Study 1

PHE-MOHS Ebola Biobank Sample Access Form

Applicants for materials from the biobank must complete and submit this application form to the PHE Head of Research Governance. Decisions to grant access should (i) help ensure that any uses of the resource are consistent with the purpose of the biobank; (ii) ensure that research projects have relevant scientific and ethical approval, and (iii) make information publicly available about uses of the PHE-MOHS biobank collection. Review of the application will be undertaken by the PHE-MOHS Ebola Biobank Governance Group (EBGG), who will consider the following eligibility criteria.

1. The proposed research is scientifically sound
2. There is sufficient funding to enable completion of the research
3. The research is designed in such a way that ensures integrity, quality and transparency
4. The research team are suitably qualified by education, training and experience.
5. The research site has adequate capacity and capability to undertake the research
6. The research has all necessary ethical and regulatory permissions.
7. The materials will be used to ensure the greatest benefit to the public
8. The research has relevance to the people of Sierra Leone, and a likelihood that the people will be able to benefit from it.

The level of scrutiny used to assess applications will be proportionate to the nature and scale of the research project, taking into account (by way of example) whether a significant amount of material is required relative to the project objectives.

In reaching its decision on a particular research application, particularly with regard to items 1 – 6 inclusive, above, the EBGG will accept the outcomes of peer review that is undertaken by a research funder. Where that is not possible or where there is doubt that an application may not satisfy any of the eight criteria, applications may be scrutinised by an independent scientific reviewer selected by the EBGG in order to ensure independence, competence and rigour of peer review

Additionally the following will be considered by EBGG:

- Quantity of material requested
- Arrangements to maintain data security and confidentiality
- Biosafety and biosecurity
- Proposed dissemination of results

When the review is complete the application will either be:

1. Approved
2. Approved pending receipt of further information/approvals within a set period of time.
3. Approved subject to a project amendment
4. Declined.

Applicants awaiting the outcome of an application for funding will receive an 'approval in principle' letter. Access will be conditional on there being sufficient research funding to use the Biobank samples for the intended purpose.

When funding is in place the research study lead should forward a copy of the funding proposal and the letter of offer to the EBGG, along with any amendments to the original project that have been required by the funder.

1 Title of Proposal

Please give a succinct descriptive title for the proposal

Are there genetic determinants of clinical phenotype and disease outcome for Ebola Virus Disease (EVD)?

2 Study team

Please list all collaborators, lead investigator named first, with institute affiliations

[redacted]

Collaborators [Section 40] [redacted]

3 Background

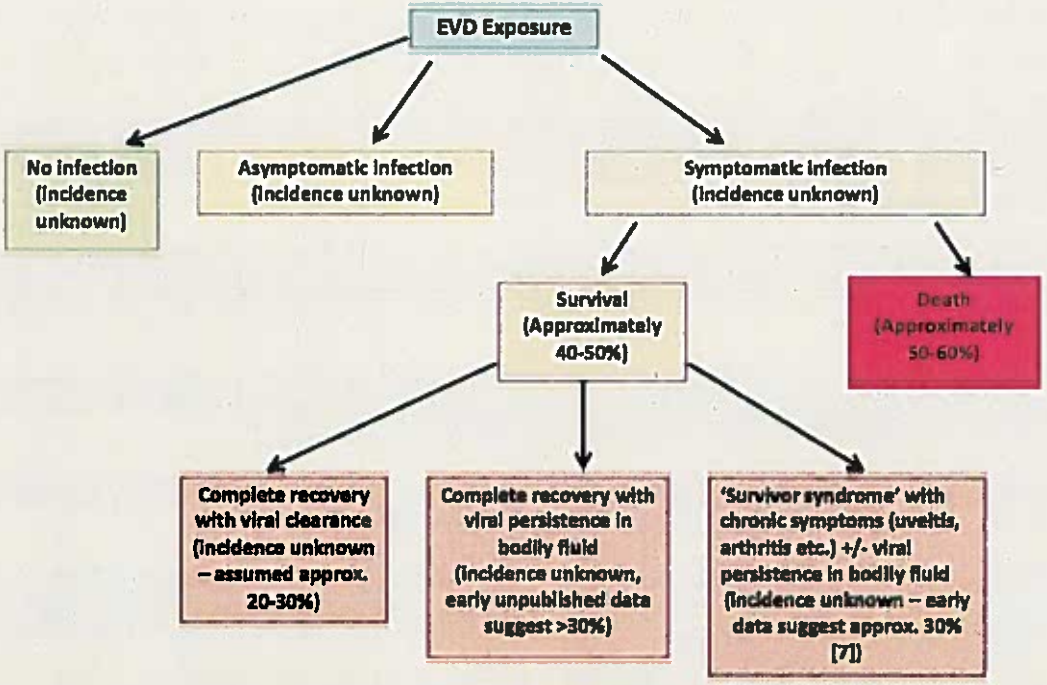
Project rationale including a brief review of relevant literature with key references (max 400 words, references additional)

In recent years there has been increasing evidence that genetic factors play an important role in susceptibility and outcome from infectious diseases. Ebola virus disease (EVD) demonstrates consistent mortality rates throughout epidemics in different locations, with high fatality even in healthy young adults[1,2]. Predisposition to severe outcomes from Ebola virus infection is unclear from an epidemiological perspective. This suggests the possibility that genetic factors may determine outcome and severity of Ebola virus disease.

Clinical variability in EVD outcome

There is growing evidence from clinical and epidemiological studies that exposure to EVD results in different outcomes, from contacts remaining uninfected, to severe, fatal infection. Among survivors, some recover fully, others have virus persistence and others develop long term complications such as uveitis or arthritis[3,4,5,6,7,8,9] (Figure 1).

Figure 1. Outcome of EVD exposure



Genetic control of susceptibility and outcome

Genetic factors play an important role in response to infection[10]. Genome wide association (GWAS), and study of familial cases have identified genetic variants underlying susceptibility, severity and outcome of many infections. For example, the principal applicants group [redacted] identified Mendelian defects causing susceptibility to mycobacterial infection[11,12]. They have also shown that common variants in complement factor H, and rare variants in TLR pathway genes determine susceptibility to meningococcal disease[13,14,15] while other genes control severity[16,17]. Rare Mendelian variants have been shown to cause viral diseases including Herpes simplex encephalitis[18], EBV lymphoproliferative syndrome[19] and fungal infections[20]. Genetic control of outcome is demonstrated in variations in CCR5 and HLA C which control progression of HIV infection[21,22,23], a finding which contributed to the development of novel therapeutic agents[21,23].

Genetic determination of Ebola

A study from the Ugandan Ebola epidemic (2000-2001) found different HLA types were associated with either fatal or non-fatal disease outcomes[24]. In mouse models genetic variation in the Niemann Pick Receptor gene (NPC1) determines outcome of EVD. Homozygous NPC1 receptor-null mice do not develop disease and heterozygous mice have mild disease[25,26,27]. These studies support the hypothesis that genetic factors control infection and outcome from EVD.

Rationale for study

EVD is currently a poorly understood disease with no proven therapeutics. Those potentially beneficial therapeutics remain limited in their accessibility to under resourced

nations. Identification of the genetic basis for susceptibility, severity and outcome of EVD is likely to identify the key immunological pathways required to resist or contain infection[28], providing further information on EVD pathogenesis and on new targets for therapy, enabling development of potential or novel therapeutics.

References

1 [redacted]

[redacted]

9 [redacted]

15 [redacted]

4 **Summary of the research with overview of methods**
Include numbers, sample types and volumes (see notes on page 3) required with statistical

justification if appropriate. Include methods of analysis distinguishing those requiring CL4 containment and those that can take place at lower containment level.

This study has been peer reviewed by the [redacted] during the process of applying for funding. It was approved and funding was granted by the [redacted] through a fellowship awarded to [redacted] At the same time it was also peer reviewed by the [redacted] during the process of applying for funding, [redacted] unding was provided prior to the final outcome of the [redacted] application process, but initial review was favourable. Ethical approval has been obtained for the study from both the Sierra Leone Ethics and Scientific Review Committee of the government of Sierra Leone [redacted] attached).

The proposed project will investigate the hypothesis that host genetic factors are key determinants of different clinical outcomes of EVD.

Aims/objectives:
To establish the following participant cohorts in Sierra Leone:

1. Patients with EVD from whom DNA and outcome data - fatal (A) or survivor (B) are available.
2. EVD contacts, classified as asymptomatically infected (C) or uninfected (D) by antibody testing.
3. Healthy controls (E) with no direct EVD exposure and antibody negative.
4. EVD survivors with complete recovery (F); viral persistence (G); "Ebola survivor syndrome"- chronic symptoms (H).

[redacted]

Experimental approach/methods *(please also see study flow chart in appendix 1 of attached study protocol)*

Pilot work

During a 1 year [redacted] fellowship awarded t- [redacted] planning visits were undertaken to establish collaborations with [redacted] and the Ministry of Health and Sanitation of Sierra Leone [redacted] A collaboration has also been established with [redacted] for access to an oral fluid antibody assay for use in cohorts C, D and E below. Further collaborations have been established with colleagues [redacted] working with the Ministry of Health in Guinea [redacted] working with the Ministry of Health and Social Welfare in Liberia. These collaborations have enabled access to samples from Guinea and Liberia in the same cohorts as listed below. This enables the potential to undertake a three-nation genome wide study of EVD in West Africa, in which the findings in each country are cross validated in the others and meta-analysed to ensure definitive findings.

Aim 1. Establish DNA cohort of fatal and non-fatal EVD cases (A, B)
If the EVD samples now at [redacted] available to us, we will extract DNA from each viable, EVD positive sample. We undertook a pilot study that demonstrated adequate DNA yield (approximately 1.5-6ug DNA/200ul sample) from both plasma and oral swab supernatant, processed and stored in a similar fashion to the samples from EVD patients at PHE labs in Sierra Leone. [redacted]

[redacted] Even if this number should fall short, there will be an adequate sample size through pooling of samples with our collaborators.

Aim 2. Establish a cohort of EVD household contacts and classify as infected (C) or uninfected (D) based on antibody status
In collaboration with [redacted] we will identify 250 EVD cases from the Port Loko, Tonkolilli and Bombali regions. Based on an estimated 4 household contacts per case, we will survey 1000 EVD household contacts with the assistance of 6 local field workers (including health care workers who are Ebola survivors), who will receive training in the conduct of the study and will be supervised by [redacted] Using a structured questionnaire developed by our collaborator [redacted] we will categorise contacts by risk of exposure, and collect oral swabs and saliva for DNA recovery and anti-EVD antibody (ELISA assay available through collaboration with [redacted] Household contacts will be categorised as asymptomatically infected or uninfected based on antibody results.

Aim 3. Establish a cohort of healthy controls with no direct EVD exposure (E)
Control households (250 households/1,000 individuals) will be recruited in the same community as EVD households. Oral fluid and saliva samples will be obtained for anti-EVD antibody and genetic analysis.

Aim 4. Establish a DNA cohort of EVD survivors - those with complete recovery (F); viral persistence (G); 'survivor syndrome' (H)
In collaboration with [redacted] survivor coordinator, [redacted] and local survivor clinics, we will obtain saliva samples from approximately 500 survivors – 100 fully recovered; 150 with viral persistence; 100 with uveitis; 100 with chronic arthritis; 50 with neurological complications. Depending on the circumstances this cohort will likely increase in size to better power the study.

Aim 5. Genomic studies
Following the fieldwork, genomic analyses will be conducted in collaboration with [redacted] Genome-wide genotyping will be undertaken in either [redacted] the survivor and fatal cohorts (A and B); as well as cohorts C-H. There are clear advantages in undertaking the genotyping for the cohorts from Liberia, Guinea and Sierra Leone at a single laboratory to avoid any batch effects and ensure consistency. Discussion between the three studies is ongoing to choose the preferred laboratory for the genotyping. Subsequently the analysis of the genotyping data will be undertaken using the bioinformatics pipeline at [redacted] and in parallel at the [redacted] Exome sequencing will also be undertaken on 50 survivors; fatal cases; asymptomatic infected cases; exposed but uninfected cases and defined survivor phenotypes and analysed using the [redacted] Exome pipeline developed for the current [redacted] coordinating investigator.

Table 1. Inclusion and exclusion criteria for study cohorts

Cohort	Inclusion criteria	Exclusion criteria
A (EVD – fatal)	Available sample & outcome data	No sample/data
B (EVD – survivor)	Available sample & outcome data	No sample/data

C (contact – asymptomatic infection)	Related household member to sample from cohort A or B	Enrolled in vaccine trial Acutely unwell at recruitment
D (contact – uninfected)	Related household member to sample from cohort A or B	Enrolled in vaccine trial Acutely unwell at recruitment
E (unexposed, uninfected)	Unexposed community member from same community as households in cohort C or D	Enrolled in vaccine trial Acutely unwell at recruitment
F (Survivor – complete recovery)	Survivor with no on going problems since discharge from treatment facility	Viral persistence in bodily fluids >1 month after negative blood PCR
G (Survivor – viral persistence)	Survivor with viral persistence in bodily fluid >1 month after negative blood PCR	On going sequelae – uveitis, arthritis etc. after discharge from treatment facility
H (Survivor – 'Survivor syndrome')	Survivor with on going sequelae – uveitis, arthritis etc. after discharge from treatment facility	History of arthritis, uveitis, neurological problems prior to EVD infection

Samples requested

We would like to access nucleic acid from all samples of patients who were positive for Ebola virus disease at any stage, both deceased patients and survivors, to form cohorts A and B. Where there is more than one sample per patient, we require nucleic acid from only one of the samples. Ideally this would be from a sample also positive for virus or where a viral sequence is already available, however optimal DNA yield is the priority and if another sample from the patient is more likely to provide a better yield, this would be preferential. Ideally we would like to receive either:

1. A minimum of 2ug of DNA from any extracted nucleic acid from all confirmed EVD cases

OR

2. Alternately if sample is being provided instead of extracted nucleic acid, we would like to obtain 220ul of sample (plasma and supernatant from swabs), from all confirmed EVD cases, from which to extract DNA.

If option 2 is granted, we would request access to the use of the [redacted] this has been discussed with [redacted] DNA can be extracted from both the plasma or oral swabs/supernatant from oral swabs currently held in the biobank [redacted] As mentioned above, a preliminary feasibility study conducted at [redacted] on healthy volunteers and mimicking the manner in which [redacted] samples were handled in Sierra Leone, demonstrated an adequate yield of DNA from sample aliquots of 200ul.

Sample handling/processing

If aliquots of samples are provided by PHE-MOHS from cohorts A and B, access will be

requested to the CL4 at [redacted] for extraction of nucleic acid. This has been discussed with [redacted] Should aliquots of nucleic acid be provided these are deemed to be safe and will be managed at CL2 level. Samples from cohorts C-H will be handled at CL2 level. While it is perhaps unclear whether virus is still circulating in some communities and individuals in Sierra Leone, all participants in cohorts C-H will be screened prior to sample collection and if acutely unwell or febrile, will be excluded from the study. The antibody ELISA's will be performed at the [redacted] laboratory, which is currently overseen by [redacted] If it is not possible to perform the ELISAs in this laboratory, they will be frozen and we will consider shipping these samples to [redacted] for analysis. The saliva samples will be extracted at [redacted] due to a high throughout automated extractor being required for the number of samples. Prior to shipment these samples will be heat treated at 60 °C for 1 hour to ensure any possible virus has been neutralised (see biosafety form – Bio1). It is not possible to heat treat the oral fluid antibody samples as this causes degradation of the antibody. The shipment of samples is approved in our ethics agreement from the Sierra Leone Ethics and Scientific Review committee, but will require further permissions in Sierra Leone prior to shipment.

Statistical analysis and power calculations

The study is powered based on the primary analysis where the minimum number of samples may be available. The primary analysis will compare cohorts A and B (fatal and non-fatal EVD). With a sample size of 600 cases (fatal) and 600 controls (survivors) and based on an assumed high prevalence of a genetic variant leading to poor outcome (mortality rates in excess of 50% suggest risk allele frequency of 0.4-0.5), the primary outcome measure is powered at >80% for a genome wide significant association ($5x10^{-8}$) to detect a risk allele with an allelic risk ratio of 1.4.

The request for approximately 2000 samples (50-60% fatal samples) is based on the assumption of an adequate DNA yield from 600 samples in each cohort. The study is powered on the assumption of 600 samples in each cohort. Less will under-power the study, however more will improve the power of the study. This calculation is based on doing a genome wide study of samples in Sierra Leone alone. With the addition of samples from Guinea and Liberia through our collaborators, our collaborative three country study will be amply powered to identify genome wide significant associations.

Data management and sharing

The project will generate considerable new data firstly on the clinical and epidemiological aspects of EVD, including the clinical spectrum and outcome, spread of infection to household contacts and comparison of infection rates within households with cases and community control households (as measured by anti-EVD antibody prevalence). Secondly it will generate a large amount of genomic data on single nucleotide polymorphisms and genetic variations within EVD cases, families and controls in Sierra Leone. It will also generate data on potential host-virus interactions demonstrating potential targets for the application of novel therapeutics.

A REDCap database will be established to house the clinical information on patients and controls in Sierra Leone, and track samples and genomic and antibody data as it is generated. The database will meet current standards of security, and ensure anonymity of all patient information. The database will be accessible by web to all collaborators with appropriate security clearance. The data will be backed up, and stored with security procedures consistent with other major international studies conducted at [redacted] GWAS and exome sequence data will be stored using procedures developed by the bioinformatics team at [redacted] A principle underlying all data generated is that it will be shared and accessible to our partners and colleagues in Sierra Leone and other

institutions. [redacted] developing a number of user friendly tools to enable genomic data to be interrogated easily through external web access to the data via a secure password.

We will publish the findings in open access journals, ensuring anonymity of participant information, and make all sequencing data accessible to investigators worldwide. All data generated will be open access from the time of publication. GWAS and microarray data will be deposited in institutional and specialised subject repositories. The data will be of value widely, particularly for researchers and trainees in host genomics and informatics. In keeping with the aims of the [redacted] the data will be available to our African partners for use in other training projects for African scientists and clinicians.

Significance
This study enables the possibility of undertaking a three-nation genome wide study of EVD in West Africa. Such a study has not been possible to undertake previously for any filoviral haemorrhagic fever due to limited sample sizes. The information that can be obtained through a genome wide study and exome sequencing of the cohorts above will advance our understanding of the pathogenesis of EVD, but also likely of other viral haemorrhagic fevers. It will potentially lead to improved clinical care and the development of novel therapeutics based on gene targets identified for severity of disease. This understanding and these therapies will be beneficial for any nation likely to suffer from epidemics of viral haemorrhagic fevers, including Sierra Leone.

5 Database variables

Please list any metadata associated with the samples required for selection of samples or interpretation of results.

We require samples from patients who were known EVD cases with positive PCR results, these can be either original or follow up, as long as the patient had a definite positive test for EVD during their illness. Ideally we would like to access any data, which accompanies the samples as we must control for ethnicity within our control groups (C,D and E). In particular we require the sample type and sample laboratory ID number to attempt to trace outcome data for each sample and also the community from which the patient arrived from. To assist with this it would be useful to know the facility from where the patient was referred and the laboratory of origin. Data on age and gender as well as symptom onset, date of hospitalisation, date tested, clinical chemistry results and viral load would also be useful (when available) in refining and separating our cohort groups and determining disease severity and confounding factors for outcome. Any data provided by PHE will be entered into the secure REDCap database we are developing.

6 Resource required and available

Please give details of funding available (or being applied for) to carry out the proposed research. Include any resource available or required for the processing of samples at PHE's CL4 laboratory

This project is fully funded by the [redacted] in the form of a [redacted] fellowship to [redacted] (award letter attached). The costs cover all field and laboratory work as well as data analysis. [redacted] through [redacted] is sponsoring the study and providing the necessary laboratory resources. The cost of extracting DNA from samples at [redacted] laboratory is included in the [redacted] funding. We are happy to extract or to receive extracted samples from [redacted] if samples are being extracted by [redacted] or both viral sequencing and host sequencing at the same time we would like to discuss methodology to ensure optimisation of host DNA extraction. Ideally if we are extracting samples at [redacted] we would like to access an automated DNA extractor

due to the number of samples to be extracted. We would also require access to the CL4 for initial processing of the samples.

Funding of the genomic sequencing and analysis has been committed to by [redacted] (letter attached).

7 Biosafety and Biosecurity

All applications must be accompanied by a biosafety and biosecurity assessment and risk management plan.

The project has been carefully screened by the biosafety department at [redacted] and given approval to proceed. Attached is the biosafety risk assessment provided to the biosafety department at [redacted]

As mentioned above all samples will be handled at CL2 level except for those samples for cohorts A and B. The handling of the samples from cohorts A and B will depend on what is provided by [redacted] If an aliquot of sample is provided this will be managed at CL4 level, which will require access to the CL4 laboratory at [redacted] for DNA extraction, as previously discussed with [redacted] If an aliquot of nucleic acid is provided this will be handled at CL2 level and stored at [redacted] until it is shipped for sequencing. All access to the CL2 laboratories and freezers at [redacted] is through secure swipe card access.

Notes

Categories of samples in Biobank

N.B. Samples are known to be EBOV positive or negative by PCR, some may be further categorised as below:

1. EBOV positive diagnostic samples (ie 1st positive sample per patient)
2. EBOV PCR positive samples taken during monitoring of treatment of those who subsequently died
3. EBOV PCR positive samples taken during monitoring of treatment of those who subsequently survived
4. EBOV PCR negative samples of those who subsequently tested positive
5. EBOV PCR negative samples of those previously testing positive
6. EBOV PCR negative patients or contacts

Database

Some further information is associated with some samples in the biobank. This includes laboratory of origin, laboratory ID number, name, age, gender, specimen type, original or

follow up sample, facility from where the patient was referred, date of hospitalisation, symptom onset, date tested, clinical chemistry results, viral load and Ebola test result.