Institution: University of Bristol		
Unit of Assessment: 1) Clinical Medicine		
Title of case study: Supercomputer-based data analysis used to develop effective Ebola containment strategies and healthcare guidance for survivors		
Period when the underpinning research was undertaken: 2011 - 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
David Matthews	Reader in Molecular Virology	2002 – present
Damian Steer	Data Research Software Engineer	2006 – present
Callum Wright	Senior HPC Systems Administrator	2006 - present
Period when the claimed impact occurred: 2014 – 2018		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact

Supercomputer-based analysis pipelines for nucleic acid sequencing data, developed at the University of Bristol, informed containment policy directed at controlling the Ebola outbreak (2013-2016) in West Africa. It was used to track real-time molecular evolution of the virus and inform effective deployment of control measures and therapeutic strategies by the World Health Organisation, Public Health England, the US Centre for Disease Control and Prevention and the government of Guinea. Successful containment reduced the human and economic costs of illness and death. Retrospective comparison of interventions and virus evolution has validated the response and informed confident management of recent and future outbreaks. Follow-up research has also informed guidance on survivor care to prevent sexual transmission. The methodology has been used in subsequent Ebola outbreaks, and the processes are currently being further developed and used for other pathogens, including COVID-19.

2. Underpinning research

Advances in genome sequencing technology have required parallel advances in computational analysis to handle, and effectively examine, the vast data sets generated. Research carried out at the University of Bristol (UoB) has been at the forefront of developing novel integrated high-throughput methods, to address these issues with a particular interest and application in respiratory viruses.

Early methods of protein identification from sequencing data relied on comparisons with a high-quality reference list. Research led by Matthews, established a novel computational technique, Proteomics informed by transcriptomics (PIT) analysis, to circumvent this. PIT collects RNA-sequencing and proteomics data simultaneously, and then uses the transcriptomic data to refine and inform the proteomic analysis [1]. In a follow-up study, PIT was used to demonstrate the potential application in a comparison of the response of bat and human cell lines to a pathogenic zoonotic virus [2]. Bats are a major reservoir of emerging infectious diseases while often remaining asymptomatic; the public health threat of bat zoonoses is evident with Ebola, Marburg, SARS, MERS, Hendra, Nipih, and COVID-19. This approach enabled the comparison of molecular responses between the species which could shed light on host pathogenesis [2].

In 2012, Public Health England (PHE) identified the need for rapid and reliable computational

pipelines to support their work in monitoring and controlling viral infections. Matthews, in collaboration with the University of Liverpool, developed these processes for PHE. A key outcome, led by UoB, was a computational pipeline that allowed accurate assembly of viral genomes and identification of nucleotide polymorphisms, from high-throughput short read Illumina sequence data. This enabled monitoring of viral evolution, for example, alterations in coat proteins which potentially render vaccines or passive antibody transfer treatments ineffective. The study used sequential passage of Ebola virus in guinea pigs. Initially the virus is not pathogenic to guinea pigs at all but after as little as five passages the virus becomes highly lethal to them. The computational pipeline developed by UoB, allowed identification of changes in the virus genome that could be associated with increasing virus pathogenicity [3]. The pipeline was then applied in real time to samples gathered during the 2013-2016 Ebola outbreak in West Africa to trace the temporal and spatial genetic evolution of the virus [4].

The next advance was to meet the demand for a tool which could provide results quickly and be deployed close to the outbreak. The MinION DNA sequencer, made by Oxford Nanopore Technologies, allowed the European Mobile Laboratories team to create a fully functioning laboratory that could collate and provide genome sequence information in rapid time. Towards the end of the outbreak in 2014/2015, MinION sequencing [4] was evaluated by PHE to demonstrate the technology was robust and would enable rapid contact tracing. Comparison of samples tested with the MinION, with the computational analysis method developed by Matthews gave PHE the confidence MinION could replace conventional lab sequencing. The time from sample to sequence was reduced from several weeks to about a day [5].

This technology was subsequently used to study sporadic flare-ups of Ebola that could have restarted the outbreak and to understand persistence of the virus in survivors. In 2016, a study of viral load in seminal fluid from a cohort of men discharged from Ebola treatment units, revealed the virus was present in the semen for nine months after they had recovered and were PCR negative for the virus in their blood [6].

3. References to the research

1) Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, **Matthews DA.** (2012). *De novo* derivation of proteomes from transcriptomes for transcript and protein identification. *Nature Methods,* 9, 1207–1211. DOI:10.1038/nmeth.2227

2) Wynne JW, Shiell BJ, Marsh GA, Boyd V, Harper JA, Heesom K, Monaghan P, Zhou P, Payne J, Klein R, Todd S, Mok L, Green D, Bingham J, Tachedjian M, Baker ML, **Matthews D,** Wang L-F. (2014). Proteomics informed by transcriptomics reveals Hendra virus sensitizes bat cells to TRAIL-mediated apoptosis. *Genome Biology,* 15, 532. DOI:10.1186/s13059-014-0532-x

3) Dowall SD, **Matthews DA,** García-Dorival I, Taylor I, Kenny J, Hertz-Fowler C, Hall N, Corbin-Lickfett K, Empig C, Schlunegger K, Barr JN, Carroll MW, Hewson R, Hiscox JA. (2014). Elucidating variations in the nucleotide sequence of Ebola virus associated with increasing pathogenicity. *Genome Biology*, 15, 540. DOI:10.1186/s13059-014-0540-x

4) Carroll MW, **Matthews DA** *et al.* (2015). Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature,* 524, 97-101. DOI:10.1038/nature14594

5) Quick *et al.* (**Matthews DA** co-author) (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature,* 530, 228-232. DOI:10.1038/nature16996

6) Sissoko *et al.* (**Matthews DA** co-author) (2017). Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. *Lancet Global Health,* 5(1): e80-e88. DOI:10.1016/S2214-109X(16)30243-1

4. Details of the impact

Due to the high mortality rate, potential transmission from person-to-person contact and the lack of approved vaccines or anti-viral therapies, Ebola virus is classified as a hazard group 4 pathogen. UoB research informed the international response to the 2013-2016 outbreak in West Africa, subsequent survivor care and public health guidance, and preparedness for future epidemics.

Informing the international response to Ebola and saving lives

In September 2014, the US Center for Disease Control and Prevention (CDC) estimated that without informed interventions or changes in community behaviour, there would be between 550,000 and 1.4 million Ebola cases in West Africa by January 2015. With a fatality rate of 25% to 90%, this would have led to many deaths.

During 2014 Matthews worked with PHE and the European Mobile Laboratory team deployed to Guinea at the epicentre of the outbreak, providing rapid yes/no diagnostic back up to medical teams on the ground. The Deputy Director of PHE's National Infection Service confirmed the importance of the ground-breaking use of real-time mass sequencing; *'Dr Matthews contributions to the sequencing work and speed of transmission mapping was pioneering – the CDC were also sequencing diagnostic viral samples of Ebola and were able to process 70 samples in 8 months. Dr Matthews evaluated 180 samples in 10 months'* [A].

Understanding of the rate of evolutionary change and transmission routes [4], provided significant extra confidence to the government of Guinea and the World Health Organisation (WHO) that the control strategies being put in place were the right ones to use. *'The targeted vaccination program deployed against Ebola was with the confidence that the virus was not aggressively evolving, that transmission was occurring from human-to-human rather than multiple cases of zoonotic origin, and for the first time the response action could be confirmed against a real-time picture of how the virus was behaving'* [A]. Evidence from Guinea that the virus had not substantially mutated suggested the current vaccine and antibody-based treatments would protect against all versions of the virus that were circulating, '*Dr Matthews' sequencing data showing no significant changes in the viral genomes as the epidemic spread gave confidence to WHO in employing the ring vaccination strategy'* [A]. This allowed the trial to go ahead which confirmed the efficacy and effectiveness of the candidate vaccine [B] and informed the expansion of the vaccination programme to Sierra Leone.

In total, 11,310 deaths out of 28,616 cases were reported as of June 2016 by WHO, significantly lower than had been originally estimated.

Ground-breaking UoB mass sequencing established during the Ebola virus outbreak (2013-2015) [3, 4], combined with validation of MinION, which allows real-time and robust genomic sequencing [5], underpinned the WHO Phase 3 surveillance strategy guidelines for Ebola response [C], and set an exemplar for future epidemics [D].

Informing Ebola survivor care and public health guidance

When a new outbreak of Ebola was declared in 2018 in the Democratic Republic of Congo, the MinION device [5] was available to rapidly confirm the pathogen and identify the strain [E]. Insights into the extended time survivors may carry the virus resulted in changes to the guidance given to men who had recovered from Ebola. Based on the evidence [including 6], a WHO

international meeting proposed 7 revisions to the WHO interim advice on preventing sexual transmission of the Ebola virus [F]. In the UK, the Advisory Committee on Dangerous Pathogens used the evidence, which was that Ebola viral RNA had been detected in seminal fluid up to 565 days after infection, to recommend a lifetime ban on any sperm or egg donation from known Ebola virus survivors, and deferral from treatment or donation from those who are known to have been in an area of active Ebola virus transmission [G].

Informing future preparedness

The ability to confirm response actions against a real-time picture has *'set the standard for all future epidemic responses, most recently the COVID-19 pandemic'* [A]. The role of innovative real-time whole genome sequencing methods to generate actionable information, with specific reference to success during the Ebola outbreak [2, 3], was highlighted by the UK Chief Medical Officer's 2016 report [H]. This report made recommendations to policy makers to improve coordination, evaluation and sharing of data to deliver the best health outcomes. With particular reference to experience gained from the Ebola epidemic (Chapter 9 – Case Study 4), the report suggested that *'This will require international standardisation of approaches to data generation and sharing'* [H]. Recommendation 16 states: *'I recommend that DH [Department of Health] works with international partners and health authorities to support systems, mechanisms and rules for rapid sharing of research and health intelligence data especially where it facilitates control of epidemic and pandemic disease threats'* [H].

As an established low cost and rapid turnaround time technology, MinION is now widely used in genomic surveillance of pathogens including Zika [D], Lassa fever [Ii], MERS-CoV [Iii], yellow fever [Iiii] and SARS-CoV-2 [J].

Zika virus, which is present in Africa, Asia, and America, was declared a Public Health Emergency of International Concern by WHO in 2016. Following the validation of MinION in the Ebola outbreak, it was again deployed in the Zika in Brazil Real-Time Analysis (ZiBRA) project, a UK-Brazil dating sharing consortium with the objective to increase genomic surveillance of Zika and understand the effect of the geographical spread of the different virus lineages on Zika-associated neurological conditions [D].

Adoption of the Chief Medical Officers recommendations [F] meant the UK was prepared to sequence the SARS-CoV-2 virus in early 2020. The COVID-19 Genomics UK Consortium [J] has performed and made publicly available more mass whole genome sequencing of SARS-CoV-2 than any other nation. The Deputy Director of the UK's National Infection service reported that *'the UK has sequenced 45% of the SARS-CoV-2 global sequencing total – the largest real-time viral genomic dataset since the 2013-2016 West African Ebola outbreak'* [A] and that *'This SARS-CoV-2 genomic surveillance is utilised by PHE and NHS Track and Trace to understand the evolution of the virus and inform the UK government's investment in vaccines to control the spread of COVID-19'* [A].

5. Sources to corroborate the impact

[A] PHE (2020). Corroborating statement – Deputy Director, National Infections Service

[B] Henao-Restrepo *et al.* (2017). Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *The Lancet,* 389, 505–518. DOI:10.1016/S0140-6736(16)32621-6

[C] WHO (2015). Guidance: Surveillance strategy during Phase 3 of the Ebola response

[D] Faria *et al.* (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine,* 8: 97. DOI:10.1186/s13073-016-0356-2

[E] Mbala-Kingebeni *et al.* (2019). Rapid Confirmation of the Zaire Ebola Virus in the Outbreak of the Equateur Province in the Democratic Republic of Congo: Implications for Public Health Interventions. *Clinical Infectious Diseases,* 68(2): 330-333. DOI:10.1093/cid/ciy527

[F] WHO (2018). International meeting on persistence of Ebola virus RNA in semen and implications for public health

[G] Advisory Committee on Dangerous Pathogens (ACDP) (2018). Ebola virus disease: ACDP guidance on sperm and egg donation

[H] Davies SC. (2016). Annual Report of the Chief Medical Officer: Generation Genome *Underpinning research [4, 5] cited in: Case Study 4 Implementing sequencing technologies during the West Africa Ebola virus outbreak (Chp.9, p.12).*

[I] (i) Kafetzopoulou *et al*. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science,* 363(6422), 74-77. DOI:10.1126/science.aau9343
(ii) Aljabr *et al*. (pre-publication 2020). Amplicon and metagenomic analysis of MERS-CoV and the microbiome in patients with severe Middle East respiratory syndrome (MERS). *bioRxiv.* DOI:10.1101/2020.11.28.400671
(iii) Faria *et al*. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. S*cience,* 361(6405), 894-899. DOI:10.1126/science.aat7115

[J] (i) Moore *et al.* (pre-publication 2020). Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. *MedRxiv.* DOI:10.1101/2020.03.05.20032011
(ii) COG UK (2020). COVID-19 Genomics UK Consortium