groups such as those people who were ill with Ebola virus and those who weren't and those people who survived Ebola virus and those who didn't. If we see a difference in the genes between these groups, we look at what role the gene has in making proteins and cells for our immune system and if this plays a role in fighting the Ebola virus.

This study is a descriptive study as it does not involve the use of any trial medicines or tests. It will describe the results found from the laboratory tests performed on the samples given.

The study will help us to understand Ebola virus more and how it interacts with the human body and our immune systems. When we understand this better it helps us to make new and better medicines to treat the infection or vaccines to prevent infection. This hopefully will mean we see less Ebola virus infection in the future or if someone does become unwell with Ebola virus infection, we will have an effective treatment we can use to cure the disease. It will benefit your community in the respect that if Ebola virus made people in your community ill again we would understand the disease better, and hopefully have a vaccine that could prevent the disease or medicines that will be able to treat the disease, improving the chances of survival.

The samples we need from you

For the study we need both people who were infected with Ebola virus and people who never showed any signs of infection with the virus. The Ministry of Health and Sanitation of Sierra Leone have already granted us access to stored samples from people who were infected with Ebola, but we need to obtain samples from those who were not known to be infected and further samples from those who survived the disease. The samples we need involve a saliva sample and a mouth swab.

How we will collect the samples

Before we take any samples, we need you to not eat or drink anything, including water, for 1 hour before the samples are collected. If your mouth is very dry and you do not have enough saliva you can take a small sip of water to help, but not a large drink.

To take the saliva sample we need you to spit into a tube that we will provide. The tube is empty, and you will need to spit into the tube up until the line marked on the tube, we will then seal the tube. In small children who are unable to spit we will use mouth sponges to collect the saliva and then we will put these into the tube instead.

For the mouth swab this is a sterile, dry, soft sponge swab on a stick that will be rubbed up against the inside of your cheek or gum for approximately 90 seconds. This swab will then be stored in a separate sealed tube.

All samples and information you give us will be completely anonymous and it will not be possible to trace the information back to you. Samples will not be stored with your name, they will be allocated a study number and all data obtained will be stored under this number. Only the fieldwork team will be aware of your name and which samples your name is linked to. All information will be stored on password protected computers in password protected files.

Risks to participants

There are no foreseeable risks to participants of the study from sample taking.

This study is a genetic study looking at human genes. We are only looking at differences in genes between groups of people, not individuals. This means we will not be able to look at samples and know if you are at risk of any other disease. We will not have any genetic information on your sample specifically and you do not need to worry that we would know something about you independent of the rest of the group. All samples and information you give us will be completely anonymous and it will not be possible to trace the information back to you. Any information we obtain from this study will be shared with you when the study is completed, if you would like to know about it.

One of the samples we will take will look at if you have immune cells that already recognise the Ebola virus; these cells are called antibodies. We do not yet understand what it means if you have antibodies to Ebola virus. We won't be able to give you these results directly as we will only be interpreting them by the study group you are in. This test is only being done to help us understand the genetic test results better.

Participation in the study

Participation in the study is voluntary. If at any time you change your mind about being part of the study, you can request to be removed from the study by contacting the study fieldwork team at the number below. If you have further questions regarding the study, please contact the fieldwork team on the number below and they would be happy to explain further.

There is no penalty for choosing not to be involved in the study.

If we are unable to obtain the information needed for the study from the sample provided you will be excluded from the study analysis. This will have no direct impact on you and will not affect you retaining your thank you gift.

The only people who will have contact with you for the purposes of the study will be Dr Nathalie MacDermott, Mr David Kargbo and the fieldwork team listed below, who you will meet when samples are collected. No one else will contact you regarding this study.

Incentives/thank you gifts

Those participants who choose to take part in the study will be provided with a bag of water and a soft drink following obtaining samples. Each person involved will receive a gift of a household or cooking item. Should a participant choose to withdraw from the study at a later date they will still retain their thank you gift.

All gifts will be distributed to participants on the same day following sample collection, where possible.

Contact details

If you have any questions related to the study, you can contact the fieldwork team on the mobile phone numbers below:

Mr Daniel Sesay
077-373-828

Mr Osman Kamara
076-799-162
077-260-070

Mr Alim Billo Bah
077-385-011

Dr Nathalie MacDermott
099-368-296

If you have concerns regarding the study or feel you have suffered an adverse consequence of taking part, you may contact the Ethics and Scientific Review Committee at any time at the address below:

Chair: Prof H Morgan
Sierra Leone Ethics and Scientific Review Committee
PCMH
Fourah Bay Road, Freetown
076-629-251

Consent form

Study of host susceptibility to Ebola virus infection

I consent to providing a saliva sample and oral swab for use in the study of host susceptibility to Ebola virus infection. I acknowledge that I have been provided with information regarding this study and I understand the information provided to me. Any questions have been answered in a manner I understand, and I am satisfied with the answer.

I am aware that my participation in the study is voluntary and I can withdraw from the study at any time by contacting the fieldwork team. I understand there will be no penalty for withdrawing from the study.

I acknowledge that I have not been coerced into taking part in this study and I do so of my own free will.

I acknowledge that if I sustain an injury or adverse consequence as a result of this study, I may notify the Sierra Leone ethics and scientific review committee to report this.

Participant name:

Signature or thumb print of participant   Date
(In children <18 years of age signature and name of guardian)

Signature and name of fieldworker obtaining consent   Date

I confirm that the information herein is written in good faith and all procedures described will be complied with, to the best of our ability.

Professor Michael Levin
Imperial College London

G. Participant Questionnaire – Household Contacts

Household contact member inventory (page 1)

Inclusion criteria: Households where a member of the household had Ebola (survived or died)

I would like to know about the people who were members of your household during the time Ebola took place. Please tell me the name of each person who was in your household at that time and who is accessible now. Please also tell me the name of anyone who was in your household at that time who died of Ebola and anyone who survived Ebola even if they are not here now. By "member" I mean a person who normally lives here and shares meals.

Community number		Head of Household full name								Dwelling #	

HH Line#	Household member name	Relation-ship to line number**	Line # related to	Sex	Age (yrs)	If <2 yrs Age (mo)	Ever had Ebola?	Died from Ebola?	Saliva sample collected	Oral fluid collected	If saliva/fluid collected place sample ID sticker below
01				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
02				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
03				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
04				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
05				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
06				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
07				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
08				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	
09				M F			No Yes DK	No Yes DK	No Yes Revisit	No Yes Revisit	

231

Socio-Economic Status: Household contacts

[Inclusion criterion: complete for households where a member of the household had Ebola]

I will start by asking some questions about your home and household at the time Ebola took place.

At the time of Ebola in this household…….

			Dwelling #
Community number			
Head of household name			
1. Was the household located in an urban or rural area at that time?	Urban		
	Rural		
2. How many rooms did your household have to live in at that time? *[Count all kinds of room but not shared rooms]*	_____ rooms		
3. How many people were living in the household at that time?	_____ adults _____ children (<18)		
4. How many rooms did your household share with other households at that time?	_____ rooms		
5. What level of access to water did your household have at that time?	Every day		
	Most days		
	Sometimes		
	Very limited access		
6. What level of access to soap did your household have at that time?	Every day		
	Most days		
	Sometimes		
	Very limited access		
7. What kind of access to a latrine did your household have at that time?	One for the household		
	Shared		
	None		
8. Was your household quarantined during the time Ebola was in your community?	Yes		
	No		

Individual questionnaire: Part A

			Dwelling #
Inclusion criteria: Each member of the household currently present excluding Ebola survivors			
[Instructions: *For children they should be answered by a proxy.*]			
1.	Interview date		
2.	Interviewer initials		
3.	Community number		
4.	Head of Household full name		
5.	Participant ID sticker		
6.	Participant Name		
7.	Who is the respondent? *[Give HH line list #]*		
8.	Ethnicity *[tribe, if mixed include both/all]*		
9.	Occupation? *[record what is said]*		
10.	Was the person an Ebola front line worker	Yes No	
11.	Did you receive a vaccine against Ebola?	Yes [→*stop interview*]	
		No	
12.	Were you ever tested for Ebola?	Yes	
		No [→*skip to 13*]	
		Don't know [→*skip to 13*]	
13.	What was the final result of your test for Ebola?	Negative	
		Positive [→*stop interview for this person*]]	
		Didn't receive result	
		Don't remember	
14.	Please think about the time after the first person with Ebola in the household got sick, through to the time you were cleared by the contact tracers. **Were you living in the household in this period?**	No [→*include as community control*]	
		Yes	
15.	[*if not a known Ebola case*] During this time period, did you experience any health problems	No	
		Yes	
16.	*If yes*: Describe those problems *[Write brief symptoms or tick box if any of the symptoms below are mentioned]*		
17.	*[When person has finished describing symptoms in their own words, ask them if they had any symptoms on this list that they have not mentioned]*	Fever	
		Fatigue (general body weakness)	
		Headache	
		Loss of appetite	
		Nausea or Vomiting	
		Abdominal pain	
		Diarrhoea	
		Blood in the stool	
		Muscle and/or joint pain/ache	
		Sore throat or pain with swallowing	
		Hiccups	

		Red eyes	
		Blurry vision	234
		Bleeding gums	
		Miscarriage	
18.	During the time of Ebola was there a baby who was born dead, sick or with a disability in your household?	Yes [describe]	
		No	

Individual questionnaire: Part B: To be completed after discussion with household.

Record the type of contact the person had with Ebola cases in the household or elsewhere including deceased possible Ebola cases while they were ill, or with the body after death. Ask probing questions if necessary to fill in gaps or check responses.　　　　1 = contact　　　0 = no contact

Participant ID sticker			
Only the highest level of contact needs to be indicated	In household	Outside household	
Level 1: Contact with the body / prepared the body			
Level 2: <u>Direct</u> contact with <u>body fluids</u> eg. blood, diarrhoea, vomit, urine. Or a baby who was breastfed by an EV+ woman			
Level 3: <u>Direct</u> close contact with <u>wet</u> case (a person with diarrhoea/vomiting/ bleeding) eg helped dress, embraced, carried, helped care, shared bed, or mother breastfeeding an EV+ child			
Level 4: <u>Direct</u> close contact with <u>dry</u> case (eg helped dress, embraced, carried, helped care, shared bed)			
Level 5: <u>Indirect</u> close contact with <u>wet</u> case (eg washed clothes, bed linen)			
Level 6: <u>Indirect</u> close contact with <u>dry</u> case (eg washed clothes, bed linen)			
Level 7: Minimal contact (eg shared meals, shared utensils)			
Level 8: No actual contact known (eg kept distance once person was symptomatic)			

Record any details that may help define the level. Record the reason for the highest level given – i.e. what made you decide that the person had contact at that level. Keep all notes with forms for the household

H. Participant Questionnaire – Community Controls

Community Control: Individual questionnaire

Inclusion criteria: Each member of the household currently present excluding Ebola survivors

[Instructions: *For children they should be answered by a proxy.*]

#	Question			
1.	Interview date			
2.	Interviewer initials			
3.	Community number			
4.	Participant ID sticker			
5.	Participant Name		Sex	M F
6.	Participant age *[if under 2 years, in months]*			
7.	Who is the respondent? *[participant/parent]*			
8.	Ethnicity *[tribe, if mixed include both/all]*			
9.	Participant related to *[only 1ˢᵗ degree relatives, write sticker # and relationship ie. child of EG-SL-CC-0052]*			
10.	Occupation? *[record what is said]*			
11.	Did you receive a vaccine against Ebola?	Yes [→*stop interview*]		
		No		
12.	Was the person an Ebola front line worker	Yes		
		No		
13.	Were you ever tested for Ebola?	Yes		
		No [→*skip to 14*]		
		Don't know [→*skip to 14*]		
14.	What was the final result of your test for Ebola?	Negative		
		Positive [→*stop interview*]		
		Didn't receive result		
		Don't remember		
15.	Did anyone in your household become sick with Ebola when you were living in the same household?	Yes [→*include with entire household as household contact]*		
		No		
16.	During the time of Ebola, did you touch the dead body of someone sick with Ebola?	Yes [→*assess risk level from list at end of questionnaire]*		
		No		
17.	During the time of Ebola, did you help to care for anyone who was sick with Ebola?	Yes [→*assess risk level from list at end of questionnaire]*		
		No		
18.	During the time of Ebola did you attend the funeral of someone sick with fever or Ebola symptoms? *[Determine if 1. prepared the body, 2. touched the body, 3. carried the body, 4. Carried the body in body-bag or 5. just observed]*	Yes [→*determine if number 1-5 and insert in right column]*		
		No		
19.	During the time of Ebola was there a baby who was born dead, sick or with a disability in your household?	Yes *[describe]*		
		No		
20.	During the time of Ebola in your community, did	Yes		

	you experience any health problems?	No		
21.	*If yes*: Describe those problems *[Write brief symptoms or tick box if any of the symptoms below are mentioned]*			
22.	*[When person has finished describing symptoms in their own words, ask them if they had any symptoms on this list that they have not mentioned]*	Fever		
		Fatigue (general body weakness)		
		Headache		
		Loss of appetite		
		Nausea or Vomiting		
		Abdominal pain		
		Diarrhoea		
		Blood in the stool		
		Muscle and/or joint pain/ache		
		Sore throat or pain with swallowing		
		Hiccups		
		Red eyes		
		Blurry vision		
		Bleeding gums		
		Miscarriage		

Community control: Risk levels

Only the highest level of contact needs to be indicated	
Level 1: Contact with the body / prepared the body	
Level 2: Direct contact with body fluids eg. blood, diarrhoea, vomit, urine. Or a baby who was breastfed by an EV+ woman	
Level 3: Direct close contact with wet case (a person with diarrhoea/vomiting/ bleeding) eg helped dress, embraced, carried, helped care, shared bed, or mother breastfeeding an EV+ child	
Level 4: Direct close contact with dry case (eg helped dress, embraced, carried, helped care, shared bed)	
Level 5: Indirect close contact with wet case (eg washed clothes, bed linen)	
Level 6: Indirect close contact with dry case (eg washed clothes, bed linen)	
Level 7: Minimal contact (eg shared meals, shared utensils)	
Level 8: No actual contact known (eg kept distance once person was symptomatic)	

Record any details that may help define the level. Record the reason for the highest level given – i.e. what made you decide that the person had contact at that level.

Socio-Economic Status: Community controls

[Inclusion criterion: complete for households where a member of the household had Ebola]

I will start by asking some questions about your home and household at the time Ebola took place. **At the time of Ebola in this household.......**

		Dwelling #	
Community number			
Head of household name			
1. Was the household located in an urban or rural area at that time?	Urban		
	Rural		
2. How many rooms did your household have to live in at that time? *[Count all kinds of room but not shared rooms]*	_____ rooms		
3. How many people were living in the household at that time?	_____ adults _____ children (<18)		
4. How many rooms did your household share with other households at that time?	_____ rooms		
5. What level of access to water did your household have at that time?	Every day		
	Most days		
	Sometimes		
	Very limited access		
6. What level of access to soap did your household have at that time?	Every day		
	Most days		
	Sometimes		
	Very limited access		
7. What kind of access to a latrine did your household have at that time?	One for the household		
	Shared		
	None		
8. Was your household quarantined during the time Ebola was in your community?	Yes		
	No		

237

I. Participant Questionnaire – Survivors

Ebola survivor sequelae: Individual questionnaire

				M F
Inclusion criteria: Ebola survivors with or without ongoing sequelae				
[Instructions: *For children they should be answered by a proxy.*]				
1.	Interview date			
2.	Interviewer initials			
3.	Participant ID sticker			
4.	Participant Name		Sex	M F
5.	Participant age [if under 2 years, in months]			
6.	Who is the respondent? [participant/parent]			
7.	Community lived in during time of Ebola			
8.	Ethnicity [tribe, if mixed include both/all]			
9.	Participant related to [only 1st degree relatives, write sticker # and relationship ie. child of EG-SL-SS-0052]			
10.	Occupation? [record what is said]			
11.	Did you have any medical problems before Ebola? [if yes describe]	Yes [describe]		
		No		
12.	Were you an Ebola front line worker?	Yes [describe]		
		No		
13.	Did you receive a vaccine against Ebola?	Yes [state if vaccinated before they had Ebola or after]		
		No		
14.	Do you know how you became infected with Ebola?	Yes		
		No		
15.	If yes, what was the mode of contact? [funeral, caring for sick relative, healthcare worker etc]			
16.	When were you sick with Ebola? [month and year]			
17.	Where were you treated?	Ebola treatment centre		
		Holding centre		
		Kerry Town Healthcare worker ETC		
18.	How long were you admitted at the treatment centre? [in days]			
19.	Did you receive any experimental treatments? [if yes, insert treatment name]	Yes		
		No		
20.	Please describe the symptoms you had when you were sick with Ebola			
21.	[When person has finished describing symptoms in their own words, ask them if they had any symptoms on this list that they have not mentioned]	Fever		
		Fatigue (general body weakness)		
		Headache		
		Loss of appetite		
		Nausea or Vomiting		
		Abdominal pain		

		Diarrhoea		
		Blood in the stool or vomit		
		Muscle and/or joint pain/ache		
		Sore throat or pain with swallowing		
		Hiccups		
		Red eyes		
		Blurry vision		
		Bleeding gums/nose		
		Bleeding from iv line site		
		Vaginal bleeding		
		Miscarriage		
22.	Since you survived have you had any continuing problems?	Yes		
		No		
23.	If yes please describe the problems			
24.	[When person has finished describing symptoms in their own words, ask them if they had any symptoms on this list that they have not mentioned, if yes tick the box to the right and insert the relevant number to clarify their symptoms]	Eye disease [1. Uveitis, 2. Cataracts, 3. Blindness]		
		Neurological problems [1. Seizures, 2. Recurrent headaches (>2/week), 3. Persistent sleep disturbance (>4 nights/week), 4. Specific paralysis, 5. Other (describe)]		
		Deafness		
		Joint problems [1. Occasional joint pain, 2. Regular joint pain limiting daily activities (>3 days/week), 3. Recurrent joint swelling (>2 episodes)]		
		Significant weight change [1. Weight loss, 2. Weight gain]		
		Recurrent infections >1 every 2 months [1. Malaria, 2. Infections requiring antibiotics, 3. Both]		
		Fertility problems [1. Problems having sexual relations (men), 2. Problems conceiving a baby (men and women)], 3. Recurrent miscarriage (women)]		
25.	Did a medical doctor diagnose you with a specific condition? [if yes list the condition, check if participant has medical record available, if yes document diagnosis recorded in medical record]	Yes		
		No		
26.	Did you receive treatment for this/these conditions? [if yes specify treatment and length of treatment, is treatment still continuing?]	Yes		
		No		
27.	Has the problem resolved, is it continuing or is it improving or worsening?	Resolved		
		Improving		
		Continuing the same		
		Worsening		

28.	Have you had your semen tested for Ebola virus? *[male participants]*	Yes		
		No	240	
29.	If yes, did the fluid contain Ebola virus?	Yes		
		No		
30.	If yes, how long did the virus stay in the fluid?	Time in months following diagnosis of Ebola =		
		Still present		
31.	Have you had fluid from your eye tested for Ebola virus?	Yes		
		No		
32.	If yes, did the fluid contain Ebola virus?	Yes		
		No		
33.	If yes, how long did the virus stay in the fluid?	Time in months following diagnosis of Ebola =		
		Still present		
34.	Have you had any other bodily fluid tested for Ebola virus?	Yes *[state type of fluid]*		
		No		
35.	How long did the virus stay in the fluid? *[specify days, weeks or months]*			

J. Standard Operating Procedures

Obtaining samples from Ebola contacts and controls in rural Sierra Leone

1. Arrival in the community
2. Personal protective equipment
3. Pre-screening questions & temperature check
4. Obtaining saliva and oral fluid swab samples
5. Obtaining blood samples
6. Transporting samples
7. Removing personal protective equipment
8. Disposal of waste
9. How to manage a spill
10. In the event of a needle stick injury/mucous membrane splash

Equipment required:
- Satellite phone & navigation system (Iridium 9575 Extreme)
- Study questionnaire – 100 copies
- Pens – 4 items
- Fine tipped indelible markers – 4 items
- Adhesive labels – 300 items
- Oragene DNA OC-500 – 75 items
- Oragene DNA OC-575 – 25 items
- Oral swabs – 100 items
- Oral swab storage receptacle – 100 items
- BD vacutainer push button blood collection set – 100 items
- 4ml yellow topped serum blood collection tubes – 100 items
- Sharps bin
- Chlorhexidine skin wipes – 150 items
- Gauze – 1 pack
- Micropore – 4 rolls
- Leak proof, sealable plastic biohazard bags & biohazard sticker – 300 items
- Medipal high level disinfectant wipes 225/canister – 2 canisters
- iM3075 Storm case – 1 item
- 80QT wheeled cooler (containing ice) – 1 item
- Heavy duty clinical waste bags – 1 roll (200/roll)
- Spray can containing Virkon tablets (to be filled with water if necessary)
- Virkon granules in Squeeze-puff container
- Safetygel shakerpot x 2
- Shovel
- Alcohol gel – 4 bottles
- Bottled water – 21 litres (3 litres/person for drinking)
- Face visor & mask – 1 pack (25/pack)
- FFP3 facemasks plus valve – 1 box

- Nitrile gloves M/L – 1 box each
- Thumbs up gowns – 1 pack
- Rubber boots – 1 pair/team member
- Multicoloured balloons
- Community incentives

1 Arriving in the community

1.1 Prior to arrival in the community/clinic to obtain samples extensive community sensitisation should have occurred and formal permission should be in place from the community leaders.

1.2 On arrival greet the community leaders, explain again the reason for the study, the samples needed and the impact on the community. Gain further permission to obtain samples

1.3 Identify location in which sample collection can occur

    1.3.1   The palava hut is usually ideal as it provides shade for people waiting and a moderate amount of privacy but is a neutral area that should be relatively free from contamination and is highly visible within the community for your safety.

1.4 Identify routes out of the community should you need to leave rapidly

1.5 Ensure vehicle is parked close to where samples are being collected but, in a location, and orientation (facing the route/road out of the community) that enables a fast departure should this be required.

1.6 Ensure the vehicle is not obstructed by other vehicles/obstacles

1.7 Notify field team leader/coordinator of exact location using GPS coordinates

1.8 Once you have departed from the vehicle and set up the equipment:

    1.8.1   Explain clearly to the community regarding the study and the samples required

    1.8.2   Address those families (previously identified) that you wish to recruit:

        1.8.2.1   Explain clearly to them regarding any risks posed to them

        1.8.2.2   Explain clearly regarding incentives to be received by families for participating in the study

        1.8.2.3   Ensure consent obtained from each participant (consent from parents should be taken for children under age 18)

        1.8.2.4   Proceed with participant recruitment as below.

2 Personal protective equipment

2.1 When collecting samples from participants where their Ebola virus status is unknown, and the procedure could generate aerosols or exposure to body secretions such as saliva/blood, the following personal protective equipment should be worn by the 'dirty' sample collector (unless the sample collector is a survivor when just gloves and rubber boots can be worn):

    2.1.1   Face visor & mask

    2.1.2   1 pair nitrile gloves

    2.1.3   Thumbs up gown
                    Rubber boots

2.2 The clean assistant should wear:

2.2.1   1 pair nitrile gloves
             Rubber boots

3   Pre-screening questions and temperature check

3.1 Prior to obtaining samples from participants whose Ebola virus status is unknown the following screening questions should be asked:

3.1.1   Are you sick with fever, vomiting, diarrhoea, body aches, tiredness or unexplained bleeding?
3.1.2   Have you been in contact with anyone known to have Ebola virus infection in the last 21 days (alive or deceased)?
3.1.3   Have you been in contact with anyone who is sick with fever, vomiting, diarrhoea, body aches, tiredness or unexplained bleeding in the last 21 days?
3.1.4   Have you attended a funeral of someone who died of unknown cause in the last 21 days?

3.2 If the answer to any of these is 'yes' samples should not be taken from the participant at this time. The participant is either excluded from the study or samples should be collected after the participant has cleared 21 days without any of the symptoms described above.

3.3 Next check the participants temperature using the hand held digital thermoscan thermometer ensuring no direct physical contact is made with the patient. The temperature should be taken in the temple region with the participant facing away from the investigator. Always ensure the thermometer and participant are shaded from the sun when measuring the temperature.

3.3.1   If the participants temperature is >37.4 ℃ or >99.4℉ samples should not be collected and the participant should be excluded from the study.

4   Obtaining saliva and oral fluid swab samples

4.1 Prior to obtaining samples:
4.1.1   Ensure study questionnaire is completed for the participant
4.1.2   Ensure appropriate personal protective equipment is worn as per section 2 above.

4.1.2.1   Samples should only be collected with 2 investigators present – one will be the 'dirty' sample collector the other will remain clean and assist the sample collector from a safe distance. Only the 'dirty' sample collector must wear the PPE described in section 2, the clean assistant should only wear rubber boots and gloves but should maintain a distance of 2 metres from the participant during sample collection when aerosols may be generated.

4.2 Provided the participant has answered 'no' to all screening questions and does not have a raised temperature on screening, proceed with obtaining samples.
4.3 Between patients always ensure that gloves have been cleaned with alcohol gel or discard using the beaking method described in section 7 below, apply alcohol gel to hands and then apply fresh gloves.
4.4 For the collection of the Oragene sample in older children (>5 years of age) and adults use the Oragene-DNA OG-500 kit:

4.4.1   Ensure participant has not eaten, drunk or brushed their teeth for approximately 1 hour before sample collection.

4.4.2   Explain procedure clearly to participant

4.4.3   Label the sample tube with the appropriate participant number.

4.4.4   Hand tube to participant and request they spit into the tube until the saliva reaches the appropriate marked level (approximately 2ml).

   4.4.4.1   If required to reach the appropriate level the participant can take a small sip of water.

4.4.5   Once an adequate sample volume is achieved seal the tube with the cap, this breaks the seal allowing stabilisation fluid to enter the tube. Remove the mouthpiece, discard in a clinical waste bag and screw the cap on the tube.

4.4.6   Place the tube in a leak proof bag with absorbent material (Safetygel granules) and seal. Then follow the instructions in section 6 below for transport of samples.

4.5 For collection of Oragene samples for children who are <6 years old or older children unwilling or unable to cooperate use the Oragene DNA OG-575 kit:

4.5.1   Ensure the child has not eaten, drunk or brushed their teeth for approximately 1 hour before sample collection.

4.5.2   Explain the procedure to the parent and child.

4.5.3   Label the sample tube with the appropriate participant number.

4.5.4   With the parent sitting and holding the child on their lap use the swabs contained in the kit to absorb saliva from the child's mouth by inserting the swab and resting against the child's cheek or under the child's tongue. When the swab is wet, remove, wring it out against the plastic lip in the receptacle and then use again to collect more saliva. Continue until the there is enough saliva to reach the marked level (approximately 0.75ml).

4.5.5   Once an adequate sample volume is achieved seal the tube with the cap, this breaks the seal allowing stabilisation fluid to enter the tube. Remove the mouthpiece, discard in a clinical waste bag and screw the cap on the tube.

4.5.6   Place the tube in a leak proof bag with absorbent material (Safetygel granules) and seal. Then follow the instructions in section 6 below for transport of samples.

4.6 For collection of oral fluid swabs in both adults and children:

4.6.1   Explain the procedure to the participant or parent and child.

4.6.2   Label the sample receptacle with the appropriate participant number.

4.6.3   Following collection of the Oragene sample, use the swab provided

   4.6.3.1 In adults hand the swab to the participant and ask them to run it along their gums for 90 seconds.

   4.6.3.2 In children hand the swab to the parent and ask them to run the swab gently along the gums of the participant for 90 seconds.

4.6.4   Ask the participant/parent to place the swab into the receptacle provided. Seal the receptacle and place in a leak proof bag with absorbent material (safetygel granules) and seal. Then follow the instructions in 6 below for transport of samples.

5   Obtaining blood samples (only if oral fluid samples not being collected)

5.1 Provided the participant has answered 'no' to all screening questions and does not have a raised temperature on screening, proceed with obtaining samples.

5.2 To obtain blood samples follow the instructions below – note **NO** sharp should be exposed/unsheathed without the necessary sharps container available in which to discard it!

    5.2.1   In addition to PPE worn for collection of saliva samples – wash gloved hands with alcohol gel, allow to dry and then apply a second pair of gloves.

    5.2.2   Label the blood tube with the appropriate participant number/label.

    5.2.3   Identify the vein and clean the overlying skin with a chlorhexidine skin swab. Discard in a clinical waste bag.

    5.2.4   Using the BD vacutainer push button blood collection set with pre-attached holder, insert needle into identified vein. Then attach 4ml yellow topped serum blood collection tube to the vacutainer system.

    5.2.5   Once the blood tube is full, remove the tube from the collection system, place in a leak proof bag with absorbent material (Safetygel granules) and seal.

    5.2.6   Press the push button on the blood collection system and the needle will retract from the vein into the plastic receptacle, ensuring a completely sharp safe system.

    5.2.7   Discard the used blood collection system into the sharps container.

    5.2.8   Discard any gauze used to stop bleeding into clinical waste bag.

    5.2.9   Follow the instructions in section 6 below for containment of samples for transport.

6   Transporting samples

6.1 All samples at the point of sampling should have been placed in a leak proof bag with absorbent material (safetygel granules) and sealed.

6.2 Following this:

    6.2.1   The clean assistant ensures the specimen bag is tightly sealed and wipes the outside of the bag with a Medipal high level disinfectant wipe. They then discard the wipe in a clinical waste bag.

    6.2.2   If the bag contains the Oragene sample, place it carefully in the iM3075 storm case (black).

    6.2.3   If the bag contains the oral fluid swab, place it carefully in the 80QT wheeled cooler (white) on top of the ice. Ensure to close the cooler lid after the sample has been placed inside.

    6.2.4   When all samples have been collected and placed in the relevant transport box place the lid on the transport box and seal and padlock shut.

    6.2.5   Samples should be transported to the University of Makeni laboratory where they will be processed/stored and prepared for onward shipment to the UK.

7   Removing personal protective equipment

7.1 Once sample collection is complete and samples are contained within their respective transport cases, personal protective equipment can be removed.

7.2 Ensure a heavy duty clinical waste bag is available for disposal of waste.
7.3 To remove personal protective equipment:

    7.3.1   Clean gloves with alcohol gel
    7.3.2   Tear apron ties on right side of body and remove apron by folding in
                   on itself, discard in waste bag
    7.3.3   Clean gloves with alcohol gel
    7.3.4   Remove mask&visor by snapping ties and pulling forwards, then
                   discard in clinical waste bag.
    7.3.5   Clean gloves with alcohol gel
    7.3.6   Remove gloves using the 'beaking method' (field team will be trained
                   in this) involving removing the gloves inside out without skin touching the outer aspect
                   of the glove. Discard in waste bag.
    7.3.7   Clean hands with alcohol gel

        7.3.8 Prior to departing from the area (clean assistant while still
                 wearing gloves):
    7.3.8.1   Ensure sharps box is sealed
    7.3.8.2   Tie waste bag and double bag waste bag, clean assistant
                   discards gloves using beaking method into outer waste bag and ties bag closed.
                   Then cleans hands with alcohol gel.
    7.3.8.3   Place all items (including samples in transport containers, see
                   section 6) in the back of the transport vehicle
    7.3.8.4   Only in the event of a spill or walking through a contaminated
                   area - spray boots or soles of shoes of investigators using
                   spray can with virkon prior to entering vehicle.

8   Disposal of waste

8.1 Remove all waste from the location of sampling unless sampling is undertaken in a designated centre for survivor clinics/sampling and the centre has its own designated high risk infection waste disposal service.

    8.1.1   Seal the sharps boxes, wipe with a medipal high level disinfectant
                 wipe, and place in the vehicle.

8.2 Disposal of waste should be as follows:

    8.2.1   Transport waste contained in sharps boxes and heavy duty waste
                 bags to the University of Makeni laboratory.

        8.2.1.1   Remove sharps boxes and discard in accordance with
                   procedures of the facility for high risk sharps disposal (most likely incineration)
        8.2.1.2   Discard waste bags in accordance with the procedures of the
                   facility for high risk dry waste management (most likely incineration).

9   How to manage a spill

9.1 In the event of a spill:

    9.1.1   Remain calm

9.1.2   Clean assistant

        9.1.2.1   Without touching the spill, sprinkle top of spill with Virkon granules from squeeze puff container
        9.1.2.2   Allow time to absorb
        9.1.2.3   Once absorbed, sprinkle with Safetygel granules
        9.1.2.4   Allow to full form into a gel
        9.1.2.5   Using the shovel and without touching the gel, ensure gel is fully on shovel and place into heavy duty clinical waste bag. Clean gloved hands with alcohol gel, then discard in clinical waste bag.
        9.1.2.6   If remnants of gel remain these should have been disinfected by both the virkon and biocidal safety gel. For extra safety apply visor & mask and thumbs-up gown, fill spray can with water and once virkon tablets dissolved, spray the gel remnants.

10   In the event of a needle stick injury/mucous membrane splash

10.1   In the event of a needle stick injury:

      10.1.1   Expose the wound
      10.1.2   With a clean glove on other hand encourage the wound to bleed
      10.1.3   Rinse thoroughly with saline wash from first aid kit (provided by clean assistant), then clean wound with chlorhexidine skin
      10.1.4   Clean both hands with alcohol gel (applied by clean assistant)
      10.1.5   Cover injury with a clean glove
      10.1.6   Proceed to remove all personal protective equipment as per section 6
      10.1.7   Obtain consent from participant to test blood sample for Ebola virus
      10.1.8   Inform field team leader
      10.1.9   If participant has positive Ebola blood test:

        10.1.9.1   Self-isolate for 21 days and monitor temperature twice daily
        10.1.9.2   If symptoms or fever develops inform field team leader for further assessment

      10.1.10   If participant has negative Ebola blood test:

        10.1.10.1   Continue with normal daily practice
        10.1.10.2   Monitor temperature twice daily for 21 days
        10.1.10.3   If symptoms or fever develop self-isolate and inform team Leader

10.2   In the event of a mucous membrane splash:

      10.2.1   Clean gloves with alcohol gel
      10.2.2   Rinse the affected area thoroughly with saline wash from first aid kit (provided by clean assistant)
      10.2.3   Do **NOT** touch the area with gloved hands
      10.2.4   Proceed to remove all personal protective equipment as per section 6
      10.2.5   Obtain consent from participant to test blood/saliva sample for Ebola virus
      10.2.6   Inform field team leader

10.2.7   If participant has positive Ebola blood test:

10.2.7.1   Self-isolate for 21 days and monitor temperature twice daily
10.2.7.2   If symptoms or fever develops inform field team leader for
              further assessment

10.2.8   If participant has negative Ebola blood test:

10.2.8.1   Continue with normal daily practice
10.2.8.2   Monitor temperature twice daily for 21 days
10.2.8.3   If symptoms or fever develops self-isolate and inform team
              leader

SOP – Sample handling and processing

1. Sample storage at source
2. Sample handling at laboratory
3. Sample processing in laboratory
4. See SOP – ELISAs

Equipment required:
1. Oragene samples
2. Oral fluid swab/sponge samples
3. Waterbath
4. Thermometer for waterbath
5. -20 freezer
6. Gloves
7. Lab coat
8. Safety goggles
9. Clinical waste bag
10. Econix Bio-bin
11. Beaker with spout (1 litre)
12. Virkon tablets
13. DNA genotek SO-1 storage boxes
14. 100 place freezer box
15. Pipette
16. Pipette tips (1000ul)
17. 1.5ml Sarstedt tubes and caps
18. Viral transport medium/buffer
19. **Medipal high level disinfectant wipes**

1. Sample storage at source

   1.1. Oragene sample receptacles will have been sealed in a plastic specimen bag with a biohazard sticker (also containing safetygel granules).
      1.1.1. These samples will be contained in the IM3075 storm case (black)
   1.2. Oral fluid sponge/swab samples will be contained in their sample receptacle and sealed in a plastic specimen bag with a biohazard sticker (also containing safetygel granules)

      1.2.1. These samples will be contained, on ice packs, in the 80QT wheeled cooler (white).

2. Sample handling at the laboratory

   2.1. All samples and sample specimen bags should be handled while wearing a lab coat and gloves.
   2.2. If opening specimen bags, eye protection (safety goggles) should be worn
   2.3. Oragene samples should be placed within the category 2 safety cabinet or isolator, whichever is available, for further processing.

      2.3.1. Upon emptying the IM3075 storm case, clean hands with 70% ethanol

2.3.2. Wipe IM3075 storm case with Medipal high level disinfectant
wipes ensuring the whole of the inner case is wiped down

2.4. Oral fluid samples should either be placed in the category 2 safety cabinet or isolator, whichever is available, for further processing or be placed directly in the relevant compartment of the -20 freezer in their specimen bag, for processing at a later date.

2.4.1. Upon emptying the 80QT wheeled cooler, clean hands with 70%
ethanol
2.4.2. Remove ice packs, wipe with Medipal high level disinfectant wipes
and place in separate compartment of -20 freezer
2.4.3. Wipe 80QT wheeled cooler with Medipal high level disinfectant
wipes, ensuring the whole of the inner case is wiped down.

3. **Sample processing in laboratory**

3.1. Oragene samples

3.1.1. While in the isolater, wipe the outer container with a medipal high
level disinfectant wipe, then discard wipe into clinical waste bag
3.1.2. Remove the large upper receptacle and replace with the
accompanying screw cap from the sample pack
3.1.3. Place receptacle in 1 litre beaker with spout containing Virkon (1x
5g tablet dissolved in 500ml water)
3.1.4. Leave to soak in Virkon for minimum 1 hour
3.1.5. Following soaking in Virkon, discard the large upper receptacle in
a clinical waste bag for later incineration.
3.1.6. Ensure the water bath is heated to 60C (verify using additional
thermometer for the waterbath)
3.1.7. Place capped Oragene sample in the removable waterbath rack
inside the isolator/category 2 safety cabinet.
3.1.8. Once the rack is full of samples remove from the isolator/safety
cabinet and place in the waterbath, ensuring rack appropriately placed for all Oragene
samples to be submerged in the water.
3.1.9. Leave Oragene samples in the waterbath for a minimum of 60
minutes (maximum 2 hours)
3.1.10. After 60 minutes remove samples from waterbath rack.
3.1.11. Place samples in the DNA genotek SO-1 storage boxes
3.1.12. Close storage box and store in the IM3075 Storm case stored in the
Laboratory at room temperature.
3.2. Oral fluid samples

3.2.1. If oral fluid samples have been placed directly into -20 freezer, ignore this step.
3.2.2. If oral fluid samples have been placed into isolator/category 2
safety cabinet for processing:

3.2.2.1. Open sample inside safety cabinet/isolator
3.2.2.2. Add 1ml viral transport medium/buffer to the sample
container
3.2.2.3. Discard pipette tip into 1 litre beaker with spout containing
Virkon (1x 5g tablet dissolved in 500ml water)

3.2.2.4.	Leave to soak in Virkon for minimum 1 hour and then discard into Econix bio-bin.

3.2.2.5.	Using the blue sponge containing oral fluid, froth the sample and then squeeze the fluid out of the sponge into the base of the container by applying pressure to the sponge against the container and twisting.

3.2.2.6.	Discard sponge into 1 litre beaker with spout containing Virkon (1x 5g tablet dissolved in 500ml water)

3.2.2.7.	Leave to soak in Virkon for minimum 1 hour and then discard into Econix Bio-bin

3.2.2.8.	Aspirate the fluid from the sample receptacle and place into a 1.5ml Sarstedt tube and seal with screw cap.

3.2.2.9.	Discard sample receptacle into 1 litre beaker with spout containing Virkon (1x 5g tablet dissolved in 500ml water)

3.2.2.10.	Leave to soak in Virkon for minimum 1 hour and then discard into clinical waste bag.

3.2.2.11.	Wipe the Sarstedt tube with a Medipal high level disinfectant wipe, then discard wipe into clinical waste bag

3.2.2.12.	Place the Sarstedt tube into the 100 place freezer box.

3.2.2.13.	Once the freezer box is full, place the box in the -20 freezer

SOP – EBOV IgG Capture Enzyme linked immunosorbent assay (ELISA)

1. Preparation
2. Procedure
3. Clear-up
4. See SOP – Sample handling and Processing V1

Equipment required for 80 samples:
1. 90 Oral fluid samples eluted in 1ml viral transport medium
2. CL2 safety cabinet
3. 1x EBOV IgG Capture ELISA kit provided by Kalon Ltd (developed by Prof Tedder, PHE)
4. 1x Fischer Scientific Multiskan FC plate reader with incubator, equipped with 450nm and 630nm filters
5. 1x Tecan plate washer
6. Gloves
7. Lab coat
8. Clinical waste bag
9. 1x Econix Bio-bin
10. 1x Beaker with spout (1 litre)
11. Virkon tablets
12. 2x 100 place freezer box
13. Pipettes (P1000, P200, P20)
14. Multichannel pipette (P300)
15. 1x 96 well V or U bottomed plate
16. 2x adhesive plate seal
17. Pipette tips (1000ul, 200ul, 20ul)
18. 1.5ml Eppendorf tubes
19. 1.5ml Sarstedt tubes and caps
20. -20 freezer
21. Medipal high level disinfectant wipes

Preparation
1. All samples will be handled with gloves and while in the CL2 laboratory disposable lab coats will be worn at all times.
2. Remove selected samples from -20 freezer and bring to room temperature in the CL2 cabinet.
3. Prepare the wash buffer by adding 1 part wash fluid concentrate to 19 parts of distilled water.
4. Prepare conjugate solution by adding one volume of conjugate concentrate to 99 volumes of conjugate diluent.
5. Label additional sample tubes for later aliquots.

Procedure
1. Using a blank 96 well V or U bottomed plate in the CL2 safety cabinet:

2. Pipette 100ul per well of the positive and negative controls to assigned wells (two wells for positive control and four wells for the negative control).
3. Spin the oral fluid samples in a centrifuge at 3000g for 1 minute to ensure a tight cell pellet in the base of the Sarstedt tube with a clarified oral fluid sample above the pellet.
4. Pipette 100ul per well of the clarified oral fluid sample into the assigned well for the sample ID number.
5. Discard the pipette tip into the 1L beaker containing Virkon solution (1 Virkon tablet per 500ml water).
6. In the CL2 safety cabinet, using the one anti-human IgG 96-well microtitre plate contained within the Kalon EBOV Capture IgG ELISA kit:
7. Using the multi-channel pipette, pipette the contents of the 96-well plate into the anti-human IgG 96-well microtitre plate ensuring the contents of the entire plate is pipetted into the ELISA plate within 10 minutes.
8. Pipette tips should be discarded in to a 1l beaker containing Virkon solution as per 2.1.4 above.
9. Discard the empty 96-well plate and pipette tips into a 1L beaker containing Virkon solution as per 2.1.4 above.
10. Cover the microtitre plate with the adhesive plate seal and wipe the whole plate with a MediPal high level disinfection wipe (taking care not to contaminate contents of plate) before removing from the CL2 cabinet or placing in the incubator.
11. Place the Tecan plate washer in the CL2 cabinet
12. Place the covered microtitre plate in the incubator in the CL2 and incubate at 37C for 60 minutes.
13. Return the sealed plate to the CL2 cabinet and unseal the plate.
14. Place the plate in the plate washer and set the plate washer to wash the wells five times with diluted wash buffer using the following programme:
15. Aspirate contents of well
16. Dispense 300ul per well of dilute wash buffer to form a meniscus
17. Leave to soak for 30 seconds then aspirate
18. Tap the wells dry using absorbent paper.
19. Using the multi-channel pipette, pipette 100ul per well of conjugate solution to the wells, discard pipette tips in an Econix biobin.
20. Re-cover the plate with a new plate seal, wipe with MediPal high level disinfection wipe taking care not to contaminate contents of plate and incubate in the Multiskan FC plate reader or the incubator in the CL2 at 37C for 120 minutes.
21. Remove the plate seal and wash the plate 5 times using the programme in 2.3.3.
22. Tap the wells dry using absorbent paper.
23. Wipe the base of the plate and any equipment being removed from the CL2 cabinet with a MediPal high level disinfection wipe taking care not to contaminate the wells of the plate.
24. It is now safe to move the rest of the procedure to the bench top next to the Multiskan FC plate reader in the CL1 laboratory.
25. Using the multi-channel pipette, pipette 100ul per well of ready to use TMB substrate, discard pipette tips into plastic jar or Econix biobin.
26. Incubate the plate in the Multiskan FC incubator or the incubator outside the CL1 at 37C for 30 minutes protecting it from strong light.
27. Using the multichannel pipette, pipette 50ul per well of the Stop solution using the same timing and sequence used to add the substrate solution. Discard pipette tips into plastic jar or Econix biobin.
28. Using the Multiskan FC plate reader insert the plate into the plate reader and read the optical density at 450nm with the reference wave length set at 630nm.

Clear up

1. Once confident of the results of the ELISA the microtitre plate can be tapped dry on an absorbent towel and discarded into a clinical waste bin.
2. The remaining clarified oral fluid from the oral fluid sample should be aliquoted into a clean, labelled 1.5ml microcentrifuge tube to separate the clarified fluid from the cell pellet. Discard pipette tips into plastic jar or Econix biobin.
3. The clarified fluid and the cell pellet should be placed in separate labelled freezer boxes and both returned for storage in the -20 freezer.
4. The 96 well plate that was placed in Virkon can be removed after soaking for minimum 60 minutes and discarded in a clinical waste bin.
5. The pipette tips discarded into Virkon should remain submerged for a minimum of 60 minutes from the last pipette tip entering the solution. Following this the solution can be drained and the pipette tips discarded into an Econix Biobin, which is then sealed in a plastic bag and discarded in the clinical waste bin.

SOP – QIAsymphony DNA extraction from DNA Genotek Oragene *saliva samples*

1. Equipment
2. Preparation
3. Procedure
4. Quantification and storage
5. Clear up

1. Equipment required
   a. QIAsymphony SP instrument
      i. QIAsymphony DSP DNA midi kit
      ii. QIAsymphony 1500ul pipette tips
      iii. QIAsymphony Sample prep cartridges
      iv. QIAsymphony rod covers
      v. Elution microtubes CL (24 x96)
   b. QIAxpert instrument
      i. QIAxpert slides (16 samples/slide)
   c. Vortex
   d. Centrifuge with buckets for *Oragene* tube size
   e. 8ml Sarstedt tubes
   f. 96 well PCR plates (-80)
   g. Metallic plate seals (-80)
   h. Virkon powder
   i. 2L plastic beakers
   j. 5L orange biobins
   k. Autoclave sacks
   l. Orange clinical waste sacks
   m. Lab coat
   n. Gloves

2. Preparation
   i. Turn on QIAsymphony and wait for it to complete initiation sequence
   ii. Load QIAsymphony SP with sufficient consumables for the run size (1-96 samples) and run scan of consumables. QIAsymphony will inform if insufficient consumables for programmed run.
      i. Ensure the QIAsymphony tip disposal receptacle is double bagged with autoclave bags.
   iii. Obtain sealed Oragene samples from Peli case and ensure they are either all OG-500 or all OG-575 (smaller) samples.
   iv. Label 96x 8ml screw cap Sarstedt tubes with appropriate linear barcode labels for samples to be run.
   v. Load samples into bucket centrifuge (max 36 samples/spin) and spin for 10 minutes at 5000g at room temperature.
   vi. Once all samples have been spun vortex each sample briefly.

     vii.   Using the CL2 safety cabinet in tissue culture (2nd floor) aliquot the saliva supernatant into the appropriately labelled 8ml Sarstedt tube taking care not to dislodge the pellet or aspirate any foreign material.

    viii.   Dispose of pipette tips in plastic beaker containing 0.1% Virkon solution and allow to soak for a minimum of 30 minutes prior to discarding in a biobin. Ensure sample is tightly sealed prior to removal from safety cabinet.

      ix.   Transport the sealed tubes in racks to the CL1 virology laboratory on the 3rd floor.

       x.   In the event of a spill ensure area is cleaned with 70% ethanol

3. Procedure
   i. QIAsymphony should be ready to use and fully loaded with consumables based on step 2i and 2ii.
   ii. On the bench load samples from racks into QIAsymphony racks for insertion into instrument, ensuring linear barcode label is visible for barcode reader to read.
   iii. Remove tube caps as the sample is loaded into the loading rack and store in appropriate receptacle
   iv. Insert one rack at a time into QIAsymphony following the on screen instructions.
   v. If running 96 samples, consider inserting further racks at a later stage once QIAsymphony is processing initial 24 samples to prevent discrepancies in distribution of cells and nucleic acid throughout sample.
   vi. Once samples loaded into QIAsymphony select appropriate *Oragene* protocol (for OG-500 select 1ml protocol, for OG-575 select 350ul protocol.
   vii. Close QIAsymphony sample loading rack and allow to complete run (for 96 samples this may take 5-6 hours).
   viii. Once run is complete:
      i. Remove sample tubes from QIAsymphony, re-cap and store remaining sample in original sample box in Peli case.
      ii. Remove 96 well Elution microtubes plate and place on bench.
   ix. In the event of a spill at any point the area should be cleaned with 70% ethanol.

4. Quantification and storage
   i. From elution microtubes plate load slides for QIAxpert (16 samples per slide) and quantify DNA concentration using QIAxpert instrument, record concentrations.
      i. Discard pipette tips into beaker of 0.1% virkon solution and leave to soak for minimum 30 minutes before discarding into orange biobin.
   ii. Using a multichannel pipette aliquot 50ul of the DNA extract from each well into the appropriately labelled 96 well PCR plate
   iii. Seal PCR plate with metallic plate seal and label seal accordingly, place plate in -80C freezer.
   iv. Re- seal the Elution microtubes plate containing the remainder of the DNA extract, ensure plate and lid are appropriately labelled and place in -80C freezer.
   v. Dispose of pipette tips into beaker containing 1% Virkon solution
      i. Permit to soak for 30 minutes and then dispose into 5L Orange biobin
   vi. Store plate of nucleic acid at -80C until shipment to genomics facility
   vii. Store Elution microtubes plate at -80C long term
   viii. In the event of a spill at any point the area should be cleaned with 70% ethanol.

5. Clear up

i.   Ensure waste removed from QIAsymphony waste disposal area, is in double bagged autoclave bag, seal bag with autoclave tape and transport to waste disposal room on 3rd floor.

ii.   Seal any full 5L orange biobins and transport to waste disposal room on 3rd floor.

iii.   Remove reagent cartridge from QIAsymphony and re-seal for storage if remaining reagent. If empty, discard cartridge in doubled autoclave bags, seal with autoclave tape and transport to waste disposal room on 3rd floor.

iv.   Remove liquid waste container from QIAsymphony and add 10g of Virkon powder to create 0.1% solution. Allow to soak for minimum 30 minutes before discarding down drain.

v.   Close all QIAsymphony doors and shut down instrument

vi.   Discard all QIAsymphony plastic waste which has not been inside the instrument into orange clinical waste bags, tie with cable tie and transport to waste disposal room on 3rd floor.

vii.   Dispose of pipette tips into 5L orange biobin once they have soaked in Virkon for 30 minutes, seal biobin and transport to waste disposal room on the 3rd floor.

viii.   Transport samples and extracted nucleic acid in sealed specimen boxes in plastic box to -80C freezer in paediatrics (nucleic acid) and Peli case in paediatrics store room (remaining sample).

ix.   Clean bench space in virology with 70% ethanol.

x.   In the event of a spill at any point the area should be cleaned with 70% ethanol.

K. Permission to use Emory University Department of Visual Medical Education Image – Email confirmation

Re: Use of Ebola virus image published in your magazine

Loftus, Mary <mary.loftus@emory.edu>
MacDermott, Nathalie E;   Konomos, Michael
Thursday, 11 April 2019 at 20:45
Show Details

Thanks so much for your beautiful note! I have no problem with it but I am copying Michael Konomos, whose visual education team produced the image, for confirmation.

--Mary Loftus
Editor, *Emory Medicine* and *Emory Health Digest*
404.727.0161

"Stories are a communal currency of humanity."—Tahir Shah, in *Arabian Nights*
http://emorymedicinemagazine.emory.edu/

From: "MacDermott, Nathalie E" <n.macdermott@imperial.ac.uk>
Date: Thursday, April 11, 2019 at 10:27 AM
To: Mary Loftus <mary.loftus@emory.edu>
Subject: Use of Ebola virus image published in your magazine

Hello,

I came across your image of the structure of Ebola virus from a 2014 edition of your magazine. I was wondering if I might have your permission to use this image in my PhD thesis on Ebola (appropriately referenced) as it is the best image I have been able to find?

On a separate note, thank you for this article. I am surprised I had not seen it before, but there were a lot of articles on Ebola around that time. I worked with Kent Brantly at the ELWA Ebola treatment facility when he contracted Ebola. I tested him for Ebola and cared for him in Liberia shortly before he was evacuated to Emory. Thank you for the care you provided him and Nancy Writebol with, I am very grateful that they are still in my life today, which is a direct result of the care they received at Emory.

Best wishes,

Nathalie

permission to use Ebola virion illustration

Loftus, Mary <mary.loftus@emory.edu>
MacDermott, Nathalie E
Thursday, 11 April 2019 at 20:45
Show Details

You replied to this message on 12/04/2019, 12:03.   Show Reply

Satyen Tripathi and Andy Matlock worked on this piece. As long as Emory gets credit ("©2014 Emory University") we are fine with them using it.

Kind regards,
Michael Konomos, MS CMI
Lead Medical Illustrator
Visual Medical Education
Emory School of Medicine
1762 Clifton Road NE, Suite 2200
Atlanta, GA 30322
office: 404.712.7466
mobile: 404.939.1439
m.konomos@emory.edu

--Mary Loftus
Editor, *Emory Medicine* and *Emory Health Digest*
404.727.0161

"Stories are a communal currency of humanity."—Tahir Shah, in *Arabian Nights*
http://emorymedicinemagazine.emory.edu/

L. Additional analysis: Transmission dynamics in communities – case studies

Transmission dynamics appear to have varied between communities and it is important to understand the factors involved in the individual circumstances of the community to understand this variation in transmission dynamics. To do this there are three specific rural communities where understanding what occurred will help to understand different factors that contributed to or assisted in halting disease transmission. Now that we have an understanding of the seropositivity rates of household contact and community controls within these communities, it helps us to further understand the dynamics which took place.

L.1 Makontha and Masuba – chief tensions

The community of Makontha is nestled within the jungle in Marampa kingdom, Port Loko district. A community of just seven dwellings and a mosque, it likely hosted a population of approximately 50-60 people prior to the epidemic, based on household sizes in this region. It had no access to running water or any formal latrines. It is a fifteen minute walk through the jungle to its sister community, Masuba which is at the end of the dirt track road, approximately one hour's drive from the Lunsar-Makeni highway. Masuba is a community that is not much bigger than Makontha, with a couple of extra houses, but does host a bore hole for water access. There is no phone signal in either community, to obtain a signal you must walk about 20-30 minutes up the dirt track from Masuba.

The population of Makontha were relatives of Masuba and had chosen to plant their own community in the jungle. The chief of Makontha was related to the chief of Masuba. The chief of Masuba did not respect the Section Chief for the region and would not routinely engage with his requests. In the autumn of 2014, a community member from Makontha visited an unwell relative in another community, Labour compound, a short walk through the jungle. Labour Compound had just started to have cases of EVD. The community member from Makontha unknowingly became infected with EVD. They returned to Makontha and became unwell. The disease gradually spread through the community, as each person died the community buried them behind the mosque. Word reached Masuba that people in Makontha were ill. When the majority of the population of Makontha had died, an uncle went to collect his unwell niece from Makontha and brought her to Masuba to care for her. Neither the chief of Makontha, nor the chief of Masuba, contacted the section chief, contact tracers, burial teams or the Ebola hotline, despite having been informed by the Section Chief of what to do. This was in part due to the limitations in cellular networks, but also because of distrust between the Local Chiefs and the Section Chief and a desire to handle things by themselves.

The entire population of Makontha died of EVD. EVD then spread through the infected niece to her family's household in Masuba, resulting in several deaths. Due to the community's desire to handle things themselves, they buried their own dead, and EVD spread through Masuba infecting three other households, killing significant numbers and wiping out one entire household. At this late stage the local chief requested assistance. Burial teams arrived to handle the dead and contact tracing was installed in the community. Subsequently Masuba was not wiped out in the same way as Makontha and the disease was contained. There are however, no known EVD survivors from Masuba.

Out of eleven household contacts recruited in Masuba one was anti-EBOV IgG positive (9%) and one had an equivocal result. Fourteen family members are reported to have died in the two households that agreed to be recruited (average of 7/household), 6 household members reported direct contact with a dead body (55%). Of 22 controls recruited from four households, one is anti-EBOV IgG positive (4.5%). This individual was likely from an affected household, but they repeatedly denied this and refused to be recruited to the study as a household contact, they were willing to be recruited as a community control however. These figures are substantially higher than the overall seropositivity figures reported for community controls and household contacts. The lack of engagement with the section chief and actors involved in the Ebola response due to distrust and previous poor relationships resulted in the loss of the entire community of Makontha and approximately a third of the population of Masuba.

Figure L-1 Makontha, Marampa Kingdom, Port Loko District, Sierra Leone. February 2017.

Figure L-2 Masuba, Marampa Kingdom, Port Loko District, Sierra Leone. February 2017. The entire family of the house on the left died of EVD. Following this the house was used as somewhere to place the sick to try and prevent disease spreading in the community, once they had decided to engage with the Section Chief. Everybody who was placed there died. This house is now uninhabited and has been claimed by the sheep and the goats.

L.2 Kontha-Bana – failure of community engagement and contact tracing

Kontha-Bana is a relatively large community in Marampa Kingdom, Port Loko district. It is on a dirt road approximately 20 minutes' drive from the town of Lunsar, which is situated on the Lunsar-Makeni highway. The dirt road passes through this community onwards toward other even more rural

communities. At the time of recruitment to the study this community was eerily quiet for a relatively large community in rural Sierra Leone, a reflection of the devastation Ebola wrought there.

In approximately September 2014 a pregnant woman absconded her quarantine in an affected area of the town of Port Loko, attempting to travel home to her village. She reached the community of Kontha-Bana but stopped there due to severe abdominal pain. As she was in the late stages of pregnancy, the women of Kontha-Bana rushed to assist her in delivering her infant. Sadly, the infant was never born and both mother and baby died that day. The mother was infected with EVD and subsequently infected a large number of the women of Kontha Bana. The first household to be affected saw the disease spread rapidly, killing all but one woman. This woman helped care for all her family members when they were sick and prepared the dead bodies of her son, husband, mother and grandmother. She never became unwell. As this was the first household affected and Ebola community sensitisation had not yet reached this community, the community believed the household had been cursed by a witch whose plane had crashed in the night (Figure L-3).

HH Line#	Household member name	Relationship to line number	Line # related to	Sex	Age (yrs)	If <2 yrs Age (mo)	Ever had Ebola?	Died from Ebola?	Saliva sample collected	Oral fluid collected	If saliva/fluid collected place sample ID sticker below
01	[redacted]			M / F	20		Yes / DK	Yes / DK	No / Yes / Revisit	No / Yes / Revisit	EG-SL-HC 0429
02	[redacted]	Husband	01	M / F	35		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
03	[redacted]	mother	01	M / F	45		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
04	[redacted]	Father	01	M / F	65		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
05	[redacted]	Grandmother	01	M / F	75		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
06	[redacted]	Uncle	01	M / F	68		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
07	[redacted]	Grandmother	01	M / F	90		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
08	[redacted]	step mother	01	M / F	75		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead
09	[redacted]	Brother	01	M / F	35		No / Yes / DK	No / Yes / DK	No / Yes / Revisit	No / Yes / Revisit	N/A dead

	None	
8. Was your household quarantined during the time Ebola was in your community?	Yes	
	No	

(handwritten response) reason for not quarantined is because of limited personnel in the town.
- 1st home of the epidemic in the section with unexplained myths about the rampant death of people in the house.
- ie a which air plane crashed at night and is the result of the rampant death.

Figure L-3 Screenshots of sections of the recruitment inventory and questionnaire for a household in Kontha Bana.

As the disease started to spread throughout the community the alarm was raised, and the military were brought in to quarantine the community and support the contact tracers. Due to the rampant spread of infection and high levels of death in the community, the military abandoned the quarantine out of fear and contact tracing was impaired. There were also few treatment facilities and inadequate bed numbers at this stage in the epidemic. Out of 63 household contacts recruited six (9.5%) are anti-EBOV IgG positive, with one equivocal result, and ten (15.9%) had contact with a dead body. Among the twelve households recruited, 62 people died (an average of 5/household) and there were six Ebola survivors. Of 41 community controls recruited from eight households, three were anti-EBOV IgG positive (7.3%), with one equivocal result. Again, these seropositivity figures are far higher than rates

demonstrated across all the communities that participants were recruited from. They indicate the impact of failed contact tracing and early case isolation. The high mortality rates also indicate the delay in EVD infected individuals receiving good supportive care. While it is difficult to be certain, it is apparent that individuals from this community seeded infection in other communities within Marampa Kingdom, a result of the failure of quarantine.

Figure L-4 The community of Kontha Bana showing the dirt road passing through it. Kontha Bana, Marampa Kingdom, Port Loko District, Sierra Leone. February 2017.

L.3 Marampa – towards a success story

Marampa is the capital of Marampa Kingdom in Port Loko district, and while rural, is a very large community. Marampa is situated towards the end of a dirt road, approximately 40 minutes' drive from the Lunsar-Makeni highway. On our arrival in Marampa to meet the community we noted both a Local Chief and a Spiritual Chief. The Spiritual chief observed our introduction to the community by the Section and Local chief and then stood up and commanded the attention of the community. He requested they all engage with our study, because the study was important, and we had risked our lives in coming to their community to conduct the study and in working in the epidemic previously. Subsequently we recruited 269 participants in Marampa. This introduction however was an insight into the leadership that had occurred during the epidemic. In Marampa we heard how the Local and Spiritual Chief had followed the instructions of the Section and Paramount Chiefs. At the time when the disease spread to Marampa, in the autumn of 2014, there was still a scarcity of treatment facilities and beds. Transport from Marampa to the treatment facilities was also limited.

To address this situation the chiefs instructed the community to identify an area the sick person in the household could be managed. Some households obtained some tarp and constructed an isolation area/shelter for the person outside of the family home. One person in the family was identified to care for the sick person until they could be transported to a treatment facility. In one part of the community an empty dwelling was used where the sick were taken and cared for by a nominated family member.

While this situation is not ideal due to the risk of exposure of the family member, it limited disease transmission within Marampa. As a result of this strategy, out of 90 household contacts recruited only two were anti-EBOV IgG positive (2.2%) and four had contact with a dead body (4.4%). Among the nine households recruited twelve people died (an average of 1.3 per household), although there were no known Ebola survivors. Twenty-eight households were recruited as community controls with a total number of participants of 179. Of these, five were anti-EBOV IgG positive (2.8%) with one with an equivocal result.

While the concept of a homebased care methodology in containing epidemics has received much negative press, the example of Marampa demonstrates that, while it does not provide optimum care for the sick, it does contain the spread of the epidemic within the community. While it should not be the goal of clinical care for EVD patients, as a stop gap measure until people can be transferred to an appropriate treatment facility, particularly from distant rural areas, this method shows promise in relation to limiting the number of people becoming infected within households and the community as a whole.

Figure L-5 Marampa, Marampa Kingdom, Port Loko District, Sierra Leone. January 2017.

L.4 Dynamics of urban communities

Community dynamics in urban settings were quite different from community dynamics in rural areas. While there was often a Headman or Headwoman for the community, the traditional chief structure was not as apparent. The communities merged from one into another and the population density was much higher. People were less dependent on subsistence agriculture, but poverty was still rife. House sizes were often smaller with smaller total family sizes, but an entire family might be squeezed into one or two rooms in order to rent out a room to tenants to boost income. Family structure also differed with less extended family present, often young adults or families had moved from a rural village to try and find employment or boost income in an urban setting. In this context older family members and sometimes children were left in rural communities. Transfer of patients to treatment facilities in urban settings was more straight forward, as was access for Ebola community sensitisation, but there still appeared to have been significant community resistance in some areas, resulting in the hiding of sick and deceased individuals initially.

L.5 Summary - Disease containment in communities

The community case studies above provide some examples of where strategies that were implemented worked and where other strategies that should have worked failed due to implementation issues. What is apparent is that community engagement, contact tracing, early isolation and treatment and safe and dignified burials should always be priorities in containing disease spread within communities. How this is addressed and implemented is crucial to the success of such measures. Gaining community trust through traditional leadership structures is essential, particularly as their support will ensure community engagement with response structures. This should not be overlooked in light of simply engaging with current political structures, which carry much less value and much more distrust in rural areas than traditional leadership structures. Where there is also dysfunction or mistrust of traditional leadership structures, actors must engage with someone considered of trust to the community, be that a spiritual leader or influential community member. In urban areas it is important to identify what the local traditional leadership structure is, as well as the political structure, to engage such individuals in the response. What is also apparent is that community engagement may well be the single most important factor in disease containment, as without it other measures will likely fail. On these grounds this ought to be one of the first measures prioritised and implemented. In areas at risk, but not yet affected by the disease, this should be implemented well before the construction of any form of treatment facility, as soon as structures are built the rumour mill begins and community resistance gains momentum.