Validating diagnostic markers that may predict the outcome of Ebola virus disease patients

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Jocelyn Ginette Pérez Lazo

October 2020

Declaration

I, Jocelyn Ginette Pérez Lazo, confirm that the work presented in this thesis is, except where indicated, the result of my own work and effort. The material contained in this thesis has not been presented, nor is currently being presented for any other degree or other qualification.

The research was carried out in the Institute of Infection, Veterinary and Ecological Sciences

Abstract

Validating diagnostic markers that may predict the outcome of Ebola virus disease patients

Jocelyn Ginette Pérez Lazo

Ebola virus disease (EVD) is one of the deadliest viral infections in humans, with case fatality rates recorded between 50 to 70%. Despite the ongoing efforts to control disease transmission using a recently licensed vaccine, no licensed treatments and prognostic tests are available that could improve the outcome of acute EVD patients and survivors. In the search for prognostic tools, different markers have been proposed. However, to date Ebola virus (EBOV) viral load measured as Ct value is currently the most reliable predictor of the clinical outcome, particularly in the context of defining survivors (Ct > 22) or fatal cases (Ct < 20). However, the viral load cannot predict accurately the clinical outcome in patients with Ct values between this range since the outcome is approximately equal between survival and a fatal infection. In a previous study 10 genes were identified by machine-learning of transcriptome data from EBOV patients as being potential prognostic markers. This study investigates whether this set of host-based response markers can predict the outcome of infection at the acute phase specially in situations where viral load gives little predictive value. Quantitative reverse transcription PCR (RT-qPCR) assays were developed for each gene transcripts and analysed in 39 clinical samples collected by the European Mobile Laboratory during the 2013-16 West Africa outbreak. A subset of the gene transcripts was used to generate machine learning models with or without EBOV Ct values to predict the outcome of a second group of 64 EVD patient samples using a blind-coded approach. The best discriminating model had a greater overall predictive accuracy (68.4%) compared to a model that only used EBOV Ct as a predictor variable (54.3%). These findings were ratified in a larger sample size, the prediction performance of the models was enhanced when a subset of the gene transcripts was combined with EBOV Ct (90-100%) than EBOV Ct alone (87%). Furthermore, a multiplex RT-qPCR assay for a subset of the gene transcripts was developed. This study proposes a novel approach that highly predicts the risk of Ebola virus disease mortality at the time of diagnosis, which could be used to speed up the triage and clinical management of patients in future outbreaks.

Acknowledgements

I would like to express my infinite gratitude to my supervisors, Prof. Julian Hiscox, Dr. Neil Blake and Prof. Lisa Ng for their guidance, valuable suggestions, and constructive criticism throughout this PhD. I am very grateful to Prof. Julian Hiscox for giving me the opportunity to work in a project about Ebola virus disease since as veterinarian I have a deep interest in zoonotic infectious diseases. The laboratory skills and knowledge I gained in this PhD will be useful in future research projects I would like to conduct in my home country, Peru.

I am also grateful to the National Fund for Scientific, Technological Development and Technological Innovation (FONDECYT)-Peru for selecting me as one of the recipients of the "Doctoral Fellowship Abroad 2016-I" to pursue a PhD in the United Kingdom.

I extend my sincere appreciation to Prof. Stephan Günther, Dr. Romy Kerber from Bernhard Nocht Institute for Tropical Medicine and Prof. Miles Carroll from Public Health England for their valuable support to this project. I am very thankful to the team in Germany for providing the clinical samples used in this study.

Finally, I would like to thank my family and fiancé for their endless love and for encouraging me to be a better researcher and human being.

Table of contents

List of figures

List of tables

List of appendices

Abbreviations

A549	Human Lung Adenocarcinoma Epithelial Cell Line
Ad5	Adenovirus type 5
ALT	Alanine aminotransferase
APO-Fas	Apoptosis antigen-Fas
AST	Aspartate aminotransferase
BOMV	Bombali virus
BDBV	Bundibugyo virus
Bp	Base pairs
BSL-4	Biosafety level 4
BVD	Bundibugyo virus disease
CatB	Cathepsin B
CatL	Cathepsin L
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CDC	US Centers for Disease Control and Prevention
cDNA	complementary DNA
CFR	Case fatality rate
CK	Creatine kinase
cRNA	complementary RNA
CRP	C-reactive protein
Ct	Cycle threshold
CTSL	Cathepsin L1
CV	Coefficient of variation
CXCL10	Interferon gamma-induced protein 10
DCs	Dendritic cells
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin
ddPCR	Droplet digital PCR
DMEM	Dulbecco's modified Eagles medium

DPBS	Dulbecco's' Phosphate Buffered Saline
DRC	Democratic Republic of the Congo
EBOV	Ebola virus
ECACC	European Collection of Authenticated Cell Cultures
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assays
EMA	European Medicines Agency
EMLab	European Mobile Laboratory
ER	Endoplasmic reticulum
ETC	Ebola Treatment Centre
EtBr	Ethidium bromide
EUA	Emergency Use Authorization
EVD	Ebola virus disease
FCS	Fetal calf serum
FDA	Food and Drug Administration
FDR	False discovery rate
GP	Glycoprotein
GrzB	granzyme B
HBEC	Human Brain-derived Endothelial Cells
HEK293	Human Embryonic Kidney 293T cells
HIV	Human Immunodeficiency Virus
HOPX	Homeodomain-only protein
HUJV	Huángjiāo virus
hMGL	human macrophage galactose- and N-acetylgalactosamine-specific C-type lectin
ICD-11	International Classification of Diseases Revision 11
ICTV	International Committee on Taxonomy of Viruses
IFAT	Indirect fluorescent antibody detection test
IFN	Interferon
IKKε	IκB kinase epsilon
IL	Interleukin

L	Large protein
LDA	Linear discriminat analysis
LLOV	LLoviu virus
LNA	Locked nucleic acid
LOD	Limit of detection
L-SIGN	Liver/lymph node-specific ICAM-3 grabbing non-integrin
MARV	Marburg virus
MBL	Human mannose-binding lectin
MCP	Monocyte chemoattractant
MCS	Multiple cloning site
M-CSF	Macrophage colony-stimulating factor
MIP-1	Macrophage inflammatory protein
MLAV	Měnglà virus
MS4A4A	Membrane-spanning 4-domains subfamily A member 4A
MVA	Modified vaccinia Ankara virus
MVD	Marburg virus disease
NHP	Non-human primates
NIF3L1	NIF3-like protein 1
NP	Nucleoprotein
NPC-1	Cholesterol transporter Niemann-Pick C1
ORFs	Open-reading frames
PACT	Protein kinase R activator
PBMCs	Peripheral Blood Mononuclear cells
PECAM	Platelet endothelial cell adhesion molecule
PLPP3	Phospholipid phosphatase 3
PY-STAT1	Tyrosine-phosphorylated STAT1
RAVV	Ravn virus
RDTs	Rapid diagnostic tests
RESTV	Reston virus
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RT-PCR	Conventional reverse transcription PCR

RT-qPCR	quantitative reverse transcription PCR
rVSV	Recombinant vesicular stomatis virus
sGP	soluble GP
ssGP	small soluble GP
sICAM	soluble intracellular adhesion molecule
SLC25A5A	ADP/ATP translocase 2
ssRNA	single-stranded RNA
SUDV	Sudan virus
SVCAM	soluble vascular cell adhesion molecule
SVD	Sudan virus disease
TAFV	Taï Forest virus
TAM	Tyro3/Axl/Mer
TBK-1	TANK-binding kinase 1
TF	Tissue factor
TGFBI	Transforming growth factor-beta-induced protein ig-h3
TIM	T cell immunoglobulin
TNF	Tumor necrosis factor
TNFR	Tumor necrosis factor receptor
TPA	Tissue plasminogen activator
TTC28	Tetratricopeptide repeat protein 28
TUBG1	Tubulin gamma-1 chain
UTRs	Untranslated regions
VCAM1	Vascular cell adhesion protein 1
VP24	Viral protein 24
VP30	Viral protein 30
VP35	Viral protein 35
VP40	Viral protein 40
vRNA	viral genomic RNA
vWF	von Willebrand factor
WHO	World Health Organization
XILV	Xīlǎng virus

Chapter 1

General Introduction

1.1 Filoviruses, an emerging and re-emerging viral threat

Filoviruses represent an important family of viruses that cause sporadic outbreaks and disease in humans. The first filovirus identified was Marburg virus (MARV) in 1967, which caused an outbreak in laboratory personnel in Marburg, Frankfurt and Belgrade as a result of handling blood, tissues and cell cultures from infected African green monkeys imported from Uganda (Siegert et al., 1967; Slenczka & Klenk, 2007). Infection resulted in a severe haemorrhagic fever with 7 deaths out of 32 cases (Slenczka & Klenk, 2007). Since then, more human pathogenic filoviruses have been discovered including Ebolaviruses such as Ebola virus (EBOV), Sudan virus (SUDV), Taï Forest virus (TAFV), and Bundibugyo virus (BDBV) (Deng et al., 1978; Le Guenno et al., 1995; Towner et al., 2008; World Health Organization, 1978). Other Ebolaviruses such as Reston virus (RESTV), Bombali virus (BOMV) and other Filoviruses such as Lloviu virus (LLOV), Měnglà virus (MLAV), Xīlǎng virus (XILV), and Huángjiāo virus (HUJV) were discovered in different animal hosts but their potential as a zoonotic threat needs to be investigated (Goldstein et al., 2018; Jahrling et al., 1990; Negredo et al., 2011; Shi et al., 2018; Yang et al., 2019).

Among all filoviruses, EBOV, which causes Ebola virus disease (EVD) in humans, has been the cause of most of the filovirus disease outbreaks in recent years, including two of the largest filovirus disease outbreaks recorded so far, the 2013-2016 outbreak in West Africa and the 2018-2020 outbreak in the Democratic Republic of the Congo (DRC). EVD has a high fatality rate, of around 50%, and while new vaccines/antivirals have been developed there has been significant drive to determine factors that influence and can be used to predict the clinical outcome of patients. To date these have not been well defined. EBOV viral load is used as a predictor of EVD outcome mainly at extremes of Ct values, where a low Ct is suggestive of fatal infection and a high Ct is suggestive of survival. Whereas, mid-range Ct values are less accurate predicting the outcome of patients. Defining a reliable prognostic tool that could stratify the mortality risk in EVD patients may help health care workers in the management of patients during the outbreaks (Hartley et al., 2017). Thus, the limited resources available could be directed to those in need of

intensive treatment. The analysis of host response genes differentially expressed during EBOV infection has identified candidate predictive markers of disease outcome (Liu et al., 2017). Future development of such markers as prognostic tools requires further analysis and validation.

1.1.1 Taxonomy and classification

1.1.1.1 Order *Mononegavirales*

The genus *Ebolavirus* from the family *Filoviridae* belongs to the order *Mononegavirales*. The name of this order derives from the Greek adjective *μόνος - monos* (alone or single) that refers to the single-stranded RNA genome, the Latin verb *negare* (to negate) referring to the negative polarity of the genome, and the suffix *virales* denoting a virus order (Kuhn et al., 2010). The order *Mononegavirales* was established in 1991 to accommodate related viruses with non-segmented, linear, single-stranded, negative-sense RNA genomes with the characteristic general gene order 3'- UTR - core proteins genes - envelope protein genes - polymerase gene -5'-UTR (Pringle et al., 1991).

1.1.1.2 Family *Filoviridae*

The name *Filoviridae* derives from the Latin noun *filum* (thread) that refers to the filamentous morphology of virions from this family, and the suffix *viridae* that denotes a virus family (Kuhn et al., 2010). Members of this family produce variously shaped, often filamentous, enveloped virions containing linear non-segmented, negative-sense RNA genomes of 15-19 kb. According to the latest report of the International Committee on Taxonomy of Viruses (ICTV) in 2020, the family Filoviridae comprise 6 genera including *Cuevavirus*, *Dianlovirus*, *Ebolavirus*, *Marburgvirus*, *Striavirus*, *Thamnovirus*, and 11 species (Kuhn et al., 2020) (Table 1.1). Research has mainly been focused in two viral species that are highly pathogenic for humans: EBOV, which is the most prominent specie of the genus *Ebolavirus* and MARV that belong to the genus *Marburgvirus*. In the latest ICTV report, one new genus and one new specie were included. The new genus *Dianlovirus* includes Měnglà

virus, which was discovered in Rousettus bats from China in 2019 and shares 32-54% nucleotide sequence identity with known filoviruses (Yang et al., 2019). A new specie of the genus *Ebolavirus,* Bombali virus was discovered in 2018 in free-tailed insectivorous bat species *Mops condylurus* and *Chaerephon pumilus* in Sierra Leone and Kenya (Forbes et al., 2019; Goldstein et al., 2018). The viral genomes of the five genera show similar architecture and are organized in the following general gene order 3'-NP-VP35-VP40-(GP)-L-5', but some genera contain additional genes.

The host range of filoviruses includes mammalian hosts for *Cuevavirus*, *Ebolavirus* and *Marburgvirus,* and piscine hosts for *Striavirus* and *Thamnovirus* (Table 1.2).

Table 1.1 ICTV-accepted taxonomy of the family Filoviridae as of March 2020.

Genus	**Species**	**Virus name (Abbreviation)**
Cuevavirus	*Lloviu cuevavirus*	Lloviu virus (LLOV)
Dianlovirus	*Měnglà dianlovirus*	Měnglà virus (MLAV)
Ebolavirus	*Bundibugyo ebolavirus*	Bundibugyo virus (BDBV)
	Reston ebolavirus	Reston virus (RESTV)
	Sudan ebolavirus	Sudan virus (SUDV)
	Tai Forest ebolavirus	Taï Forest virus (TAFV)
	Zaire ebolavirus	Ebola virus (EBOV)
	Bombali ebolavirus	Bombali virus (BOMV)
Marburgvirus	*Marburg marburgvirus*	Marburg virus (MARV)
		Ravn virus (RAVV)
Striavirus	*Xilang striavirus*	Xīlǎng virus (XILV)
Thamnovirus	*Huangjiao thamnovirus*	Huángjiāo virus (HUJV)

Table 1.2 Host range and geographic location of members of the family Filoviridae

Genus	Specie	Host range	Geographic location	Reference
Cuevavirus	LLOV	Schreiber´s long-fingered bats (*Miniopterus schreibersii*)	Spain, Northeast Hungary	(Kemenesi et al., 2018; Negredo et al., 2011)
Dianlovirus	MLAV	Rousettus bats	China	(Yang et al., 2019)
Ebolavirus	EBOV, SUDV, RESTV	Frugivorous and insectivorous bats, humans.	Central Africa, China, Philippines, Bangladesh, Singapore and India	(De Nys et al., 2018; Dovih et al., 2019; Jayme et al., 2015; Laing et al., 2018; Leroy et al., 2005; Olival et al., 2013; Taniguchi et al., 2011; Yuan et al., 2012).
	RESTV	Cynomolgus macaques, pigs, humans.	Philippines, China	(Barrette et al., 2009; Hayes et al., 1992; Jahrling et al., 1990; Marsh et al., 2011; Pan et al., 2014; Rollin et al., 1999)
	BOMV	Free-tailed insectivorous bats (*Mops condylurus, Chaerephon pumilus)*	Sierra Leone, Kenya	(Forbes et al., 2019; Goldstein et al., 2018)
Marburgvirus	MARV, RAVV	Egyptian fruit bats (*Rousettus aegyptiacus*), humans.	Uganda, Democratic Republic of Congo, Kenya, Zambia, Gabon, South Africa and Sierra Leone	(Amman et al., 2020; Kajihara et al., 2019; Swanepoel et al., 2007; Towner et al., 2009).
Striavirus	XILV	Striated frogfish (*Antennarius striatus*)	China	(Shi et al., 2018)
Thamnovirus	HUJV	Greenfin horse-faced filefish (*Thamnaconus septentrionalis*)	China	(Shi et al., 2018)

1.1.2 EBOV genome, virion and life cycle

The negative sense ssRNA genome of EBOV is approximately 19kb in size and encodes seven structural proteins including the nucleoprotein NP, polymerase cofactor VP35, matrix protein VP40, glycoprotein GP, transcriptional activator VP30, nucleocapsid-associated protein VP24 and the polymerase L. The GP gene also encodes two non-structural proteins known as soluble GP (sGP) and small soluble GP (ssGP) (Hume & Mühlberger, 2019) (Figure 1.1 Part A). The viral genome is encapsidated by NP and associated with L, VP35 and VP30 to form the ribonucleoprotein (RNP) complex. VP40 and VP24 surround the complex and are wrapped in a lipid envelope derived from the host cell during the virion egress. GP is inserted in the virion membrane and protrudes from the surface. GP consists of two subunits GP_1 and GP_2. (Hoenen et al., 2011)(Figure 1.1 Part B).

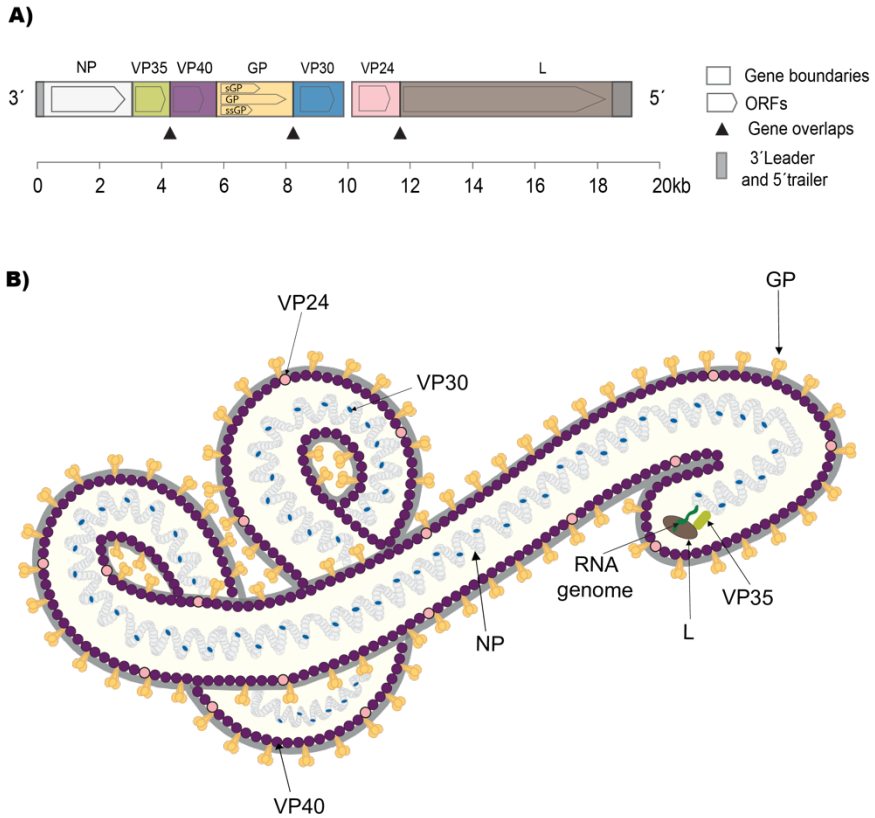


Figure 1.1 Schematic representation of the EBOV genome and particle. *A) Schematic diagram of EBOV genome. The genes are shown as boxes, gene overlaps are marked by black triangles, the bars represent the open-reading frames (ORFs) that encode each viral protein. B) Representation of the viral particle. The RNA genome is encapsidated by NP, VP35, VP30 and L. VP40 and VP24 are matrix proteins. GP trimers are studded in the viral membrane.*

The life cycle of EBOV begins with the cell entry (Figure 1.2). The virion attaches to the host cell surface through interaction of the GP_1 subunit with various host cell factors such as T cell immunoglobulin and mucin domain (TIM) proteins e.g. TIM1; TAM receptor tyrosine kinases (Ax1, Tyro3, Mer); C-type lectin receptors including dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN) and its homolog L-SIGN, human mannose-binding lectin (MBL), and human macrophage galactose- and N-acetylgalactosamine-specific C-type lectin (hMGL) (Alvarez et al., 2002; Brudner et al., 2013; Kondratowicz et al., 2011; Takada et al., 2004). Following attachment, the viral particle is taken up via macropinocytosis into endolysosomes (Aleksandrowicz et al., 2011; Saeed et al., 2010). In the late endosome, GP is cleaved by cathepsin B (CatB) and cathepsin L (CatL) (Chandran et al., 2005). The acidification of the endosome (low pH) induces the interaction of GP_1 with the host protein, cholesterol transporter Niemann-Pick C1 (NPC-1), which is necessary for the fusion of the viral envelope with the endosomal membrane (Carette et al., 2011). The fusion facilitates the release of the RNP complex (mainly RNA + NP protein) into the cytoplasm where the transcription and replication of the negative sense ssRNA genome takes place (Muhlberger, 2007). The primary transcription utilises viral RNP complex proteins brought into the cell, and viral mRNAs are translated into proteins in the cytosol by host ribosomes. The newly synthesized viral proteins facilitate a secondary transcription and genome replication (Hoenen et al., 2019). For ebolaviruses, the proteins NP, VP35, VP30 and L are necessary for transcription (Mühlberger et al., 1999). The replication involves copying the viral genomic RNA (vRNA) into a full-length positive sense antigenome (cRNA) which then serves as a template for the synthesis of progeny negative sense genomes (Muhlberger, 2007). Mature RNP complexes condense into a packaging form and are transported to the plasma membrane, where budding of the viral particle is mediated by VP40 (Jasenosky et al., 2001). GP is transported to the cell surface via the secretory pathway, where it is post-translationally modified by the addition of *O*-linked and *N*-linked glycans and furin-mediated cleavage into the mature GP_1 and GP_2 subunits; then is incorporated into the budding particles (Feldmann et al., 1994; Volchkov et al., 1998)

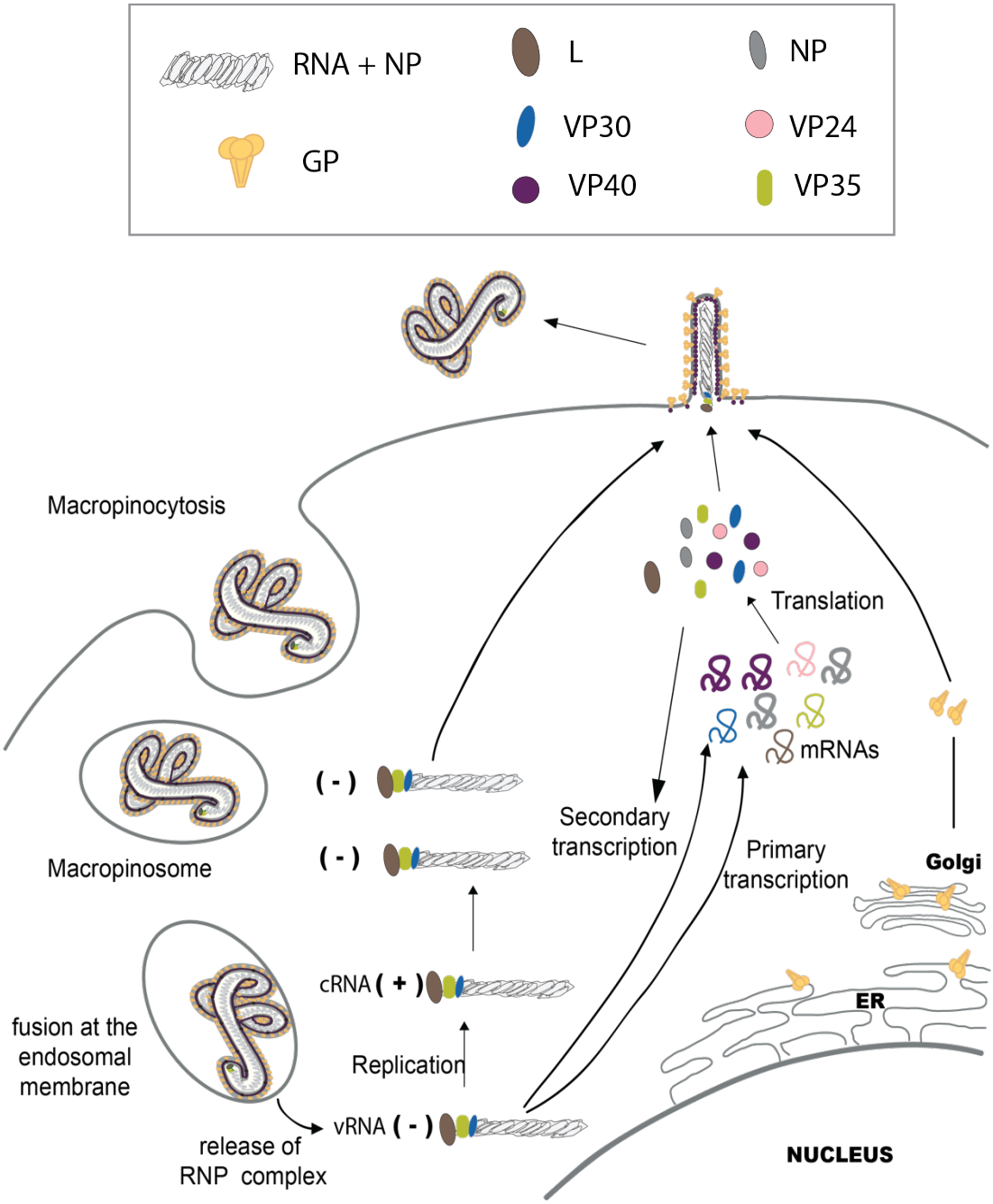


Figure 1.2 EBOV life cycle. *Binding of the viral particle triggers the uptake by macropinocytosis, the fusion with the endosomal membrane results in the release of the RNP complex to the cytosol. The genome is replicated and transcribed into mRNAs, viral proteins are then translated and transported to the plasma membrane for budding of new viral particles. The secretory pathway is composed of the endoplasmic reticulum (ER), Golgi apparatus, lysosome and cell membrane.*

1.2 Ebola Virus Disease (EVD)

According to the latest WHO International Classification of Diseases Revision 11 (ICD-11) of 2018, the term EVD is restricted to the disease only caused by EBOV (specie *Zaire ebolavirus*). For other members of the genus *Ebolavirus* that have caused more than one reported human infection, the diseases are named according the identified viral specie, e.g. Bundibugyo virus cause Bundibugyo virus disease (BVD) and Sudan virus cause Sudan virus disease (SVD). In the case of patients infected with ebolaviruses who show an unusual clinical presentation the term used is "Atypical Ebola disease". For patients infected with ebolaviruses that are not covered by EVD, BVD or SVD the term used is "Other specified Ebola disease" e.g. the disease caused by TAFV. Finally, the term "Ebola disease, virus unspecified" is used for patients who are suspected to be infected with ebolaviruses but the virus has not been identified (Kuhn et al., 2019).

The use of "haemorrhagic fever" was also discouraged by the ICD-11 and should no longer be used for EVD or any disease associated to filoviruses since infected individuals do not always show haemorrhage (Kuhn et al., 2019).

1.2.1 Epidemiology of EVD

EBOV was first discovered in 1976 during an outbreak in the Équateur Province in DRC (formerly Zaire) in Yambuku village which is near the Ebola river (that gave the virus name) (World Health Organization, 1978)(Figure 1.3). A total of 318 cases (280 deaths and 38 survivors) with a high case fatality rate (CFR) of 88% was recorded (Table 1.3). The index case was a man who showed malaria-like symptoms and received an injection of chloroquine at the Yambuku Mission Hospital. Subsequent cases either received contaminated injections at the same hospital or had contact with an infected individual (World Health Organization, 1978).

Since the first EVD outbreak until 2009, sporadic outbreaks have been recorded in the DRC, Gabon and the Republic of the Congo. During this period, the largest

outbreaks recorded were in 1995 in Kikwit and surrounding areas (315 cases and 81% CFR), and the 2007 outbreak in Luebo and Mweke (264 cases, 71% CFR), both in the DRC (Khan et al., 1999; World Health Organization, 2007). Of note, in many of the outbreaks during this period, the index cases were in contact with dead monkeys, suggesting this as the source of the infection. For example, the EVD epidemic in Gabon in 1994-1995 (152 cases, 60% CFR) started in gold mining villages situated near a rainforest where deaths of monkeys were reported (Georges et al., 1999). Likewise, the 1996 outbreaks in Gabon started when people got ill after butchering or had contact with dead chimpanzees (Georges et al., 1999). The 2001-2002 outbreak that occurred in the border of Gabon and the Republic of the Congo recorded 124 cases with 78% CFR, and it was also associated to hunting or contact with dead monkeys (gorillas and chimpanzees) (World Health Organization, 2003).

The most significant outbreak that has occurred to date is the 2013-2016 outbreak in the West Africa in countries such as Guinea, Liberia and Sierra Leone. This is currently the largest EVD outbreak so far recorded, with a total of 28,616 cases and 11,310 deaths (World Health Organization, 2016b). Additionally, 36 cases (with 15 deaths) were recorded in other countries (Mali, Senegal, USA, France, Germany, Netherlands, Nigeria, Norway, Spain, Italy, UK, Switzerland) as infected individuals travelled outside of main epicentre of the outbreak. The average CFR of this outbreak was 39.5%, however there were high fatalities rates reported in Guinea (66.7%) and Liberia (45%) in comparison to Sierra Leone (28%). The index case was a 2 year-old boy who died in December 2013, in Guéckédou, Guinea (Baize et al., 2014). Deep sequencing and phylogenetic analysis of EBOV genomes from patient samples supported a single introduction of the virus into the human population, with subsequent spread of the infection via human-to-human contact (Carroll et al., 2015; Dudas et al., 2017; Tong et al., 2015). The spillover into humans is thought to have occurred as a zoonotic transmission from bats (Baize et al., 2014).

While the outbreak in West Africa was happening, another small outbreak was seen in 2014 in the Équateur Province in western DRC (69 cases, 71% CFR). This outbreak was not genetically nor epidemiologically related to the one in West Africa. Sequence

analysis indicated that the EBOV isolated from patients was related to the one that caused the 1995 outbreak in Kikwit, DRC (Maganga et al., 2014).

More recently additional EVD outbreaks have occurred in the DRC. Two small outbreaks have been observed, first in 2017 in Likati, north DRC (8 cases, 50% CFR), and then in 2018 again in the Équateur Province in western DRC (54 cases, 61% CFR) (World Health Organization, 2018). However, also initiating in 2018 the second largest EVD outbreak so far recorded. This outbreak was first reported in Kivu province in eastern DRC, then spread to Ituri province, and 4 cases were confirmed in Uganda (Ilunga Kalenga et al., 2019). The outbreak lasted for 2 years and was declared over by WHO on June 25, 2020 with a total of 3,470 cases and 66% CFR (World Health Organization, 2020b). For this outbreak, molecular epidemiological analysis also revealed a single spillover of EBOV to the human population and the subsequent human-to-human transmission. The index case reported was an individual from Masimbembe (Mangina health zone), but the source of the spillover is unknown (Mbala-Kingebeni et al., 2019).

A new outbreak, and 11th in the DRC, was announced on 1 June, 2020 after the detection of cases in Mbandaka in Équateur Province of western DRC. As of 10 October, 128 cases including 53 deaths (41.4% CFR) have been recorded (World Health Organization, 2020e). The genetic sequence analysis showed that this is a new spillover. The virus circulating in the Équateur Province in western DRC is different from the one that caused the 2018-2020 outbreak in eastern DRC and from the previous one that hit the same region in 2018. An investigation is ongoing to determine the source of the new outbreak (World Health Organization, 2020a). Details of all EVD outbreaks are shown in Table 1.3

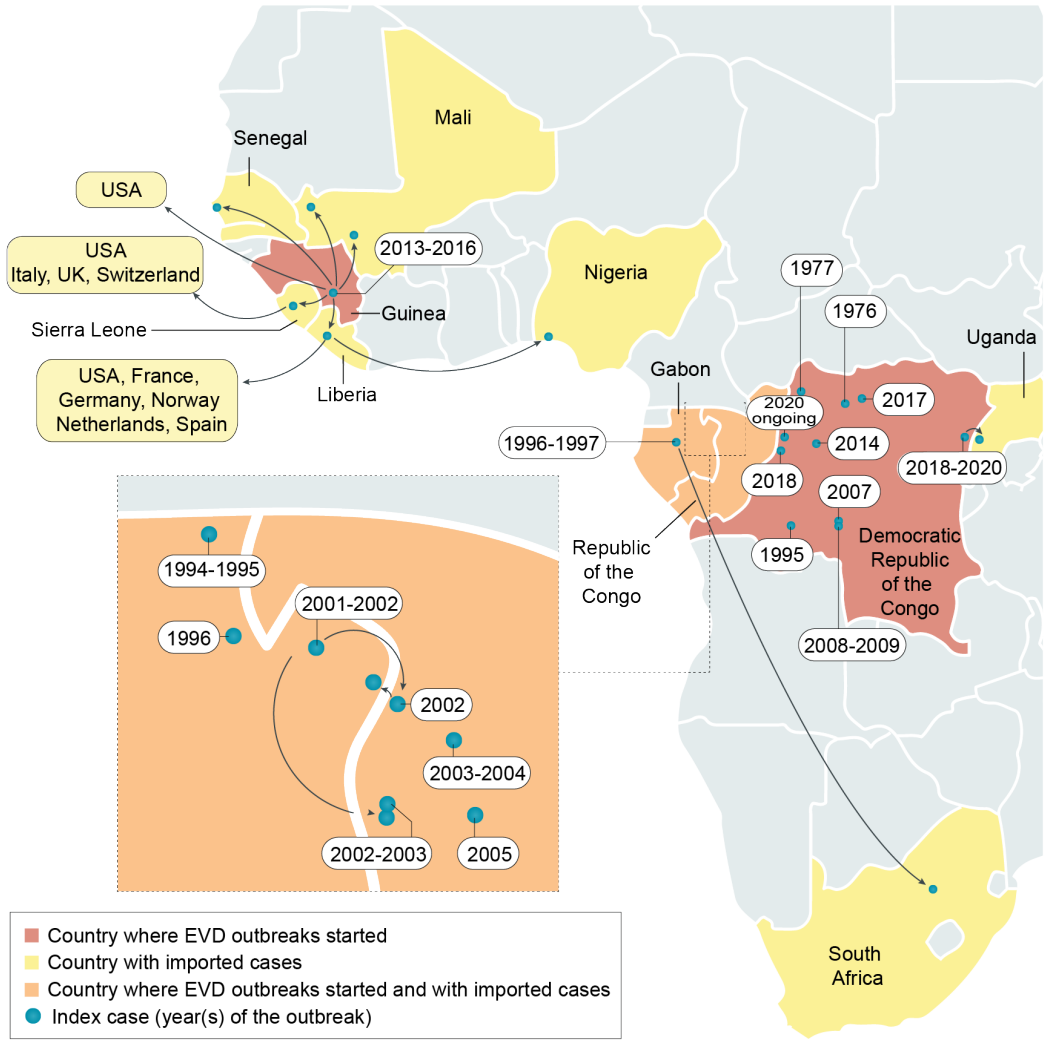


Figure 1.3 Geographic locations and dates of EVD outbreaks up to 2020. *The map shows the location and years of all the reported EVD (specie Zaire ebolavirus) outbreaks since the first outbreak in 1976 to the current outbreak in the Équateur Province of the DRC. Modified and updated from Jacob et al. (2020).*

Table 1.3 Details of EVD outbreaks with number of cases and case fatality rate

Country (year)	Number of cases	Case fatality rate (%)
DRC* (1976)	318	88.1
DRC (1977)	1	100.0
Gabon (1994-1995)	52	61.5
DRC (1995)	315	81.0
Gabon (1996)	31	67.7
Gabon (also exported to South Africa) (1996-1997)	62	74.2
Gabon, Congo (2001-2002)	124	78.2
Congo (2003, Jan-Apr)	143	90
Congo (2003, Nov-Dec)	35	83
Congo (2005)	12	83
DRC (2007)	264	71
DRC (2008-2009)	32	47
Guinea, Sierra Leone and Liberia (also exported to Mali, Senegal, USA, France, Germany, Netherlands, Nigeria, Norway, Spain, Italy, UK, Switzerland) (2013-2016)	28652	39.5
DRC (2014)	69	71
DRC (2017)	8	50
DRC (2018)	54	61
DRC (also exported to Uganda) (2018-2020)	3470	66
DRC (2020, June-present)	128	41.4

*DRC = Democratic Republic of the Congo
Modified and updated from Jacob et al. (2020)

1.2.2 Transmission of EVD

EVD is considered to be a zoonotic infection and outbreaks are initiated either by the spillover from an EBOV-infected natural reservoir to humans, or through an intermediate host such as non-human primates (NHP). Bats (mainly frugivorous species) are considered to be the natural reservoirs of EBOV since EBOV-antibodies and viral nucleotide sequences have been detected in different bat species; however the virus has not yet been isolated from bats in natural conditions (De Nys et al., 2018; Leroy et al., 2005; Olival et al., 2013).

An important evidence that bats are the natural reservoir of some filoviruses is the experimental infection of MARV in Egyptian rousette bats that demonstrated viral replication and horizontal transmission (Schuh et al., 2017). MARV shedding was detected in oral, rectal and urine samples from inoculated bats. In contact bats, MARV RNA was detected in oral and rectal swabs in the early study phase (0-56 days post infection), but viremia and seroconversion only in the late study phase (7-8 months post infection) with a rapid decline of IgG levels (Schuh et al., 2017). Similar to other reported experimental viral infections in bats (i.e Nipah virus) no clinical signs of disease was observed (Halpin et al., 2011). This feature may be related to differences in viral tissue tropism and host immune response in bats that allows them to harbour many viruses. Bats have different interferon and antiviral ISG strategies which can also vary between species (Omatsu et al., 2008; Zhou et al., 2011). They also have mechanisms to control excessive inflammation e.g. suppression of TNF-α expression, a less effective stimulator of interferon genes and an antibody response that potentially promote incomplete viral clearance (Banerjee et al., 2017; Schuh et al., 2019; Turmelle et al., 2010; Xie et al., 2018)

Lately, EBOV genetical material and EBOV-antibodies have been found in a Greater long-fingered bat (*Miniopterus inflatus*) in Liberia. This was the first identification of EBOV in a bat in West Africa, however, it has not yet been confirmed if the strain is the same as that which caused the West Africa outbreak. These results have yet to be published (Kupferschmidt, 2019).

Prior EBOV outbreaks in Central Africa have been associated with deforestation and bushmeat hunting, where human cases have been linked to contact with and consumption of EBOV-infected chimpanzees, gorillas and duikers that may have been directly infected by the natural reservoir (CDC, 2020; Giesecke, 2014; Rouquet et al., 2005).

After the spillover, the virus is transmitted between humans by direct contact or by contact with infected tissues, body fluids or contaminated fomites. Infectious EBOV has been isolated from blood, saliva, urine, semen, breast milk, aqueous humor (Bausch et al., 2007; Vetter et al., 2016). Molecular assays have also detected EBOV RNA in amniotic fluid, cerebrospinal fluid, conjunctiva/tears, skin swabs, vaginal fluid, sweat and stool (Bausch et al., 2007; Keita et al., 2019; Vetter et al., 2016). Sexual contact has been considered as the likely route of transmission in some EVD cases reported during the 2013-2016 West Africa outbreak (Den Boon et al., 2019).

1.2.3 EVD case definition

Currently there is no EVD case definition globally applied (Huizenga et al., 2019). Multiple case definitions were used during the West Africa outbreak, that were variations of the EVD case definition developed by WHO. Guinea followed the WHO suspected EVD case definition and similar versions with slight variations were used by Sierra Leone and Liberia (shown in detail in Table 1.4) (Huizenga et al., 2019; Reaves et al., 2014). However, these case definitions are not 100% sensitive (detecting true EBOV-positive individuals) or 100% specific (detecting true EBOV-negative individuals). Thus, an analysis of the WHO case definition applied in Guinea showed a sensitivity of 68.9% and 49.6% specificity (Hsu et al., 2018).

Table 1.4 Case definition for a suspected case during the 2013-2016 EVD outbreak

WHO suspected EVD case definition*	Sierra Leone national suspected case definition	Liberia suspected case definition
Any person, alive or dead, suffering or having suffered from a sudden onset of high fever and having contact with an individual with suspected, probable or confirmed EVD or a dead or sick animal; OR	Temperature > 38.0 and 3 or more of the following symptoms: abdominal pain, diarrhea, difficulty breathing, difficult swallowing, fatigue, headache, hiccups, loss of appetite, muscle/joint pain, nausea/vomiting, unusual bleeding	History of acute fever and 3 or more of the following symptoms: headache, nausea, vomiting, diarrhea, intense fatigue, abdominal pain, general muscular or joint pain, difficulty swallowing, difficulty breathing, or hiccups; OR
Any person with sudden onset of high fever and at least three of the following symptoms: headache, lethargy, anorexia or loss of appetite, aching muscles or joints, stomach pain, difficulty swallowing vomiting, difficulty breathing, diarrhea, hiccups; OR		Fever with acute clinical symptoms or signs of hemorrhage (epistaxis, conjunctival injection, petechia, hematemesis, hematochezia, melena); OR
Any person with inexplicable bleeding; OR		Death of a person with this history; OR
Any sudden, inexplicable death		Any unexplained death

*Case definition for a suspected case during an EVD outbreak (to be used by mobile teams, health stations and health centres)

1.2.4 Pathogenesis of EVD

EBOV infections in humans are acquired by contact with infected body fluids, tissue or fomites likely through breaks in the skin or mucosal surface (Dowell et al., 1999). Although most human cells can be infected, the primary target cells are mononuclear phagocytes (Kupffer cells, macrophages, microglia) and dendritic cells (DCs) (Bray & Geisbert, 2005; Geisbert et al., 2003). EBOV GP interacts with macrophages and DCs through host factors such as DC-SIGN, Mer, integrin αV and NCP1 (Alvarez et al., 2002; Dahlmann et al., 2015; Rogers et al., 2019). The infected primary target cells can migrate from the initial site of infection and also virus can disseminate to other sites such as regional lymph nodes, liver, spleen, adrenal glands via blood/lymphatics (Geisbert et al., 2003) (Figure 1.4). *In-vivo* studies in NHP showed the rapid spread of

infection to regional lymph nodes, liver, spleen, lung, bone marrow, where high viral titers were detected at only 2-3 days post infection (Geisbert et al., 2003).

In-vivo EBOV infection in animal models and data from EVD in humans indicates a massive release of pro-inflammatory cytokines and chemokines including IL-1β, IL-6, IL-8, IL-10, MCP-1, MIP-1α, and TNF α (Caballero et al., 2016; Wauquier et al., 2010). It has also been shown that there is a strong upregulation of genes related to cytokine/chemokine signalling in response to EBOV infection of macrophages *in-vitro* (Wahl-Jensen et al., 2011). MCP-1, and MIP-1α can recruit more macrophages to the area of infection, allowing EBOV to infect more cells (Geisbert et al., 2003). IL-10 also seems to enhance the virus entry in human macrophages (Stantchev et al., 2019).

EBOV infection induces an aberrant DCs maturation, which supress DCs functions required to induce an adaptive immune response i.e. antigen presentation to T lymphocytes and T- cell proliferation (Bosio et al., 2003). Infection also induces apoptosis of human CD4+ and CD8+ T cells which has been correlated to the clinical outcome of EVD patients. Lymphocyte populations are drastically lower in fatal cases compared to survivors (Wauquier et al., 2010). The apoptosis of T lymphocytes may be the result of an abortive infection in these cells. EBOV is capable of entering T lymphocytes and producing viral RNAs and proteins but is unable to release infectious virus due to the presence of an unknown cellular restriction factor that inhibits viral replication (Younan et al., 2019).

The release of oxygen free radicals like nitric oxide by the infected macrophages trigger the apoptosis of NK cells, tissue damage and loss of vascular integrity (Geisbert et al., 2003). Release of TNFα from infected macrophages can also increase endothelial permeability (Hensley et al., 2002). The breakdown of endothelial barriers causes oedema and hypovolemic shock. Early transcriptional changes in the liver and adrenal gland indicate a downregulation of genes involved in metabolism, coagulation and blood pressure (Jankeel et al., 2020). This is associated with thrombocytopenia and increase plasma levels of liver enzymes.

Later stages of the infection indicate inflammation, cell death, upregulation of genes associated with cardiac injury and vasodilation. This leads to circulatory failure, hypotension, shock and multi-organ damage which is typical of EVD (Jankeel et al., 2020)

1.2.4.1 EBOV immune evasion

EBOV inhibits the intrinsic and innate cell antiviral defence, allowing efficient virus replication in host cells (Jacob et al., 2020). $GP_{1,2,}$ antagonizes the antiviral function of the cellular response factor Tetherin (Kaletsky et al., 2009). The sGP seems to have an anti-inflammatory role reducing the production of pro-inflammatory cytokines by macrophages and protecting the integrity of the endothelium (Bradley et al., 2018; Wahl-Jensen et al., 2005). It also acts as a decoy antigen by binding the antibodies against $GP_{1,2}$, and diverting the humoral response towards sGP (Mohan et al., 2012). The EBOV VP35 protein blocks the induction of type I IFN by binding double-stranded RNA and interacting with cellular proteins such as TANK-binding kinase 1 (TBK-1), IκB kinase epsilon (IKKε), and Protein kinase R activator (PACT) (Luthra et al., 2013; Prins et al., 2009). VP35 also supress microRNA-direct silencing in host cells, this alters the cellular environment and enhances EBOV replication (Zhu et al., 2012). VP24 inhibits IFN-α/β and IFN-γ signalling by impairing the nuclear accumulation of tyrosine-phosphorylated STAT1 (PY-STAT1) (Reid et al., 2006). VP40 may have a role in viral persistence, by inducing apoptosis of host immune cells (Pleet et al., 2017)

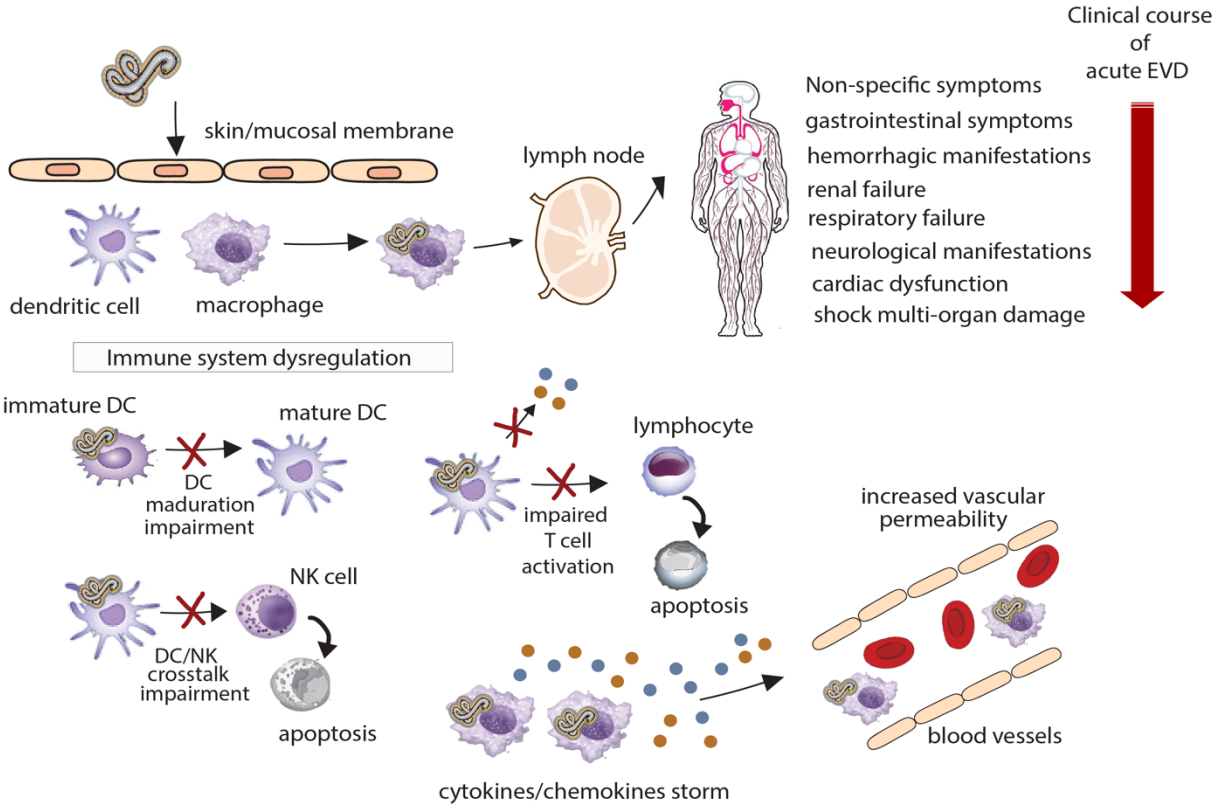


Figure 1.4 Pathogenesis of the Ebola virus disease. *EBOV particles enter the body through injuries or contact via mucosal membranes. The primary targets cells are macrophages and dendritic cells. Infected cells migrate to regional lymph nodes while new virions are produced. The immune system dysregulation leads to systemic infection. EVD typically starts as a non-specific viral syndrome, followed by gastrointestinal symptoms and some patients go into shock and systemic inflammatory response.*

1.3 Clinical manifestations, diagnosis of EVD, and molecular epidemiology during an ongoing outbreak

1.3.1 Clinical manifestations of acute EVD

The mean incubation period (time from infection to symptom onset) varies depending on the route of transmission. The mean incubation period reported following percutaneous transmission is 5.86 ± 1.42 days and following person-to-person transmission or contact with infected animals is 7.34 ± 1.35 days. When all routes of transmission are taken into account the mean incubation period reported is 6.22 ± 1.57 days (Velasquez et al., 2015). Data from the 2013-2016 West Africa outbreak shows that early symptoms of EVD (0-3 days since symptom onset) includes fever (up to 40°C), malaise, fatigue and body aches. During the first seven days the disease progresses to severe gastrointestinal symptoms (nausea, vomiting, diarrhoea). This dehydration causes hypotension and vascular leakage. A fraction of patients show improvement by day 10 of illness. During the terminal phase (generally between 7-12 days) tissue hypoperfusion and vascular leakage lead to multiple organ dysfunction syndrome and shock. Some patients also develop neurologic symptoms (≥ 10 days). Most deaths occurred between 7-12 days of illness (Chertow et al., 2014).

Asymptomatic infection has not been adequately studied, previous serological studies reported very different estimates of the proportion of asymptomatic infection among the contact persons between 3% and 46% (Bower & Glynn, 2017; Dean et al., 2016; Glynn et al., 2017; Leroy et al., 2000). This difference could be attributed to the use of different assays, cross-reactions with non-Ebola viruses, and the heterogeneity of the study populations. A recent retrospective study in a long study population from Guinea found 8.33% EBOV seropositives in 216 paucisymptomatic contact persons and 3.32% seropositives in 1174 asymptomatic individuals (Diallo et al., 2019). The most common symptoms among paucisymptomatic contacts were headache, fatigue and fever (Diallo et al., 2019). Risk factors associated with seropositivity in asymptomatic persons were

participation in burial rituals, contact with blood or vomit of EVD individuals (Diallo et al., 2019).

1.3.2 Diagnosis of acute EVD

The gold standard method to confirm the presence of EBOV is viral isolation in cell culture (mainly Vero E6 African Green monkey kidney cells), but this technique requires biosafety level 4 (BSL-4) containment facilities, and is restricted to research and public health laboratories. Over time, different tests have been developed for the clinical diagnosis of EVD in humans during an outbreak e.g. serological tests that detect host antibodies against the virus, molecular tests that detect viral RNA sequence and rapid antigen detection tests.

Serological tests such as indirect fluorescent antibody detection test (IFAT) and enzyme-linked immunosorbent assays (ELISA) have been used in investigations of past outbreaks to detect current or prior infection with EBOV. However, these tests showed a limited utility for the diagnosis of acute EVD due to the variable onset of the humoral response to EBOV. The IgM response is not evident before 16 days after symptom onset, and during the 2013-2016 West African outbreak IgM or IgG levels have been shown to be low or mostly absent in patients with fatal outcome (Colavita et al., 2019). This suggest that these tests are not adequate for the diagnosis of acute EVD but are a useful tool for seroprevalence studies at the population level (Macneil et al., 2011; Wauquier et al., 2009).

Molecular tests have proved to have better utility in the diagnosis of acute EVD since the viral load rises during the first days of infection. Conventional reverse transcription PCR (RT-PCR) test for EVD, that comprises a PCR amplification of L, GP and NP genes, was the first molecular test developed and evaluated in clinical samples from the 1995, 1996 outbreaks and used for clinical diagnosis in the 2000 SUDV outbreak in Uganda (Formenty et al., 2006; Sanchez et al., 1999; Towner et al., 2004). Real-Time RT-PCR was then developed for EVD diagnosis. This showed greater specificity than conventional RT-PCR due to the use of sequence specific probes and generated faster results (Towner et al., 2004). The viral RNA copy number derived

from real-time RT-PCR was shown to correlate well with quantification of viral load by plaque assays (Towner et al., 2004). Furthermore, low Ct values indicating high viral load were associated with higher mortality (Hunt et al., 2015; Kerber et al., 2019; Towner et al., 2004).

In early EVD outbreaks, the diagnostic testing has been carried out in international reference laboratories and WHO collaborating centres for the diagnosis of viral haemorrhagic fevers (Broadhurst et al., 2016). This caused a delay in getting the test results and the implementation of measures to control the outbreaks. To reduce the time of diagnosis, field molecular diagnostic laboratories started to be deployed in outbreak settings e.g. during the SUDV outbreak in 2000, and EVD outbreaks in 2003, 2007 and 2012 (Grolla et al., 2012; Grolla et al., 2005; World Health Organization, 2001).

During the 2013-2016 West African outbreak nearly 40 field molecular diagnostic laboratories from different countries were deployed. Due to the unprecedented size and spread of the outbreak in West Africa there was an urgent need for development of appropriate diagnostics test that could be used in the outbreak setting. Thus, WHO and the Food and Drug Administration (FDA) evaluated to issue an Emergency Use Authorization (EUA) status to EVD diagnostic tests. The FDA issued EUA to 11 diagnostic tests and WHO to 7 tests. Table 1.5 shows all the approved diagnostic tests for EVD outbreaks. The RealStar Filovirus Screen Real Time RT-PCR kit (Altona Diagnostics, Hamburg, Germany) was the first one that received the EUA authorization. This kit detects a broad range of filoviruses including EBOV, SUDV, TAFV, BDBV, RESTV, MARV, RAVV in plasma samples. The RealStar Ebolavirus kit based on the former kit and optimized for EBOV detection also received EUA (Cnops et al., 2019).

All Ebola cases (probable and suspected) are referred to a designated Ebola Treatment Centre (ETC) or appropriate health care facility where health workers safely collect the appropriate samples. For the diagnosis of EVD in live patients, whole blood samples are collected in EDTA tubes. Oral swabs are collected from deceased

patients or in situations where blood collection is not possible (e.g. children). Swab samples from live patients is not recommended since has low sensitivity in the real time RT-PCR (World Health Organization, 2014, 2015a). The workflow of the field molecular diagnostic laboratories implemented in the ETCs during the West African outbreak included sample inactivation in a glove box, manual RNA extraction and performance of the real time RT-PCR assays (Rieger et al., 2016). Since these procedures could take hours, automated PCR platforms that integrates these steps have been developed to reduce the time to results. One example is the Gene Xpert Ebola assay (Cepheid) that was developed during the West African outbreak and obtained the EUA status from WHO and FDA. This platform detects EBOV NP and GP genes. The performance evaluation of this test in residual diagnostic samples from the 2013-2016 outbreak showed high sensitivity (100%) and specificity (99.5-100%) on whole blood and buccal swab samples, resulting in better performance than the real time RT-PCR used in field laboratories (82% sensitivity and 100% specificity for the RealStar Ebolavirus kit) (Rieger et al., 2016; Semper et al., 2016). This platform has started to be used during the 2018 outbreak in DRC. The implementation of this test in DRC was faster since the platform was already installed in multiple laboratories in the country for the diagnosis of Human Immunodeficiency Virus (HIV) and tuberculosis (Albert et al., 2016; Butler, 2018).

The PCR-based diagnosis requires that patients have to attend a health centre where blood samples can be taken and transported to laboratory. Thus, the average delay from symptom onset to diagnosis in the West African outbreak was about 5 days (WHO Ebola Response Team, 2014). As a transmission can occur during this time, to speed up the diagnosis, point of care rapid diagnostic tests (RDTs) have been developed.

Among the developed RDTs, the OraQuick Ebola Rapid Antigen Test (OraSure Technologies, Pennsylvania, USA) is the only one that currently has both authorizations by WHO and FDA. On October 2019, the FDA granted marketing authorization. This test is an *in-vitro* diagnostic single-use immunoassay for the qualitative detection of antigens (VP40) within the *Ebolavirus* genus and does not

differentiate between viruses. This test is intended to be used in venipuncture or fingerstick whole blood from symptomatic patients consistent with EVD and oral fluids from recently deceased individuals suspected to have died of EVD (Butler, 2018).

1.3.3 Molecular epidemiology

During an outbreak, genome sequencing is important for characterizing the viral agent and determine the evolutionary rate. For the outbreak surveillance, a real-time sequencing approach that generate fast results can help to understand the transmission dynamics and assist the epidemiological response (Carroll, 2019). Thus, a genome surveillance system using MinION (a portable genome sequencer) was designed and implemented in Guinea during the 2013-2016 West African outbreak (Quick et al., 2016). Using a combination of 11 amplicons > 97% of the EBOV genome was reliably amplified, and the bioinformatics approach was validated with clinical samples previously sequenced on Illumina by Carroll et al. (2015) (Table 1.5). Following this system, sequencing results could be obtained within a day and were useful for estimating the substitution rate and provide information about the circulating lineages (Quick et al., 2016).

The real-time genomic surveillance can also support the investigation of transmission chains. Thus, it was very useful in the investigation of the resurgence of EVD in Guinea in February 2016 that started by the sexual transmission of the virus from an EVD survivor who showed EBOV in seminal fluid 531 days after onset of disease (Diallo et al., 2016). It was also valuable for detecting frequent transmissions across the Guinea-Sierra Leone border during the 2013-2016 West Africa outbreak (Quick et al., 2016). Unfortunately, this approach has not been used during the recent outbreaks in the DRC.

Table 1.5 EBOV diagnostic tests for EVD outbreaks

Test (manufacturer)	Test type	Target	Samples	Sensitivity	Specificity	Viruses detected
Rapid viral antigen detection tests						
Dual Path Platform (DPP) Ebola antigen system (Chembio)[a]	Immunochromatographic lateral flow assay	VP40	Venous whole blood (EDTA), venous plasma (EDTA) and capillary fingerstick whole blood	Qualitative; less sensitive than PCR; requires confirmatory testing	From limited data, does not cross-react with other ebolaviruses	EBOV
OraQuick Ebola rapid antigen test (OraSure Technologies)[b,c]	Immunochromatographic lateral flow assay	VP40	Oral fluid and whole blood	97.1% (from oral fluid from deceased individuals); LLOD: 53 ng per ml for whole blood samples and 106 ng per ml for oral fluid	98–100% from venous whole blood samples; 99.1–100% from oral fluid from deceased individuals	BDBV, EBOV and SUDV; does not differentiate between ebolaviruses
SD Q Line Ebola Zaire Ag test (SD Biosensor)[b]	Immuno-precipitation lateral flow assay	$GP_{1,2}$, NP and VP40	Plasma, serum and whole blood	84.9% for whole blood and plasma	99.7% for whole blood and plasma	EBOV
PCR-based tests						
Ebola real-time RT-PCR kit (Liferiver Bio-tech)[b]	Fluorescent real-time RT-PCR	Nucleic acids from ebolaviruses	Serum, body fluid and urine	LLOD: 23.9 copies of viral genome per reaction	Not available	Ebolaviruses
EZ1 test (DOD)[a]	Real-time TaqMan RT-PCR with fluorescent reporter dye detected at each PCR cycle	EBOV nucleic acids	Whole blood and plasma	Qualitative; LLOD: 100–1,000 pfu per ml depending on live or inactivated EBOV isolate and cycler used	100%; no cross-reactivity with other ebolaviruses or marburgviruses	EBOV

FilmArray NGDS BT-E (BioFire)[a]	Fluorescent nested multiplex RT-PCR	EBOV nucleic acids	Whole blood, plasma and serum	LLOD: 1,000 pfu per ml or 4.36×10^3 genome equivalents[d] per ml for live virus	EBOV; no cross-reactivity with other ebolaviruses or marburgviruses	EBOV
FilmArray Biothreat-E (BioFire)[a]	Fluorescent nested multiplex RT-PCR	EBOV nucleic acids	Whole blood and urine	95% detection rate confirms LOD; LOD: 6×10^5 pfu per ml using γ-irradiated EBOV	89–100% using whole blood samples, depending on the study population (Sierra Leone and UK)	EBOV
Idylla Ebola virus triage test (Biocartis)[a]	Qualitative real-time RT-PCR with fluorescent reporter dyes generated upon amplification of cDNA	EBOV and SUDV nucleic acids	Whole blood and urine	97% positive agreement compared with a non-reference standard; LLOD: 465 pfu per ml or 178 copies per ml	100% for EBOV	EBOV and SUDV
LightMix Ebola Zaire TIB MolBio with Lightcycler (Roche)[a]	Qualitative real-time RT-PCR with fluorescent reporter dye detected at each PCR cycle	EBOV nucleic acids	Whole blood	95% positive agreement compared with a non-reference standard; LLOD: 4,781 pfu per ml	100% for EBOV	EBOV
Ebola virus NP real-time RT-PCR (ThermoFisher (CDC))[a]	Qualitative real-time RT-PCR with fluorescent reporter dye detected at each PCR cycle	EBOV *NP* RNA	Whole blood, serum, plasma and urine[e]	99.80%; LLOD: 600–700 $TCID_{50}$ copies per ml	100% for EBOV	EBOV

RealStar Ebolavirus RT-PCR kit (Altona Diagnostics)[a,b]	Real-time RT-PCR with fluorescent dye-labelled probes to detect PCR amplicons	Nucleic acids from ebolaviruses	Plasma	82%; LLOD: 1 pfu per ml	100% for EBOV	Ebolaviruses
EBOV VP40 real-time RT-PCR (CDC)[a]	Real-time RT-PCR with fluorescent dye-labelled probes to detect PCR amplicons	EBOV *VP40* RNA	Whole blood, serum, plasma and urine[e]	LLOD: 400–600 $TCID_{50}$ per ml from whole blood; 250–600 $TCID_{50}$ per ml, depending on body fluid sample and extraction method used	100% for EBOV	EBOV
Gene Xpert Ebola (Cepheid)[a,b]	Real-time RT-PCR with fluorescent signal from probes for quality control	EBOV *NP* and *GP* nucleic acids	Whole blood and oral fluids	100%; LLOD: 232.4 genomic copies per ml	99.5% from whole blood; 100% from oral fluid	EBOV
Real-Time Sequencing						
MinION (Oxford Nanopore Technologies, UK)[f]	Genome sequencing that employs a targeted RT-PCR	EBOV genome	Whole blood, serum, resuspended swab, resuspended urine	A combination of 11 amplicons reliably amplified > 97% of the EBOV genome	The validation process showed no false positive variant calls	EBOV

CDC = US Centers for Disease Control and Prevention, DOD = US Department of Defense, EBOV = Ebola virus, GP = glycoprotein, LLOD = lower limit of detection, LOD = limit of detection, NP =nucleoprotein, pfu =plaque-forming units, RT-PCR = PCR with reverse transcription, SUDV = Sudan virus, TCID50 = 50% tissue culture infective dose (concentration at which 50% of cultured cells are infected with a diluted solution of viral fluid), VP40 = viral protein 40. [a]Emergency use authorization approved by the US Food and Drug Administration (FDA). [b]Emergency use authorization approved by the WHO. [c]Approved by the FDA.; [d]Genome equivalents are calculated by converting the length of a genome in base pairs to micrograms of RNA. [e]Should not be the only specimen tested. [f]Used for real-time genomic surveillance during the West Africa outbreak (Quick et al., 2016). Table adapted from Jacob et al. (2020).

1.4 Therapeutics for EVD

1.4.1 Vaccines

A wide range of potential EVD vaccines were in development during previous decades, however they have not been tested in human trials since few sporadic outbreaks occurred. It was only in response to the unprecedent 2013-2016 EVD epidemic in West Africa that the development and evaluation of these vaccines was urgently accelerated. The vaccines in advance development are listed in Table 1.6. As no vaccine was commercially licensed during the 2013-2016 EVD epidemic, three of these candidate vaccines were used on compassionate basis using a ring-vaccination strategy to attempt to control the outbreak. This involved the vaccination of contacts of infected individuals and contacts of those contacts. The candidate vaccines used all expressed the Zaire EBOV GP and were the recombinant vesicular stomatitis virus-Zaire Ebola virus (rVSV-ZEVOV-GP), a replication-defective chimpanzee adenovirus serotype 3 vector vaccine (ChAd3-EBO-Z), and an adenovirus type 26-vectored vaccine (Ad26.ZEBOV).

The rVSV-ZEVOV-GP is administered as a single dose and has demonstrated to be protective against lethal EBOV challenge in rodents and NHP (Marzi et al., 2015; Wong et al., 2014). It was found to be immunogenic for humans in phase 1 trials (Agnandji et al., 2016). It was also demonstrated to be highly effective in a phase 3 trial during the outbreak in Guinea under the ring vaccination strategy (Henao-Restrepo et al., 2017). This vaccine and ring vaccination strategy was also implemented to control the recent EVD outbreaks in DRC including the vaccination of front-line health care workers. The WHO preliminary analysis of the use of this during the 2018-2020 DRC outbreak estimated an efficacy of 97.5% (95% CI: 95.8-98.5%)(World Health Organization, 2019b). A total of 303,905 people were vaccinated with rVSV-ZEVOV-GP during the 2018-2020 outbreak (World Health Organization, 2020c). In the ongoing outbreak in Équateur Province in DRC, a total of 27,303 people have als been vaccinated with rVSV-ZEVOV-GP (World Health Organization, 2020d).

The vaccine ChAd3-EBO-Z has previously shown to protect NHP against a lethal challenge with EBOV (Stanley et al., 2014). It is administered as a single dose and was found safe and immunogenic for humans in phase 1 trials (Tapia et al., 2016). The results of phase 1 trials also showed that boosting with a modified vaccinia Ankara virus (MVA) expressing Zaire EBOV GP confers long-term protection, a regimen that could be useful in front-line workers who need extended protection (Tapia et al., 2016). This candidate vaccine was evaluated in parallel with the rVSV-ZEVOV-GP during the 2013-2016 West African outbreak as part of the Partnership for Research on Ebola Vaccines I Liberia I (PREVAIL I) as part of a phase 2 clinical trial. Lower antibody response was observed in the recipients of ChAd3-EBO-Z versus rVSV-ZEBOV-GP after 1 month (71% vs 84%) and 12 months of vaccination (64% vs 80%) (Kennedy et al., 2017). A phase 3 clinical trial with this vaccine was not conducted due to the declining cases in Liberia at the time.

Another promising vaccine candidate is the adenovirus type 26-vectored vaccine encoding EBOV GP (Ad26.ZEBOV) boosted by a MVA-vectored vaccine encoding glycoproteins from EBOV, SUDV, MARV and TAFV nucleoprotein (MVA-BN-Filo). It was also found safe and immunogenic in phase 1 trials (Milligan et al., 2016). An immune response as early as 14 days was observed after the primary vaccination with Ad26.ZEBOV, and was elevated after boosting with MVA-BN-Filo (Milligan et al., 2016). This combination confers a durable immunity for at least 360 days (Mutua et al., 2019). During the 2018-2020 DRC outbreak a total of 20,339 people received the primary vaccination and 9,560 of them received the booster (World Health Organization, 2020c).

In November 2019, the European Medicines Agency (EMA) granted the market authorisation to the US pharmaceutical company Merck to produce the rVSV-ZEVOV-GP vaccine, hence the product can be stockpiled and distributed widely. In December 2019, the FDA also approved this vaccine for prevention of EVD. This vaccine also obtained a license in February 2020 for use in four African countries: DRC, Burundi, Ghana and Zambia. In May 2020, EMA also granted the market authorization to Johnson & Johnson company to produce Ad26.ZEBOV and MVA-BN-Filo.

Table 1.6 Overview of the EVD vaccines with advanced development

Vaccine	Leading company or institution and country of origin	Vector	Administration	Ebola component and glycoprotein*	Authorizations
Recombinant VSV-ZEBOV	Merck (USA)	VSV	Single dose	Ebola virus, Kikwit strain (1995)	WHO prequalification, EMA marketing authorization, FDA approval. License in Democratic Republic of the Congo, Burundi, Ghana, Zambia
ChAd3- EBO-Z- with or without MVA-BN-Filo	GlaxoSmithKline (UK) and, for MVA-BN-Filo, Bavarian Nordic (Denmark)	Chimpanzee adenoviral serotype 3 or MVA	Single dose or heterologous prime-boost regimen	Ebola virus, Mayinga strain (1976)	
Ad26.ZEBOV with MVA-BN-Filo	Johnson & Johnson (USA), and MVA-BN-Filo from Bavarian Nordic (Denmark)	Human adenoviral serotype 26 or MVA	Heterologous prime-boost regimen	Ebola virus, Mayinga strain (1976)	EMA marketing authorization
Ad5-ZEBOV	Academy of Military Medical Sciences and CanSino Biologics (China)	Human adenoviral serotype 5	Single dose or homologous prime-boost regimen	Ebola virus, Makona strain (2014)	Licensed in China
GamEvac-Combi	Gamalei Scientific Research Institute of Epidemiology and Microbiology (Russia)	VSV and Ad5-vectored vaccine	Heterologous prime-boost regimen	Ebola virus, Makona strain (2014)	Licensed in Russia

VSV = vesicular stomatitis Indiana virus; MVA = modified vaccinia Ankara virus; Ad5 = human adenoviral serotype 5. * The year the strain (from which the glycoprotein was derived) was isolated is given in brackets. Modified and updated from Malvy et al. (2019).

1.4.2 Treatment

Without a specific treatment , management of EVD patients consist of the provision of supportive care that includes fluid and electrolyte replacement, and treatment for concurrent infections (e.g. antimalarials, antibiotics) (World Health Organization, 2016a). There are no licensed treatments for EVD but some investigational therapeutic agents are administered through the EUA and under compassionate use protocols. These approaches can be classified into directed antiviral therapies or host directed therapies. The directed antiviral therapies include nucleoside and nucleotide analogues e.g. favipiravir and remdesivir (both RNA polymerase inhibitors) and nucleic acid-based drugs e.g. TKM-130803 (a small interfering RNA lipid nanoparticle directed against the gene products encoding the EBOV proteins L and VP35). Host directed therapies or immunotherapeutics are ZMapp (a mixture of three monoclonal antibodies that target the EBOV GP), MAb114 (a single human monoclonal antibody agent) and REGN-EB3 (a mixture of three human IgG1 monoclonal antibodies).

Some of these investigational treatments were tested during the West African EVD epidemic. Most studies during this outbreak were non-randomize trials that did not draw conclusions on the efficacy of these treatments, for instance experimental treatment with favipiravir and TKM-130803 (Dunning et al., 2016; Sissoko et al., 2016). Similarly, the results of a randomized controlled trial evaluating the efficacy of ZMapp did not meet the pre-specified statistical threshold for efficacy (Prevail II Writing Group et al., 2016).

During the 2018-2020 EVD outbreak in the DRC, most patients admitted to the ETUs received ZMapp, MAb114, remdesivir or REGN-EB3 under the Monitored Emergency Use of Unregistered and Investigational Interventions (MEURI) protocol or through the Pamoja Tulinde Maisha “Together Save Lives” (PALM) randomised controlled trial (Mulangu et al., 2019). The PALM study found that patients treated with a combination of standard care and either MAb114 or REGN-EB3 had higher probability of survival than those treated with ZMapp or remdesivir. Despite receiving either

MAb114 or REGN-EB3, 34% of all patients and 63% of patients that had high viral load died. Hence, there was not a successful conclusion of this trial (Mulangu et al., 2019). Other therapeutic interventions such as the transfusion of whole blood, plasma, serum from convalescent individuals have not demonstrated significant improvement in survival of treated patients (van Griensven et al., 2016).

1.5 Predictive markers of EVD outcome

It was during the 2013-2016 EVD pandemic in West Africa that EBOV research intensified. This outbreak led to the accelerated development of vaccines and therapeutics as described above, and the analysis of different cohort of patients from Guinea, Sierra Leone and Liberia gave insights into EVD progression and prognostic factors associated with fatality. Until the success of the vaccines was established, it was necessary to consider prognostic/predictive markers of the clinical outcome and these areas were investigated in parallel to the work on candidate vaccines and therapeutics. The 2013-2016 pandemic showed that prognostic markers are necessary for distinguishing high-risk patients to support the clinical decision of health workers especially in settings with limited resources (Crowe et al., 2016). This is also necessary for the evaluation of the efficacy of therapeutic strategies that address the amelioration of the systemic inflammatory response and organ failure and not just viral replication (Iversen et al., 2020). To date, the proposed prognostic/predictor markers of the EVD outcome are mainly viral load, place of residence, age, gender, clinical markers, laboratory markers and co-infection with malaria.

1.5.1.1 Viral load

High viral load is currently considered the strongest predictor of EVD fatality. It has been associated with a fatal outcome in previous outbreaks of SUDV (Sanchez et al., 2004; Towner et al., 2004). Likewise, studies mainly from the 2013-2016 West African EVD outbreak indicate that patients with low EBOV Ct values (high viral loads) at the time of diagnosis are more likely to die (Hunt et al., 2015; Kerber et al., 2019). A study

of EVD-confirmed patients from Guinea reported that the probability of death increased depending on the level of viremia (viral load) during the first week after symptom onset, from 21% in patients with low viremia (< $10^{4.4}$ copies/ml), to 53% in patients with intermediate viremia (between $10^{4.4}$ and $10^{5.2}$ copies/ml) and 81% in those with high viremia (≥ $10^{5.2}$ copies/ml)(Faye et al., 2015). Viremia remains higher in fatal cases than survivors from day 2 to day 13 after the onset of symptoms. The peak of viremia for survivors is reached earlier than in fatal cases (day 5 versus day 7) and then declines, whereas patients who died maintained a high viral load until death (Kerber et al., 2019; Lanini et al., 2015; McElroy et al., 2016).

Currently, there is no standard Ct value to be used as a prognostic marker, since the Ct values obtained depends on the diagnostic assay used and differ between studies (Crowe et al., 2016). Thus, in different studies and locations the Ct values linked to fatal cases versus survivors was variable (Crowe et al., 2016; Fitzpatrick et al., 2015; Haaskjold et al., 2016). The Ct values can also vary depending on the type of sample used for diagnosis. For example, the median Ct values obtained from swabs samples of people that died in the community was 3.4 Ct units higher than the admission Ct values obtained from blood samples of fatal hospital cases (Kerber et al., 2016). Most studies during the 2013-2016 West Africa outbreak used a semi-quantitative RT-PCR that does not include a standard curve to determine the viral load. Hence, the measured Ct values might not be universally applicable and the viral load is only an approximation. Besides, the Ct values obtained in survivors and fatal cases can be within the same range (between 22 to 25) making difficult the correct classification of patients in each risk group (Crowe et al., 2016; Fitzpatrick et al., 2015; Haaskjold et al., 2016).

1.5.1.2 Place of residence

Another risk factor linked to a fatal outcome is the place of residence. Patients living in rural areas in Africa are more likely to die compared to individuals living in the capital area where there is more access to healthcare, more availability of medical supplies and healthcare personnel (Furuse et al., 2017). Patients that acquired EVD

during the West Africa outbreak and were repatriated to USA or Europe had also better access to medical care in their home countries (McElroy et al., 2016)

1.5.1.3 Age

During the West African outbreak, EVD cases have been reported in people aged less than 1 to more than 100 years, with the incidence of cases lower in children younger than 16 years compared to adults (WHO Ebola Response Team et al., 2015). However, most studies indicate that young children and elderly are the groups at higher risk of death due to EVD. In a large study of cases from Guinea, Liberia and Sierra Leone conducted by the WHO Ebola Response Team a higher CFR was reported in children 4 years old or younger compared to children between 10 to 15 years of age (WHO Ebola Response Team et al., 2015). The high probability of death in children younger than 5 years of age was also found when different cohorts of patients from the same outbreak were analysed (Faye et al., 2015; Kerber et al., 2016; Kerber et al., 2019). Likewise, adults older than 45 years of age was also a high-risk group during the 2013-2016 West Africa outbreak (Crowe et al., 2016; Faye et al., 2015; Fitzpatrick et al., 2015; Kerber et al., 2016; Qin et al., 2015; Schieffelin et al., 2014; Tong et al., 2015; WHO Ebola Response Team et al., 2015). This finding was similar to previous data obtained in past EVD outbreaks, where high fatality rates in older adults was also reported (Sadek et al., 1999).

Possible explanations for the high fatality rates observed in young children and older adults or the elderly may be an insufficient immune response, an immature immune system in children, comorbidities in elderly individuals, and both groups may be more vulnerable to severe dehydration. For these patients, an early diagnostic testing and appropriate treatment is necessary (Furuse et al., 2017).

1.5.1.4 Gender

Studies in small cohorts of EVD patients from West Africa and one study from the 1995 EVD outbreak in DRC did not find a significant difference in fatality rate between males and females (Crowe et al., 2016; Li et al., 2016; Qin et al., 2015; Sadek et al., 1999). However, the WHO Ebola Response Team evaluated a larger sample size of probable and confirmed EVD cases from Guinea, Liberia and Sierra Leone during the West Africa outbreak and found that female patients were significantly less likely to die than male patients (63.0% versus 67.1%) (WHO Ebola Response Team, 2016). This difference was also observed when only confirmed EVD cases were evaluated. It was thought that there was not a significant difference in the number of exposures to a sick person between female or male (WHO Ebola Response Team, 2016). The reason for this difference is still unknown and needs to be studied, but a possible explanation would be the divergent effects of androgens and oestrogens in the immune system (Kovats, 2015; Trigunaite et al., 2015). In terms of risk of community infection, it was reported that male patients have spent longer time (~ 12 hours) in the community than females while they were symptomatic (WHO Ebola Response Team, 2016). This information may be important to consider during the application of outbreak control measures.

1.5.1.5 Clinical markers

During the West African outbreak, the symptoms more associated with fatality were diarrhoea, confusion, conjunctivitis, weakness, dizziness, fever, headache, extreme fatigue, difficulty swallowing, vomiting, mental symptoms, bleeding and loss appetite (Fitzpatrick et al., 2015; Li et al., 2016; Qin et al., 2015; Schieffelin et al., 2014; Tong et al., 2015). Diarrhoea was the most frequent symptom reported in fatal cases from all these studies, 90% of patients from a cohort in Sierra Leone who had diarrhoea died (Schieffelin et al., 2014). Large volumes of watery diarrhoea (5 or more litres/day) could persist up to 7 days (Chertow et al., 2014). Bleeding manifestations such as bleeding from mouth, puncture sites and bloody faeces are also reported in fatal cases although it is not frequent. Severe pain that require opiate treatment was

also seen in moribund patients (Haaskjold et al., 2016). A prompt support treatment may reduce the likelihood of death in patients with these symptoms. To reduce the massive gastrointestinal loss, a treatment with oral anti-emetics, anti-diarrheal therapy and rehydration with oral electrolyte solutions seems beneficial in patients who are hypovolemic, not in shock and are able to self-care. For hypovolemic patients who are not in shock but are unable to self-care, intravenous fluid and electrolyte therapy is necessary (Chertow et al., 2014; Li et al., 2016)

1.5.1.6 Laboratory markers

High levels of creatinine, blood urea nitrogen (BUN), alanine aminotransferase (ALT), aspartate aminotransferase (AST), C-reactive protein (CRP), low calcium levels, hyperkalaemia, have been associated with fatal outcome (Hunt et al., 2015; Kerber et al., 2019; Schieffelin et al., 2014). Fatal cases also report higher haemoglobin concentration, haematocrit, median white cell count, lymphocyte count, granulocyte count and lower platelet count than survivors (Hunt et al., 2015). The increased elevation of BUN, creatinine and potassium values over time in patients who died indicates an aggravated renal function (Kerber et al., 2019; Schieffelin et al., 2014). Indeed, acute kidney injury was more common in fatal cases than survivors, and severe acute kidney injury is considered a strong risk factor for mortality but not only restricted to late disease stages since it was also seen in patients with a stage 1 disease (Hunt et al., 2015). ALT and AST were 5 times the normal level in patients more likely to die. In EVD patients with severe hepatitis, AST was raised up to > 15 times the normal values (Hunt et al., 2015). Although it was noted that ALT, AST and creatine kinase (CK) tend to normalize before death (Kerber et al., 2019). The levels of these liver enzymes along with creatinine, carbon dioxide, BUN, became normal in patients who recovered from the disease (Schieffelin et al., 2014). Individuals that received favipiravir treatment and had a fatal outcome presented alterations in the glucose levels (hyperglycemia, hypoglycaemia), hyperkalemia, increase levels of creatinine, BUN, bilirubin, CRP, CK, lower levels of calcium and albumin, and increase ALT, AST values. These alterations were similar but less prominent in untreated EVD patients who survived (Kerber et al., 2019).

1.5.1.7 Co-infection with Malaria

An increase in mortality has been observed in patients co-infected with *Plasmodium falciparum*, a parasitic protozoan that causes malaria in humans. A higher proportion of co-infection with this parasite was reported in children younger than 15 years of age in Guinea (Kerber et al., 2016; Kerber et al., 2019). This co-infection increased the CFR in young children (5-14 years old) by 20%, probably due to an insufficient immune response (Kerber et al., 2016). Similarly, another study reported that mortality was higher in patients co-infected with malaria than infected only with EBOV (Waxman et al., 2017). In contrast, other studies suggest that *Plasmodium* infection can confer protection to EBOV infection due to the induction of M1 polarization of tissue macrophages through the production of IFNγ (Rogers et al., 2020).

1.5.1.8 Host genetics

Individual people can differ in their susceptibilities to different viral infections since it is determined by host genes, the environment and their interactions. Genome-wide association studies (GWAS) link the genetic variation with viral infections. For instance, GWAS detected polymorphisms in humans that affect the phenotype upon infection with different viruses e.g. HIV-1, Hepatitis C, Dengue, Influenza A virus (Dang et al., 2014; van Manen et al., 2012; Zignego et al., 2014; Zuniga et al., 2012). In experimental EBOV studies, polymorphisms in the NPC1 receptor could reduce the affinity of EBOV-NPC1 interaction in some bat cell lines and recombinant inbred mice infected with mouse-adapted EBOV showed distinct disease phenotypes ranging from resistant to lethal disease likely due to distinct Tie2 polymorphisms (Ng et al., 2015; Rasmussen et al., 2014). In EVD patients, a study investigated the relationship between killer cell immunoglobulin-like receptors (KIRs) - human leukocyte antigen (HLA) combinations and the clinical outcome (Wawina-Bokalanga et al., 2021). KIR AA haplotype was more frequent in fatal cases, while KIR BB haplotype in survivors. It was also found that KIR genes 2DL2, 2L5 and 2DS4-003 were associated with disease outcome. From them, KIR 2DS4-003 and 2DL5 genes were significantly more

common in fatal cases than survivors and 2DL2 was suggested to play a protective role in EVD (Wawina-Bokalanga et al., 2021). These genetic differences may be involved in the overwhelmed innate immune response observed in EVD; however, further GWAS in larger cohorts of patients are recommended.

1.5.1.9 Host markers

Different potential host markers that correlate with EVD severity have been identified by statistical or machine learning analysis of data obtained from testing clinical samples using luminex-based assays or RNA sequencing. Patients with severe EVD have increased levels of biomarkers linked to innate immune dysregulation with prolonged elevation of proinflammatory cytokines and chemokines. These included IL-6, IL-8, IL-10, interferon gamma-induced protein 10 (CXCL10/IP-10), macrophage colony-stimulating factor (M-CSF), monocyte chemoattractant protein (MCP)-1, MCP-2, interleukin-1 receptor antagonist (IL-1RA), macrophage inflammatory protein (MIP)-1β, MIP-1α, tumor necrosis factor receptor type I (TNFR-I) and TNFα (McElroy et al., 2016). The known roles of such proteins can be correlated to the pathogenesis seen in patients. TNFα may be involved in the endothelial activation and as a consequence the activation of coagulation. Cytokines and chemokines are produced by monocytes or macrophages indicating that antigen-presenting cells are important during EVD. The rise of cytokines level is correlated with the days before death similar to septic shock syndrome (Reynard et al., 2019). The coagulation pathway imbalance is caused by the increase of D-dimer, von Willebrand factor (vWF), tissue factor (TF), thrombomodulin that are release from activated or damaged endothelial cells. High levels of TF indicates clinical co-agulopathy. Although other studies have not found it associated to fatal outcome (McElroy et al., 2014). The endothelial dysfunction is caused by high levels of platelet endothelial cell adhesion molecule (PECAM)-1, P-selectin, soluble intracellular adhesion molecule (sICAM)-1, soluble vascular cell adhesion molecule (sVCAM)-1. High levels of granzyme B (GrzB), fractalkine/CX3CL1 cause lymphocyte cytotoxicity and homing (Kerber et al., 2018; McElroy et al., 2016). High viremia can also stimulate more T-cell activity, and CD8 T cells can contribute to tissue damage (McElroy et al., 2015). Soluble TNF receptor II (TNF-RII), soluble VEGF

receptor 1 (sVEGFR1), tissue plasminogen activator (TPA) also increases in fatal cases, whereas epidermal growth factor (EGF) decreases (Kerber et al., 2018). The transcriptomic analysis of white blood cells in fatal cases also corroborates the upregulation of pathways involved in the innate immune response and apoptosis (Reynard et al., 2019).

Patients with moderate EVD have increased levels of mediators related to immune activation and control such as apoptosis antigen-Fas (APO-Fas), IFN- β, IL-29, IL-5, sFas ligand (sFasL), and TNFR-II. The immune response is controlled by the antiviral interferons (IFN- β, IL-29 or IFN-λ). sFasL may activate apoptosis or inflammatory cells, and TNFRII control the IL-6 mediated inflammation (McElroy et al., 2016). EGF and normal T-cell expressed and secreted (RANTES) also increase in survivors and remain unchanged in fatal cases; the former is a protein involved in gastric protection, and the latter activates and recruits T cells to peripheral tissues (Kerber et al., 2018). In survivors, there is also a decline in the expression of D-dimers, TPA and thrombomodulin indicating control of endothelial integrity and coagulation (Kerber et al., 2018). The expression of GrzB, IL-10, IL-1RA, IL-6, IL-8, IP-10, M-CSF, MCP-1, PECAM-1, sICAM-1, sTNF-RII, sVEGFR1, TNFα also decreases (Kerber et al., 2018). Although these mediators were reported differentially expressed between survivors and fatal cases, they did not show strong power for outcome prediction in the ROC analysis (Kerber et al., 2018).

1.6 Machine learning models for predictive biomarkers of EVD outcome

Artificial intelligence is a discipline within computer science that deals with the simulation of intelligent behaviour in computers and has been significantly improved over the last decades (Lo Piano, 2020). Machine learning is a branch of artificial intelligence that involves the development and application of computer algorithms that improve their own performance through experience (training) (Mitchell, 1997). The machine learning methods follows three stages: algorithm design, learning and testing. These algorithms are autonomous and self-sufficient when perform the learning function demonstrating the capacity to learn from data without the need to

be explicitly programmed (Lo Piano, 2020). It has been applied in different fields such as agriculture, education, finance, medicine and many others (Bahrammirzaee, 2010; Bauer et al., 2019; Beam & Kohane, 2018; Richens et al., 2020). Within the area of genomics machine learning algorithms have been recently applied to identify the most discriminative biomarkers of patient mortality in infectious diseases (Davi et al., 2019; Hu et al., 2020; Jong et al., 2016; Liang et al., 2020; Yan et al., 2020).

In EBOV, few studies have developed machine learning models that could predict the clinical outcome of patients. A combination of different variables such as viral load, age, time to presentation, clinical symptoms and body temperature from EVD patients were used by Colubri et al. (2019) to develop a machine learning model. This model showed 74% predictive accuracy when was externally validated against two independent datasets. Transcriptome data of peripheral blood taken from acute-EVD patients was used by Liu et al. (2017) to identify a set of genes that could discriminate between survivors and fatal cases using different machine learning algorithms. The best model showed 92% predictive accuracy and identified a panel of ten genes that may be potential predictive markers of EVD outcome. These genes were Vascular cell adhesion protein 1 (VCAM1), Homeodomain-only protein (HOPX), Tubulin gamma-1 chain (TUBG1), Phospholipid phosphatase 3 (PLPP3), Membrane-spanning 4-domains subfamily A member 4A (MS4A4A), Transforming growth factor-beta-induced protein ig-h3 (TGFBI), Tetratricopeptide repeat protein 28 (TTC28), NIF3-like protein 1 (NIF3L1), ADP/ATP translocase 2 (SLC25A5), and Cathepsin L1 (CTSL) (Liu et al., 2017). The description of these genes is shown in Table 1.7.

Table 1.7 Details of the 10 genes selected as potential markers to predict the EVD outcome*

Gene/Protein	Function
Vascular cell adhesion protein 1 (VCAM1)	Important in cell-cell recognition. Appears to function in leukocyte endothelial cell adhesion and signal transduction.
Homeodomain-only protein (HOPX)	Atypical homeodomain protein which does not bind DNA and is required to modulate cardiac growth and development. May act as a tumor suppressor.
Tubulin gamma-1 chain (TUBG1)	Tubulin is the major constituent of microtubules. The gamma chain is found at microtubule organizing centers such as the spindle poles or the centrosome. This protein is required for microtubule formation and progression of the cell cycle.
Phospholipid phosphatase 3 (PLPP3)	It is a membrane glycoprotein localized at the cell plasma membrane. Catalyzes the conversion of phosphatidic acid to diacylglycerol. Involved in integrin-mediated cell-cell adhesion in angiogenesis.
Membrane-spanning 4-domains subfamily A member 4A (MS4A4A)	May be involved in signal transduction as a component of a multimeric receptor complex.
Transforming growth factor-beta-induced protein ig-h3 (TGFBI)	Plays a role in cell adhesion. May play a role in cell-collagen interactions.
Tetratricopeptide repeat protein 28 (TTC28)	During mitosis, may be involved in the condensation of spindle midzone microtubules, leading to the formation of midbody.
NIF3-like protein 1 (NIF3L1)	May function as a transcriptional corepressor through its interaction with COPS2, negatively regulating the expression of genes involved in neuronal differentiation.
ADP/ATP translocase 2 (SLC25A5)	Catalyzes the exchange of cytoplasmic ADP with mitochondrial ATP across the mitochondrial inner membrane.
Cathepsin L1 (CTSL)	Important for the overall degradation of proteins in lysosomes. Plays a critical role in normal cellular functions such as protein turnover, antigen processing, bone remodeling. Important protein for cardiac morphology and function

* Information taken from UniPROT Consortium.

1.7 Thesis aims

In the last EVD outbreaks in DRC, the ring vaccination strategy using the recently licensed rVSV-ZEVOV-GP vaccine has shown to be effective for controlling disease transmission. However, there are still no licensed treatments for EVD and no prognostic markers that can support the clinical decision of health workers in the triage of patients e.g., for the evaluation of different treatment strategies. Furthermore, given the difficulties and unlikelihood of widespread EBOV vaccination, the continuing risk of emergence and re-emergence of EBOV in Africa or other filovirus diseases, and the high fatality rates of EVD reported in the recent outbreaks, highlight the importance of identifying biomarkers that could accurately predict the clinical outcome of EVD patients. Previously, potential biomarkers have been proposed as putative predictors of EVD outcome, but they have not been validated in clinical samples or are less accurate than viral load. The viral load, represented as a Ct value, is still the strongest predictor of the outcome particularly in the context of defining survivors (Ct > 22) or fatal cases (Ct < 20). However, the viral load cannot predict accurately the clinical outcome in patients with Ct values between this range since the outcome is approximately equal between survival and a fatal infection. Given that EVD is a disease of the host response, a previous study from our group on patient samples from the 2013-2016 West Africa outbreak suggested that the host response can help delineate the clinical outcome and a set of host markers were identified as potential predictors of the outcome (VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28, NIF3L1, SLC25A5, and CTSL). Thus, this thesis investigates whether this set of host-based response markers can predict the outcome of infection at the acute phase specially in situations where viral load gives little predictive value. To achieve this, the objectives of this thesis are:

- Developing quantitative reverse transcription PCR (RT-qPCR) assays that can measure the transcript abundance of the 10 candidate predictive genes.

- Determine the transcript abundance of each gene by RT-qPCR analysis in blood samples from EVD patients collected by the European Mobile Laboratory (EMLab) in Guinea during the West Africa outbreak.

- Developing machine learning classification models using the transcript abundance of the candidate genes as predictor variables and determine the predictive accuracy of these models.

Chapter 2

Materials and Methods

2.1 Cell culture and Peripheral Blood Mononuclear cells (PBMCs)

Human Embryonic Kidney 293T cells (HEK 293) and Human Lung Adenocarcinoma Epithelial Cell Line (A549) were obtained from the European Collection of Authenticated Cell Cultures (ECACC). Human Brain-derived Endothelial Cells (HBEC) and PBMCs were kindly provided by the Brain Infections Group, University of Liverpool.

293T and A549 cell monolayers were grown under high humidity incubation (37°C, 5% CO_2) with DMEM (Sigma-Aldrich) supplemented with 10% (v/v) heat-inactivated FCS (Sigma-Aldrich), 100 units/ml penicillin (Sigma-Aldrich), 100 µg/ml streptomycin (Sigma-Aldrich) and 2 mM L-glutamine (Sigma-Aldrich). For maintenance, cells were routinely cultured in 10 cm tissue culture dishes and passaged when confluent by discarding the growth media, washing the cells with 6 ml of Dulbecco's' Phosphate Buffered Saline (DPBS) solution (Sigma-Aldrich). The wash solution was removed and 2.5 ml of 1x Trypsin-EDTA solution (0.05% Trypsin, 0.02% EDTA) (Sigma-Aldrich) was added. Cells were then incubated at 37°C for 2-5 minutes to dislodge cells from the surface. Cells were then resuspended by adding 7.5 ml of growth medium to the dish.

2.2 Human samples and Ethics Statement

Five healthy lab donors (aged from 29 to 36 years old) with no previous history of travel to Africa were recruited from the on-going study "Immune responses to infection of white blood cells from healthy people". The research protocol was approved by the Ethics Committee of the Institute of Infection and Global Health, University of Liverpool (RETH000685). Informed consent was obtained from each subject. Venous blood was collected aseptically from each individual in standard ethylenediaminetetraacetic acid (EDTA) tubes (4 ml) by a trained phlebotomist.

Stored RNA samples extracted from whole blood collected from acute EVD patients were kindly provided by Stephan Günther' group from Bernhard Nocht Institute for Tropical Medicine (Hamburg, Germany). These samples were collected by the EMLab

unit in Guinea during the outbreak of Ebola in 2014 and 2015 within the project "Ebola virus disease-correlates of protection, determinants of outcome and clinical management" (EVIDENT, Proposal #666100). The clinical samples were received and tested in three different groups: a) 15 samples used for a pilot study (representing 6 fatal cases and 9 survivors), b) 39 samples with known clinical outcome (20 survivors and 19 fatal cases), c) 64 samples (34 fatal cases and 30 survivors) for which at the time of analysis in this study the clinical outcome was not provided so that the study could be carried out blind.

2.3 RNA extraction

2.3.1 RNA extraction from human cell lines and PBMCs

Total RNA extraction including DNase treatment was carried out using RNeasy Plus Mini Kit (Qiagen). For the human cell lines, 600 µl of buffer RLT Plus (β-Mercaptoethanol added) was used to disrupt the pellet cells (1 x 10^7 cells). The next steps followed the manufacturer's protocol for human cell lines with the addition of 80 µl DNase incubation mix (10 µl DNase I stock solution mixed with 70 µl Buffer RDD) (Qiagen) applied directly to the RNeasy Mini spin column and incubated at room temperature for 15 min before the washing steps.

For the PBMCs pellet sample (2.2 x 10^7 cells), 600 µl of buffer RLT (supplemented with β-ME) was added to the cells, and further steps followed the protocol for isolation of PBMCs with RNeasy Mini kit (Qiagen). Likewise, 80 µl DNase incubation mix was also added before the washing steps.

In both cases, RNA was eluted in 30 µl nuclease free water (Qiagen) and the concentration was measured using NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). The integrity of the RNA was verified by electrophoresis in a 1.5% agarose gel (Invitrogen), prepared with 0.5X TBE buffer (GeneFlow); run at 100V for 20 min and visualized with GeneFlash UV transilluminator (Syngene, Cambridge, UK).

2.3.2 RNA extraction from control and clinical samples

Blood samples from healthy donors were processed immediately after collection in a biological safety cabinet following the protocol of RNA extraction from human whole blood using the QIAamp RNA Blood Mini Kit (Qiagen).

Clinical samples were handled at the European Mobile Laboratory (EMLab) unit deployed in Guinea. The EBOV in the blood samples was first inactivated by adding 560 µl of Buffer AVL (Qiagen) and 560 µl of 100% Ethanol while still in a glove box. After inactivation, samples were handled on the bench following appropriate standard laboratory safety precautions (Kerber et al., 2016). Viral RNA was extracted from 50 µl of whole blood collected in EDTA tubes using the QIAamp Viral RNA Mini kit (Qiagen).

The RNA from control and clinical samples was quantified using Qubit RNA HS Assay Kit (Thermo Fisher Scientific), and read by Qubit Fluorometer (Thermo Fisher Scientific), then diluted with RNA storage solution (Thermo Fisher Scientific) to 2 ng/µl to be tested by RT-qPCR.

2.4 Primer design

Primers in this study were designed to target conserved regions of all available transcript sequences visualized from the transcriptome database AceView (mRNAs from GenBank or RefSeq, and single pass cDNA sequences from dbEST and Trace) of each gene of interest. The primers were designed using NCBI/Primer-BLAST and RealTimeDesign™ software (Biosearch Technologies, Novato, USA).

Several default parameters were considered during the primers design (Table 2.1). The specificity of the primers was then checked by NCBI-Primer BLAST tool (http://wwwncbi.nlm.nih.gov/tools/primer-blast/). All primer sets were synthesized by Eurofin Genomics (Ebersberg, Germany).

Table 2.1 Default criteria for designing primers for RT-qPCR assays*

Requisites	Primers
GC content	30-80% (Gene expression: 50-60 %)
Calculated primer Tm	55-60°C; Tm of primers should not differ > 2°C
Runs of identical nucleotides	Maximum 3 (No G bases)
Primer length	15-30 bp
PCR product length	50-150 bp
Primer-dimers, hairpins	Avoid
3' end rule	Maximum two G or C in the last 5 bp

*Table adapted from Rodríguez et al. (2015)

2.5 Probes design

Probes were designed by RealTimeDesign™ (Biosearch Technologies, Novato, USA) according to the parameters indicated in Table 2.2

Table 2.2 Default criteria for designing probes for RT-qPCR*

Requisites	Primers
GC content	30-80% (Gene expression: 35-65%)
Calculated probe Tm	68-70°C (8-10°C above Tm of primers)
Runs of identical nucleotides	Maximum 3 (No G bases)
Probe length	15-30 bp
PCR product length	50-150 bp
Primer-dimers , hairpins	Avoid
Autoquenching	No G on the 5' end
GC ratio	C > G

*Table adapted from Rodríguez et al. (2015)

The probes were dually labelled with fluorescent dye (HEX, TxRd, CY5 or FAM) on the 5' end and compatible black hole quencher (BHQ 1, BHQ 2 or BHQ 3) on the 3' end. Locked nucleic acids (LNA) were included in some probe sequences. The probes were manufactured by Sigma-Aldrich (Dorset, UK)

2.6 Construction of plasmids

The plasmid constructs were designed and obtained by either inserting an amplified PCR product of the target sequence into vectors by TA cloning or by synthesis of recombinant plasmids by GeneArt Gene Synthesis service (Thermo Fisher Scientific).

2.6.1 Reverse transcription

RNA from human cell lines and PBMCs was converted into complementary DNA (cDNA) using RevertAid H Minus Reverse Transcriptase as follows: 1 µg of total RNA was mixed with 0.5 µg Oligo(dT)15 Primer (Promega) in a 0.5 ml PCR tube. The reaction volume was made up to 10 µl with nuclease-free water (Qiagen) and incubated at 70°C for 5 minutes for priming and chilled on ice. To each reaction mix, 4 µl of 5X Reverse Transcription Buffer, 2 µl PCR nucleotide mix (Promega) (a premixed solution containing nucleotides dATP, dCTP, dGTP and dTTP at a concentration of 10 mM in water at pH 7.5), 0.5 µl RNasin Ribonuclease Inhibitor (Promega), and 1.5 µl nuclease free water were added and incubated at 37°C for 5 minutes. The reaction was stopped and 1 µl RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific) was added to the mixture, incubated at 42°C for 1 hour followed by 10 minutes at 70°C to stop the reaction.

A minus Reverse Transcription control "-RT" (containing all the mixture components except the reverse transcriptase) was included to examine genomic DNA contamination. The cDNA was subsequently diluted with 80 µl nuclease-free water (Qiagen) to get a final volume of 100 µl and stored at −20°C until further use.

2.6.2 End-point Polymerase Chain Reaction

The PCR reactions were carried out on a PxE 0.5 thermal cycler (Thermo Fisher Scientific) using the following conditions: 2 µl cDNA dilution, 0.75 µM each primer, 1X Reddy Mix PCR Master Mix (Thermo Fisher Scientific) in 20 µl final reaction volume. The following cycling conditions were used: initial denaturation step at 94°C

for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, elongation at 72°C for 30 sec and a final extension at 72°C for 10 min. PCR products were separated in a 2% agarose gel (Invitrogen) (prepared with 0.5X TAE buffer), run at 100V for 30 min and visualized by ethidium bromide (EtBr) staining (Sigma-Aldrich) with GeneFlash UV transilluminator (Syngene, Cambridge, UK).

PCR products were purified from the agarose gel by excision of bands and using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

2.6.3 Cloning and Sequencing

Two different approaches were used for cloning the target sequences into vectors:

- Cloning into pGEM-T and pGEM-T Easy vectors: Purified PCR products were cloned into pGEM-T and pGEM-T Easy vectors (Promega) which have 3′-T overhangs at the insertion site for the ligation of the PCR product, and T7 and SP6 RNA polymerase promoters for further *in-vitro* transcription Figure 2.1.

  Transformation was performed in competent Escherichia coli strain DH5α cells (Agilent Technologies), which were grown on LB agar plates supplemented with 100 µg/ml ampicillin. Colonies were screened by PCR using 10 µl of a single colony diluted in 50 µl nuclease-free water, 0.25 µM M13 Forward and M13 Reverse primers, 1X Reddy Mix PCR Master Mix (Thermo Fisher Scientific) in a final reaction volume of 20 µl. Cycling conditions were as follow: initial denaturation step at 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec, extension at 72°C for 30 sec and final extension at 72°C for 10 min. PCR products were run on 1.5% agarose gel (Invitrogen) stained with EtBr at 100V for 30 min and visualized using GeneFlash UV transilluminator (Syngene, Cambridge, UK).

Only clones containing the expected insert size were selected for plasmid purification. Plasmids were isolated using Plasmid Maxi Kit (Qiagen), eluted in 300 µl of nuclease free water (Qiagen) and sequenced at Source Bioscience (Nottingham, England). The sequences were then validated by BLAST analysis on NCBI.

- Cloning into pcDNA3.1(+): Plasmid constructs were designed using the freely accessible SnapGene Viewer program. The coding sequence for each gene transcript were inserted into pcDNA3.1(+) downstream the T7 polymerase promoter necessary for *in-vitro* transcription Figure 2.1. The purified plasmids containing the synthesized genes were ordered from GeneArt Gene Synthesis (Thermo Fisher Scientific).

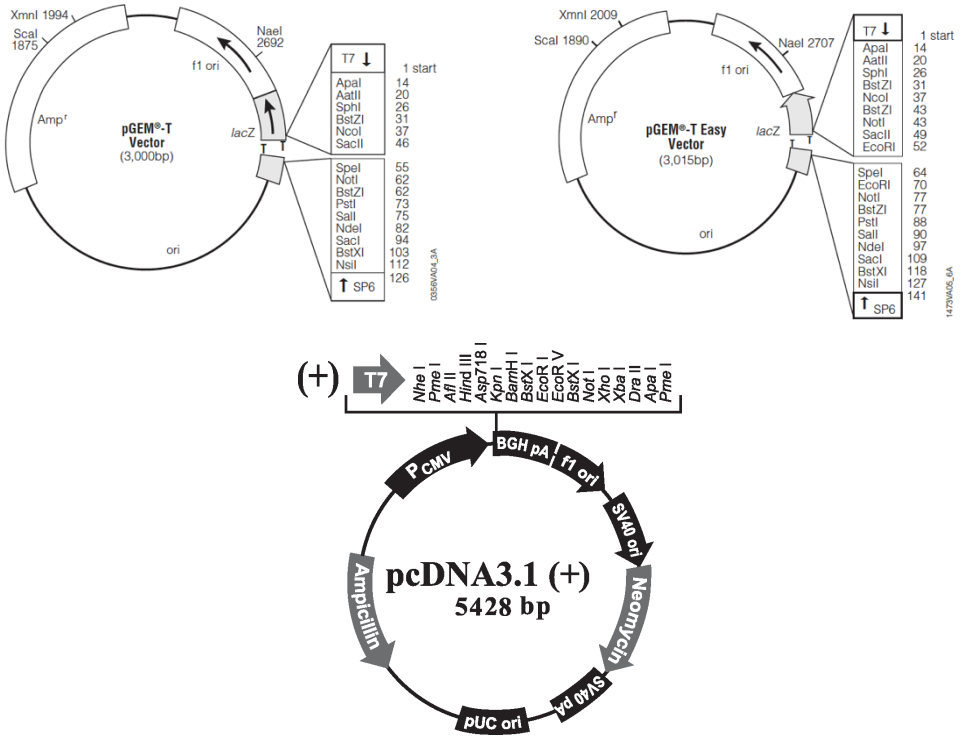


Figure 2.1 Map of the vectors used for cloning the target gene sequences in this study. *A) pGEM-T vector and B) pGEM-T easy vector contains a T7 and SP6 RNA polymerase promoters flanking the multiple cloning site (MCS). C) pcDNA 3.1(+) vector have the MCS in the forward orientation (+) and T7 promoter.*

2.7 Synthesis of RNA transcripts

2.7.1 Linearization of DNA plasmid

Linearization of 5 µg DNA plasmid was performed using the digestion protocol by restriction enzyme SalI-HF, NcoI-HF or NotI-HF (New England Biolabs). The amount of restriction endonuclease was 1 unit/50 µl reaction. The 10X reaction buffer volume, time, and temperature of incubation and to stop the reaction for each enzyme was in accordance with the manufacturer's instructions. The clean-up of linear DNA was performed using the Wizard DNA Clean-Up System (Promega).

2.7.2 *In-vitro* transcription

The transcription reaction of 1 µg linearized DNA was carried out following the standard protocols for *in-vitro* transcription using MEGAshortscript™ T7 Kit (Thermo Fisher Scientific) incubated at 37°C for 4 hours, MAXIscript™SP6 Transcription Kit (Thermo Fisher Scientific) incubated at 37°C for 1 hour or MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific) incubated at 37°C for 4 hours. DNA template from the *in-vitro* transcribed RNA (cRNA) was removed by digestion with 1 µl TURBO DNase I (Thermo Fisher Scientific) at 37°C for 15 minutes, and the purification of transcripts was carried-out using the MEGAclear™ Transcription Clean-Up Kit (Thermo Fisher Scientific). Then, cRNA transcripts were eluted in 100 µl of RNA storage solution (containing 1 mM Sodium Citrate at pH 6.4) (Thermo Fisher Scientific). The quantity and quality of RNA were assessed using NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific).

The length of the purified cRNA transcripts was confirmed by gel electrophoresis analysis. The *in-vitro* transcripts and RiboRuler Low Range RNA Ladder or RiboRuler High Range RNA Ladder (Thermo Fisher Scientific) were first denatured mixing equal volumes of the 2X RNA loading dye and RNA samples, heated at 70°C for 10 min, chilled on ice and then loaded in 2% native agarose gel (Invitrogen) stained with SYBR

safe DNA Gel stain (Thermo Fisher Scientific). Bands were separated at 100V for 45 min and visualized with GeneFlash UV transilluminator (Syngene, Cambridge, UK).

2.8 Primers optimization by Real-Time qPCR

Primer's concentration and optimal annealing temperature was evaluated by SYBR green based Real-Time qPCR. The cDNA (4 µl) obtained from converting RNA extracted from human PBMCs was added as a template to a reaction mixture of iTaq Universal SYBR Green Supermix (2X) (Bio-Rad) and 100 or 200 nM of each primer in a final reaction volume of 12 µl. The reaction was run on the Corbett Research RG-6000 Real Time PCR. The following thermal cycling parameters were used: 1 cycle of polymerase activation, DNA denaturation at 94°C for 3 min, and 40 cycles of denaturation at 94°C 15 sec, and annealing extension at 60°C for 30 sec.

The melt curve analysis was carried out at 65°C-95°C with an increment of 0.5°C at 5 sec/step. The qPCR products were also visualised on a 2% agarose gel (Invitrogen), at 100V for 45 min and visualized by EtBr staining (Sigma-Aldrich) with GeneFlash UV transilluminator (Syngene, Cambridge, UK).

2.9 Generation of a standard curve for RT-qPCR assays

The measurement of RNA copies of the *in-vitro* transcripts was determined using a previous formula developed by Fronhoffs et al. (2002). The formula gives the molecules per µl (N), if the concentration of the cRNA (C) is known in relation to the fragment size (K) multiplied by a factor derived from the molecular mass and the Avogadro constant:

$$N \text{ (molecules per µl)} = \frac{C \text{ (cRNA µg/µl)}}{K \text{ (fragment size/b)}} \; 182.5 \times 10^{13}$$

Ten-fold serial dilutions of RNA transcripts were prepared using RNA storage solution (containing 1 mM sodium citrate, pH 6.5 +/- 0.1 buffer) (Thermo Fisher Scientific) to prevent RNA degradation.

To construct the standard curve, dilutions of the *in-vitro* RNA standards spanning 1x 10^8 to 1 x 10^4 copies/µl or 1x 10^7 to 1 x 10^3 copies/µl were tested in triplicate by the SYBR Green I- based one step RT-qPCR. For the multiplex RT-qPCR, dilutions spanning 1x 10^6 to 1 x 10^2 copies/µl were used in triplicate. The standard curve was plotted as the mean Ct values against the log cRNA copy number.

2.10 SYBR Green I-based RT-qPCR assays

The optimized conditions for the one-step quantitative reverse transcription PCR (RT-qPCR) assay for each gene were as follow: 5 µl of RNA added to a reaction mixture of 1X iTaq™ Universal SYBR® Green reaction mix (Bio-Rad), 1X iScript Reverse transcriptase (Bio-Rad), 100 nM of primer forward, 100 nM of primer reverse in a final reaction volume of 20 µl. The assay was run on the CFX 96™ real time PCR machine (Bio-Rad). The following thermal cycling parameters were used: 1 cycle of RT for 10 min at 50°C followed by 1 cycle of polymerase activation and DNA denaturation at 94°C for 1 minute, and 40 cycles of denaturation at 94°C for 10 minutes and annealing/extension at 60°C for 30 minutes. After amplification, a melting curve analysis program was used to verify the authenticity of the PCR products by their specific melting temperatures (Tm) according to the instrument documentation.

2.11 Multiplex Probe-based RT-qPCR assay

The optimized conditions for the one-step multiplex RT-qPCR were as follows: 5 µl of RNA added to a reaction mixture consist of: 1X Quantifast Multiplex RT-PCR Master mix (without ROX) (Qiagen), 400 nM of each primer forward and reverse, and 200 nM of each probe and 0.25 µl of QuantiFast RT mix (Qiagen) in a final reaction volume of 25 µl. The reaction was run on the Rotor-Gene Q 5plex Platform (Qiagen) with the

following cycling conditions: reverse transcription at 50°C for 20 minutes, PCR initial heat activation at 95°C for 5 minutes, 45 cycles of 2 step-cycling (denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 30 seconds). The data acquisition was performed during the annealing step on 3 channels: yellow, orange, green.

2.12 Performance evaluation of RT-qPCR assays

2.12.1 Linearity of RT-qPCR

The efficiency and dynamic range of the respective singleplex assays were established by amplifying five 10-fold dilutions of RNA transcripts of each gene. Each dilution was amplified in triplicate in three separate runs.

2.12.2 Analytical sensitivity

To determine the limit of detection (LOD) of the RT-qPCR assays; 10-fold dilutions of *in-vitro* transcripts containing 1×10^{11} down to 1×10^{0} RNA copies/µl were tested in duplicate.

2.12.3 Assay specificity

To investigate the specificity of the SYBR Green I-based one step RT-qPCR assay, a melting curve analysis was performed followed by agarose gel analysis of the amplified products. Likewise, amplified products from the multiplex RT-qPCR assay were evaluated by agarose gel analysis.

2.12.4 Assay repeatability

The repeatability (intra- and inter-assay variance) of the SYBR Green I-based one step RT-qPCR assay and the multiplex RT-qPCR were assessed using the 10-fold dilutions of *in-vitro* RNA transcripts. The assays were performed in triplicate for the selected dilutions within a single run to evaluate the intra-assay variance, and in three

independent runs to evaluate the inter-assay variance. The standard deviation (SD), and coefficient of variation (CV) were separately calculated using the mean Ct values.

2.13 Statistical analysis

The transcript abundance of each gene measured by RT-qPCR was described with the mean and the standard error of the mean (SEM). Observations with missing values were excluded from the calculations. The nonparametric Mann-Whitney *U* test was used to determine the statistical difference in the transcript abundance of each gene between survivors and fatal cases using the GraphPad Prism version 8.3.0 (328). Kernel density estimation was used for visualising the distribution of the potential biomarkers using the *ggplot2* package in R open source software version 3.6.0 (2019-04-26)(R Core team, 2019; Wickham, 2016).

2.14 Machine Learning Analysis

All the following analyses were performed using R open source software version 3.6.0 (R Core team, 2019). To measure the importance of each feature (age, EBOV Ct and gene transcripts) the Boruta algorithm was implemented using the *boruta* package (Kursa & Rudnicki, 2010). The *caret* package was used to spot check the performance of a diverse set of machine learning algorithms (linear, non-linear, trees and ensemble algorithms) in the dataset. In this analysis the repeated cross validation option was used to resample the data and $k = 10$ was tested.

Prior to building the classification models such as Support Vector Machine and k-Nearest Neighbours, the subjects with missing values were removed from the analysis.

Accuracy was the metric used to evaluate the classification models and is provided in the *caret* package (Kuhn, 2008). The overall approach for evaluating the accuracy of the predictive models was first splitting the sample dataset in two parts: the 70% training set which was used to build the model and the 30% set (unseen data) used

for evaluating the performance and predictive accuracy of the classifiers. The *caret* package and built-in functions were used to pre-process the training data (using *scale* and *center* transforms), train each model and generate 10-fold cross validation sets to tune model hyperparameters (Kuhn, 2008). The *rose* package was also included during the pre-process of training data to deal with imbalanced classes (unequal number of survivors and fatal cases) (Lunardon et al., 2014). Plots showing the confusion matrix between the EVD outcome predicted by the machine learning models and the actual outcome of the patients were created using the *ggplot2* package.

Chapter 3

Optimization of singleplex quantitative reverse transcription PCR (RT-qPCR) for identification and measurement of prognostic host markers of EVD

3.1 Introduction

While potential benefits of interventions such as drugs and vaccines to reduce the burden of infectious diseases are commonly estimated, the life-saving impact of new diagnostic and prognostic technologies remains overlooked (Ghani et al., 2015). Notwithstanding, diagnostics have proved to be fundamental for containing infectious disease outbreaks. In the past, a poor diagnostic preparedness has contributed to significant delay in the identification of the causative agents of different infectious disease outbreaks such as EVD, Lassa fever, Zika, Yellow fever (Berkley, 2018; Hamblion et al., 2018; Lowe et al., 2018; Perkins et al., 2017). A poor diagnostic preparedness was also seen in the early period of the current COVID-19 pandemic. PCR testing was not widely available and only limited to patients who were admitted to hospital with COVID-19 symptoms and only later extended to symptomatic health-care workers (Pallett et al., 2020). This delay in diagnosis led to the rapid spread of COVID-19 (Rong et al., 2020). In the case of EVD, during the 2013-2016 West African outbreak, it took nearly 3 months for health officials and their international partners to identify EBOV as a causative agent (World Health Organization, 2015b). By that time, the outbreak spread rapidly within Guinea and was imported into Liberia and Sierra Leone. The improvement of diagnostics by an accelerated development of field-deployable molecular assays like RT-PCR showed to be vital not only for an earlier case detection but also for patient management and the eventual containment of the outbreak (Kelly-Cirino et al., 2019).

While diagnostic tests of EVD have been improved over the years to the current use of automated PCR platforms and the development of rapid diagnostic tests, less attention has been paid to developing prognostic tests. These tests could be useful tools for patient management specially in outbreak settings with limited resources. For example, during the peak of the 2013-2016 West African outbreak, the healthcare facilities overpassed their capacity and clinicians had to make difficult decisions about how to triage patients and family expectation regarding the disease outcome (Crowe et al., 2016). Due to the absence of a prognostic test that determines the probable clinical outcome of EVD in the patient, clinicians made decisions considering the viral

load levels (determined by Ct values of RT-PCR assays) as a predictor of the clinical outcome. As was mentioned in Chapter 1, viral load is currently the most reliable and strongest predictor of the clinical outcome of acute EVD since viremia is correlated to the severity of the disease. Although the quantification of viral load is able to stratify patients into risk groups (survivors or fatal outcome), especially when Ct value is high or low (i.e. high viremia levels are found in EVD fatalities compared to survivors), it is hard to predict the outcome of patients with mid-range Ct values (i.e. values between Ct 20 and Ct 22). In these cases there is a 50/50 outcome (death/survival) and therefore the viral load Ct value could not be used as a predictor of outcome (Liu et al., 2017). Thus, development of prognostic tests that do not solely rely on the EBOV Ct values at the time of diagnostic testing would be valuable for improving the management of patients during an outbreak.

In the search for other potential predictors of outcome, some studies have investigated the factors that may influence the disease outcome such as place of residence, age, gender, clinical markers (diarrhoea, fever, headache, bleeding manifestations among other symptoms), laboratory markers (high levels of BUN, ALT, AST, CRP, hypocalcemia, hyperkalaemia, among others) and co-infection with *Plasmodium falciparum* (described in detail in Chapter 1). Although these parameters or a combination of them have been recommended for predicting the clinical outcome, a threshold laboratory value in these approaches need to be set up (Reisler et al., 2017). Besides, these studies have been conducted retrospectively from small cohorts of EVD patients from the 2013-2016 West Africa outbreak and require further validation with independent clinical data from different EVD treatment sites (Schieffelin et al., 2014; Yan et al., 2015; Zhang et al., 2015).

Up to now, little attention has been paid to the identification of host genes that could be useful for predicting the outcome in EVD outbreaks. As mentioned in Chapter 1, a previous study from our group using RNAseq analysis investigated the transcriptomic changes in peripheral blood samples from acute patients collected during the 2013-2016 outbreak in Guinea. Ten genes, whose abundance was different between survivors and fatal cases were identified as potential predictors based on machine

learning approaches applied to whole transcriptomic datasets (Liu et al., 2017). These were: VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28, NF3L1, SLC25A5, and CTSL (Liu et al., 2017). Such sequencing approaches are not field deployable and can have long turnaround times. Therefore, identifying a small subset of gene whose mRNAs can act as markers may allow a multiplex assay based on RT-qPCR.

To investigate whether these candidate predictive genes are suitable for analysing EVD clinical outcome, it was first necessary to generate RT-qPCR assays for the individual gene transcripts. An additional gene, Interferon-stimulated gene 15 (ISG15), was included later in this work, since it was found significantly upregulated during the acute phase of EVD (Caballero et al., 2016; Greenberg et al., 2020). Thus, the aim of this chapter was to develop and optimize RT-qPCR assays that could measure the transcript abundance of these genes in EVD clinical samples.

3.2 Overview of the RT-qPCR assays development

The generation of the RT-qPCR assays for each of the candidate predictive gene transcripts consisted of three steps. The flowchart shown in Figure 3.1 outlines these steps. In the first step, primers were designed to target the conserved regions of the transcript sequences for each gene. The RNA from human cell lines and PBMCs was converted into cDNA by reverse transcription to be used as a template for end-point PCR to test the designed primers. Successfully amplified PCR products were inserted into pGEM-T by TA cloning. A recombinant plasmid was isolated and sequenced to verify the inserted sequences. This plasmid was then linearized and used as a template for *in-vitro* transcription from either the SP6 or T7 promoter. The RNA copy number of the *in-vitro* transcripts were then measured by the formula developed by Fronhoffs et al. (2002).

In the second step, RT-qPCR assays were optimized for each of the candidate predictive gene transcripts. This involved primer optimization and the standardization of the protocol to quantify the copy number or abundance of mRNA transcribed by each gene by the absolute standard curve method. This method used

of serial dilutions of the *in-vitro* transcripts to create a standard curve. The sensitivity and linearity were estimated from the standard curves. RT-qPCR assays with good efficiency were further used in the next step, and those with poor efficiency were re-optimized.

In the third step, to demonstrate the performance of the RT-qPCR assays, the assays were evaluated in control samples (5 healthy donors) and a small cohort of clinical samples (n = 15) collected during the 2013-2016 outbreak in Guinea by the EMLab that included acute EVD patients with survival (n= 9) or fatal outcome (n = 6).

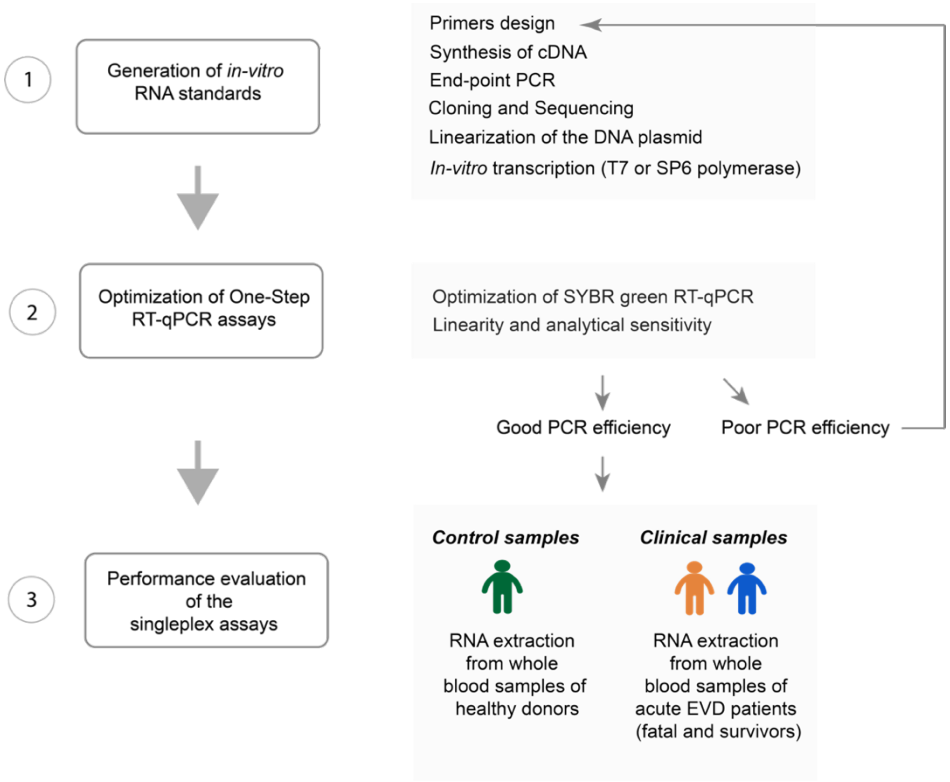
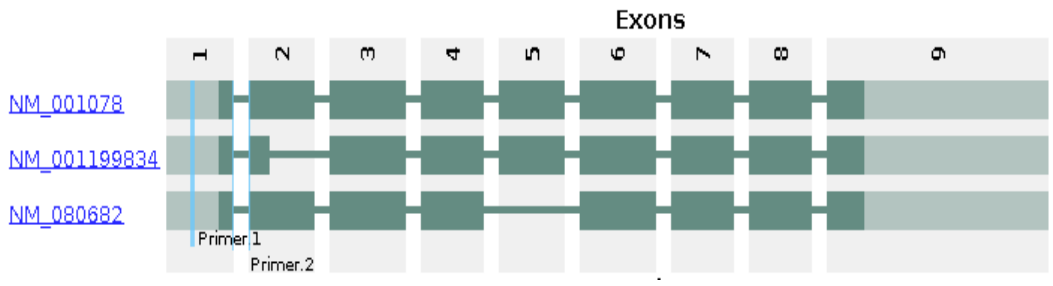


Figure 3.1 Flowchart of the generation of RT-qPCR assays for each candidate predictive gene transcripts. *The flowchart shows three processing steps: 1) generation of RNA standards that involves different procedures (primer design, synthesis of cDNA, end-point PCR, cloning and sequencing, linearization of DNA plasmid, in-vitro transcription), 2) optimization of the RT-qPCR assays using the absolute standard curve method, and 3) performance evaluation of the RT-qPCR assays in samples from healthy donors (green), and samples from EVD survivors (blue) and fatal EVD patients (orange).*

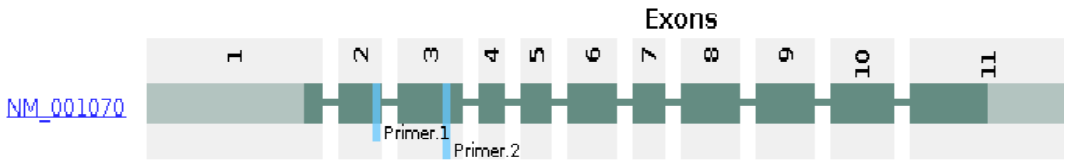
3.3 Design of primer sets

Primers were designed for the set of ten candidate predictive genes. In order to avoid genomic DNA (gDNA) amplification and its detection by real-time PCR, forward and reverse primers were designed in different exons separated by a single intron. All primers were designed in-house, except primers for VCAM1 that were taken from Aquila et al. (2017). These VCAM1 primers targeted two continuous exons, with the reverse primer spanning an exon-exon junction since this enables the amplification and detection of RNA sequences only. For the design of primers, the target exons for each gene were selected based on the alignments of all available mRNA transcript sequences of each gene from the NCBI's reference sequence database (RefSeq) (http://www.ncbi.nlm.nih.gov/RefSeq/). Primers were designed using the NCBI Primer-BLAST algorithms. The aim was to select primers which would amplify a product between 75-200 bp, with GC content between 50-60 % and Tm 50-65 °C. Figure 3.2 shows the known exon-intron structures for known transcripts of each gene, with location of the selected primers shown. Table 3.1 shows the sequence of the primer sets.

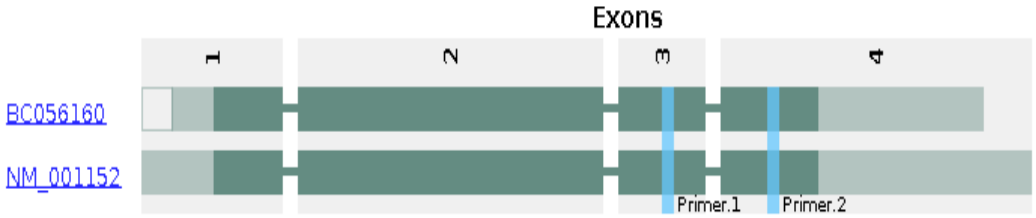
VCAM1



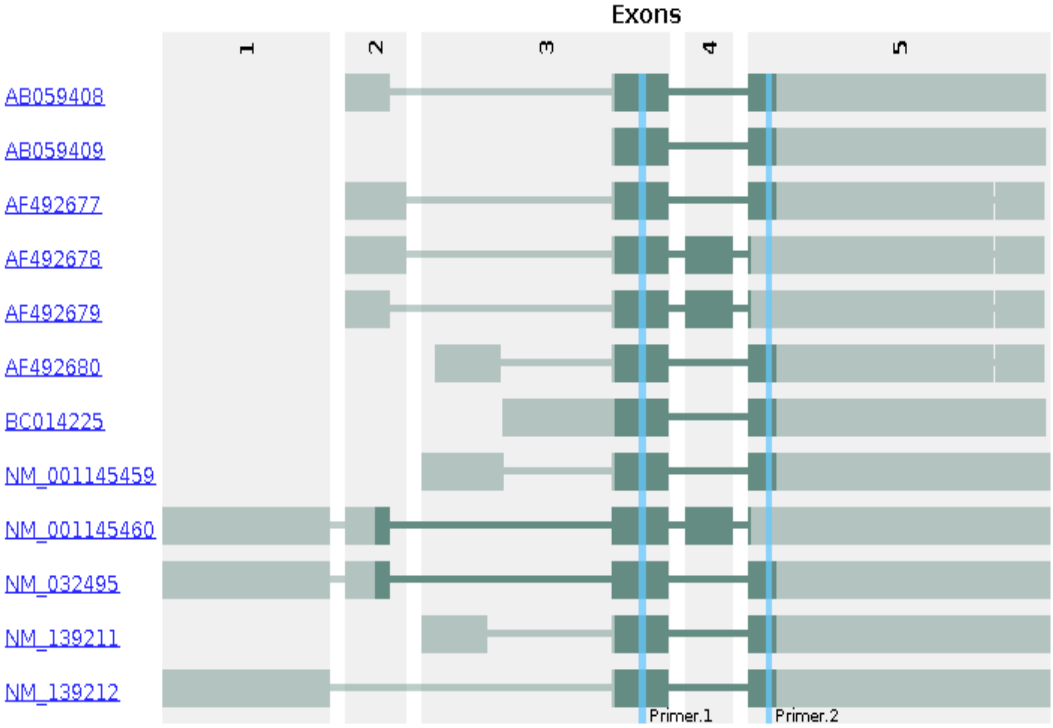
TUBG1



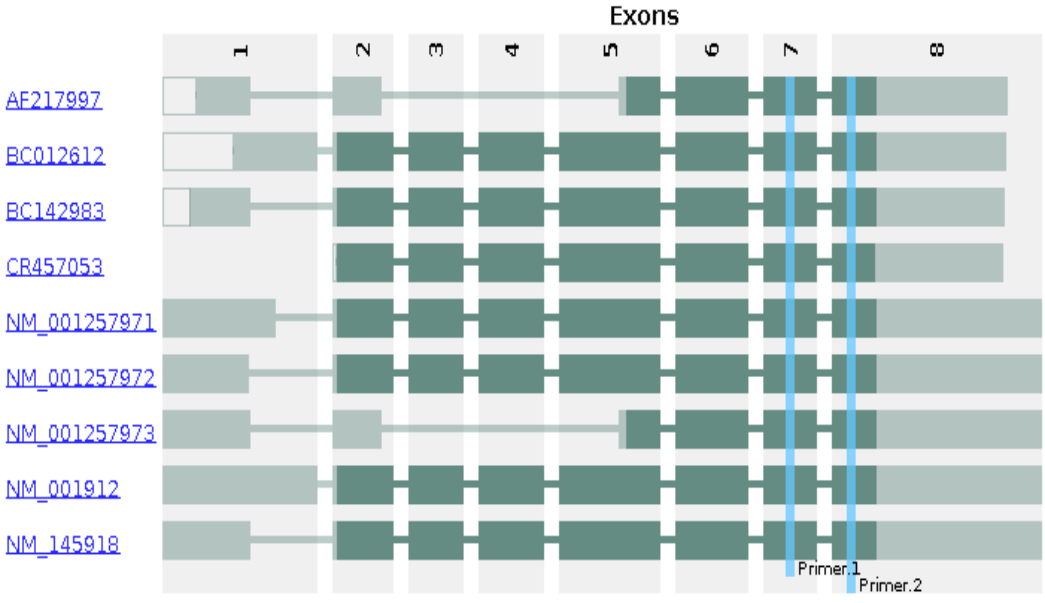
SLC25A5



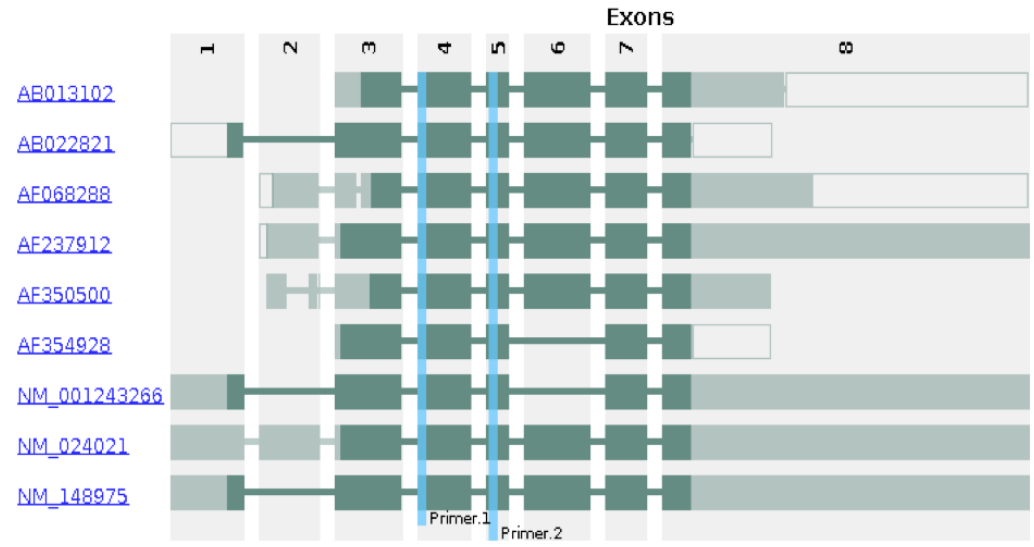
HOPX



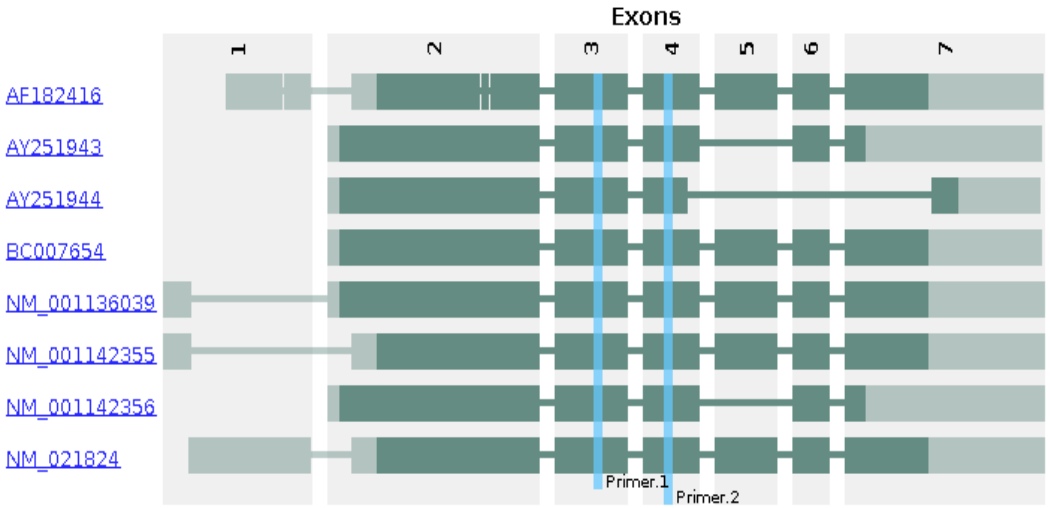
CTSL1



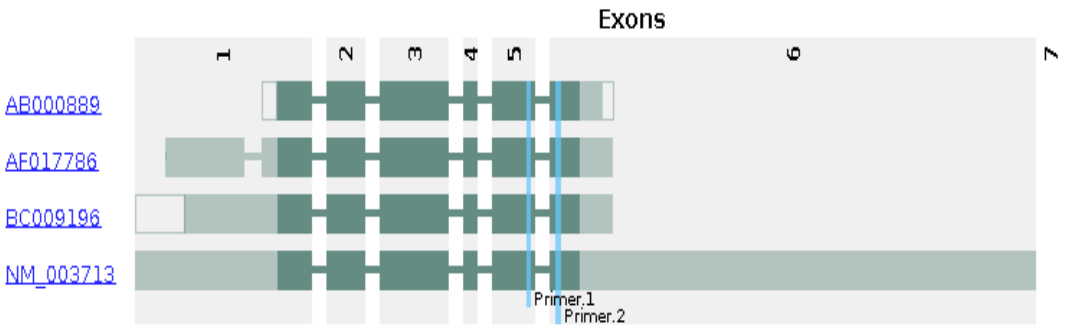
MS4A4A



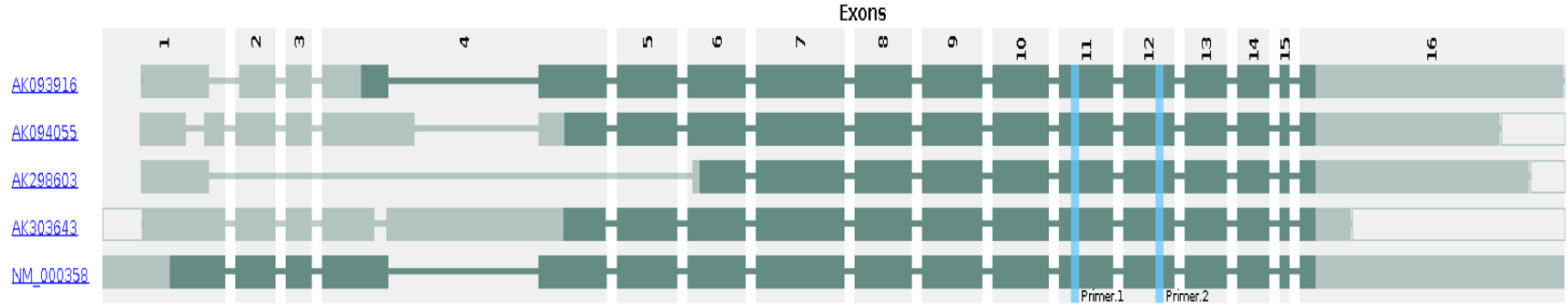
NF3L1



PLPP3



TGFBI



TTC28

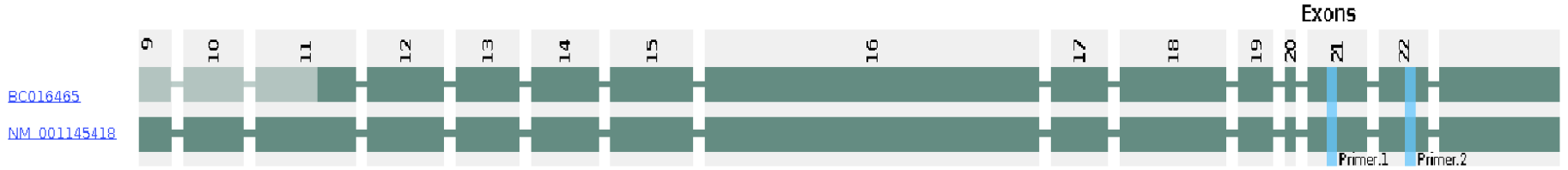


Figure 3.2 Location of the designed primer sets. *Graphic displaying the known RefSeq-annotated transcript and splice variants of the gene identified by accession number (text in blue). The boxes represent the exons, the lighter green regions are untranslated regions (UTRs), and dark green areas are the coding regions. The connecting lines are the introns. The light blue vertical lines show the target locations of the RT-qPCR primers. Primer 1 is the label for primer forward and Primer 2 the label for primer reverse. The figure was created using the tool Primer-check from SpliceCenter (Ryan et al., 2008).*

Table 3.1 Sequences of initial primers designed to test the ten candidate predictive genes transcripts*

Gene	NCBI Accession number	Sequences 5'- 3'	Amplicon length (bp)	Nucleotide position
VCAM1	NM_001199834.1	For: GGTATCTGCATCGGGCCTC Rev: TAAAAGCTTGAGAAGCTGCAAACA	193	105- 297
HOPX	NM_032495.5	For: CAACAAGGTCGACAAGCACC Rev: GACGGATCTGCACTCTGAGG	151	652-802
TUBG1	NM_001070.4	For: CCGCAAGGACGTCTTTTTCTAC Rev: TCCTCCTCCATGTTCCGACA	160	536-695
PLPP3	NM_003713.4	For: AGGATTTGCTCAAGGAGCCC Rev: AGGGAGAGCGTCGTCTTAGT	84	1342-1425
MS4A4A	NM_148975.2	For: GTGCAGATTCTGACTGCCCT Rev: CTGCAATTGACAAGGATCCTGAA	157	340-496
TGFBI	NM_000358.2	For: GACTGACGGAGACCCTCAAC Rev: AGAGACTTTAGCCGCACCAG	199	1738-1936
TTC28	NM_001145418.1	For: GACTTGGCATCAGCTCCCAT Rev: GAAGTGCACAGTCCGTCGAT	165	5765-5929
NF3L1	NM_001136039.2	For: ACACCCAAGACCTGGACAAA Rev: TCTACCACCTGCATCAAAGCC	145	615-759
SLC25A5	NM_001152.4	For: GCCGGGTTGACTTCCTATCC Rev: ACATTGGACCATGCACCCTT	164	786-949
CTSL1	NM_001257973.1	For: CTGGTGGTTGGCTACGGATT Rev: CTCCGGTCTTTGGCCATCTT	122	610-731

*All primers were designed in-house, except VCAM1 which was taken from Aquila et al. (2017)

3.4 Construction of plasmids

Before the construction of plasmids containing sequences of the gene transcripts for the generation of *in-vitro* RNA standards, it was first important to check the performance of the primer sets listed in Table 3.1. For this, the primers were first tested by end-point PCR using cDNA synthesized from RNA extracted from available human cell lines: 293T, and A549. The primers amplified the expected PCR products, depending on the specific gene expression in cell lines. Thus, TUBG1, TTC28, PLPP3, NF3L1, SLC25A5, CTSL amplified in 293T and A549 cell lines. HOPX amplified only in 293T cells, and TGFBI only in A549 cells. VCAM1 and MS4A4A did not amplified in any of these cell lines, therefore these were tested in RNA extracted from HBEC cell line kindly provided by the Brain Infections Group of the University of Liverpool. Only VCAM1 amplified in HBEC. MS4A4A amplified in RNA extracted from human PBMCs that were also provided by the Brain Infections Group.

For the construction of plasmids, first an end-point PCR was performed using the primer sets of each gene transcript and their specific cell line as a template. The PCR products were then inserted into plasmids vectors (pGEM-T and pGEM-T Easy vector) by TA cloning. To confirm the sequence and integrity of the PCR products all plasmids were sequenced. Results were compared with the nucleotide databases in the NCBI GenBank using the BLAST tool and confirmed 100% identity with the transcript reference sequences for each gene. The orientation of the inserted sequences was also checked using SnapGene Viewer software (Chicago, USA).

3.5 Generation of *in-vitro* RNA standards

To measure the transcript abundance of each gene, standard curves were created for the RT-qPCR assays. *In-vitro* RNA standards (cRNA) were used in the construction of the standard curves. For the cRNA synthesis, the purified plasmids were linearized by restriction enzyme digestion with either SalI or NcoI (pGEM-T and pGEM-T vector map is shown in Figure 2.1 in Chapter 2). NF3L1, VCAM1, HOPX, TUBG1, PLPP3, and CTSL plasmids were linearized with SalI while MS4A4A, TGFBI, TTC28, and SLC25A5

plasmids were linearized with NcoI. The target sequences were then transcribed by *in-vitro* transcription using either a T7 promoter (for VCAM1, PLPP3, TUBG1, CTSL, NF3L1, HOPX) or a SP6 promoter (for MS4A4A, TGFBI, TTC28, SLC25A5). This was determined by the orientation of the PCR product to ensure sense strand transcripts were generated for constructing a standard curve for mRNA quantification. The transcribed RNA was treated with DNase to digest the plasmid, and no DNA bands were observed in all cRNA by agarose gel analysis. A 260/280 ratio of approximately 2.0 was seen for all *in-vitro* transcripts indicating a good nucleic acid purity. The size of cRNA products were confirmed by gel electrophoresis analysis. To analyse cRNA in a native/non-denaturing agarose gel, the RNA ladder and cRNA sample were first denatured as described in Chapter 2. Figure 3.3 shows the cRNA products as single bands with no significant degradation. High intensity bands represent 500ng of RNA loaded into the gel, faint bands like HOPX are due to loading of less than 500ng because the RNA concentration obtained after the clean-up of transcripts was low (Table 3.2). The copy numbers of the cRNAs are listed in Table 3.2 and were calculated based on its concentration and molecular weight using a published formula developed by Fronhoffs et al. (2002) that was described in Chapter 2.

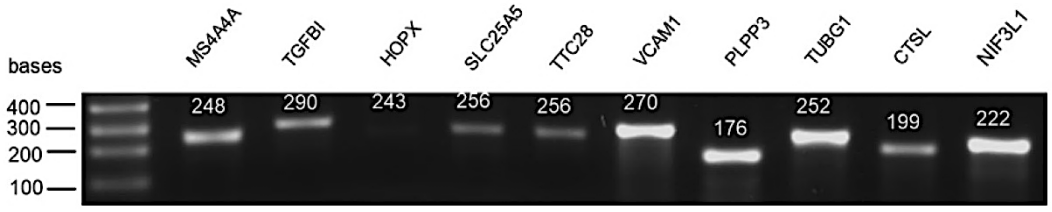


Figure 3.3 Agarose gel electrophoresis of in-vitro transcripts. *The RNA transcripts for the 10 genes and RNA ladder were first denatured and loaded in a 2% native agarose gel and visualized by staining with ethidium bromide. Numbers above the RNA bands are the expected sizes (bases).*

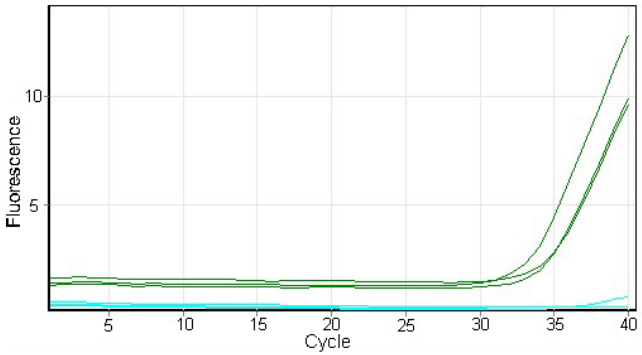
Table 3.2 Quantification of the in-vitro transcripts

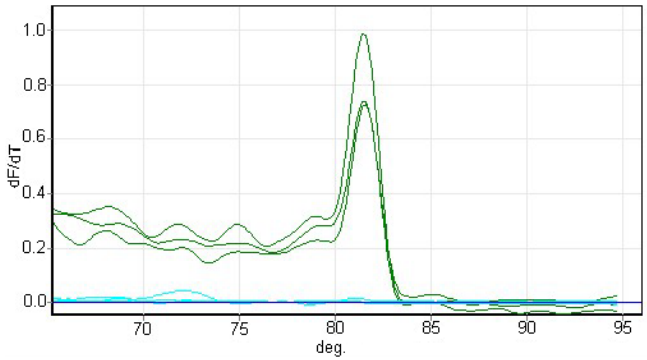
Gene	Length of cRNA (bases)	cRNA concentration ng/µL	OD 260/280	OD 260/230	Quantification of cRNA standard N (molecules per µL)
VCAM1	270	341.3	2.16	1.52	23×10^{11} N
HOPX	243	22.5	2.28	1.00	2×10^{11} N
TUBG1	252	262.1	2.07	1.99	19×10^{11} N
PLPP3	176	177.6	2.09	1.08	19×10^{11} N
MS4A4A	248	108.05	2.30	1.61	8×10^{11} N
TGFBI	290	67.4	2.19	1.24	4×10^{11} N
TTC28	256	48.2	2.12	1.18	3×10^{11} N
NIF3L1	222	311.7	2.17	1.78	25×10^{11} N
SLC25A5	256	57.0	2.14	1.08	4×10^{11} N
CTSL	199	166.25	2.18	1.38	16×10^{11} N

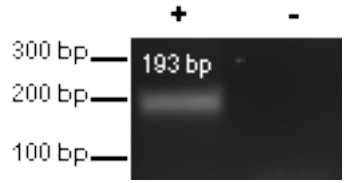
3.6 Primers testing and optimization in real time qPCR

Although the primers sets used have been successful to amplify products by end-point PCR (section 3.4), it was necessary to evaluate their performance in real time quantitative PCR (qPCR) that use SYBR green chemistry. The concentration and optimal annealing temperature for each primer set was optimized by SYBR green real time qPCR. This method does not include a reverse transcription step, thereby cDNA is used as a template. The primers sets were tested on cDNA synthesized from human PMBCs and run in triplicate. Different primers concentrations (100 and 200 nM) and annealing temperatures (55°C and 60°C) were initially evaluated. For each primer set, the optimal primer concentration and annealing temperature was 100nM and 60°C, respectively. The data shown in Figure 3.4 represents the amplification curves, melt curves and the qPCR product resolved on an agarose gel for each of the 10 primer sets. A single peak is shown in the melt curve analysis that correspond to the single band observed on the agarose gel. The melt curve analysis did not identify non-specific products such as a primer dimers that could have co-amplified and show a different peak with different melting temperature. This was confirmed by the agarose gel analysis of qPCR products that did not show additional bands.

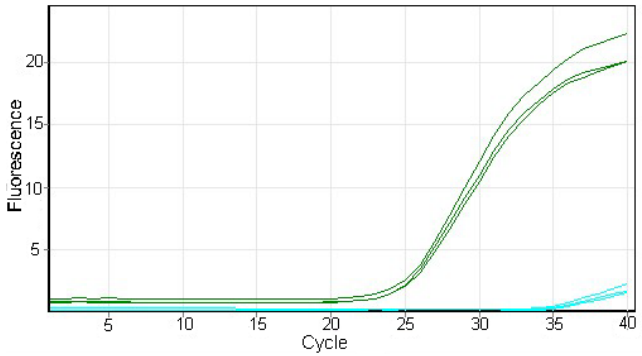
VCAM1

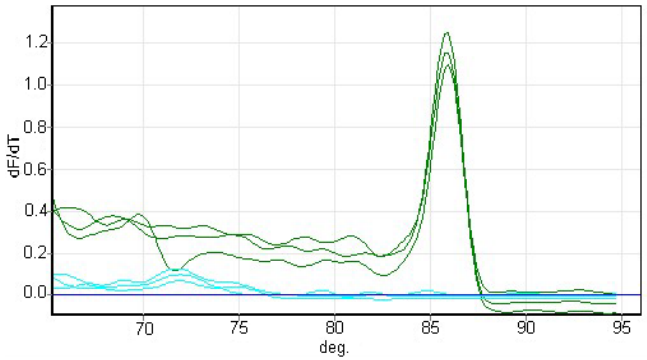

Fluorescence
Cycle

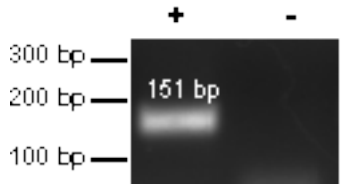

dF/dT
deg.


+
-
300 bp
200 bp
100 bp
193 bp

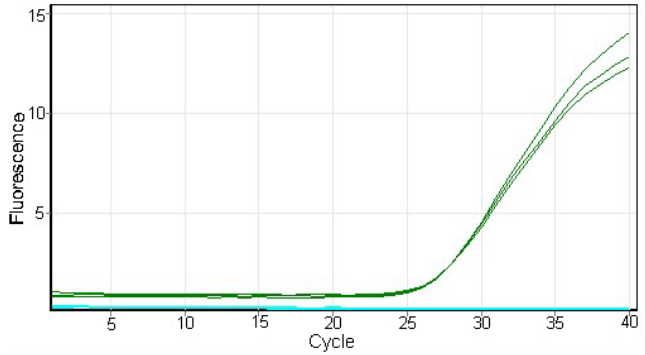
HOPX

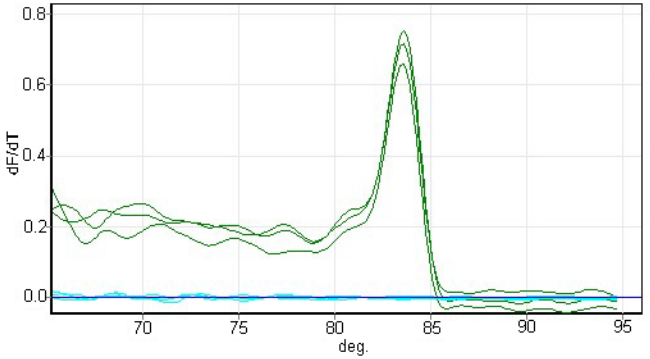

Fluorescence
Cycle


dF/dT
deg.


+
-
300 bp
200 bp
100 bp
151 bp

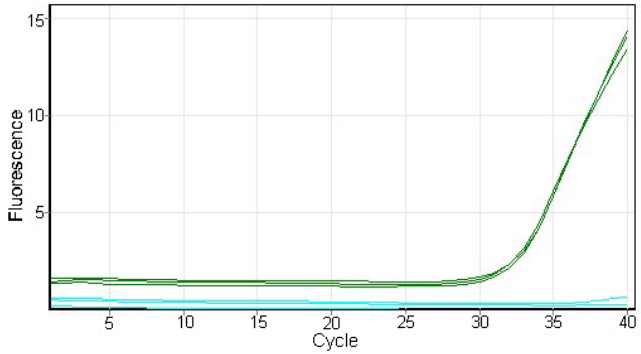
TUBG1

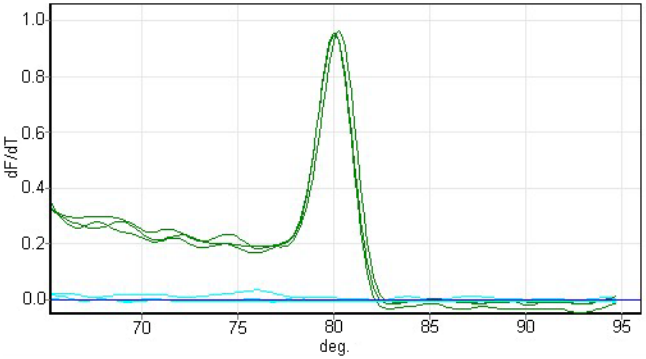

Fluorescence
Cycle
15
10
5
5
10
15
20
25
30
35
40

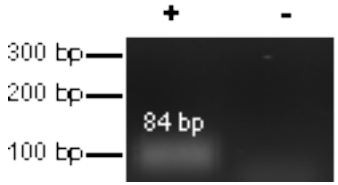

dF/dT
deg.
0.8
0.6
0.4
0.2
0.0
70
75
80
85
90
95


+
-
300 bp
200 bp
100 bp
160 bp

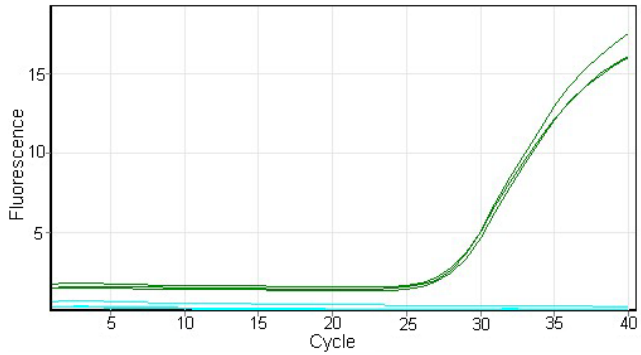
PLPP3

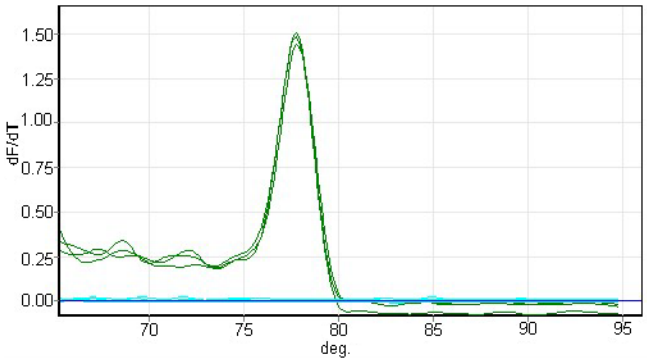

Fluorescence
Cycle
15
10
5
5
10
15
20
25
30
35
40


dF/dT
deg.
1.0
0.8
0.6
0.4
0.2
0.0
70
75
80
85
90
95

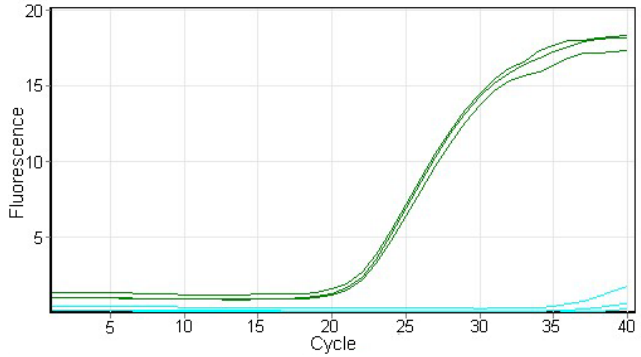

+
-
300 bp
200 bp
100 bp
84 bp

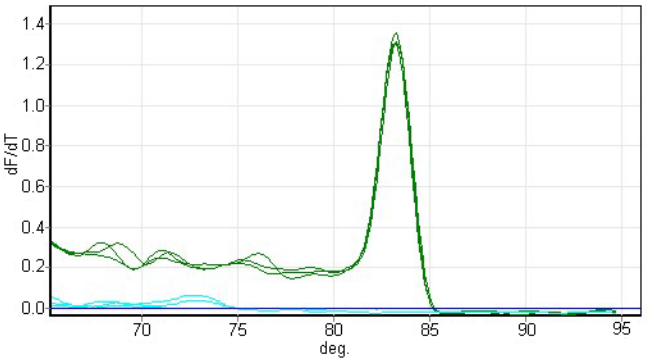
MS4A4A

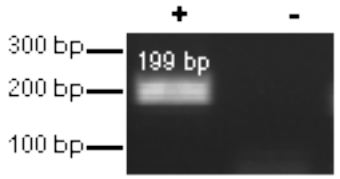




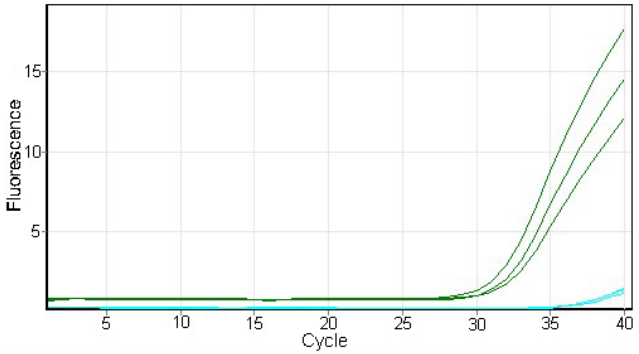
TGFBI

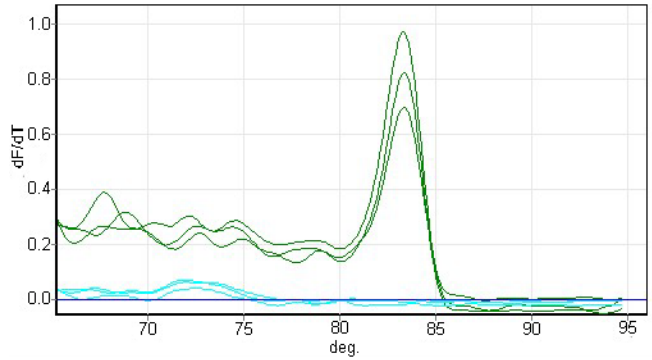






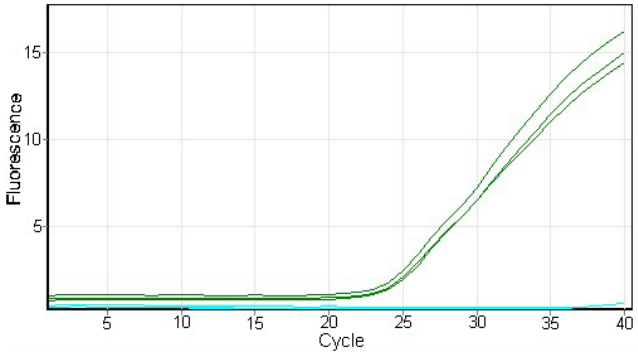
TTC28

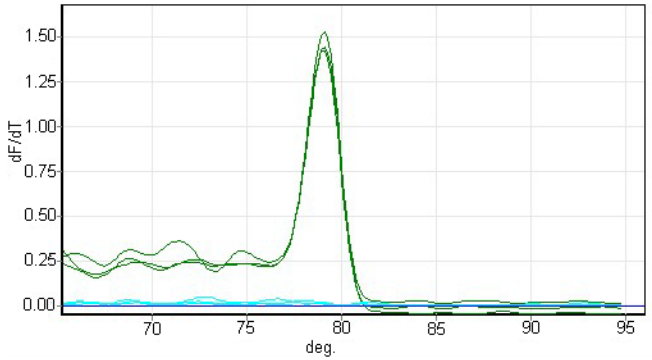

Fluorescence
Cycle

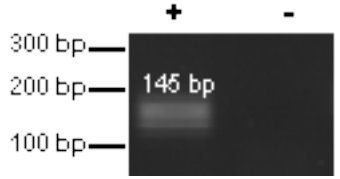

dF/dT
deg.


+
-
300 bp
200 bp
100 bp
165 bp

NF3L1


Fluorescence
Cycle


dF/dT
deg.


+
-
300 bp
200 bp
100 bp
145 bp

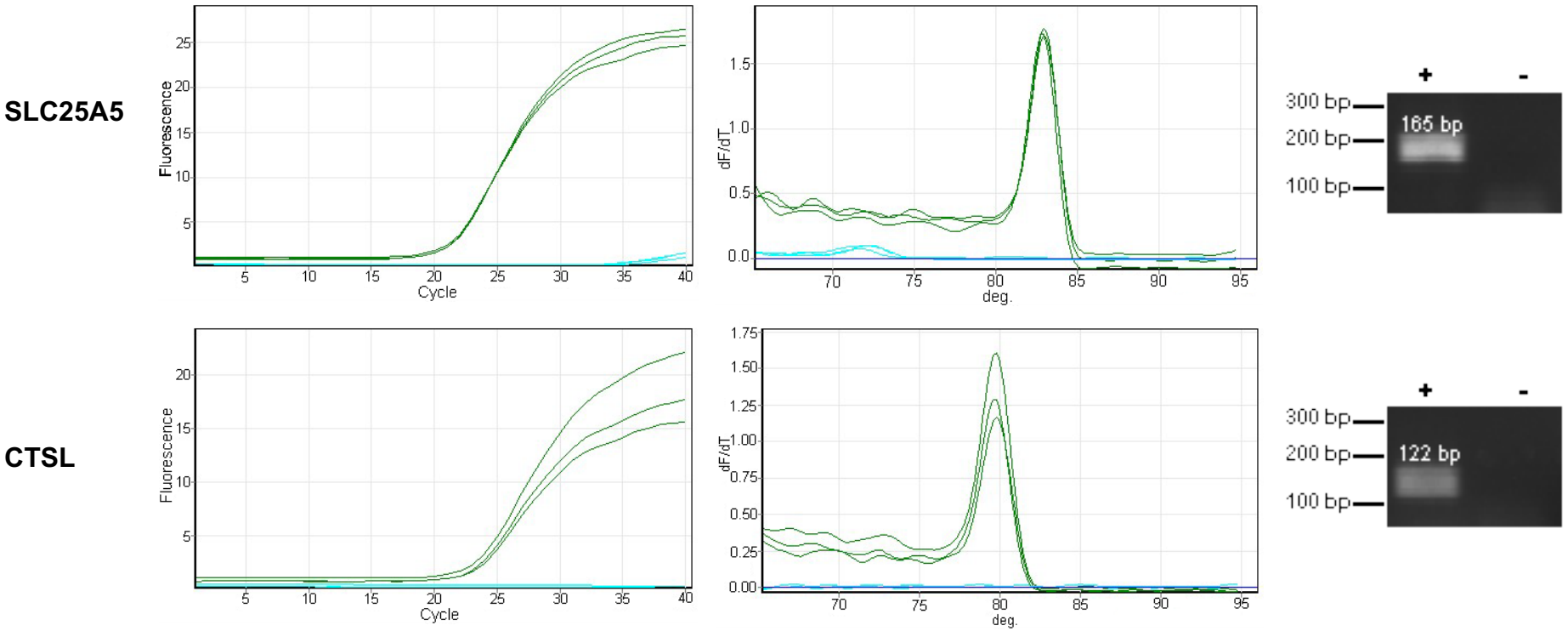


Figure 3.4 Analysis of the primer sets by Real-Time qPCR. *Left panel shows the amplification curve of cDNA tested in triplicate. Right panel shows the melting curve of the amplified products. Also shown is the gel analysis of the qPCR products of one of the triplicates; lane 1 shows a single band of the amplified product and lane 2 shows the negative control (no template added). Numbers above the DNA bands are the expected product sizes (bp).*

3.7 Application of primer pairs in one-step RT-qPCR

As the primer pairs were to be used to detect transcript levels direct in RNA from clinical samples rather than cDNA, a reverse transcription step would be required to be incorporated into the assay. Thus, the optimized conditions used in real time qPCR were further evaluated in a one-step quantitative reverse transcription PCR (RT-qPCR) assay. For this, SYBR Green based one-step RT-qPCR assays were performed using 10-fold serial dilutions of the synthesized cRNA from each candidate gene transcript as a template to establish standard curves. Each standard curve was generated by plotting the Ct values versus serial dilutions of the cRNA templates tested in triplicate ranging from 10^3 to 10^7 copies/µl. The assays detected the target sequences of each gene transcript. Reproducible amplification curves were obtained from the dilution series and replicates of cRNA used to create standard curves (Figure 3.5).

3.8 Performance evaluation of the RT-qPCR assays

To evaluate the performance of the primer pairs in the RT-qPCRs the following parameters were assessed: linearity, efficiency, analytical sensitivity (limit of detection) and precision (repeatability and reproducibility), as recommended in the minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines (Bustin et al., 2009).

3.8.1 PCR efficiency and linearity

The amplification efficiency was estimated from the slope of the standard curve with the mathematical formula $E = 10^{-1/slope}-1$ (Bustin et al., 2009). Although the RT-qPCR assays for 8 of the gene transcripts, VCAM1, TUBG1, PLPP3, MS4A4A, TTC28, NF3L1, SLC25A5A and CTSL1 showed a good correlation coefficient (r^2 value) of 0.99, not all RT-qPCR assays showed a high PCR efficiency (Figure 3.5 and Table 3.3). RT-qPCR assays for 3 genes: CTSL, SLC25A5 and NF3L1 obtained PCR efficiencies greater than 90%, which were within the acceptable range of efficiency expected for RT-qPCR assays (E = 90% to 105% ; Slope = -3.2 to -3.5) (Johnson et al., 2013; Svec et al., 2015).

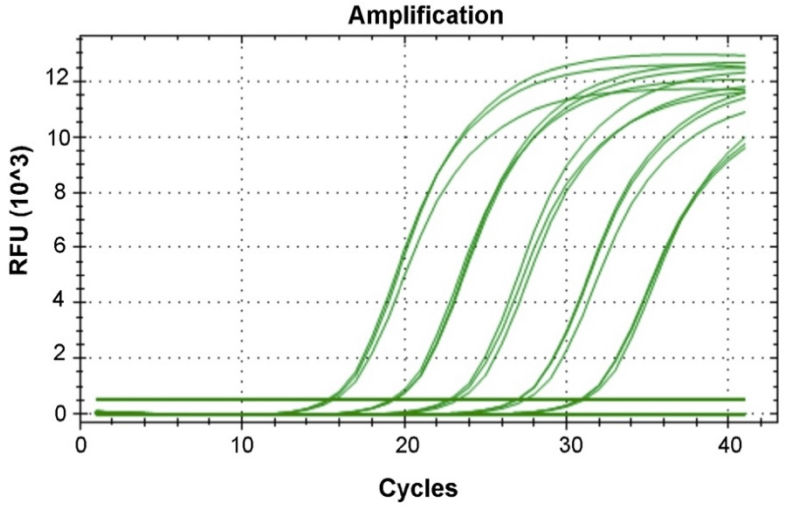
The efficiency of RT-qPCR assays for VCAM1, TUBG1, MS4A4A and TTC28 were lower than 90%. These results suggested that the assays for VCAM1, TUBG1, MS4A4A, and TTC28 had poor primer design or suboptimal reaction conditions. Therefore, the assays were modified by redesigning the primers for the gene transcripts that had a low PCR efficiency such as VCAM1, TUBG1, MS4A4A, TTC28. At this point of the study, the efficiency of RT-qPCR assays for HOPX and TGFBI primer sets was not determined since it was decided to include these genes in the group of genes for primer re-design. Likewise, SLC25A5A was also included in these groups of genes to find primer sets that have higher PCR efficiency. To re-design primer sets, it was necessary to construct new plasmids with longer gene sequence inserts so more alternative primer sets could be tested to check the PCR efficiency (see section 3.9)

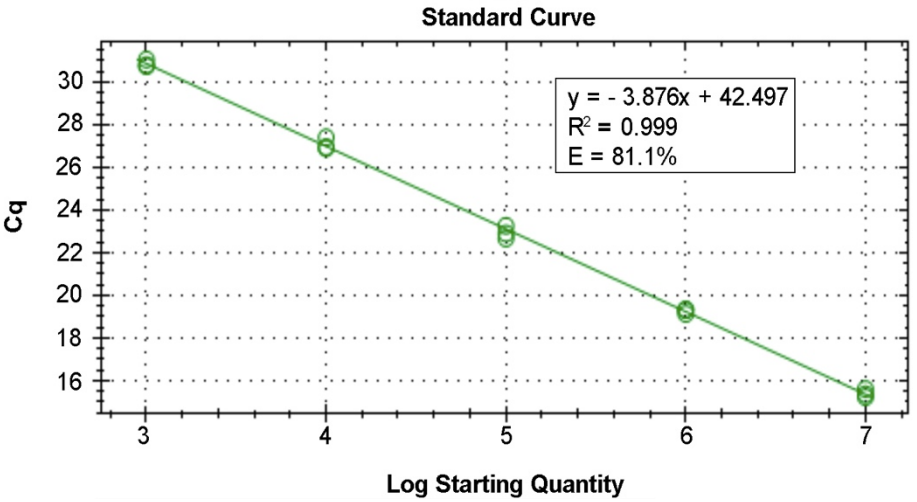
Table 3.3 Performance evaluation of the RT-qPCR assays

Gene	qRT-PCR product size (bp)	Slope	r^2 value	PCR Efficiency (E)
VCAM1	193	-3.876	0.999	81.14%
HOPX	151	ND	ND*	ND
TUBG1	160	-3.820	0.999	82.7%
PLPP3	84	-4.221	0.998	72.54%
MS4A4A	157	-3.725	0.999	85.54%
TGFBI	199	ND	ND	ND
TTC28	165	-3.623	0.999	88.81%
NF3L1	145	-3.466	0.998	94.32%
SLC25A5	164	-3.55	0.996	91.28%
CTSL	122	-3.475	0.999	94%

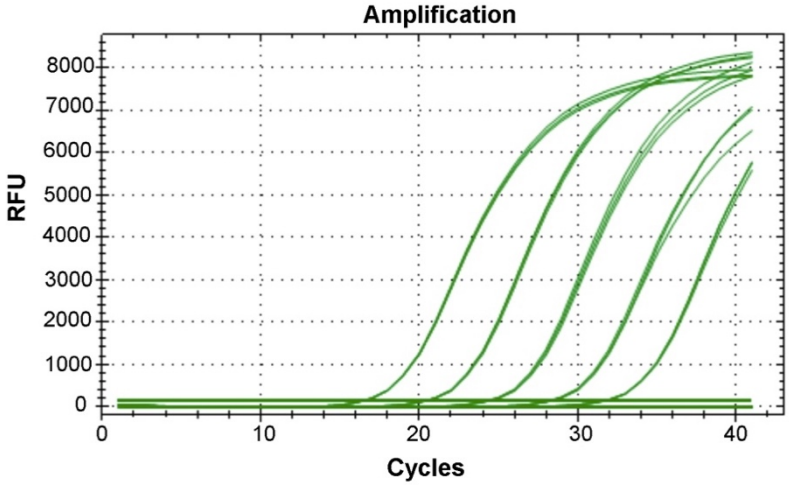
*ND: not determined.

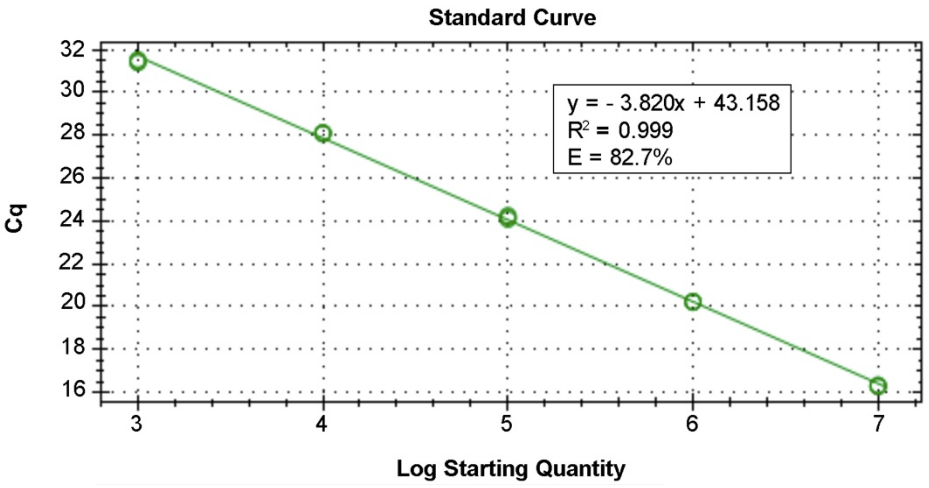
VCAM1


Amplification
RFU (10^3)
Cycles

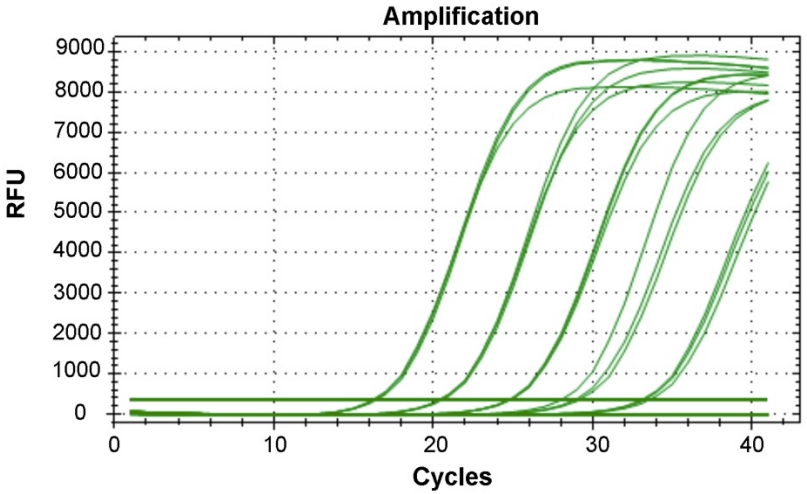

Standard Curve
y = - 3.876x + 42.497
R² = 0.999
E = 81.1%
Cq
Log Starting Quantity

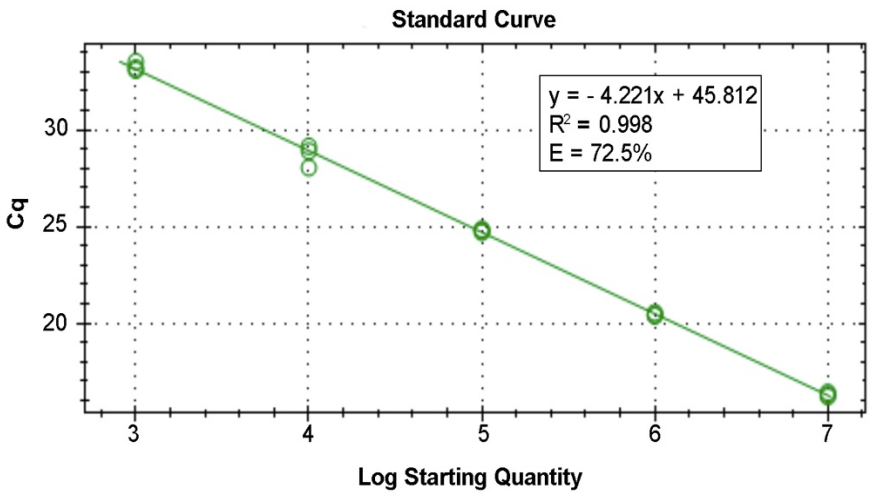
TUBG1


Amplification
RFU
Cycles

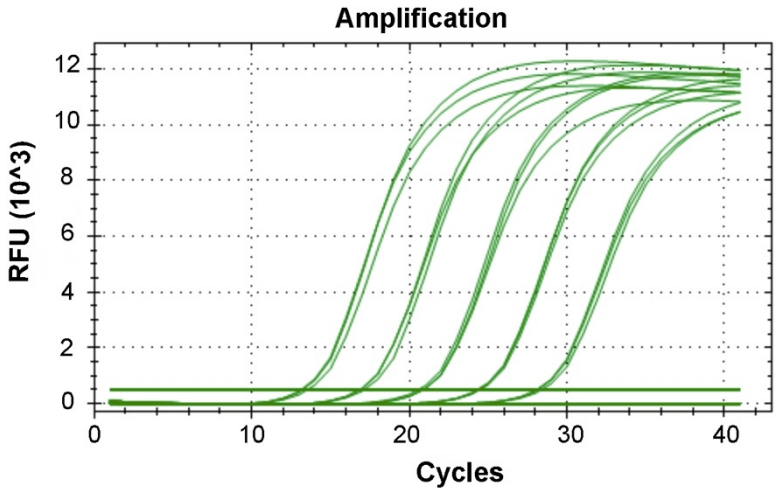

Standard Curve
y = - 3.820x + 43.158
R² = 0.999
E = 82.7%
Cq
Log Starting Quantity

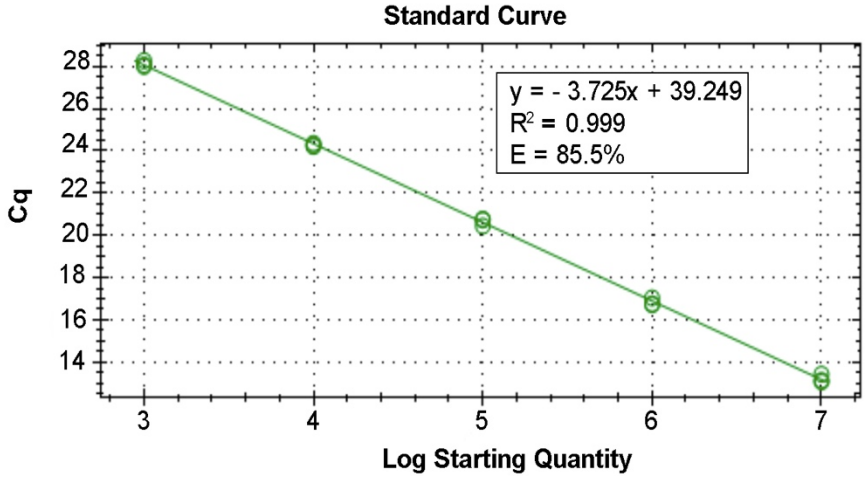
PLPP3



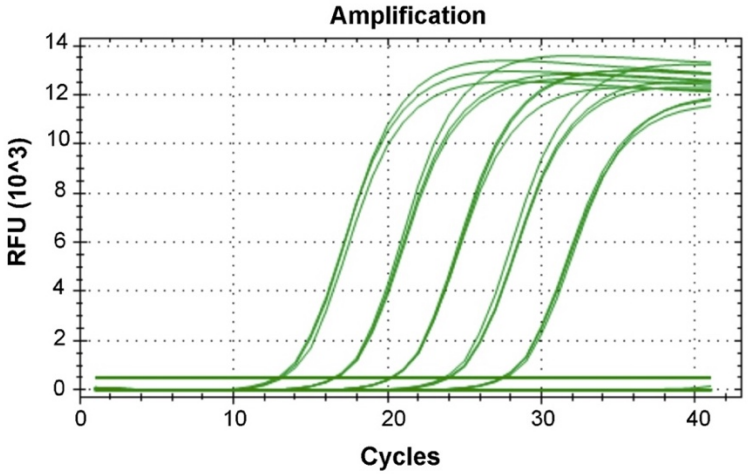


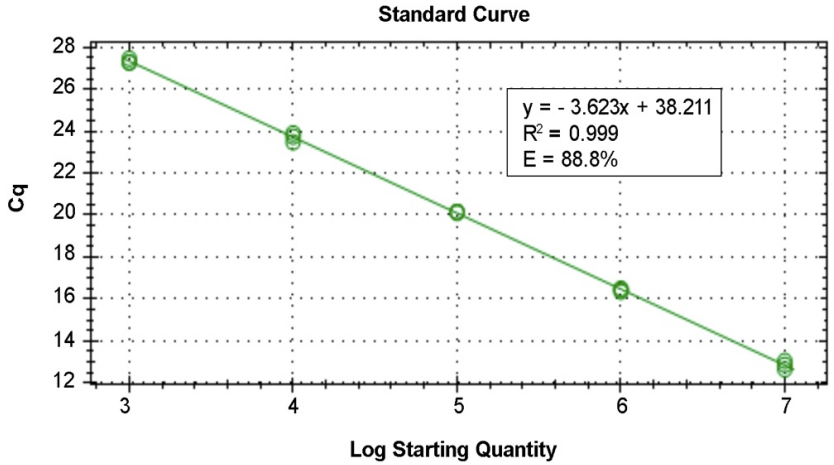
MS4A4A



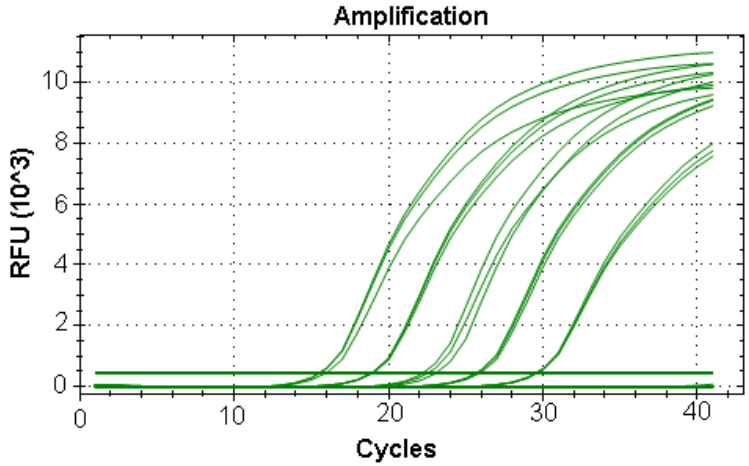


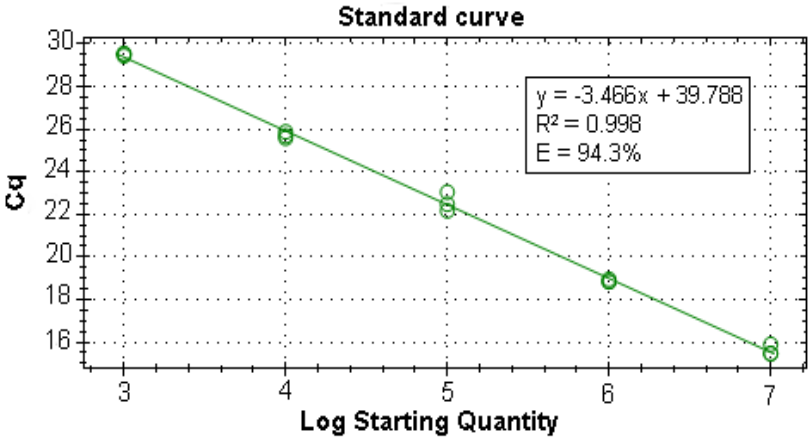
TTC28


Amplification
RFU (10^3)
Cycles

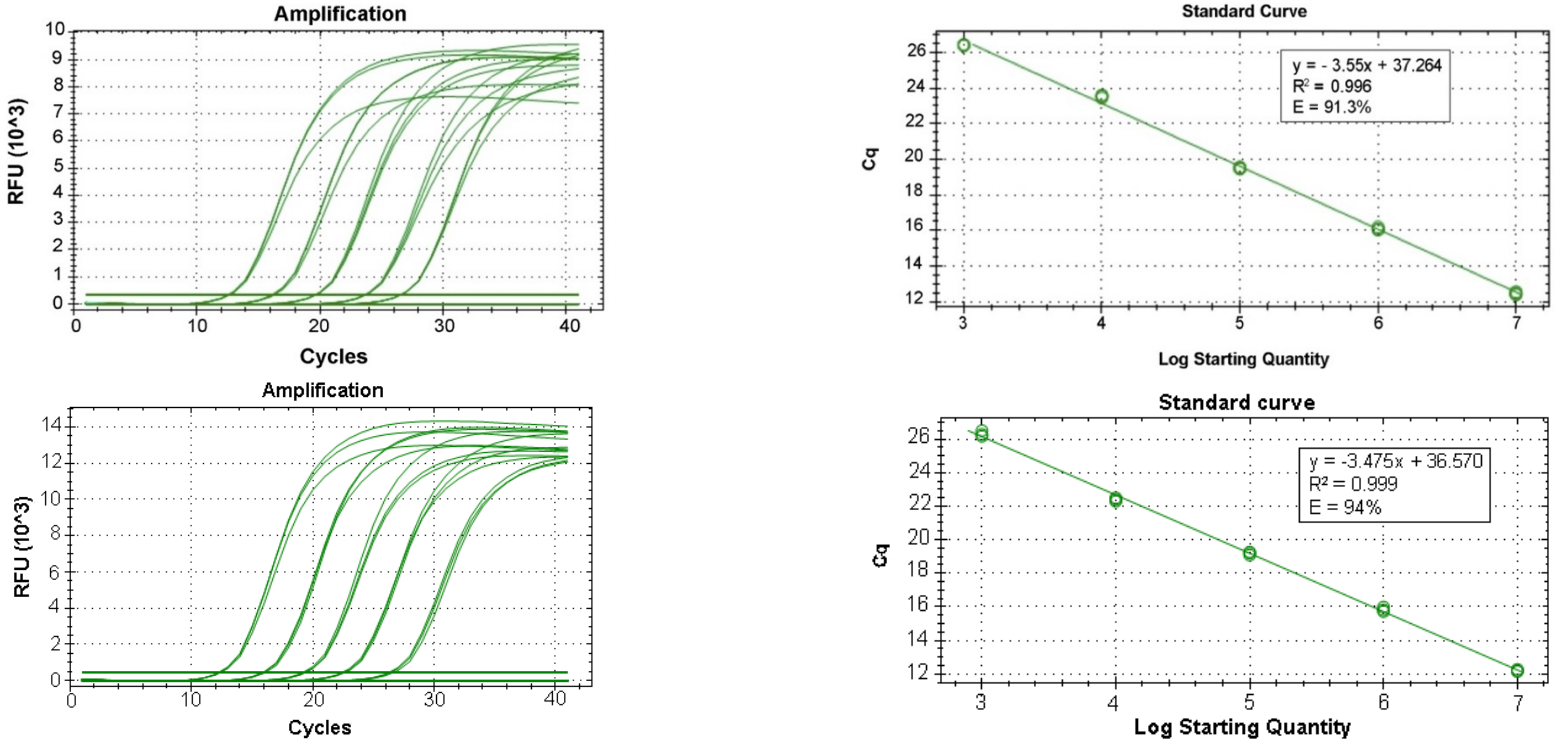

Standard Curve
y = - 3.623x + 38.211
R² = 0.999
E = 88.8%
Cq
Log Starting Quantity

NF3L1


Amplification
RFU (10^3)
Cycles


Standard curve
y = -3.466x + 39.788
R² = 0.998
E = 94.3%
Cq
Log Starting Quantity

SLC25A5



CTSL

Figure 3.5 Linearity and PCR efficiency of the RT-qPCRs for 8 of the candidate predictive genes transcripts*. Left panels show the amplification chart of the five cRNA standards for each gene, with dilutions ranging from 10^3 to 10^7 copies/µl and tested in triplicate. Right panels show the standard curves created with these dilutions. The mean Cq values are plotted against the log of concentration of the target gene copies/reaction. The PCR efficiency for each target was calculated using the slope of the calibration curve with the formula $E = 10^{-1/slope}-1$. E = PCR efficiency, R^2 = correlation coefficient. HOPX and TGFBI were not evaluated.*

3.9 Re-designing primer sets

The data shown in section 3.8 suggested poor primer performance for VCAM1, TUBG1, MS4A4A, and TTC28. Therefore, it was necessary to re-design primer sets for these gene transcripts. However, the short length of the synthesized *in-vitro* transcripts generated from the cloned PCR products for each of the 10 genes (each less than 300 bases) (Table 3.2) limited the ability to design alternative primer sets. One approach was tested, where primers sets were modified by additional of the relevant single nucleotide (data not shown). However, the PCR efficiency did not improve. Thus, it was decided to clone longer regions of the target gene transcripts into a relevant plasmid vector. This would allow generation of multiple primer pairs for testing to maximise the potential for identifying primers that comply with the requirements for RT-qPCR assays.

At this point in the study, it was decided to include the gene ISG15 amongst the set of potential predictive gene transcripts. Transcriptome analysis of peripheral blood in human and EBOV-infected NHP have shown that interferon-stimulated genes (ISGs) are strongly upregulated in the early stage of infection (Caballero et al., 2016; Liu et al., 2017). From these genes, a high expression of ISG15 was observed in EBOV-infected NHP and distinguished between survivors and anticoagulant-treated NHP that did not survive EBOV infection, with lower expression values in survivors (Caballero et al., 2016; Garamszegi et al., 2014). Likewise, a greater abundance of ISGs transcripts was observed in acute-fatal compared to acute-survivor patients from Guinea (Liu et al., 2017). Thus, it was felt that this would be appropriate to test in our study.

3.9.1 Construction of new plasmids

A sequence of approximately 1 kb was selected for the gene transcripts: VCAM1, TUBG1, MS4A4A, TTC28, HOPX, TGFBI, and SLC25A5 to be inserted into a plasmid vector. The new plasmid constructs were synthesised by GeneArt Gene Synthesis (https://www.thermofisher.com/order/geneartgenes/projectmgmt). From the

available vectors in GeneArt Gene Synthesis, the expression vector pcDNA3.1(+), which contains the T7 promoter, was selected (see Figure 2.1).

3.9.2 Generation of new *in-vitro* RNA standards

To be able to generate *in-vitro* RNA transcripts the new circular plasmids were linearized with NotI, which is downstream of the gene insert (Figure 3.6). Linear DNA was then transcribed with T7 RNA polymerase in the presence of rNTPs until the polymerase falls off at the end of the template. Long RNA transcripts were synthesized for the genes, more than 1Kb length for VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28 and SLC25A and 727 bases for ISG15.

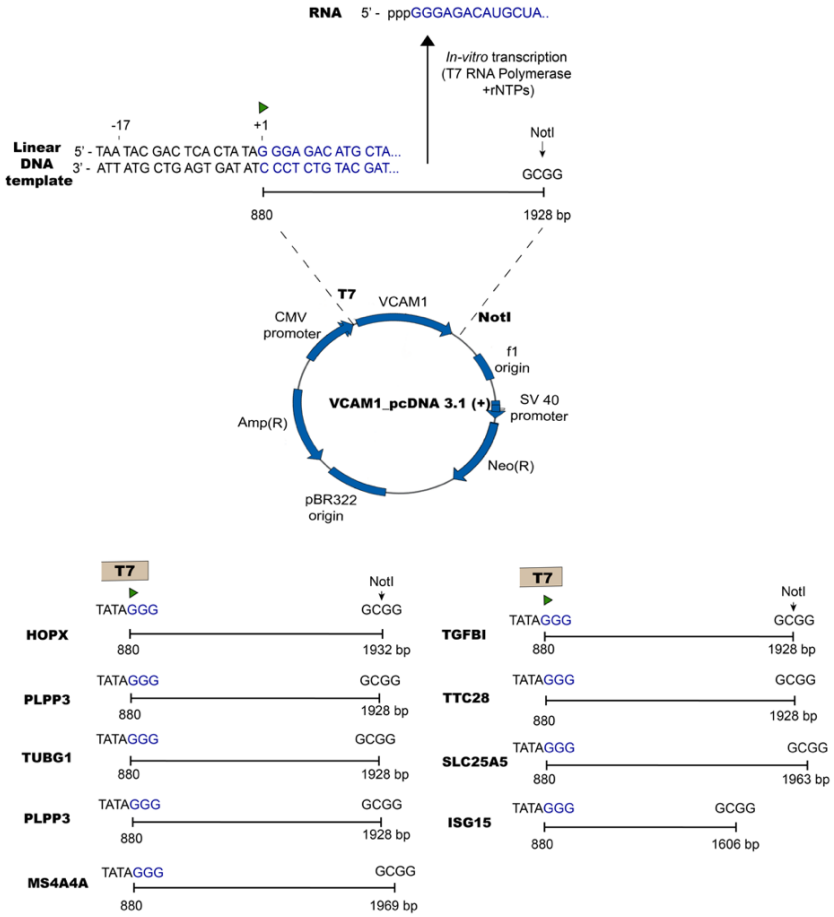


Figure 3.6 Schematic representation of the generation of new in-vitro transcripts. *The upper panel shows a circular plasmid with the insert of interest (VCAM1 sequence) cloned between a T7 promoter and a unique restriction site. The plasmid is linearized with NotI and transcribed from the promoter to yield multiple RNA transcripts terminated by "running-off" the template. The T7 promoter covers a sequence that range from -17 to + 6 and +1 is the first nucleotide incorporated into the transcription product. Lower panel shows graphs representing the transcribed sequence of each gene. Figure adapted from Beckert and Masquida (2011).*

The *in-vitro* transcripts were first treated with DNase to remove plasmid DNA. To analyse the size and integrity of the cRNA, samples were resolved in a native/non-denaturing agarose gel, as described in Chapter 2. No DNA bands were observed in all cRNA standards, and only a single band was observed corresponding to the expected size of the transcript (Figure 3.7). The slight smeared appearance may be attributed to specific sized incomplete transcripts. The concentration of the cRNA standards and the copy number, calculated using the formula previously described, are shown in Table 3.4.

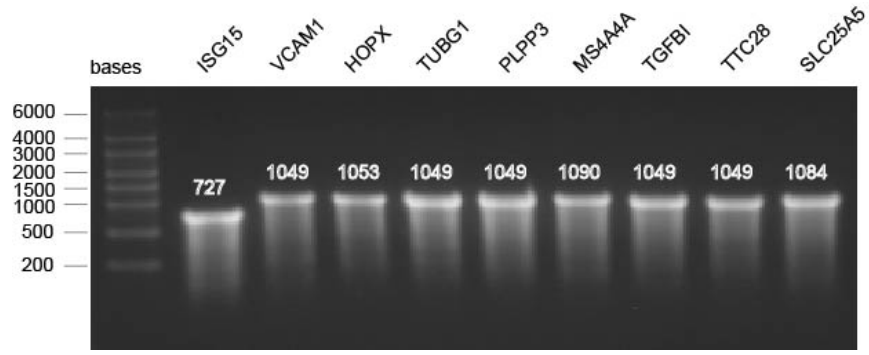


Figure 3.7 Agarose gel analysis of in-vitro transcripts generated from pcDNA3.1(+) constructs containing the large gene fragments. *The gel shows the new long RNA transcripts generated for ISG15, VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28 and SLC25A. The RNA ladder and cRNA standards were denatured and loaded in a 1% native/non-denaturing agarose gel and visualized by staining with SYBR safe gel stain. Numbers above the RNA bands are the expected sizes (bases).*

Table 3.4 Quantification of the new in-vitro RNA standards

Gene	Length of cRNA (bases)	cRNA concentration ng/µL	OD 260/280	OD 260/230	Quantification of cRNA standard N (molecules per µL)
ISG15	727	994.9	2.16	2.05	3×10^{12} N
VCAM1	1049	1191.45	2.26	2.24	21×10^{11} N
HOPX	1053	1248.4	2.31	2.37	21×10^{11} N
TUBG1	1049	1150.15	2.12	2.01	21×10^{11} N
PLPP3	1049	1347.95	2.23	2.28	23×10^{11} N
MS4A4A	1090	1147.75	2.30	1.61	18×10^{11} N
TGFBI	1049	1315.05	2.20	1.63	23×10^{11} N
TTC28	1049	1197.1	2.18	2.15	21×10^{11} N
SLC25A5	1084	1149.7	2.28	2.01	19×10^{11} N

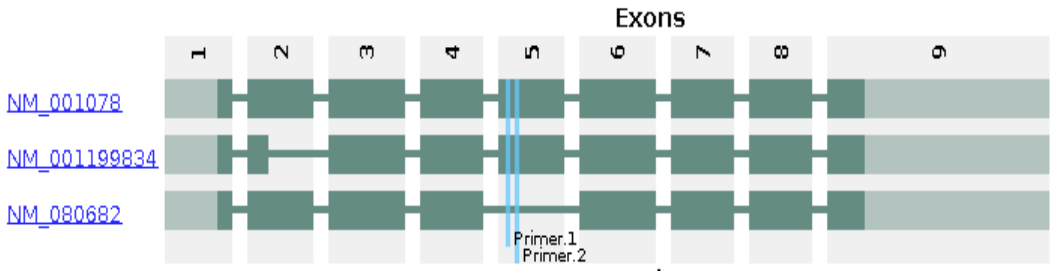
3.9.3 Re-design of primer sets for the RT-qPCR assays

For the new *in-vitro* transcripts, new primer sets were designed using RealTimeDesign™ software (Biosearch Technologies, Novato, USA). The aim was that primers could detect alternative transcripts and splicing variants of the target genes to avoid missing any gene expression. At least 3 primer sets were designed for each candidate predictive gene transcripts. All primer sets were evaluated by RT-qPCR using the *in-vitro* RNA transcripts. The primer sets that had a PCR efficiency lower than 90% and a correlation coefficient that was not within the acceptable range between 0.99 and 0.999, were not considered for further procedures (Table 3.5). Only the primer sets that showed an optimal assay efficiency (> 90%) and correlation coefficient were chosen for further use. These selected primer sets are listed in Table 3.6 and the target location of these primers is shown in Figure 3.8. The selected primers sets for VCAM1, HOPX, MS4A4A, SLC25A5 and ISG15 span within the same exon. Forward primers for TUBG1, TTC28 and TGFBI crosses an intron-exon boundary. Forward and reverse primers for PLPP3, CTSL and NF3L1 are located in different exons.

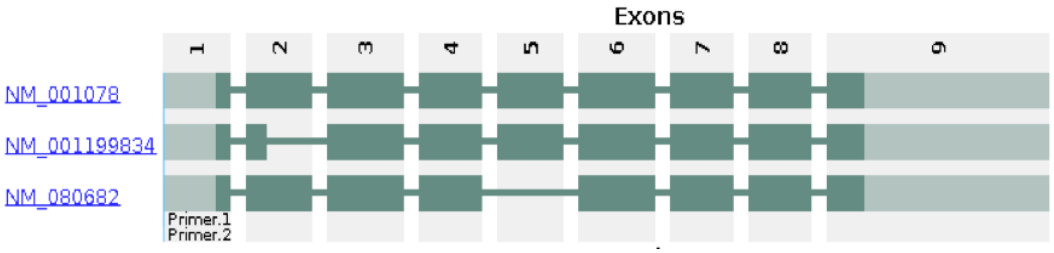
Table 3.5 New primer sets designed that had poor performance

Gene	Sequence 5'- 3'	Slope	r^2 value	PCR Efficiency (E)	Location
VCAM1_a	For: GCTCAGATTGGAGACTCAGTCA Rev: AGAGGGCTGTCTATCTGGGTTCTC	-3.730	1.000	85.4%	Within exon 5
VCAM1_b	For: TTTGACAGGCTGGAGATAGACT Rev: TCAATGTGTAATTTAGCTCGGCA	-3.703	0.999	86.2%	Within exon 3
VCAM1_c	For: CAGTAAGGCAGGCTGTAAAAGA Rev: TGGAGCTGGTAGACCCTCG	-3.630	0.998	88.6%	exon 3, exon 4
VCAM1_d	For: TGGGAAGCCGATCACAGTCAAG Rev: TCTATCTCCAGCCTGTCAAATGGG	-3.578	1.000	90.3%	Within exon 3
VCAM1_g	For: GGTGACTCCGTCTCATTGACT Rev: CAGAGGGCCACTCAAATGAATC	-3.897	1000	80.6%	Exon2,exon 3
HOPX_b	For: GACAAGCACCCGGATTCCA Rev: GTCTGTGACGGATCTGCACTC	-4.018	0.999	77.4%	Exon 3, exon 4
HOPX_c	For: AGTGGCGGCGCTCAGAAG Rev: TGCCTGCCATCTCCTTAGTC	-3.816	0.999	82.8%	Within exon 4
HOPX_e	For: GGCAGGCATTGACAGCTTC Rev: AGCAGGACAGCTAGCAATGGAA	-3.730	0.999	85.4%	Within exon 4
HOPX_i	For: TCCCAGTAGCAATTTGCTTTCATA Rev: GTGTTCCCTTCTCCCTCATATTATC	-3.794	0.999	83.5%	Within exon 4
TUBG1_a	For: CCCAACCAGGACGAGATGA Rev: CTGCGTCAGCCTCTTGAGT	-3.911	0.996	80.2%	Within exon 6
TUBG1_b	For: AGCTGGTGTCTACCATCATGT Rev: CGTAGTGAGAGGGGTGTAGC	-3.755	0.998	84.6%	Exon7/8, exon 8
MS4A4A_a	For: TGTGGAAAGGATTGCAAGAGAAG Rev: TCCCATGCTAAGGCTCATCAG	-3.739	0.999	85.1%	Exon 2, exon 3
MS4A4A_b	For: ACCATGCAAGGAATGGAACAG Rev: TTCCCATGCTAAGGCTCATCA	-4.007	1.000	77.7%	Exon 2, exon 3
MS4A4A_d	For: TGCTCTGTTGTACCCCTGGT Rev: AAACCTCATTAAGTGGTGTGGG	-3.657	0.998	87.7%	Exon 6, exon 7
MS4A4A_e	For:GGTCCGAGGTAGTCTAGGAATGAAT Rev: CGCCAAGCTAAATGTGTTGATTAAG	-3.708	1.000	86.1%	Exon 4/5, exon 5
MS4A4A_f	For: CAGCATCTCCCACACCACTT Rev: TCTGGAGCATTTGTCTGTTGA	-3.630	0.998	88.6%	Within exon 7
TGFBI_b	For: GAAGGGAGACAATCGCTTTAGC Rev: TGTAGACTCCTTCCCGGTTGAG	-3.700	0.998	86.3%	Exon 12, exon 13
TGFBI_c	For: TGGGAGATGCCAAGGAACT Rev: GCCTCCGCTAACCAGGAT	-3.644	0.999	88.1%	Exon 13, exon14
TGFBI_d	For: GTCCATGTCATCACCAATG Rev: CTCCTGTAGTGCTTCAAG	-4.026	0.999	77.2%	Exon 15, exon 17
SLC25A5_b	For: AGTTTTGGCTCTACTTTGCAGG Rev: GGCCCTTAATCCCATCAGATTTG	-3.840	0.998	82.1%	Within exon 2
SLC25A5_d	For: GTGTCTGTGCAGGGTATTATCATC Rev: GGATCCGGAAGCATTCCCTTTG	-3.658	0.999	87.7%	Within exon 2
ISG15_a	For: GAGAGGCAGCGAACTCATC Rev: CAGGGACACCTGGAATTCGTT	-3.690	0.998	86.6%	Exon 1, exon 2
ISG15_c	For: GACGACCTGTTCTGGCTGAC Rev: CGCAGGCGCAGATTCATG	-3.659	0.999	87.6%	Within exon 2

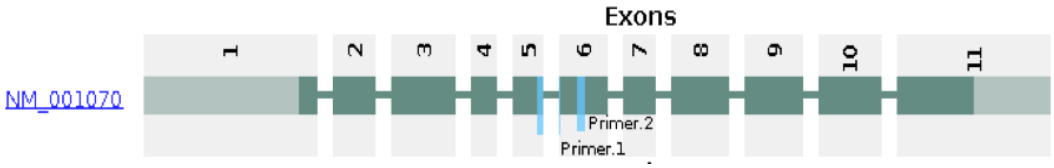
VCAM1



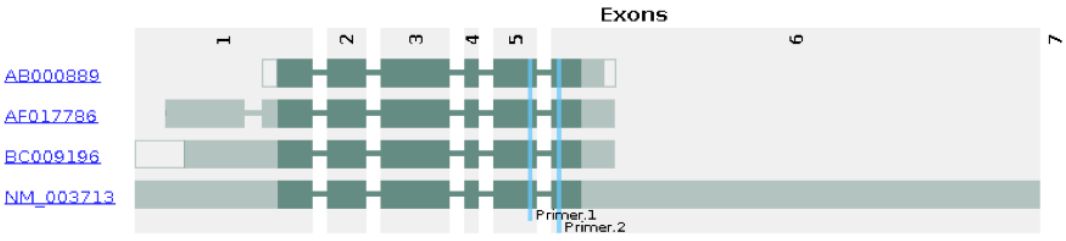
HOPX



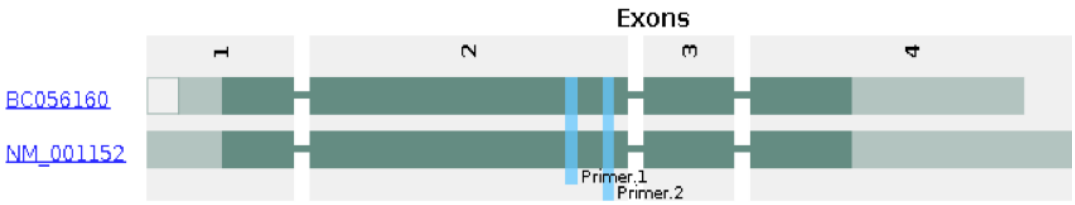
TUBG1



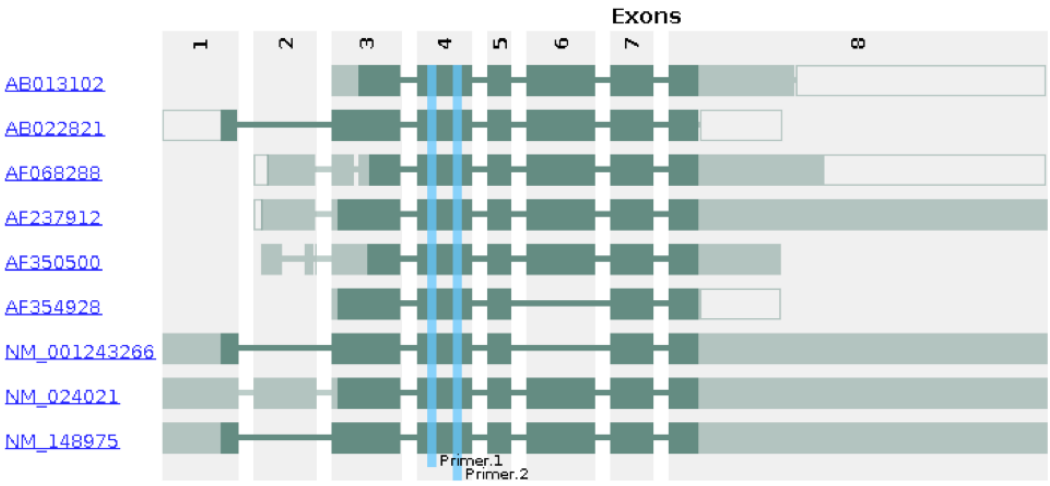
PLPP3



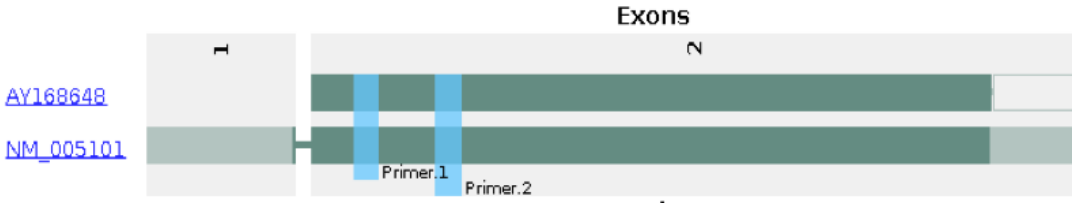
SLC25A5



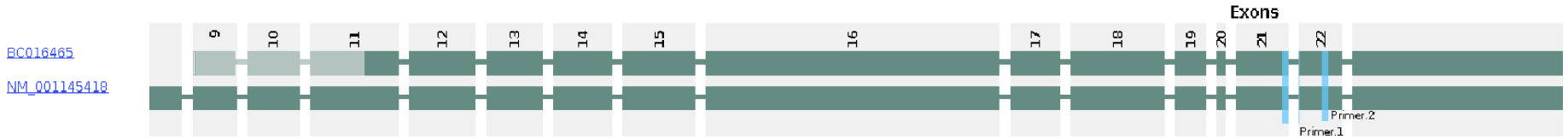
MS4A4A



ISG15



TTC28



TGFBI

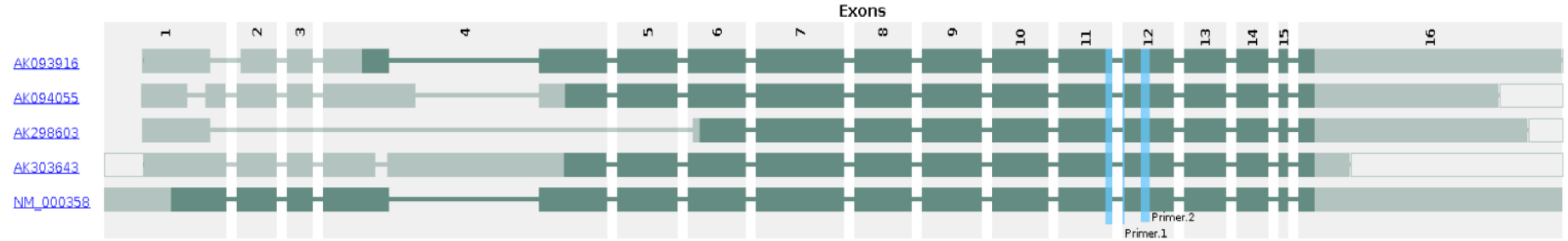


Figure 3.8 Location of the new primer sets. *The figure shows the location of the new primer sets designed for different gene transcripts that demonstrated a good PCR efficiency. The known RefSeq-annotated transcript and splice variants of each gene are identified by accession number (text in blue). The green boxes represent the exons, the lighter green regions are UTRs, and dark green areas are the coding regions. The connecting lines are the introns. The light blue vertical lines show the target locations of the RT-qPCR primers. Primer 1 is the label for primer forward and Primer 2 the label for primer reverse. The figure was created using the tool Primer-check from SpliceCenter (Ryan et al., 2008).*

Table 3.6 Final list of primer sets for the 11 candidate predictive gene transcripts

Gene	NCBI Accession number	Sequences 5'- 3'	Amplicon length (bp)	Nucleotide position
VCAM1	NM_001199834.1	For: CCCGGATTGCTGCTCAGAT Rev: TCACAGCCCATGACACTACA	61	997-1057
HOPX	NM_032495.5	For: CTCAGAGTGCAGATCCGTCA Rev: GCGGAGGAGAGAAACAGAGA	88	784-871
TUBG1	NM_001070.4	For: GAACGGCTGAATGACAGGTATC Rev: GACCACCACATCGCTCATCTC	87	861-947
PLPP3	NM_003713.4	For: GGCAGGATTTGCTCAAGGA Rev: GGGAGAGCGTCGTCTTAGTCT	86	1339-1424
MS4A4A	NM_148975.2	For: TGAGCCTTAGCATGGGAATAACA Rev: CCCGATATACACGGAAATAGGGT	80	362-441
TGFBI	NM_000358.2	For: ACGGAGCAGACTCTTGGGAGATG Rev: CCTCCGCTAACCAGGATTTCATC	84	1823-1906
TTC28	NM_001145418.1	For: TCCTGGCAGCTCTAGGTTT Rev: GCACAGTCCGTCGATTAGC	90	5835-5924
NF3L1*	NM_001136039.2	For: ACACCCAAGACCTGGACAAA Rev: TCTACCACCTGCATCAAAGCC	145	615-759
SLC25A5	NM_001152.4	For: TGATGGGATTAAGGGCCTGTA Rev: GAAGTAGGCGGCTCGGTAG	76	617-692
CTSL*	NM_001257973.1	For: CTGGTGGTTGGCTACGGATT Rev: CTCCGGTCTTTGGCCATCTT	122	610-731
ISG15	NM_005101.3	For: GGCAACGAATTCCAGGTGTC Rev: ACGCCGATCTTCTGGGTGATC	80	141-220

*Primers sets designed for NF3L1 and CTSL remained from the previous list since they showed a high PCR efficiency.

3.9.4 Performance evaluation of the new primer pairs in RT-qPCR assays

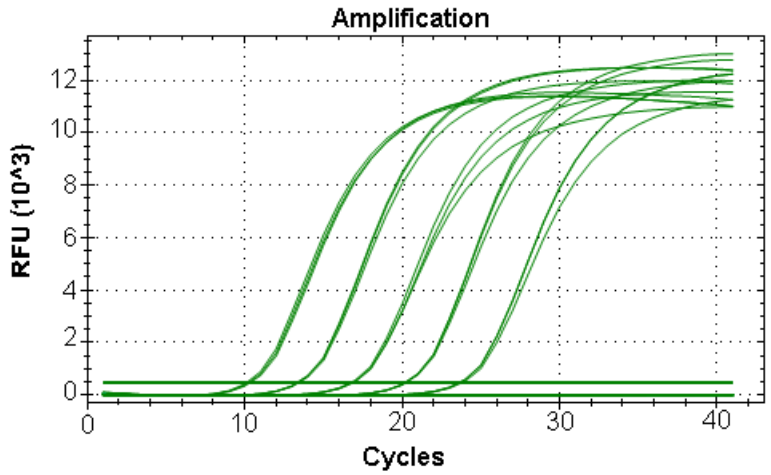
3.9.4.1 PCR efficiency and linearity

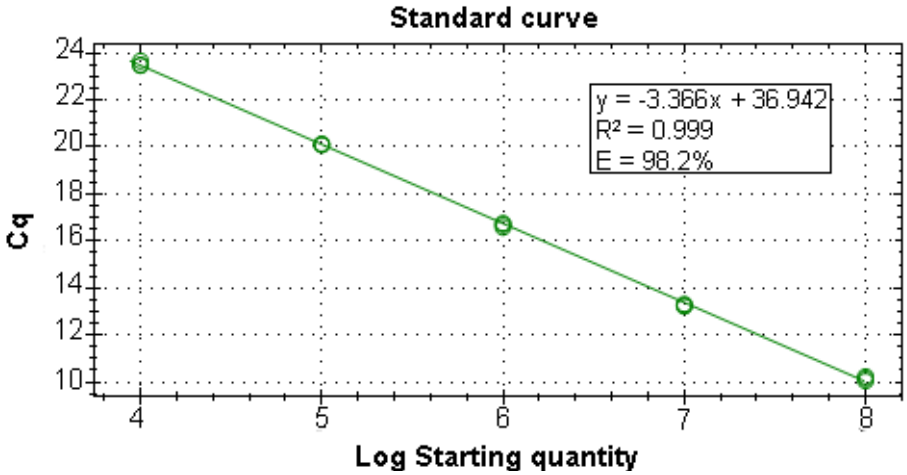
The linearity of the new assays, which is the dynamic range over which a reaction is linear, was established by amplifying three technical replicates and 5 logs of template concentration. To evaluate the linearity, the correlation coefficient (r^2) value was evaluated. As shown in Table 3.7, the r^2 was greater than 0.99 in the assays for all primer pairs, suggesting a good confidence correlating the Ct value and the logarithm of the standard RNA copy number. The PCR efficiencies were also calculated and are shown in Figure 3.9. The results show an overall high PCR efficiency for all genes (> 90.0%) with the exception of HOPX. PCR efficiency for HOPX was 86.7%. Again, with the exception of HOPX, the slope of the standard curves were in the acceptable range, between -3.3 to -3.6. The primers sets for HOPX were kept despite the low efficiency to proceed to the next steps that includes the evaluation of the optimized RT-qPCRs in human samples.

Table 3.7 Performance evaluation of the new optimized RT-qPCR assays

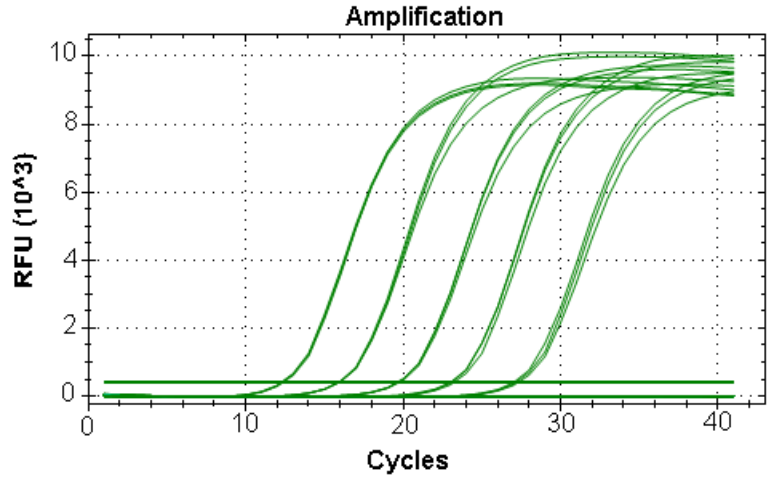
Gene	Length of cRNA (b)	qRT-PCR product size (bp)	Slope	r^2 value	PCR Efficiency (E)
VCAM1	1049	61	-3.367	0.999	98.2%
HOPX	1053	88	-3.689	0.999	86.7%
TUBG1	1049	87	-3.421	0.999	96.0%
PLPP3	1049	86	-3.387	1.00	97.3%
MS4A4A	1090	80	-3.586	0.999	90.1%
TGFBI	1049	84	-3.390	0.999	97.2%
TTC28	1049	90	-3.366	0.999	98.2%
SLC25A5	1084	76	-3.531	0.998	92.0%
ISG15	727	80	-3.428	0.998	95.8%

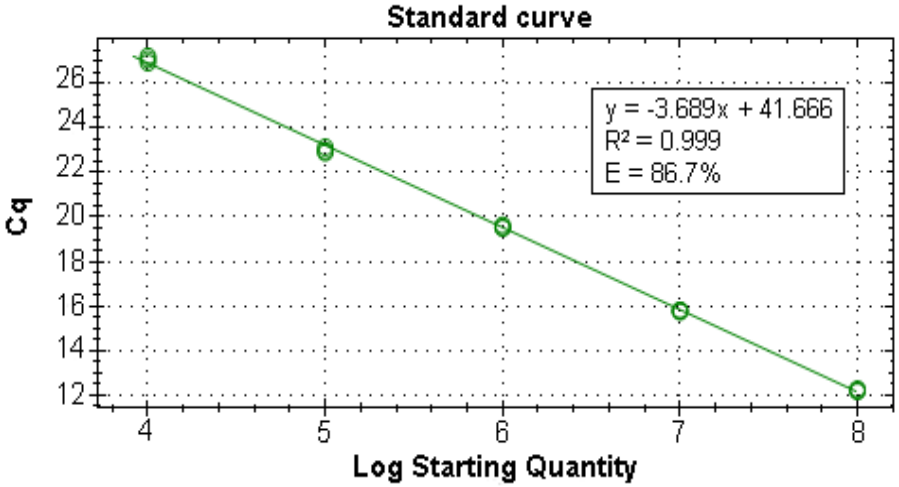
VCAM1


Amplification
RFU (10^3)
Cycles

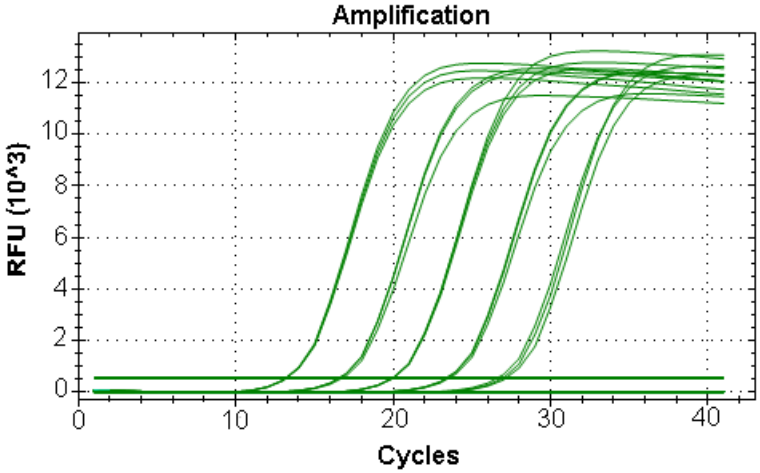

Standard curve
y = -3.366x + 36.942
R² = 0.999
E = 98.2%
Cq
Log Starting quantity

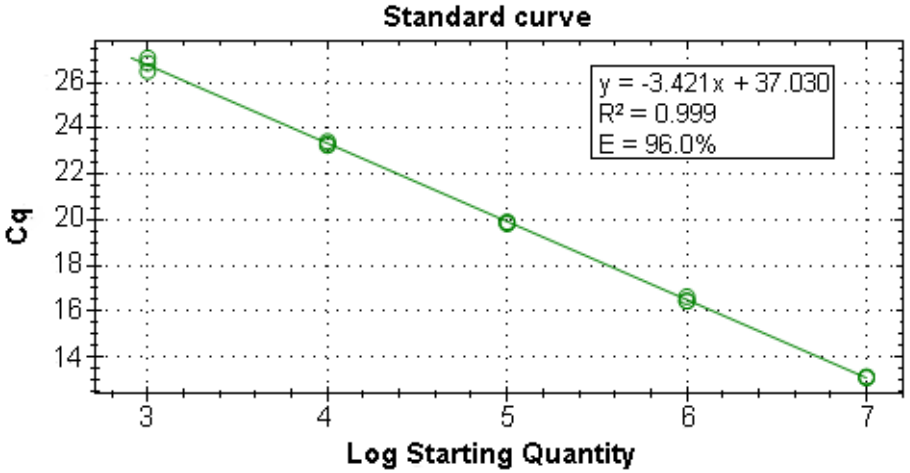
HOPX


Amplification
RFU (10^3)
Cycles

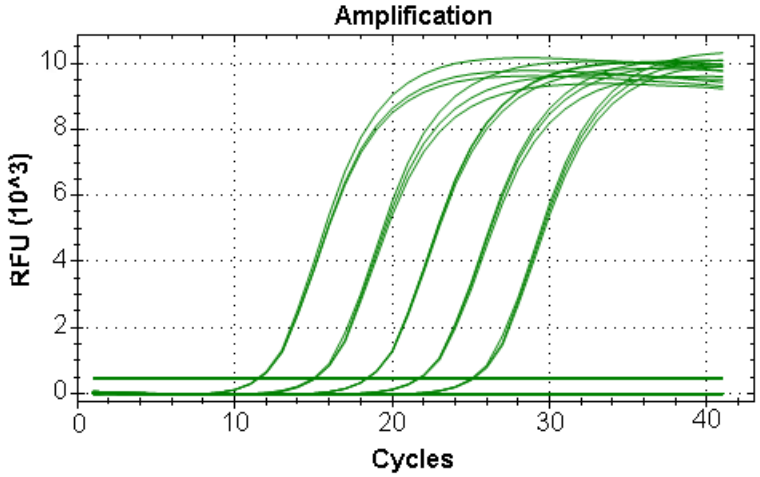

Standard curve
y = -3.689x + 41.666
R² = 0.999
E = 86.7%
Cq
Log Starting Quantity

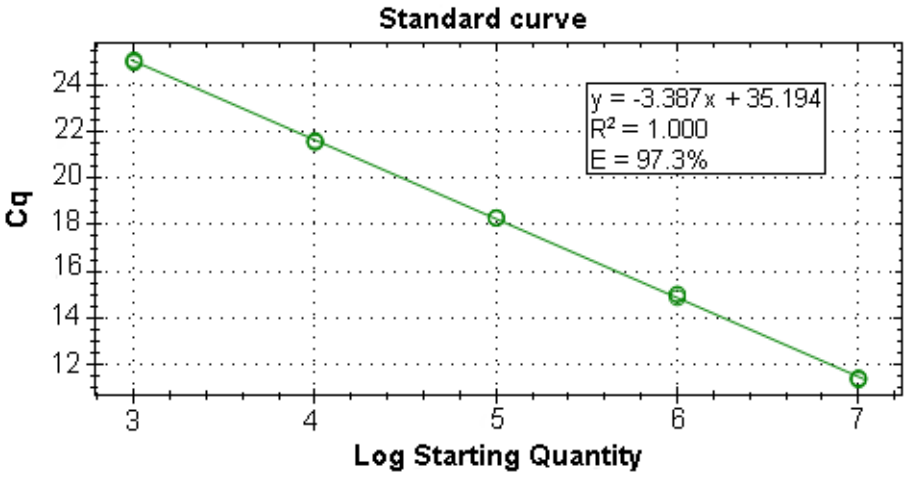
TUBG1


Amplification
RFU (10^3)
Cycles

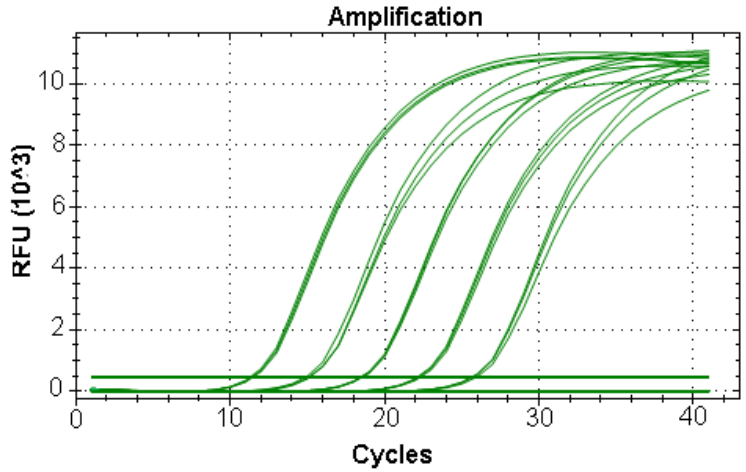

Standard curve
y = -3.421x + 37.030
R² = 0.999
E = 96.0%
Cq
Log Starting Quantity

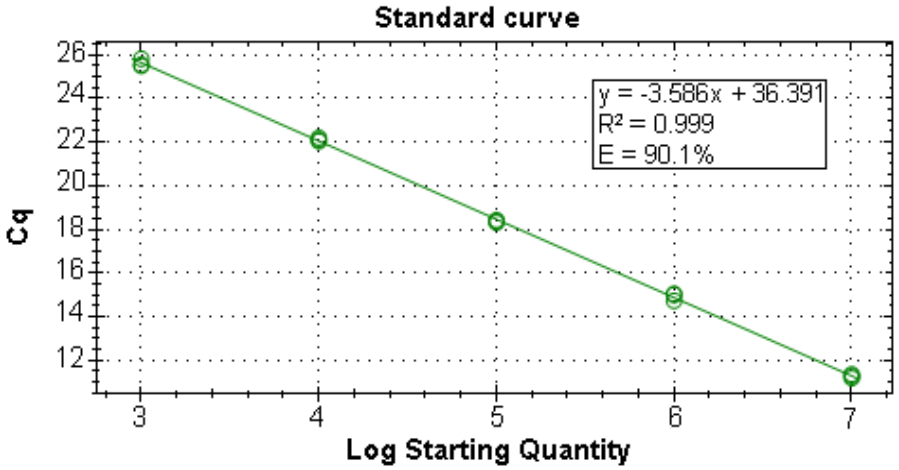
PLPP3


Amplification
RFU (10^3)
Cycles

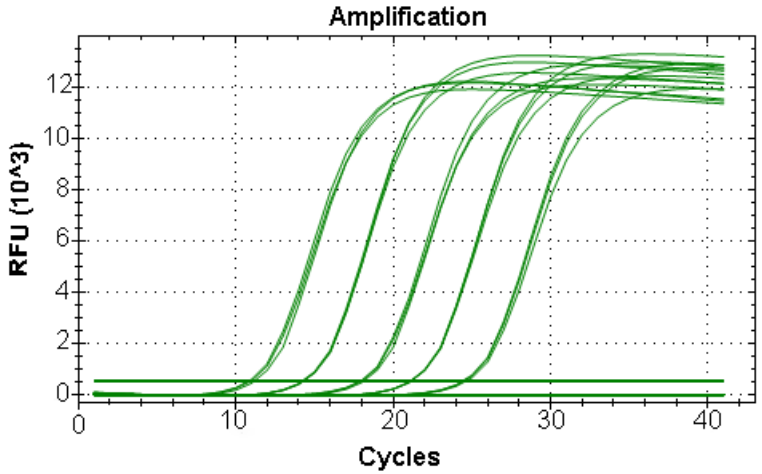

Standard curve
y = -3.387x + 35.194
R² = 1.000
E = 97.3%
Cq
Log Starting Quantity

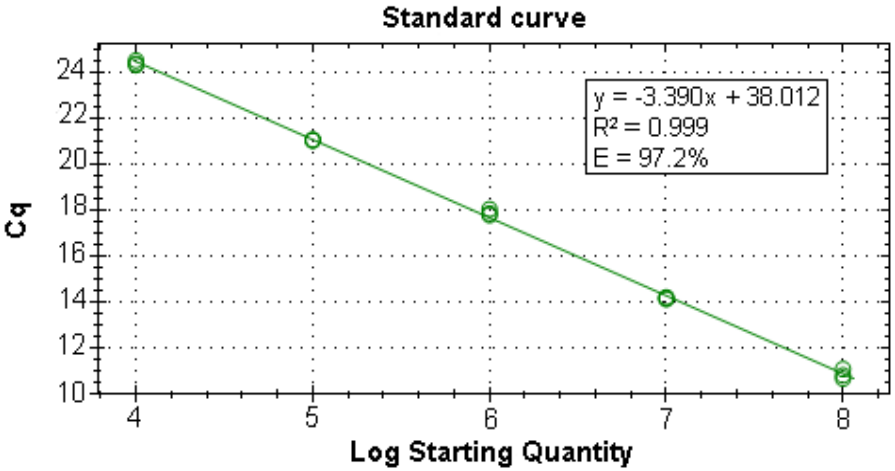
MS4A4A



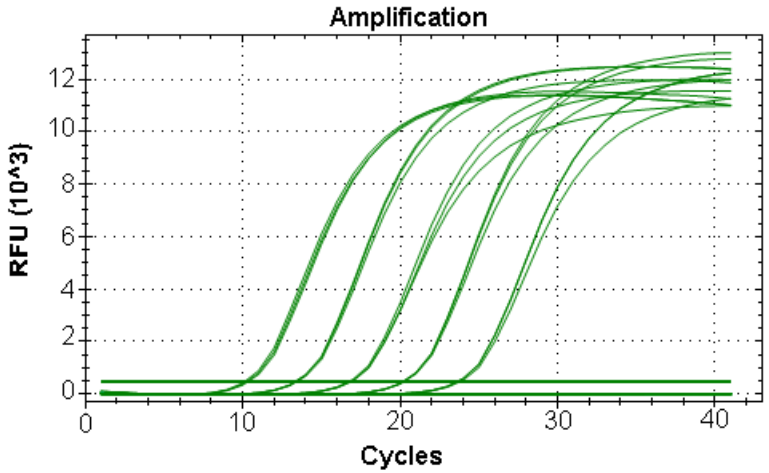


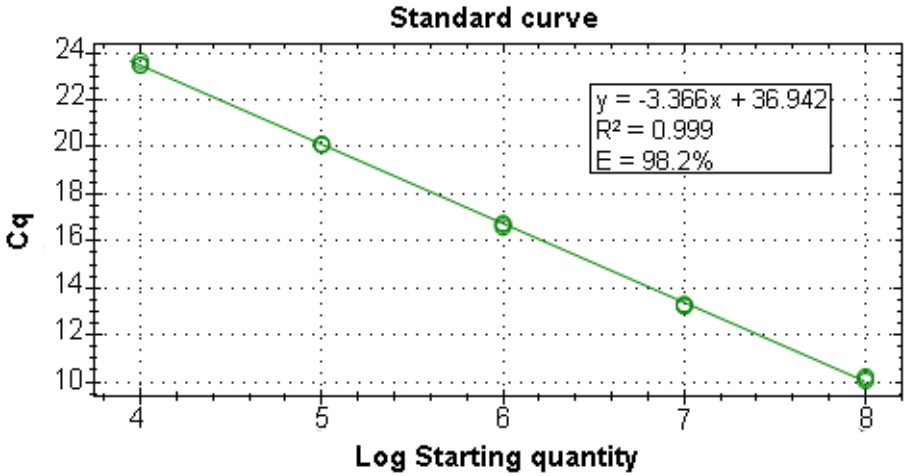
TGFBI



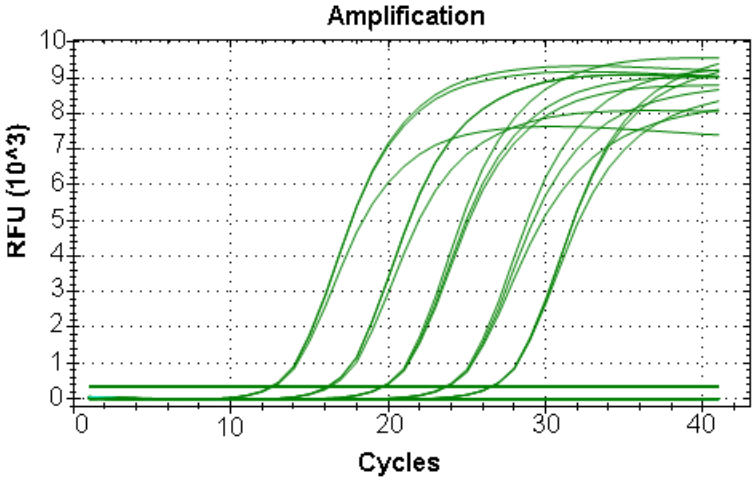


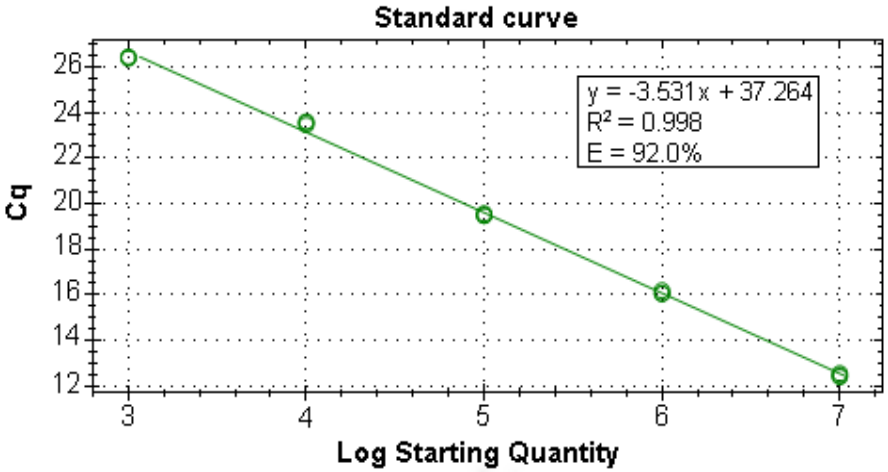
TTC28





SLC25A5





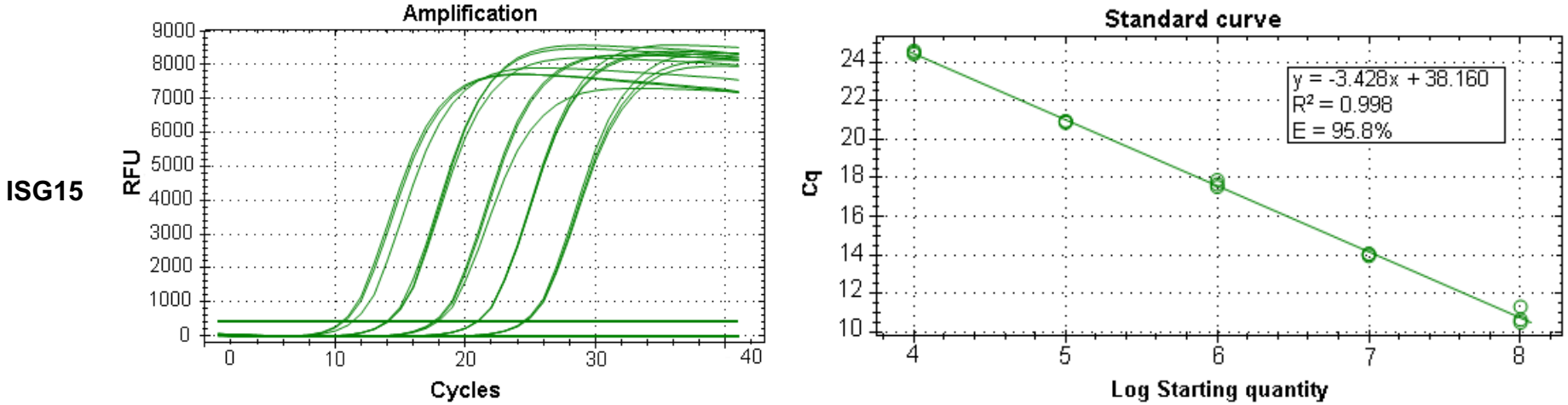
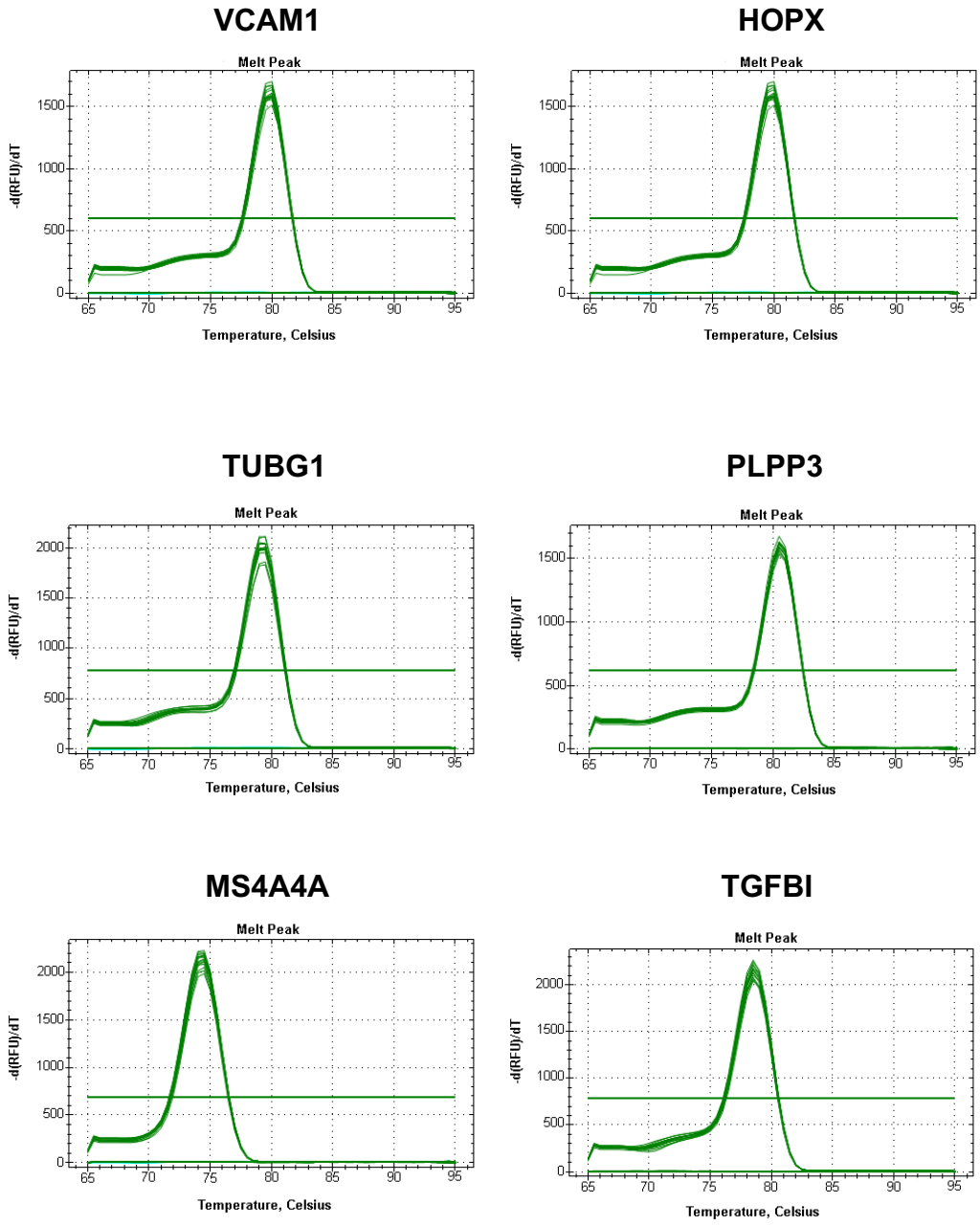


Figure 3.9 Linearity and PCR efficiency of the new optimized assays. *Left panels show the amplification chart of the five cRNA standards for each gene, with dilutions ranging from 10^3 to 10^7 or 10^4 to 10^8 copies/µl and tested in triplicate. Right panels show the standard curves created with these dilutions. The mean Cq values are plotted against the log of concentration of the target gene copies/reaction. The PCR efficiency for each target was calculated using the slope of the calibration curve with the formula $E = 10^{-1/slope}-1$. E = PCR efficiency, R^2 = correlation coefficient.*

3.10 Specificity of the optimized RT-qPCR assays for the 11 gene transcripts

The specificity of the assays was evaluated by melt curve analysis, which can detect non-sequence specific products that may have co-amplified and produce melt curves with a different melting temperature. The melt curve analysis of the optimized RT-qPCR assays for the 11 genes detected single peaks, and no non-specific products or primer dimers were observed (Figure 3.10). Thus, indicating the amplification of specific amplicons.



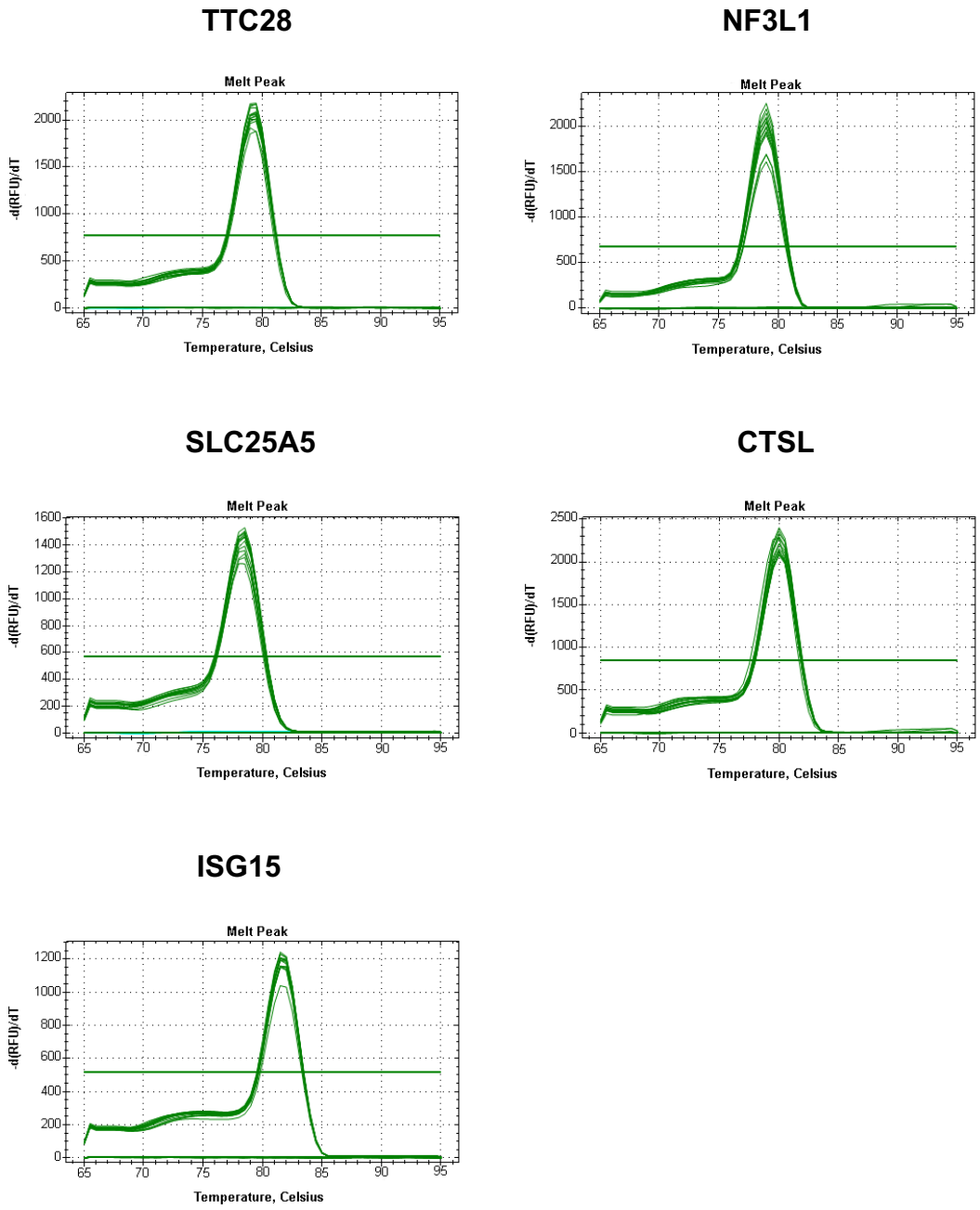


Figure 3.10 Melting curve analysis of all the optimized RT-qPCR assays. *The figures show the melting peaks of the RT-qPCR products obtained from the cRNA dilutions used to construct the standard curve. A single peak is visible for all primer pairs and genes tested.*

3.11 Analytical sensitivity (Limit of detection) of the optimized RT-qPCR assay

The Limit of Detection (LOD) was only determined for the VCAM1 assay as a representation of the set of 11 genes that had the same RT-qPCR conditions. This was done by testing the diluted *in-vitro* transcripts from 10^8 to 10^0 in duplicate by RT-qPCR. The minimum detection limit of the RT-qPCR assay was 10 RNA copies/µl, the 10^0 dilution did not amplify, and no band was observed in the 10^0 dilution when the PCR products were analysed by agarose gel electrophoresis (Figure 3.11).

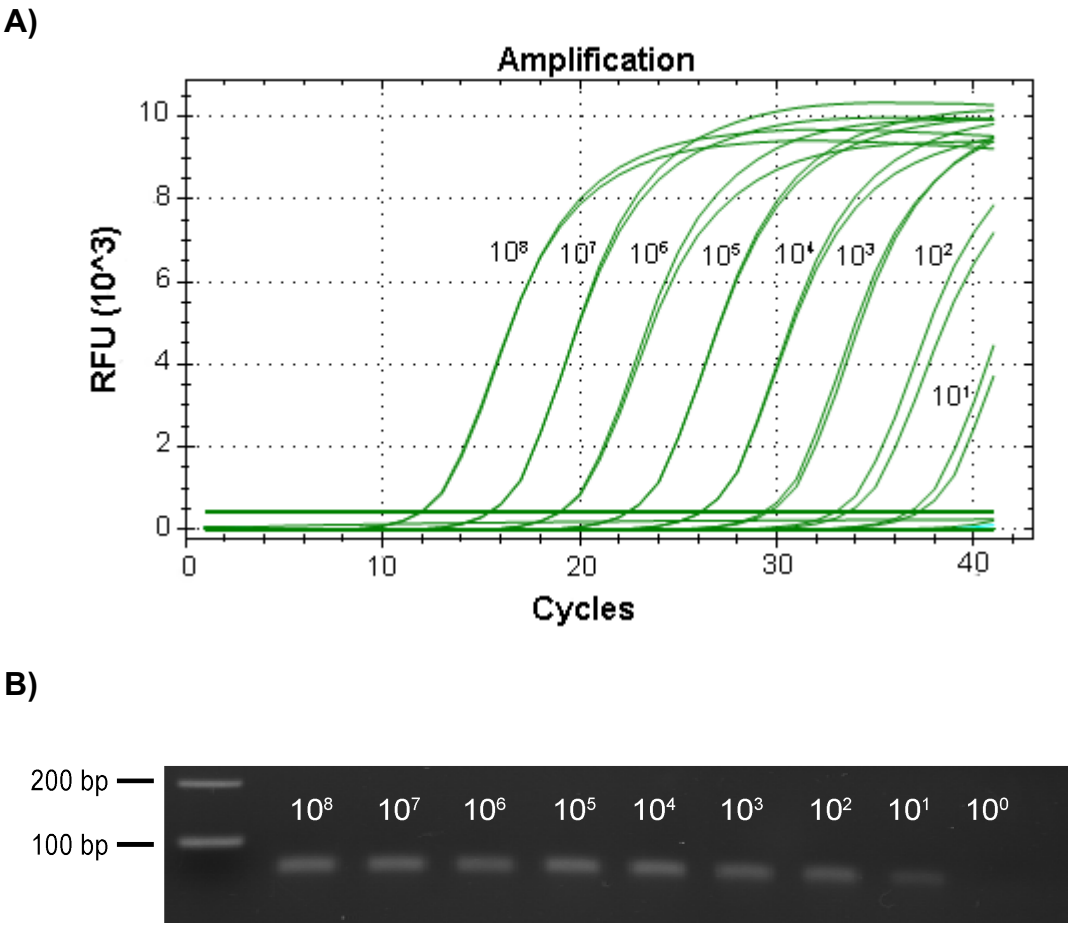


Figure 3.11 Sensitivity of the SYBR Green RT-qPCR assay for VCAM1 transcript. *A) Chart showing the serial dilutions of VCAM1 cRNA ranging from 10^8 to 10^0 copies/µl amplified by RT-qPCR. The lowest dilution that amplified was 10 copies. Negative control (no template added) is shown in light blue and did not amplified. B) Agarose gel analysis of the RT-qPCR products; the lowest dilution detected was 10 copies.*

3.12 Precision of the optimized RT-qPCR assays for the 11 gene transcripts

The precision parameter was evaluated in two ways: intra-assay variation and inter-assay variation. The intra-assay variability was assessed by analysing the Ct values obtained from 3 replicates at each standard dilution run in the same plate. Good repeatability of the optimized assay was observed since low levels of coefficient of variation (CV) were obtained at every RNA concentration in the standard curves (highest CV was 4.03% and standard deviation was 0.44) (Table 3.8). Likewise, the inter-assay variability was evaluated from the Ct values obtained in three different runs, thereby in different plates. The highest CV obtained was 11.70% and standard deviation of 1.15 (Table 3.9).

Table 3.8 Intra-assay variability of the optimized singleplex RT-qPCRs for the 11 gene transcripts

RT_qPCR assay		Copy number of cRNA standards					
		1.00E+08	1.00E+07	1.00E+06	1.00E+05	1.00E+04	1.00E+03
VCAM1	Ct values mean±SD	ND†	16.17±0.25	19.38±0.08	23.04±0.01	26.36±0.11	29.52±0.02
	CV*	ND	1.55%	0.40%	0.03%	0.41%	0.06%
HOPX	Ct values mean±SD	12.23±0.06	15.79±0.03	19.56±0.07	23.00±0.13	27.07±0.14	ND
	CV	0.46%	0.22%	0.35%	0.56%	0.52%	ND
TUBG1	Ct values mean±SD	ND	13.11±0.02	16.51±0.11	19.87±0.04	23.32±0.09	26.81±0.29
	CV	ND	0.16%	0.68%	0.20%	0.37%	1.09%
PLPP3	Ct values mean±SD	ND	11.4±0.04	14.97±0.08	18.3±0.02	21.58±0.04	25.03±0.06
	CV	ND	0.35%	0.54%	0.08%	0.19%	0.22%
MS4A4A	Ct values mean±SD	ND	11.28±0.08	14.93±0.17	18.37±0.05	22.11±0.08	25.62±0.17
	CV	ND	0.71%	1.17%	0.27%	0.37%	0.65%
TGFBI	Ct values mean±SD	10.87±0.21	14.18±0.04	17.90±0.13	21.04±0.02	24.39±0.10	ND
	CV	1.93%	0.25%	0.72%	0.10%	0.42%	ND
TTC28	Ct values mean±SD	10.14±0.09	13.26±0.05	16.68±0.09	20.11±0.01	23.55±0.11	ND
	CV	0.89%	0.35%	0.55%	0.09%	0.46%	ND
NF3L1	Ct values mean±SD	15.61±0.25	18.88±0.06	22.57±0.45	25.72±0.15	29.51±0.06	ND

	CV	1.61%	0.34%	1.98%	0.59%	0.19%	ND
SLC25A5	Ct values mean±SD	ND	12.48±0.08	16.1±0.08	19.51±0.05	23.54±0.07	26.41±0.03
	CV	ND	0.61%	0.47%	0.25%	0.31%	0.11%
CTSL	Ct values mean±SD	ND	12.21±0.06	15.85±0.14	19.19±0.09	22.40±0.11	26.32±0.17
	CV	ND	0.51%	0.86%	0.48%	0.47%	0.64%
ISG15	Ct values mean±SD	10.82± 0.44	14.02±0.06	17.70±0.18	20.91±0.04	24.51±0.09	ND
	CV	4.03%	0.46%	1.03%	0.18%	0.36%	ND

*CV = Coefficient of variation, †ND = not detected

Table 3.9 Inter-assay variability of the optimized singleplex RT-qPCRs for the 11 gene transcripts

RT_qPCR assay		Copy number of cRNA standards					
		1.00E+08	1.00E+07	1.00E+06	1.00E+05	1.00E+04	1.00E+03
VCAM1	Ct values mean±SD	ND†	16.14±0.37	19.29±0.19	22.81±0.23	26.30±0.05	29.59±0.34
	CV	ND	2.30%	0.98%	1.00%	0.20%	1.15%
HOPX	Ct values mean±SD	12.36±0.18	16.11±0.45	19.51±0.07	23.04±0.06	27.12±0.07	ND
	CV	1.43%	2.77%	0.36%	0.25%	0.26%	ND
TUBG1	Ct values mean±SD	ND	13.27±0.20	16.72±0.30	20.18±0.32	23.63±0.36	27.05±0.21
	CV	ND	1.48%	1.81%	1.59%	1.54%	0.76%
PLPP3	Ct values mean±SD	ND	11.62±0.70	14.98±0.70	18.43±0.70	21.78±0.62	25.49±0.82
	CV	ND	6.04%	4.64%	3.80%	2.89%	3.22%
MS4A4A	Ct values mean±SD	ND	12.00±0.01	15.93±0.26	19.70±0.12	23.29±0.10	26.97±0.28
	CV	ND	0.10%	1.61%	0.58%	0.43%	1.05%
TGFBI	Ct values mean±SD	10.11±0.89	13.67±0.58	17.26±0.65	20.68±0.52	24.22±0.42	ND
	CV	8.75%	4.25%	3.77%	2.50%	1.71%	ND
TTC28	Ct values mean±SD	9.79±1.15	13.34±0.51	16.81±0.50	20.41±0.44	23.76±0.52	ND
	CV	11.70%	3.82%	2.99%	2.18%	2.20%	ND
NF3L1	Ct values mean±SD	15.77±0.45	19.17±0.42	22.71±0.48	26.20±0.47	29.70±0.33	ND
	CV	2.87%	2.18%	2.11%	1.80%	1.13%	ND
SLC25A5	Ct values mean±SD	ND	11.74±0.65	15.35±0.72	18.71±0.69	22.54±0.88	25.65±0.67
	CV	ND	5.55%	4.68%	3.69%	3.89%	2.63%

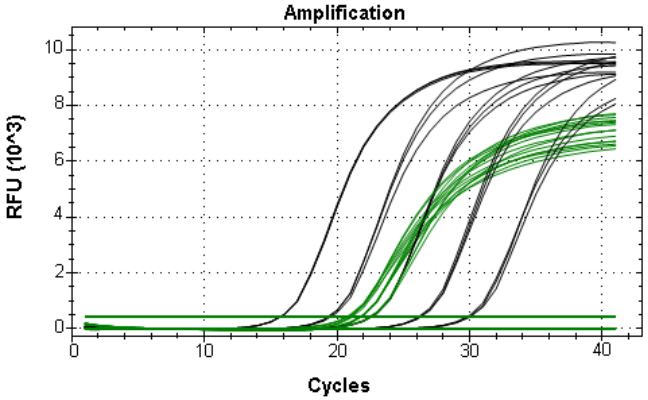
CTSL	Ct values mean±SD	ND	12.01±0.30	15.49±0.34	19.03±0.23	22.34±0.17	26.07±0.27
	CV	ND	2.53%	2.20%	1.19%	0.77%	1.04%
ISG15	Ct values mean±SD	10.64±0.16	14.00±0.16	17.59±0.20	20.98±0.25	24.54±0.19	ND
	CV	1.51%	1.11%	1.13%	1.20%	0.78%	ND

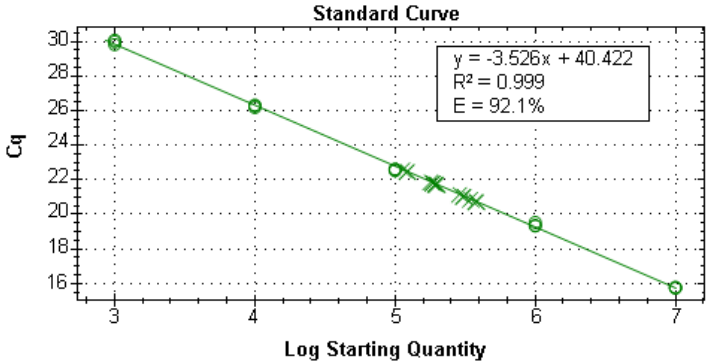
CV = Coefficient of variation, †ND = not detected

3.13 Performance evaluation of all the optimized assays in control samples

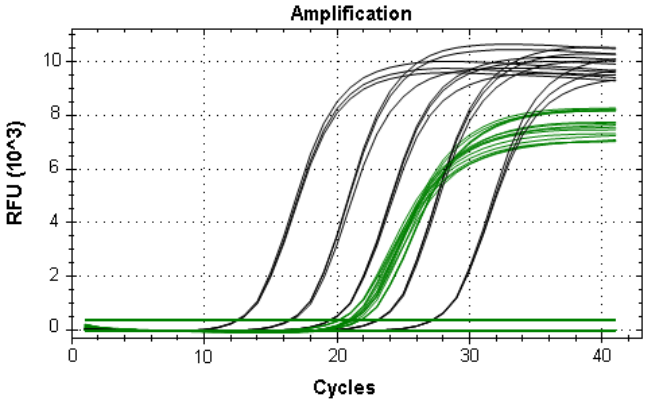
Blood samples from five healthy donors were obtained to evaluate the performance of the RT-qPCR assays using RNA derived from samples using a similar methodology as was used in the field to collect RNA from EVD patients. Thus, RNA was extracted directly from whole blood using a spin-column procedure and the transcript abundance of the 11 genes measured by RT-qPCR using the primer and conditions defined in section 2.10. The samples from each of the five donors was tested in triplicate and the RNA copy number for each gene was obtained from the extrapolation of the Ct into the standard curve (Figure 3.12). In all the assays, the amplification curves for the *in-vitro* transcript dilutions were tight, indicating a minimum sample to sample variation in all the assays. Likewise, the data obtained for each replicate of the control samples was consistent. The variation of the Ct values was less than 1 cycle in both *in-vitro* transcripts dilutions and control samples. The PCR efficiency was high (> 90%) in the RT-qPCR assays for VCAM1, TUBG1, PLPP3, TGFBI, TTC28, NF3L1, SLC25A5, CTSL and ISG15 with slopes within the acceptable range of -3.2 to -3.5. This is consistent with previous results shown in Table 3.7 and Table 3.3 for NF3L1 and CTSL. The PCR efficiency for HOPX was 89.5 % and for MS4A4A was 88.6% with slopes of -3.6. This PCR efficiency does not differ much from previous results where PCR efficiency for HOPX was 86.7% and for MS4A4A was 90.1% (shown in Table 3.7). This difference can be attributed to the slight variation between Ct values in some *in-vitro* transcripts and control samples, however still less than 1 cycle. A good correlation coefficient (> 0.99) was obtained in the RT-qPCR assays for all the gene transcripts.

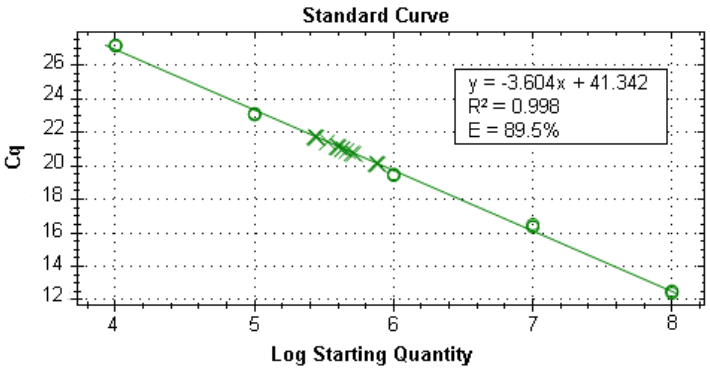
VCAM1


Amplification
RFU (10^3)
Cycles

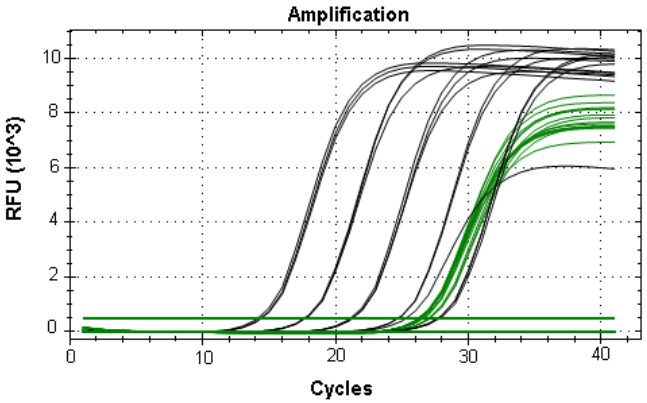

Standard Curve
y = -3.526x + 40.422
R² = 0.999
E = 92.1%
Cq
Log Starting Quantity

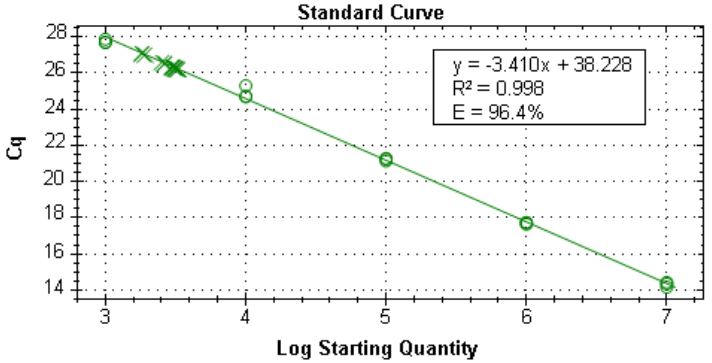
HOPX


Amplification
RFU (10^3)
Cycles

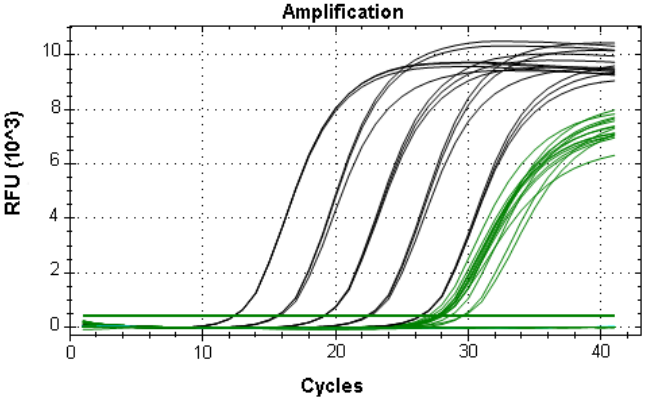

Standard Curve
y = -3.604x + 41.342
R² = 0.998
E = 89.5%
Cq
Log Starting Quantity

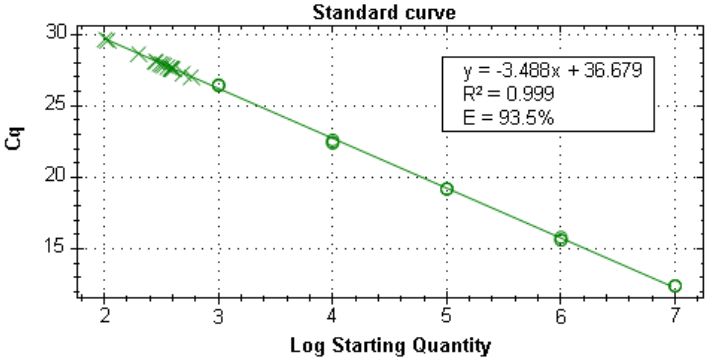
TUBG1


Amplification
RFU (10^3)
Cycles

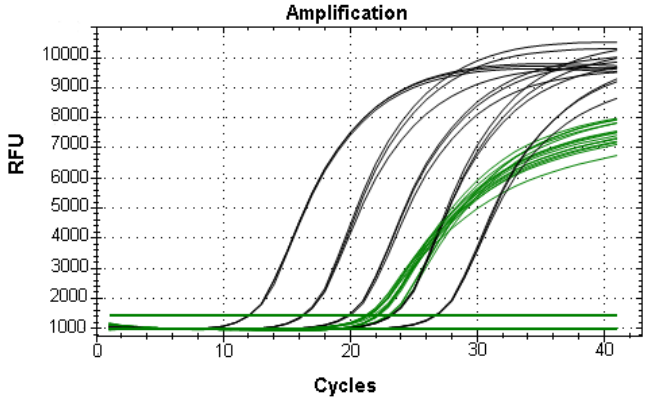

Standard Curve
y = -3.410x + 38.228
R² = 0.998
E = 96.4%
Cq
Log Starting Quantity

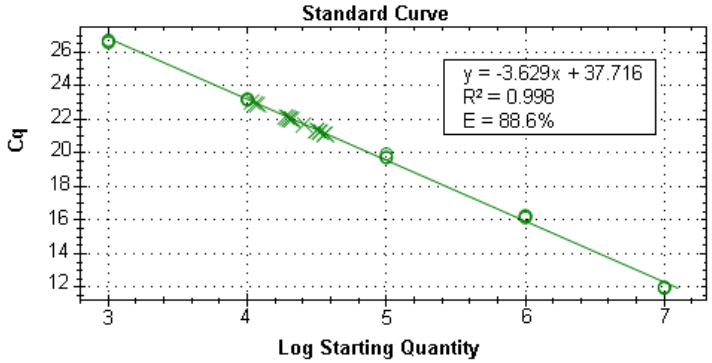
PLPP3


Amplification
RFU (10^3)
Cycles

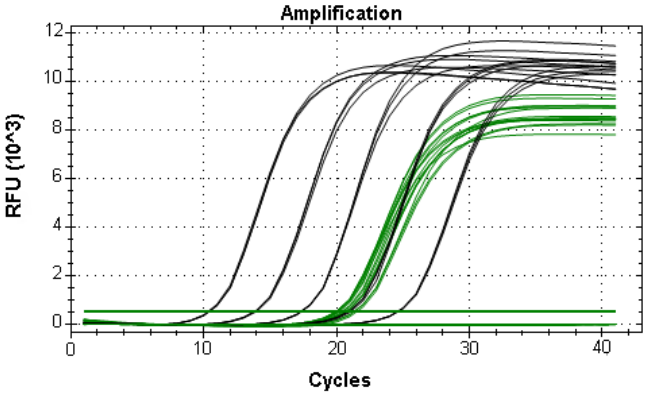

Standard curve
y = -3.488x + 36.679
R² = 0.999
E = 93.5%
Cq
Log Starting Quantity

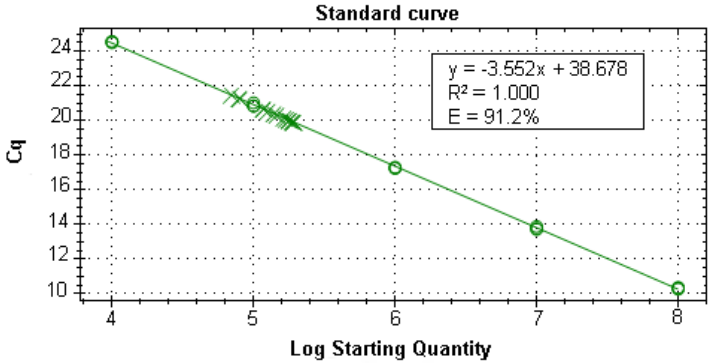
MS4A4A



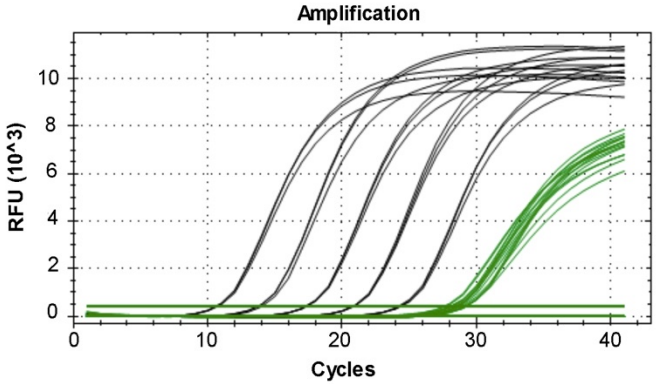


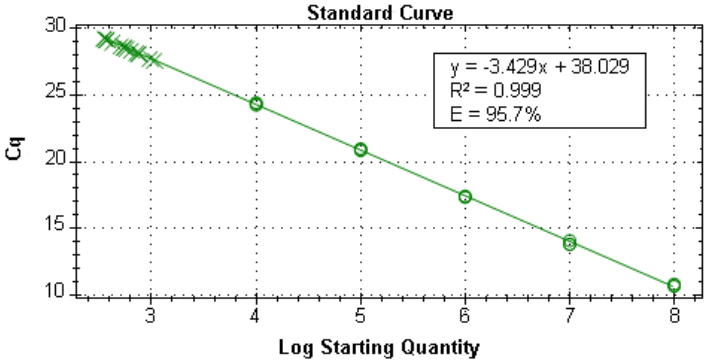
TGFBI



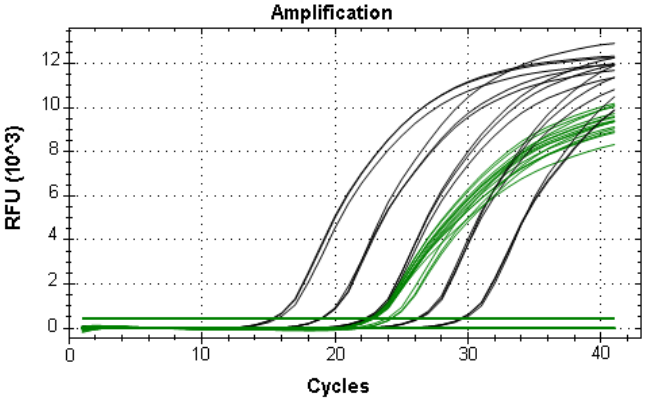


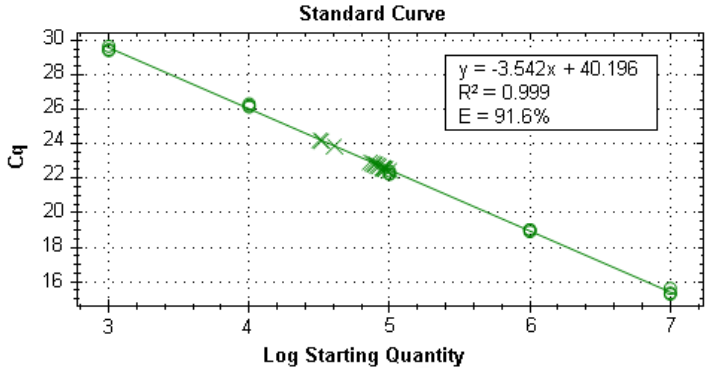
TTC28


Amplification
RFU (10^3)
Cycles

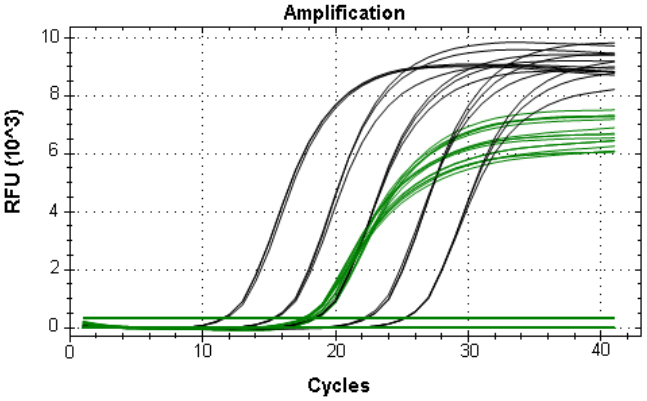

Standard Curve
y = -3.429x + 38.029
R² = 0.999
E = 95.7%
Cq
Log Starting Quantity

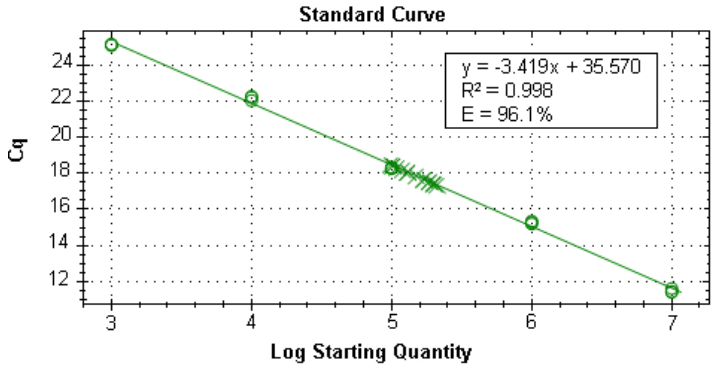
NF3L1


Amplification
RFU (10^3)
Cycles

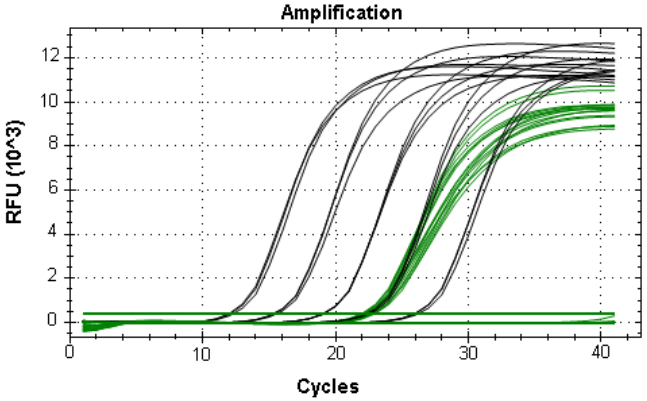

Standard Curve
y = -3.542x + 40.196
R² = 0.999
E = 91.6%
Cq
Log Starting Quantity

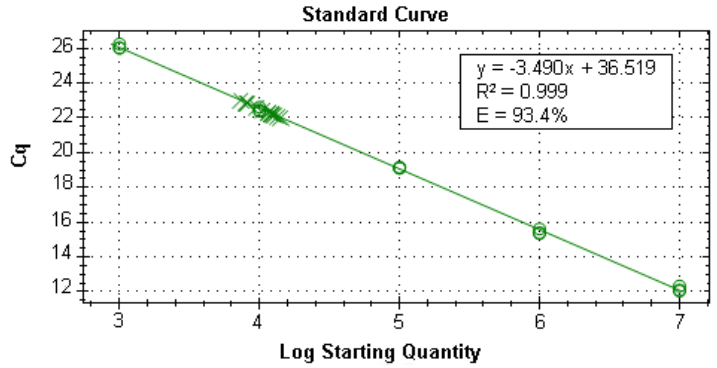
SLC25A5





CTSL





ISG15

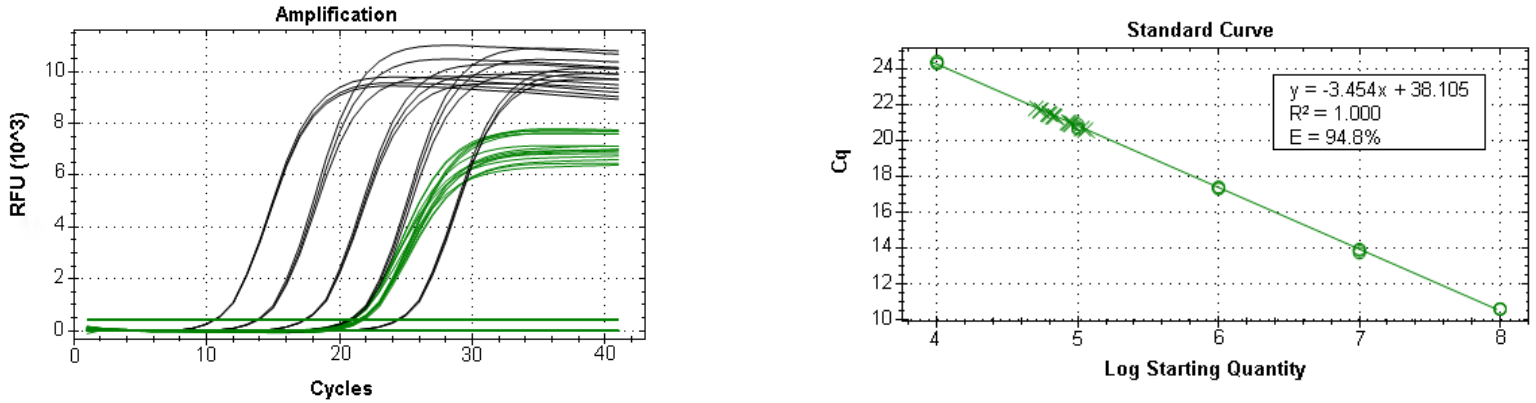


Figure 3.12 Determining the transcript abundance of the 11 genes in control samples by RT-qPCR. *RNA extracted from whole blood of 5 healthy donors was used as a template in each PCR reaction. Left panel shows the amplification chart of the RT-qPCR assays, in each run the in-vitro standards were tested in triplicate (black curves), control samples (green curves) and negative controls (no template added) were tested in duplicate. The right panel shows the standard curve used to determine the transcript abundance of each gene in the control samples by the extrapolation of the Cq values. The circles represent in-vitro RNA standards, and X represent the control samples. E= PCR efficiency, R^2 = correlation coefficient.*

The mean copy number and Ct value from each gene transcript was obtained from calculating the mean of the five donors. The highest transcript abundance was found in HOPX (466291 copies/µl, Ct = 21.01) and the lowest transcript abundance was found in PLPP3 (329.7 copies/µl, Ct = 28.04). The mean copy number and Ct values for all the genes transcripts are shown in Table 3.10. The mean copy number for each gene transcript is also summarized in Figure 3.13.

Table 3.10 Mean RNA copy number and Ct value of the predictive genes in control samples

Gene	Mean RNA copy number	Mean Ct value
VCAM1	234697 (95% CI: 110728 - 358666)	21.60 (95% CI: 20.76 - 22.44)
HOPX	466291 (95% CI: 239224 - 693357)	21.01 (95% CI: 20.28 – 21-73)
TUBG1	2773 (95% CI: 2075 - 3471)	26.52 (95% CI: 26.10 – 26.93)
PLPP3	329.7 (95% CI: 181.6 – 477.8)	28.04 (95% CI: 27.13 – 28.95)
MS4A4A	24236 (95% CI: 12139 – 36333)	21.93 (95% CI: 21.04 – 22.81)
TGFBI	142969 (95% CI: 86216 – 199722)	20.44 (95% CI: 19.74 – 21.15)
TTC28	601.1 (95% CI: 318.1 – 884.1)	28.59 (95% CI: 27.90 – 29.28)
NF3L1	76297 (95% CI: 46530 – 106065)	22.99 (95% CI: 22.21 – 23.78)
SLC25A5	151556 (95% CI: 103843 – 199269)	17.91 (95% CI: 17.41 – 18.40)
CTSL	11012 (95% CI: 8250 – 13774)	22.45 (95% CI: 22.04 – 22.85)
ISG15	75409 (95% CI: 49904 – 100914)	21.30 (95% CI: 20.80 – 21.80)

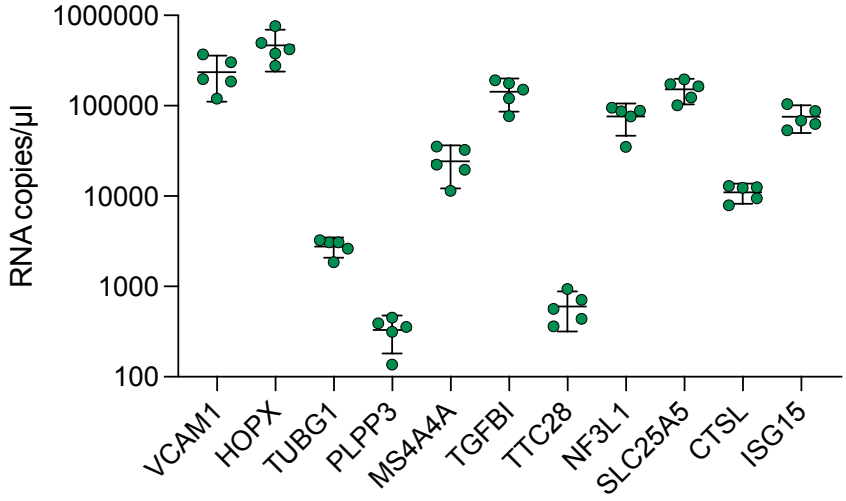


Figure 3.13 Transcript abundance of the 11 candidate predictive genes in control samples. *The figure shows the transcript abundance of each gene detected by RT-qPCR in 5 healthy donors. A green dot represents an individual, and bars show the mean, and the 95% confidence interval.*

3.14 Performance evaluation of the optimized assays in clinical samples (pilot study)

The next step was to apply and test these primers sets using RNA derived from clinical samples from EVD patients. We used leftover samples collected by the EMLab deployed in Guinea that were currently available in the lab, therefore there was not enough RNA sample to analyse all genes. Only VCAM1 was analysed in these samples. A total of fifteen clinical samples (9 survivors, and 6 fatal cases) were tested in duplicate to evaluate the performance of the optimized VCAM1 RT-qPCR assay and are listed in Table 3.11.

Table 3.11 Details of the clinical samples used for the pilot study

Sample ID	EBOV Ct value	Outcome	RNA concentration (ng/µl)
EM_0596	13.71	fatal	30.4
EM_080094	14.14	fatal	7.34
EM_080103	14.22	fatal	4.40
EM_075288	14.32	fatal	5.66
EM_076636	14.42	fatal	18.7
EM_080073	14.46	fatal	4.20
EM_076821	23.96	survivor	19.00
EM_076807	23.05	survivor	5.40
EM_076778	24.89	survivor	7.60
EM_076777	23.01	survivor	9.28
EM_075882	23.84	survivor	11.40
EM_075872	21.69	survivor	6.94
EM_075909	20.72	survivor	4.32
EM_075898	22.01	survivor	4.80
EM_076603	24.64	survivor	10.5

Figure 3.14 shows the transcript abundance of VCAM1 obtained from testing the clinical samples by RT-qPCR and is compared to the RNA copy number obtained from control samples in the previous section. From this data, we can see that fatal cases have a tendency to have higher transcript abundance of VCAM1 compared to survivors and non-infected individuals. There was a significant difference in the VCAM1 transcript copy number between controls and fatal cases ($p = 0.0173$) but not between controls and survivors ($p = 0.1898$) nor survivors and fatal cases ($p = 0.4559$). The mean RNA copy number calculated from the five control samples was 234697 (95% CI: 110728-358666) copies/µl and mean Ct value was 21.60. The mean RNA copy number calculated from all survivors was 408703 (95% CI: 259066-558339) copies/µl with a mean Ct value of 21.42 (95% CI: 20.73 – 22.10), and from all fatal cases the mean RNA copy number was 503675 (95% CI: 343586-663764) copies/µl with a mean Ct value of 20.95 (95% CI: 20.42- 21.48).

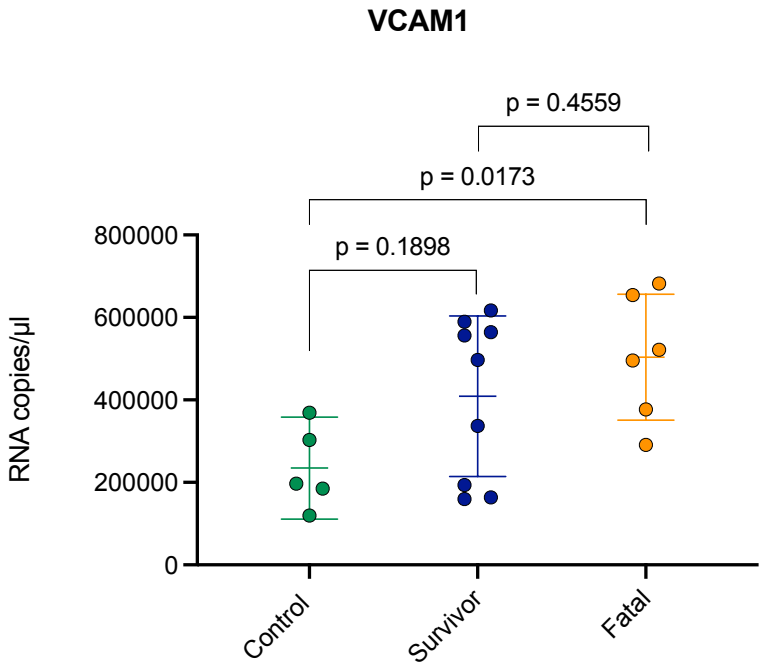


Figure 3.14 Evaluation of the optimized RT-qPCR assay for VCAM1 transcript in clinical samples. *The transcript abundance of VCAM1 was measured by RT-qPCR and compared between three groups of subjects: controls (green dots, n = 5), EVD survivors (blue dots, n= 9) and EVD fatalities (orange dots, n = 6). The bars show the mean Ct value and the 95% confidence interval. The differences in the transcript abundance between the groups were evaluated via the non-parametric Mann-Whitney U test.*

3.15 Discussion

This chapter has described the development and optimization of RT-qPCR assays for quantifying the transcript abundance of 11 potential EVD prognostic markers (VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28, NF3L1, SLC25A5, CTSL and ISG15). The absolute quantification was the method used for determining the transcript abundance of each gene. For this, a standard curve was constructed with serially diluted *in-vitro* RNA standards of known concentration. Thus, the concentration of unknown samples (healthy donors or EVD acute patients) could be determined by the interpolation of its Cq into the standard curve.

Although most of the first primers sets (except NF3L1 and CTSL) designed for the RT-qPCR assays did not show a good PCR efficiency, the latter designed primer sets showed a better performance. Thus, all the optimized assays developed in this study showed to be a sensitive method since most assays had an efficiency near 100%. The high PCR efficiency of these assays suggest that the Cq values are representative of the target concentration. Likewise, the assays demonstrated to have a good repeatability, since most CV values were less than 10% which is the recommended value for RNA measurements (Kralik & Ricchi, 2017; Sanders et al., 2014).

VCAM1, ISG15, MS4A4A, TGFBI, PLPP3, HOPX and CTSL have been proposed as a prognosis marker of different types of cancer (Chen et al., 2016; Huang et al., 2013; Lin et al., 2017; Pan et al., 2020; Ruan et al., 2014; Vishwakarma et al., 2017; Zhu et al., 2015). However, none of the 11 genes of this study has been reported before as a prognostic marker for EVD. Notwithstanding, some of these genes have been associated to severity of other human infectious diseases. For instance, increased levels of VCAM1 were found in SARS-CoV-2, HIV, and Hepatitis C viral infection but also associated with disease severity in SARS-CoV-2 infection (Kukla et al., 2009; Melendez et al., 2008; Tong et al., 2020). Soluble VCAM1 (sVCAM1) is also increased in Dengue virus infection (Koraka et al., 2004). A study of EVD patients from the West African outbreak found increased levels of sVCAM1 in fatal cases and/or decreased levels in survivors (Kerber et al., 2018). In contrast, other study only found increased

levels of sICAM but normal levels of sVCAM1 in a small cohort of EVD patients that were evacuated to the USA (McElroy et al., 2016).

ISG15 has been suggested as a potential prognostic marker of Respiratory Syncytial viral infection since it was found upregulated in acutely ill patients and correlated with viral load (Do et al., 2017). In EBOV studies, ISG15 and other Interferon-stimulated genes were strongly upregulated over the course of the disease in NHP infected with EBOV, Lassa, and Marburg virus (Caballero et al., 2016; Greenberg et al., 2020). *In-vitro* EBOV studies showed that ISG15 inhibits the viral replication by conjugation (ISGylation) of the viral proteins affecting the viral replication machinery and inhibiting the virus release (Durfee et al., 2010; Okumura et al., 2008).

There is little or no published information about the expression of the other candidate predictive genes (MS4A4A, TUBG1, TGFBI, CTSL, PLPP3, TTC28, HOPX, NIF3L1, SLC25A5) during viral infections in humans. MS4A4A has been found upregulated in infants with acute Respiratory Syncytial viral infection (Fjaerli et al., 2006). TGFBI was upregulated in NHP infected with Lassa virus (Malhotra et al., 2013). CTSL was found indispensable for the entry of EBOV and SARS-CoVs (Zhou et al., 2016).

The next chapter examines the transcript abundance of the 11 gene transcripts in a larger group of samples from EVD acute patients (survivors and non-survivors) and assesses the predictive value of these genes as prognostic markers of the disease outcome.

Chapter 4

Determining the predictive value of the candidate predictive gene transcripts in EVD patients

4.1 Introduction

A fast and accurate clinical assessment of disease severity is vital during an outbreak. Since early intervention is associated with improved outcome, identifying patients that are at imminent risk of serious illness or death would ensure that these patients receive immediate medical attention (Liang et al., 2020). Prognostic biomarkers that can distinguish the likely clinical outcome of patients would be a useful tool to relieve the clinical burden, support clinical management, logistical planning, and potentially reduce the mortality rate, especially in resource-poor settings (Yan et al., 2020). Thus, for the identification of potential predictive markers of disease severity, machine learning approaches have been increasingly used in different viral diseases such as Dengue, Influenza, Respiratory syncytial virus infections and the ongoing COVID-19 pandemic (Davi et al., 2019; Hu et al., 2020; Jong et al., 2016; Liang et al., 2020; Yan et al., 2020).

In EBOV, studies have focused on investigating the risk factors associated with EVD mortality such as age, sex, clinical symptoms, viral load, and co-infections. These have used statistical models and the development of scoring systems (Schieffelin et al., 2014; Yan et al., 2015; Zhang et al., 2015). Only a few recent studies have applied machine learning approaches to identify biomarkers that predict the outcome of EVD patients. A previous study from our group identified genes that correlated with outcome using classification models such as Support Vector Machine, Random Forest and Paired Gene profiling based on the transcriptome data of peripheral blood taken from acute-EVD patients (Liu et al., 2017). The predictive accuracies of these models were 85%, 89% and 92%, respectively (Liu et al., 2017). Other machine learning models combining variables such as viral load, age, body temperature, time to presentation, and clinical symptoms (bleeding, dyspnea, dysphagia, jaundice asthenia, weakness, and diarrhoea) had predictive accuracies around 74% when the models were externally validated against two independent datasets (Colubri et al., 2019). Thus, there are currently no predictive models for EVD that could be used in the field to assist health care workers.

The machine learning approach follows a design-learn-test process which is the way to test an algorithm that can successfully learn to discriminate the clinical outcome of the patients (Libbrecht & Noble, 2015). In this chapter, discriminative models were built since the study was focused on discriminating survivors and fatal cases using predictor variables, and not the interpretation of the relationship between these variables as occurs with regression models. This chapter aims to assess the predictive value for EVD outcome using the transcript abundance of the candidate predictive genes described in Chapter III. First, the transcript abundance of the candidate predictive genes, as quantified by singlepex SYBR Green-based RT-qPCRs was analysed to determine if there was a difference in abundance of these transcripts between survivors and fatal cases. After this, using the predictor variables EBOV Ct values and/or abundance of gene transcripts, supervised machine learning algorithms were used to train samples from acute EVD patients whose outcome was known. To evaluate the performance of the process, the trained models were used to predict the outcome of a blinded set of samples from acute EVD patients.

4.2 Results

4.2.1 Analysis of a cohort of EVD patients with known outcome

Using 39 clinical samples with known disease outcome (20 survivors and 19 fatal cases), collected by the EMLab during the outbreak in Guinea, the transcript abundance of VCAM1, ISG15, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28, NF3L1, SLC25A5, and CTSL was determined using the singleplex RT-qPCRs described in Chapter III. Due to the low volume and concentration of some of the samples (shown in Appendix Table 1) it was not possible to analyse each transcript abundance for all samples. The first two gene transcripts that were tested were VCAM1 and ISG15. Priority was given to VCAM1, since the abundance of the VCAM1 transcript was found to be significantly different between survivors and fatal cases in a previous transcriptomic analysis of peripheral blood from acute EVD patients (Liu et al., 2017). Besides, a significant difference in the kinetic expression of soluble VCAM1 was also reported between survivors and fatal cases in a cohort of patients from the 2013-

2016 outbreak in Guinea (Kerber et al., 2018). ISG15 was also given priority since showed an early and strong upregulation during the course of EVD (Caballero et al., 2016; Greenberg et al., 2020).

The clinical samples were tested in duplicate, and 10 ng was used per RT-qPCR. For VCAM1, a total of 39 samples (20 survivors and 19 fatal cases) were tested, ISG15 and CTSL were tested in 32 samples (15 survivors and 17 fatal cases), TUBG1 was tested in 31 samples (14 survivors and 17 fatal cases), MS4A4A was tested in 24 samples (9 survivors and 15 fatal cases), TTC28 was tested in 16 samples (4 survivors and 12 fatal cases), TGFBI was tested in 12 samples (2 survivors and 10 fatal cases), PLPP3 was tested in 10 samples (1 survivors and 9 fatal cases), NF3L1 and SLC25A5A were tested in 7 samples (1 survivor and 6 fatal cases), and HOPX was tested only in 4 fatal cases. Control samples (5 healthy donors) were also included in each PCR run and tested in duplicate.

The transcript abundance of all genes quantified by RT-qPCRs in the 39 clinical samples and the 5 control samples are presented in Appendix Table 2. The distribution of the transcript abundance of each gene per individual is shown in Figure 4.1. From this data, we can see a similar distribution of each gene transcript between control samples. The transcript abundance of VCAM1, ISG15, CTSL, and MS4A4 looks higher in EVD patients compared to non-infected individuals. The transcript abundance of TGFBI looks lower in EVD survivors and fatal cases compared to control samples. To analyse the difference in the transcript abundance of each gene between survivors and fatal cases, each gene was evaluated separately. The mean transcript abundance of each gene in survivors and fatal cases was calculated and were as follow: VCAM1 (78649 vs 230357 RNA copies/µl), ISG15 (41338 vs 54719 RNA copies/µl), CTSL (3069 vs 2946 RNA copies/µl), TUBG1 (562.2. vs 323.5 RNA copies/µl), MS4A4A (3560 vs 14841 RNA copies/µl), TTC28 (92.59 vs 53.85 RNA copies/µl), TGFBI (3755 vs 239.1 RNA copies/µl), PLPP3 (33.24 vs 121.5 RNA copies/ µl), NF3L1 (4479 vs 2829 RNA copies/µl), SLC25A5 (18252 vs 48713 RNA copies/µl), HOPX (57574 RNA copies/µl).

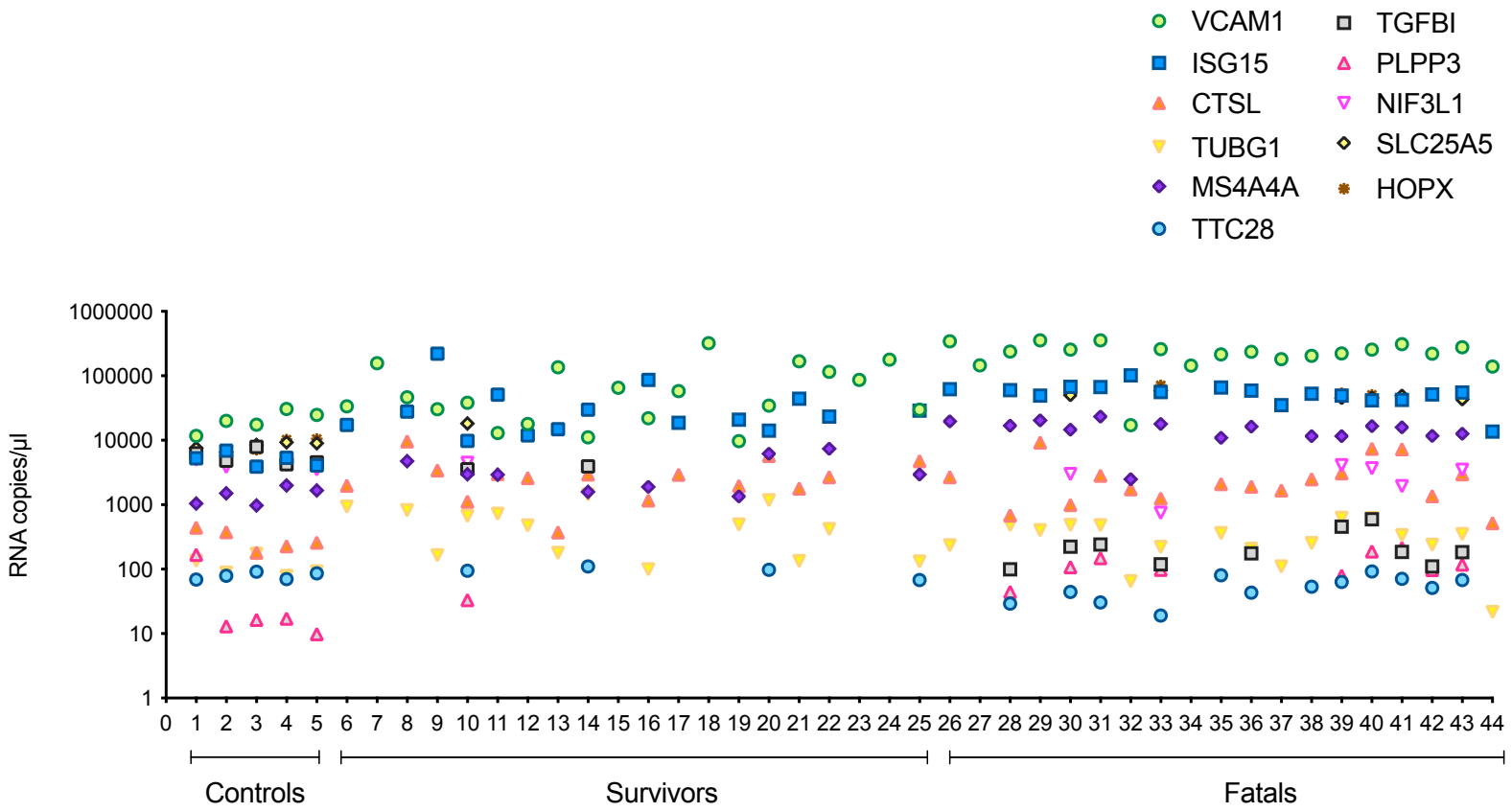
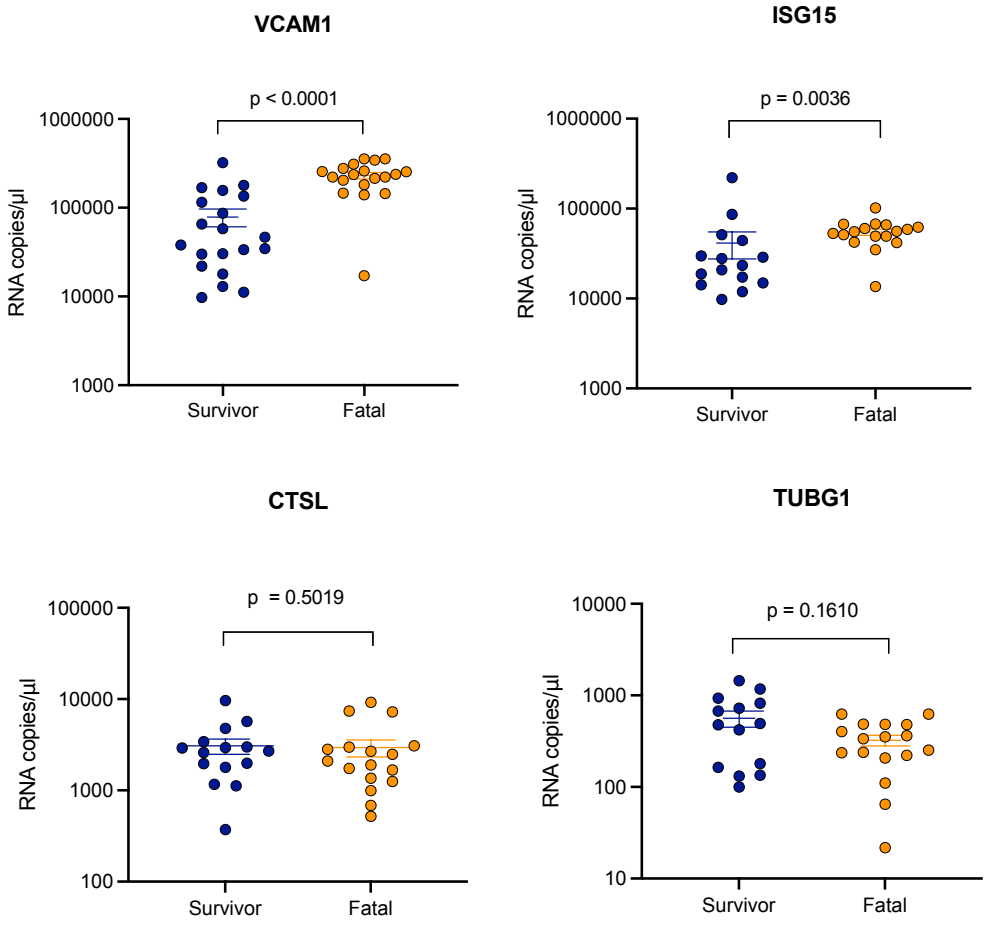
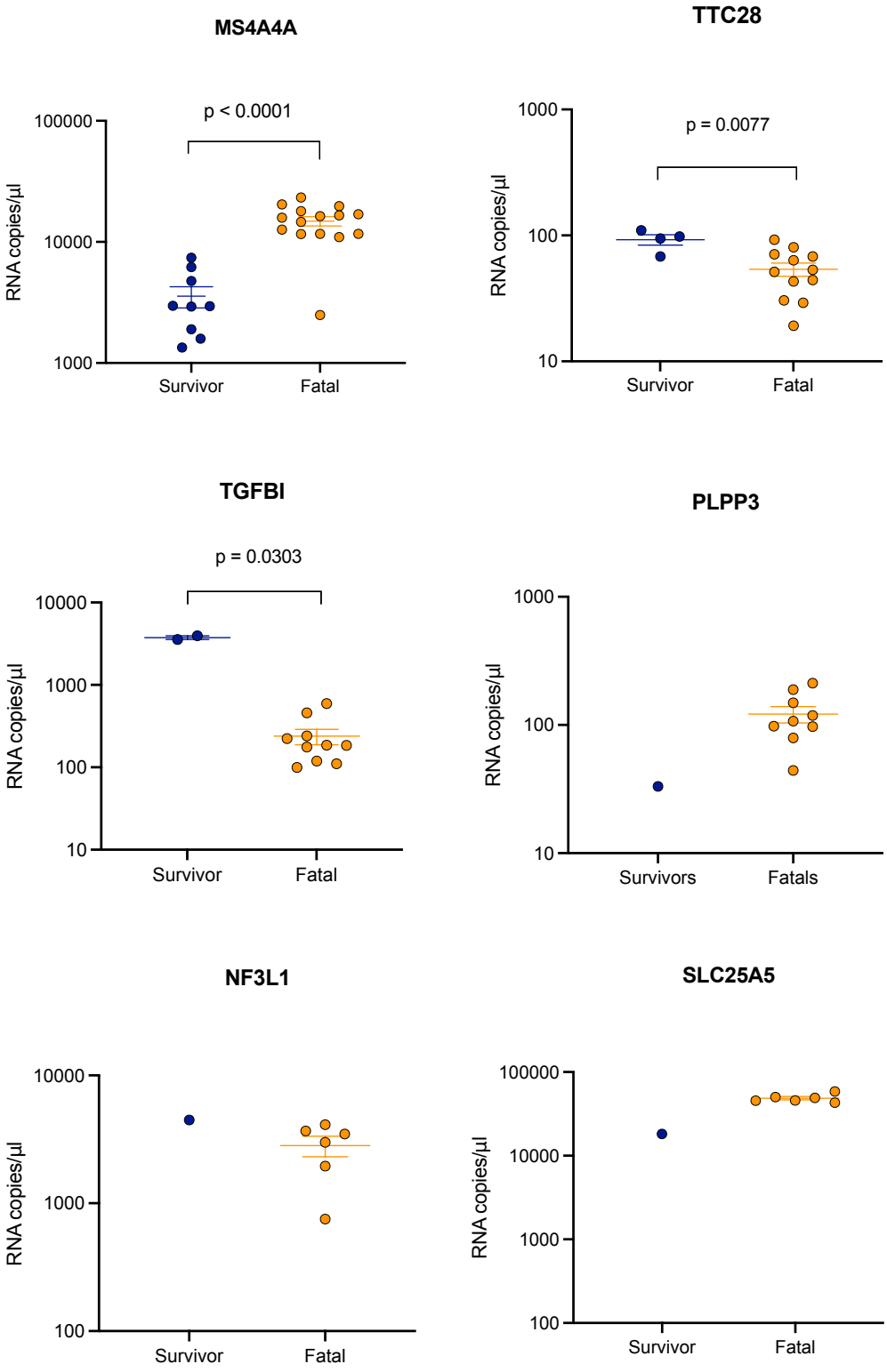


Figure 4.1 Distribution of the transcript abundance of all the candidate predictive genes in controls and clinical samples. *The transcript abundance of each of the 11 genes were obtained by RT-qPCRs from 39 EVD patients (20 survivors and 19 fatal cases) and 5 healthy donors. Each symbol represents a specific gene transcript. The human samples are numbered.*

The non-parametric Mann-Whitney *U* test was used to determine the statistical difference in the transcript abundance of each gene between survivor and fatal cases. As seen in Figure 4.2 there was a significant difference in the transcript abundance of VCAM1 ($p < 0.0001$), ISG15 ($p = 0.0036$), MS4A4A ($p < 0.0001$), TTC28 ($p = 0.0077$) and TGFBI ($p = 0.0303$) between survivor and fatal groups. The fatal cases had higher transcript levels of MS4A4A (4 times more), VCAM1 (3 times more) and ISG15 (2 times more) compared to survivors. Whereas fatal cases had lower transcript levels of TTC28 (2 times less) and TGFBI (16 times less) than survivors. No significant differences were found for CTSL and TUBG1. Statistical analysis for PLPP3, NF3L1, SLC25A5 and HOPX was not possible since the data was only available from one survivor for each gene transcript (due to limiting RNA). These findings suggest that VCAM1, ISG15, MS4A4A, and possibly TTC28 and TGFBI (low number of samples analysed) may be good predictors of the clinical outcome.




MS4A4A
p < 0.0001
100000
10000
1000
RNA copies/µl
Survivor
Fatal
TTC28
p = 0.0077
1000
100
10
RNA copies/µl
Survivor
Fatal
TGFBI
p = 0.0303
10000
1000
100
10
RNA copies/µl
Survivor
Fatal
PLPP3
1000
100
10
RNA copies/µl
Survivors
Fatals
NF3L1
10000
1000
100
RNA copies/µl
Survivor
Fatal
SLC25A5
100000
10000
1000
100
RNA copies/µl
Survivor
Fatal

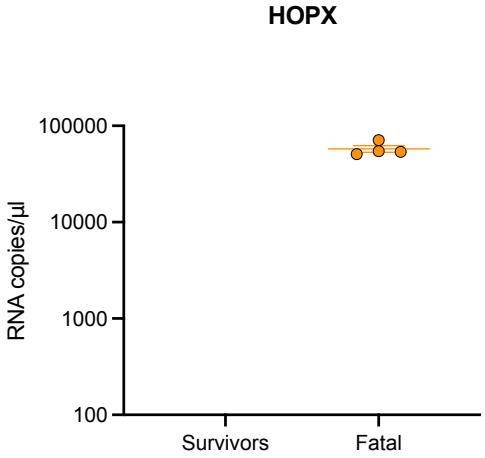


Figure 4.2 Transcript abundance of the candidate predictive genes tested by RT-qPCRs in EVD patients with known outcome. *The RNA copy number of the candidate gene transcripts in the clinical samples were quantitated by RT-qPCRs using the standard curve method. The clinical samples were tested in duplicate and the mean RNA copy number per individual was obtained by plotting the mean Ct value obtained from both replicates against the standard curve created with in-vitro RNA dilutions of known concentration. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences in the transcript abundance between survivor and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

The EBOV Ct values and other relevant information about the 39 clinical samples were provided by the team in Germany (Bernhard Nocht Institute for Tropical Medicine) and is presented in Appendix Table 3. The distribution of the transcript abundance of each gene by EBOV Ct values shows a clear separation between survivors and fatal cases (Figure 4.3). This was confirmed by the statistical analysis that showed a significant difference between survivors and fatal cases ($p < 0.0001$). The mean Ct values in survivors was 28.53 (minimum = 25.18, maximum = 34.57), and for fatal cases was 17.43 (minimum = 15.12, maximum = 20.98). This suggested that EBOV Ct value was a strong predictor of the clinical outcome as has been proposed before (Kerber et al., 2016; Liu et al., 2017).

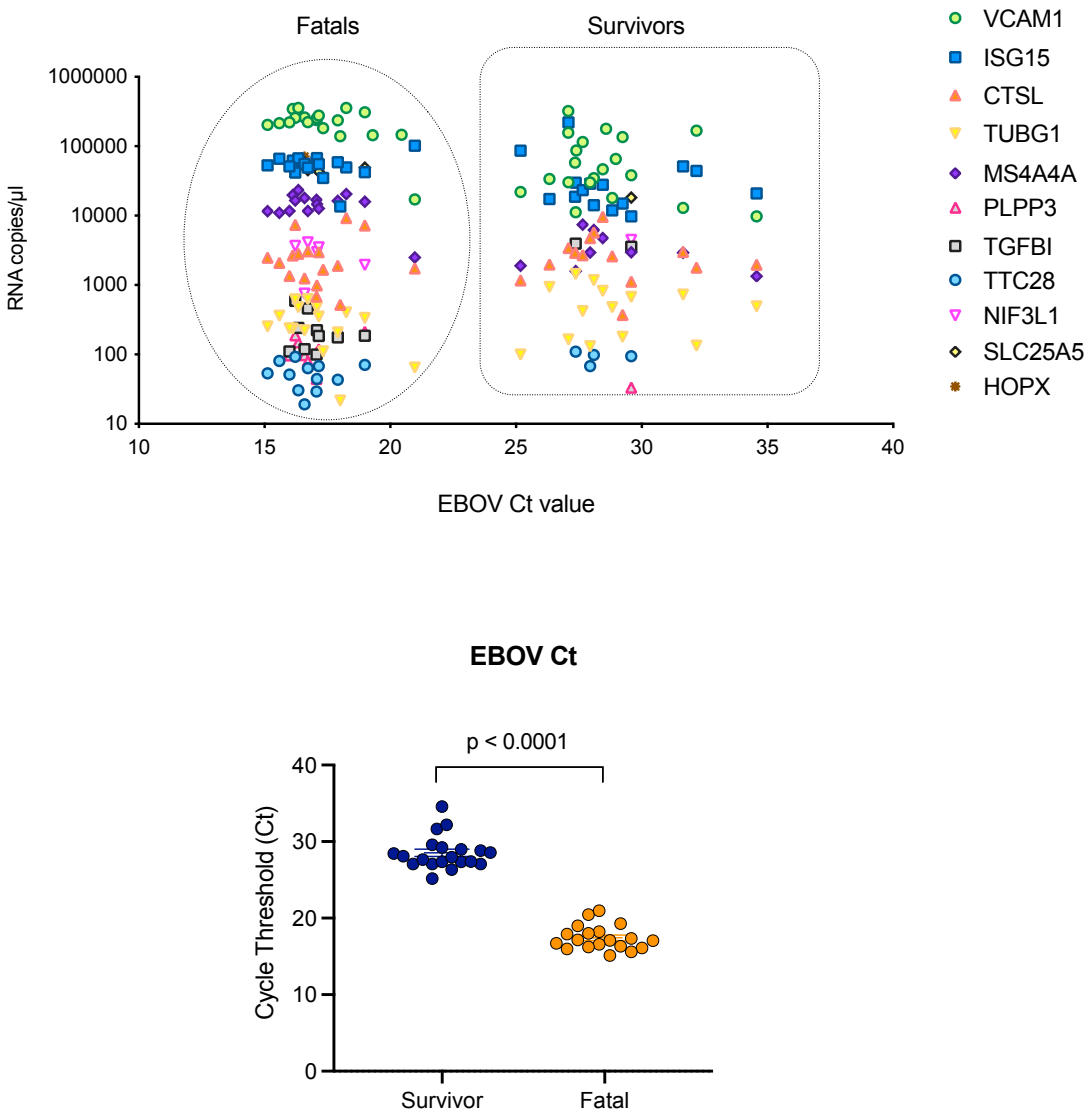


Figure 4.3 Transcript abundance of all the candidate predictive genes and EBOV Ct values recorded from 39 EVD patients. *The upper panel shows the RNA copy number obtained by RT-qPCRs from each of the 11 gene transcripts distributed by EBOV Ct values recorded from the 39 EVD patients. Each symbol represents a specific gene transcript. The lower panel shows the distribution of the EBOV Ct value by clinical outcome. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences in the EBOV Ct values between survivors and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

4.2.2 Developing machine learning classification models

The previous section showed a significant difference in the transcript abundance of VCAM1, ISG15, MS4A4A between survivors and fatal cases, therefore these gene transcripts were selected to try to develop a machine learning model. TTC28 and TGFBI were not included since these genes were tested in a low number of samples (16 and 12 samples, respectively). One gene more, TUBG1, was also included in the group of selected gene transcripts because there was more data available for this gene (31 samples) similar to VCAM1 (39 samples), ISG15 (32 samples) and MS4A4A (24 samples) which was necessary for developing machine learning models.

To develop machine learning models that could classify clinical samples into survivors or fatal cases the workflow used is shown in Figure 4.4. This was a binary classification model since there were two classes of outcome: survivor and fatal. Models were developed in four steps. The first step was pre-processing of the raw data. This step involved the selection of meaningful features or attributes for building the model. Some machine algorithms also required that the selected attributes are on the same scale which was achieved by transformation or normalization of the data. To determine whether the machine learning model performs well on the training set but also performs well in new data, the entire dataset was split in two: training dataset (70%) and testing dataset (30%). The second step involved using the training dataset to train and optimize the machine learning model (including model selection, cross-validation methods, performance metrics and hyperparameter optimization which is called “tuning”). The testing dataset was the “unseen data” which was hold until the end to evaluate the final model (third step). As a fourth step, the models could be used to predict the outcome of a new independent dataset.

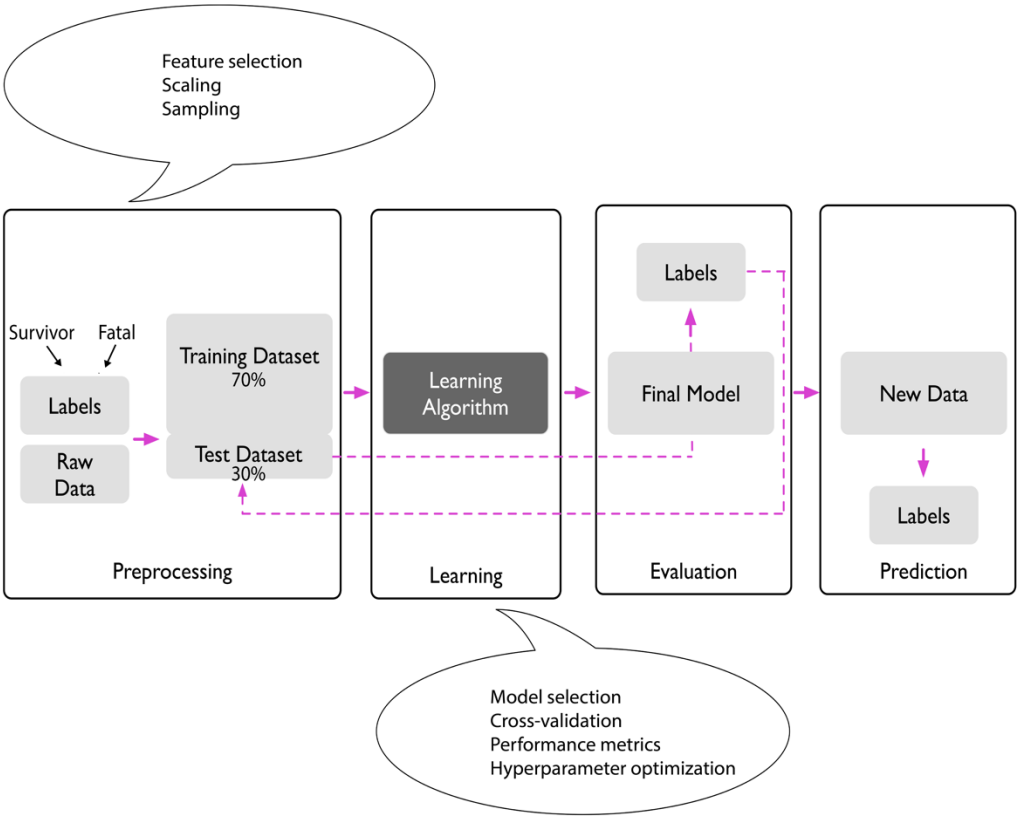


Figure 4.4 Workflow for building and evaluating the classification models in this study*. The workflow consisted of four steps: pre-processing of the raw data, learning or training of the machine learning algorithms, evaluation of the classification models and prediction of the outcome in a new independent dataset. Diagram modified and adapted from Raschka and Mirjalili (2017).*

4.2.2.1 Pre-processing of the data for building the classification models

For building classification models, the raw data used was the transcript abundance of VCAM1, ISG15, MS4A4A and TUBG1 obtained by RT-qPCRs from the cohort of 39 EVD patients with known outcome (section 4.2.1). After the removal of samples with missing values on any of the four gene transcripts, only the data from 24 patients (9 survivors and 15 fatal cases) was used for building the classification models.

To train the most accurate possible classifier only attributes that would contribute most to the accuracy of the model were selected. Thus, the attributes VCAM1, ISG15, MS4A4A and TUBG1 were evaluated using the Boruta algorithm, which is a feature selection algorithm that works as a wrapper built around a random forest (Kursa & Rudnicki, 2010). This means that the attributes compete with a randomized version

of them or called “shadow attributes”. Then, the importance of each original attribute is extracted from the random forest model and only attributes that are above a given threshold of importance are kept. The measure of importance is known as the Z score and the threshold is defined as the highest attribute importance recorded among the shadow attributes. From Figure 4.5, the Boruta analysis identified VCAM1, ISG15, MS4A4A and TUBG1 as relevant variables for the classification of the clinical outcome. This is because the Z scores in these variables were higher than the scores obtained from the random variables the algorithm created (“shadows attributes” of all the attributes tested: VCAM1, ISG15, MS4A4A and TUBG1).

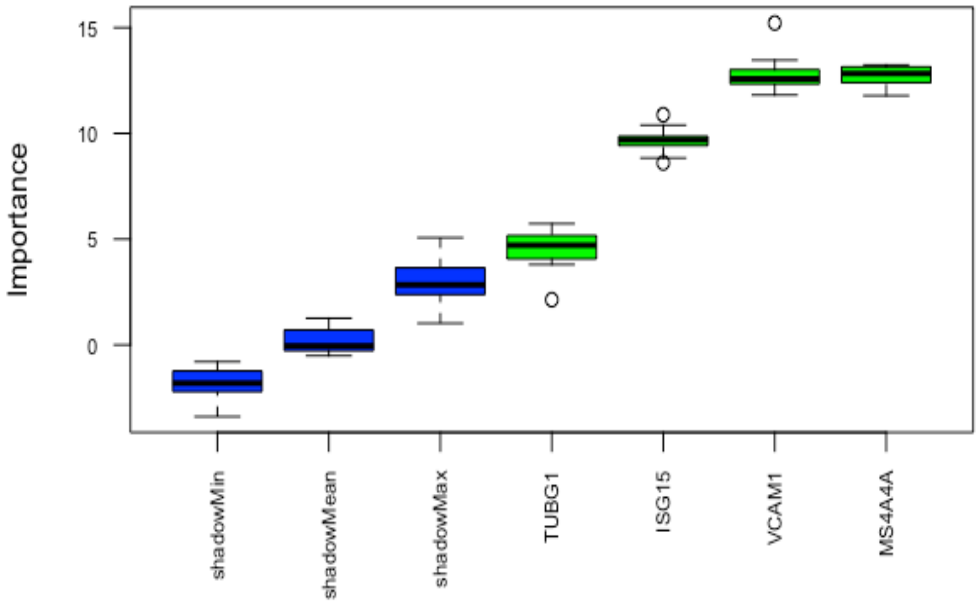


Figure 4.5 Boruta result plot for the selected candidate predictive gene transcripts in 24 EVD patients with known outcome. *The Boruta algorithm adds randomness to the data by creating shuffled copies (“shadows attributes”) of all the attributes (VCAM1, ISG15, MS4A4A and TUBG1), and runs a random forest classifier on the extended data to compute Z scores. It finds the maximum Z score among shadow attributes (MZSA) and assign a hit to every attribute that scored better than MZSA. At every iteration, checks whether an attribute has a higher Z score than the MZSA and consider them as “important” or as “unimportant” if the Z score is lower than MZSA, and removes them permanently. The algorithm stops after a limit of 500 random forest runs. The y-axis shows importance in terms of Z-scores. Blue boxplots correspond to minimal, average and maximum Z score of a shadow attribute. Green boxplots represent Z scores of the confirmed important attributes. Box plot features are as follows: circles are outliers, dashed-lines “whiskers” are 1.5 times the interquartile range, rectangle is the first and third quartiles, and horizontal bar is the median.*

To visualize the distribution of the data from the 24 patients with known outcome and detect the presence of outliers and the relationship between the attributes (gene transcripts) a scatterplot matrix was created (Figure 4.6). The histograms showed that ISG15 was normally distributed, TUBG1 had a positive skewed distribution and MS4A4A and VCAM1 were not normally distributed.

Although the four candidate predictive gene transcripts were considered relevant attributes for classification of the outcome in the Boruta analysis, it did not rule out the possibility that some of them may be correlated variables (Kursa & Rudnicki, 2010). Because the presence of highly correlated variables (named as "multicollinearity") may affect the interpretation of machine learning models, we wanted to examine whether these variables were highly correlated or not. To investigate this, a Pearson correlation analysis was performed. From Figure 4.6 we can see that VCAM1 and MS4A4A had a strong positive correlation (0.97), a linear relationship. There was also a moderate negative correlation between ISG15 and TUBG1 (-0.58) and no significant correlation was observed between the other pairs. Although, the strong relationship of VCAM1 and MS4A4A is not yet understood, it might suggest that are redundant variables.

To eliminate or reduce multicollinearity two approaches were possible: removing one of the highly correlated variables from the model or use non - linear algorithms such as Support Vector Machine or Random Forest which are less sensitive to multicollinearity, and demonstrated a good predictive performance even when correlated variables were included (Farrell et al., 2019; Neter et al., 1996). Non-linear algorithms were chosen.

Because this study aimed to predict the outcome of EVD patients rather than discriminating between EBOV infected and not infected individuals, the data obtained from testing the four candidate predictive gene transcripts in 5 control samples (healthy donors) were not included in the models. The distribution of the transcript abundance of VCAM1, ISG15, TUBG1 and MS4A4A when control samples were included did not show a clear separation between survivors and fatal cases

(Figure 4.6). This suggested that including control samples in the models may influence and bias the prediction.

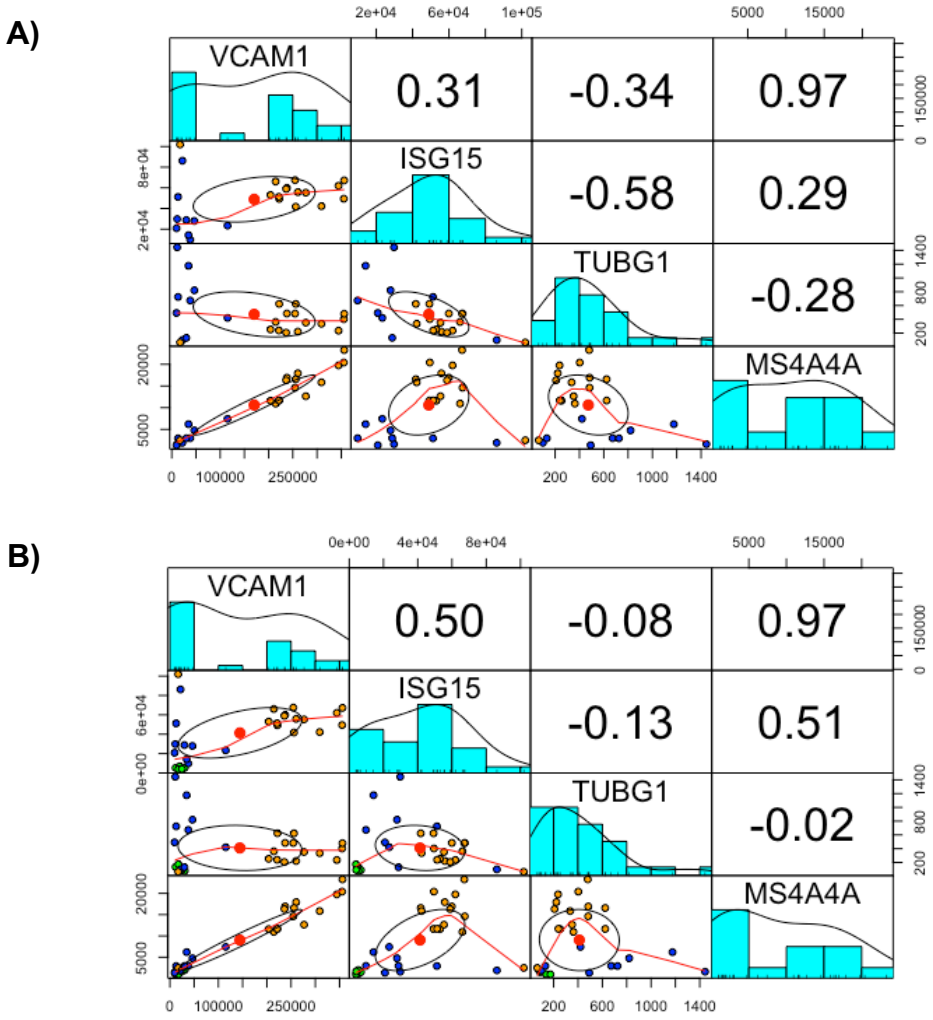


Figure 4.6 Scatterplot matrix of the selected candidate predictive gene transcripts for the EVD patients dataset with known outcome. *For each of the numerical variables (transcript abundance of host biomarkers) a histogram is shown. Pairwise scatterplots (below the diagonal) show each combination of the host biomarkers. A) Data from 9 survivors (blue dots) and 15 fatal cases (orange dots) from the cohort of EVD patients with known outcome. Some variables help to classify better survivors and fatal cases but, in some combinations, there is still an overlap. B) A scatterplot matrix including the data of 5 healthy donors (green dots) does not show a clear separation between survivors and control samples. The upper panel above the diagonal shows pairwise Pearson's correlation coefficients obtained from the combination of host biomarkers.*

4.2.2.2 Training Support Vector Machine classification models

The scatterplots in Figure 4.6 showed that the distribution of the transcript abundance of VCAM1, ISG15, TUBG1 and MS4A4A between survivors and fatal cases were not linearly separable so linear classifiers such as Logistic regression or Linear Discriminant analysis (LDA) were not suitable to use. This suggested that non- linear algorithms may perform better since are more flexible with the data and less prone to be affected by outliers (see Figure 4.7). For building the models, the radial SVM algorithm was selected since it can maximize the margin of separation between the survivors and fatal cases.

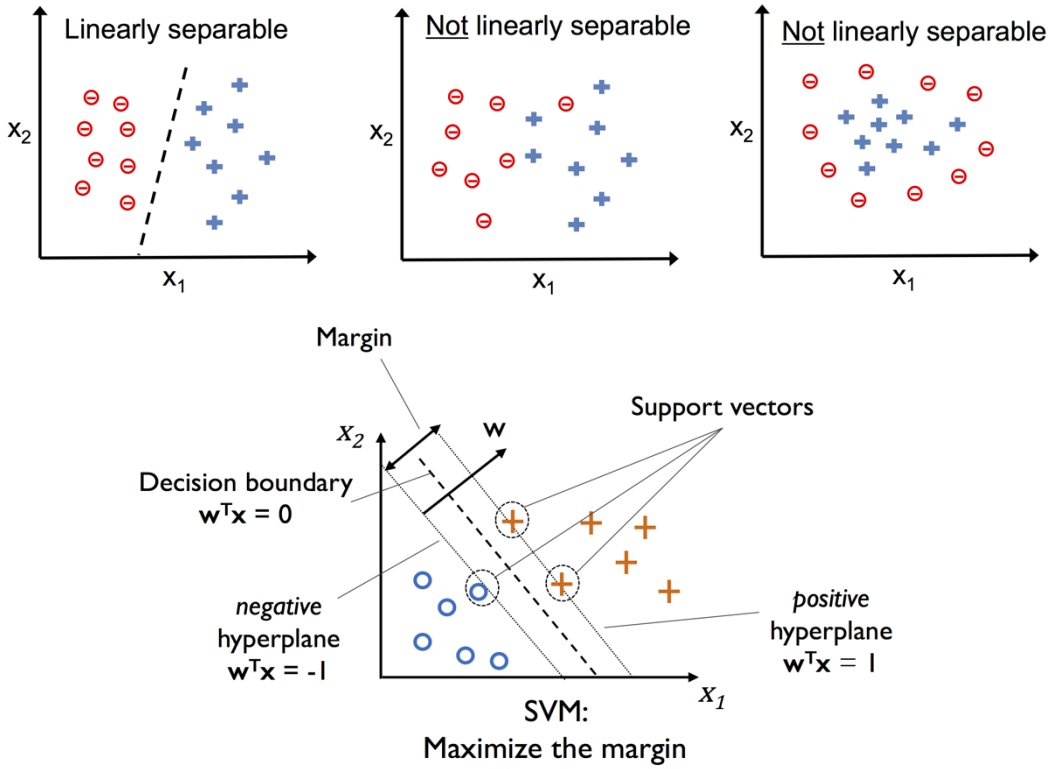


Figure 4.7 Graphical representation of linear, non-linear separable problems and the SVM algorithm. *The upper panel shows a representation of linear and non-linear separable problems. The first means that the dataset can be classified in two categories using a single line. On the contrary, the latter indicates that the dataset requires a non-linear line for separating them into their respective categories. The lower panel shows an illustration of the Support Vector Machine algorithm. The main objective of SVM is to maximize the margin, which is the distance between the separating hyperplane referred as a "decision boundary" and the training samples that are closest to this hyperplane called "support vectors". Figures were taken from Raschka and Mirjalili (2017)*

In the R software, the data from the 24 patients with known outcome was first randomly split into a training dataset 70% (18 observations consisting of 7 survivors and 11 fatal cases) and a testing set 30% (6 observations consisting of 2 survivors and 4 fatal cases) following the first step of the workflow (Figure 4.4).

To avoid bias in the prediction model due to an imbalance dataset (unequal number of survivors and fatal cases in the training dataset); the training dataset was rebalanced to have an equal number of fatal and survivor cases before training the model using the *rose* package in R software. This is a smoothed bootstrap-based technique which reduce the effects of an imbalanced distribution of classes by generating artificial balanced samples (Lunardon et al., 2014).

SVM models were built with data from the training set with different combinations of the variables VCAM1, ISG15, MS4A4A and TUBG1. For each model, the resampling method used was the *k*-fold cross validation (indicated in the second step of the workflow in Figure 4.4). This is a function that can be used within the *caret* package in R software. This function split the training dataset into 10 subsets or folds (called *k*-subsets), so 1 subset was held-out while the model was trained on all other subsets, and the process repeated until the overall accuracy was determined in order to maximize the model performance (Figure 4.8). After this, the algorithm was tuned with the optimal hyperparameters to find the best model in each combination of variables.

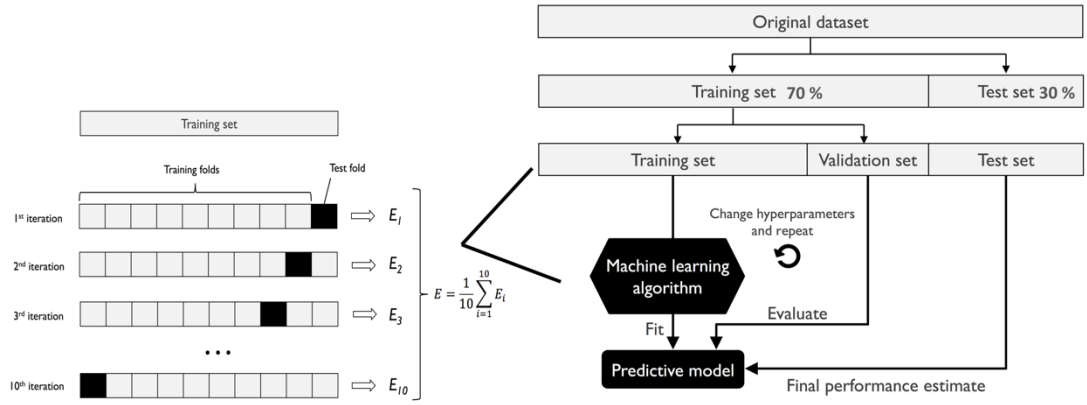


Figure 4.8 K-fold cross validation method used for resampling the training dataset. The graph in the right illustrates the concept of holdout cross-validation. *After splitting the entire dataset in training set 70% and testing set 30%, the training set was split in a validation set to repeatedly evaluate the performance of the model after training different parameter values. This was done through the k-fold cross validation function (left graph) which randomly split the training dataset into k folds without replacement, where k-1 folds are used for model training and one-fold is used for performance evaluation. This process repeated k = 10 times, so k models were obtained, and the average performance of these models was calculated. This method was used for model tuning since allows to find the optimal hyperparameter values for selecting the best model. Graphs modified and adapted from Raschka and Mirjalili (2017).*

4.2.2.3 Models evaluation in the testing dataset and prediction of the outcome of an independent "blinded" cohort of patients

As indicated in the third and fourth step of the flowchart in Figure 4.4, the final models built in section 4.2.2.2 with data from the training set (18 observations consisting of 7 survivors and 11 fatal cases) were first evaluated to predict the outcome of patients from the testing set (6 observations consisting of 2 survivors and 4 fatal cases) and then to predict the outcome of an independent cohort of samples.

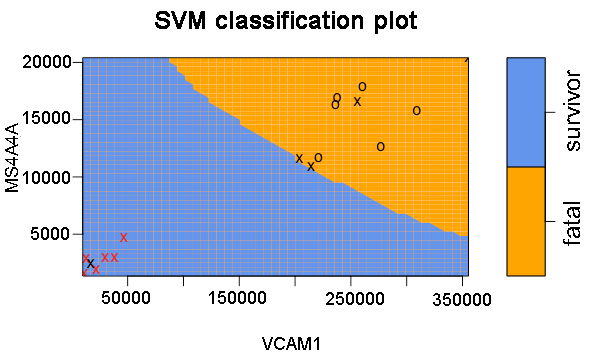
The independent cohort of samples belonged to 64 EVD patients collected by the EMLab deployed in Guinea. To evaluate the final models in this cohort of patients, a blind experiment was carried out. For this, all the relevant information from these samples was withheld by the team in Germany (Bernhard Nocht Institute for Tropical

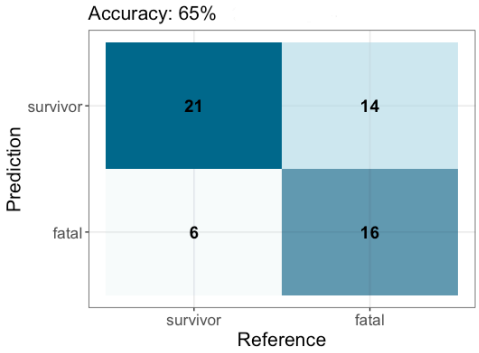
Medicine), this included EBOV Ct, outcome, sex, age, treatment, and malaria co-infection. The transcript abundance of VCAM1, ISG15, MS4A4A and TUBG1 from these samples was quantified using the optimized one-step RT-qPCRs. These assays are described in detail in section 2.10. Not all clinical samples were able to be tested for the 4 gene transcripts due to the low RNA concentration volume. Only sample RNA_16 did not amplify to any of the genes, suggesting that the RNA was degraded or very low. The RNA concentration of the 64 EVD patients is presented in Appendix Table 4 and the transcript abundance of each of the selected genes obtained by RT-qPCRs is presented in Appendix Table 5. After the removal of samples with missing values, a total of 57 samples had complete data of the 4 gene transcripts and was used to predict the clinical outcome of these patients using the final models.

The outcomes predicted by SVM models for the 57 patients of the blinded cohort were first reported to the team in Germany and they provided us the percentage accuracy of our models. Different models were tried consecutively to obtain the best percentage accuracy. To visualize the performance of each SVM model, the outcome of the patients was unblinded to graph confusion matrices (Figure 4.9). The SVM model built with 4 variables: VCAM1, ISG15, MS4A4A and TUBG1 was 100% accurate predicting the outcome of the patients from the testing set but had a lower prediction accuracy in the “blinded” cohort of samples, correctly classifying the outcome in only 65% (95% CI, 0.5113 - 0.7709) of the patients. This model correctly classified most of the survivors 21/27 (77.8%) but was less accurate classifying the fatal cases 16/30 (53.3%). To determine if reducing the number of variables improved the prediction percentage, the VCAM1 data was removed from the model since VCAM1 and MS4A4A are correlated variables and the inclusion of both might be affecting the model. The SVM model combining MS4A4A, ISG15 and TUBG1 had a higher accuracy 67% (95% CI, 0.5294 - 0.786) compared to the model that also included VCAM1. This model could predict almost all survivors 25/27 (92.6%) but only half of fatal cases 13/30 (43.3%).

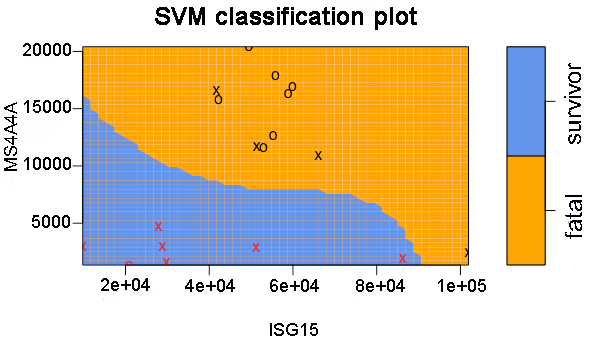
Training models that target only one gene transcript did not improve the performance. Thus, models that included a single variable either VCAM1, ISG15 or MS4A4A were 100% accurate predicting the outcome of patients from the testing set but were less accurate predicting the outcome of the "blinded" cohort of patients. The model for TUBG1 was less accurate predicting the clinical outcome of patients from the testing set, it correctly classified 67% of the patients. From the four candidate predictive gene transcripts, MS4A4A and TUBG1 had the best performance when were evaluated alone in the "blinded" cohort of patients. The MS4A4A model was 61.4% (95% CI, 0.4757 - 0.74) accurate and correctly classified all survivors 27/27 (100%) but most of the fatal cases were not correctly classified 8/30 (26.7%). The TUBG1 model had a similar overall accuracy of 61.4% (95% CI, 0.4757 - 0.74), and correctly classified most of the fatal cases 27/30 (90%) but most of the survivors were misclassified 8/27 (29.6%). Based on these results, MS4A4A seemed to be a good predictor of survival while TUBG1 a predictor of fatality. To determine whether MS4A4A and TUBG1 could perform better when they were used together, a model with both variables was trained. This model did not show a better accuracy 61.4% (95% CI, 0.4757 - 0.74) compared to the models built with each single variable. This model predicted almost all survivors 24/27 (88.8%) and 19/30 (36%) of fatal cases. Models with a single variable such as VCAM1 had a lower performance accuracy 49.1 % (95% CI, 0.3563 - 0.6271) predicting half of the fatal cases 15/30 (50%) and half of the survivors 13/27 (48.1%). Likewise, ISG15 model had an accuracy of 44% (95% CI, 0.3074 - 0.5764), and predicted 18/30 (60%) of fatal cases and 7/27 (25.9%) of survivors. Other models with different combinations of variables did not show a better performance when were tested in the "blinded" cohort of patients.

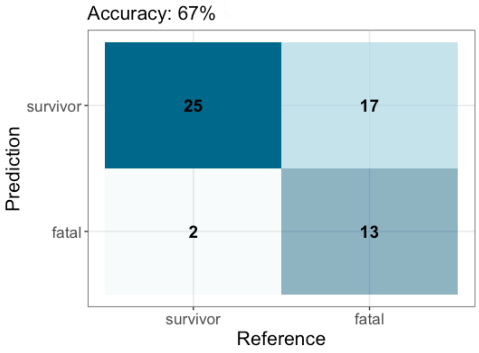
A) VCAM1 + ISG15 + MS4A4A + TUBG1



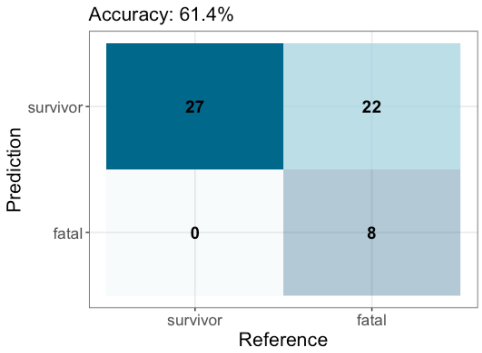


B) MS4A4A + ISG15 + TUBG1

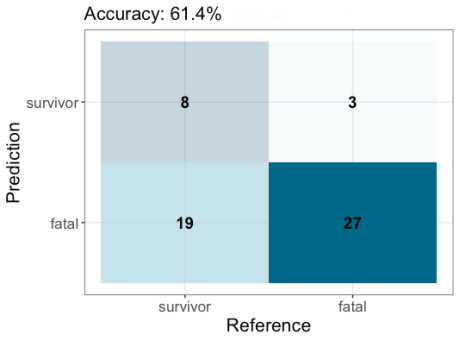




C) MS4A4A



D) TUBG1



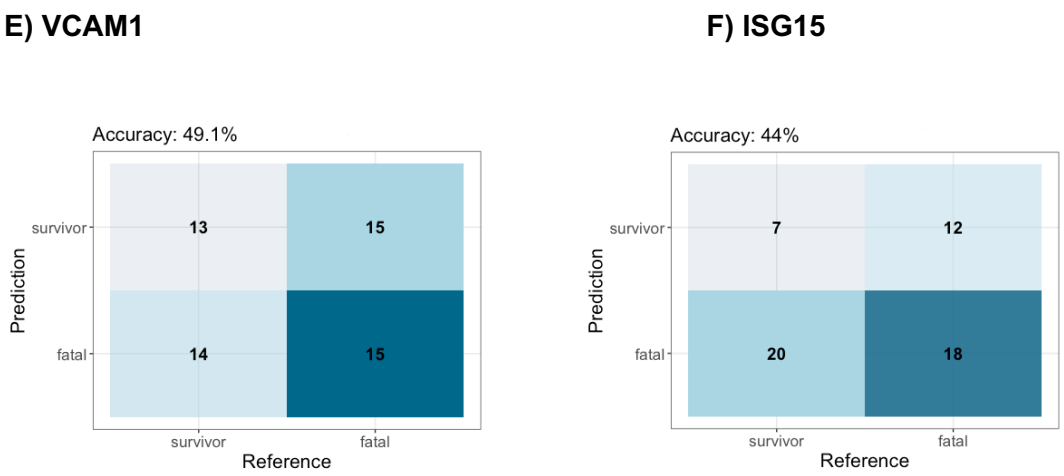


Figure 4.9 SVM models and confusion matrix showing models performance in 57 samples from the "blinded" cohort of patients. *For models with a combination of 2 or more variables a classification plot was created to visualize the data of the trained model. The classification plots show only 2 dimensions. The decision boundaries are created for both outcomes and are coloured in orange (fatal) and light blue (survivor). The support vectors are represented by crosses and the data points are the circles (in both cases red is for survival outcome and black for fatal outcome). Confusion matrix plots are presented for better assessment and interpretation of the classifier performance. The confusion matrix reports the predicted and actual outcome. The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

In an attempt to improve the performance of the SVM models, the EBOV Ct values for the coded samples were unblinded and used as an additional predictor variable. The EBOV Ct values and other relevant information about these cohort of patients are presented in Appendix Table 6. EBOV Ct value was always gathered as part of the diagnostic assay by the European Mobile Laboratory in West Africa.

New SVM models were built including the EBOV Ct values, for this the EBOV Ct values were first included in the training dataset. Section 4.2.1 has shown that EBOV Ct values may be a strong predictor of the outcome. To check the distribution of the data including the EBOV Ct values in the 24 patients with known outcome (dataset that was split in training and testing set), we created a scatterplot matrix (Figure 4.10). The correlation between the variables; EBOV Ct values, VCAM1, ISG15, MS4A4A and TUBG1 was evaluated by Pearson correlation analysis. From the Figure

we can see that EBOV has a positive skewed distribution and have a strong negative correlation with VCAM1 (-0.85) and MS4A4A (-0.81), and a moderate negative correlation with ISG15 (-0.58) and positive correlation with TUBG1 (0.42).

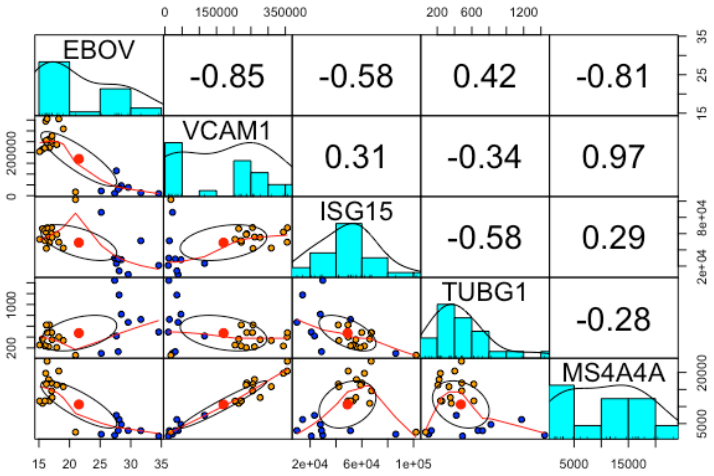


Figure 4.10 Scatterplot matrix of EBOV Ct and candidate predictive gene transcripts for the 24 EVD patients dataset with known outcome. *The plot shows the data from 9 survivors (blue dots) and 15 fatal cases (orange dots) from the cohort of EVD patients with known outcome. Histograms are shown for each predictor variable (EBOV Ct, VCAM1, ISG15, MS4A4A, TUBG1). Pairwise scatterplots (below the diagonal) show each combination of the host biomarkers and EBOV Ct. The upper panel above the diagonal shows pairwise Pearson's correlation coefficients obtained from the combination of the predictor variables.*

Different SVM models were trained combining variables EBOV Ct, VCAM1, ISG15, MS4A4A, TUBG1 and are shown in Figure 4.11. The model that combined these five variables was 100% accurate for predicting the outcome of the testing set (6 observations consisting of 2 survivors and 4 fatal cases), but 65% (95% CI, 0.5113-0.7709) accurate predicting the outcome of the 57 patients from the "blinded" cohort of samples. This model classified correctly almost all fatal cases 28/30 (93.3%), but only 9/27 (33.3%) of survivors.

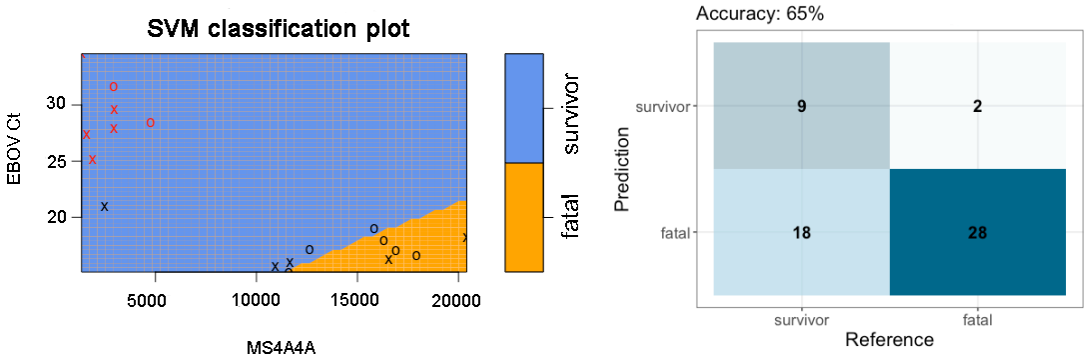
The model that excluded only VCAM1 was more accurate 68.4% (95% CI: 0.5476 – 0.8009) than the model that excluded MS4A4A 61.4% (95% CI,0.4757 - 0.74). Both models predicted the same proportion of fatal cases 28/30 (93.3%). The model that combined EBOV Ct, MS4A4A, ISG15 and TUBG1 was more accurate classifying the

survivors (11/27, 40.7%) than the model that combined EBOV Ct, VCAM1, ISG15 and TUBG1 (7/27, 25.9%). Models that combined EBOV Ct with a single candidate predictive gene transcript did not show a better performance (accuracies ranged from 53 to 56%). Interestingly, the model trained with only EBOV Ct values correctly predicted the outcome of only 54.3% (95% CI, 0.4066 - 0.6764) of the patients from the "blinded" cohort. EBOV Ct showed to be a good predictor of mortality, classifying correctly 28/30 (93.3%) of fatal cases but a poor predictor of survival, classifying correctly only 3/27 (11.1%) of survivors.

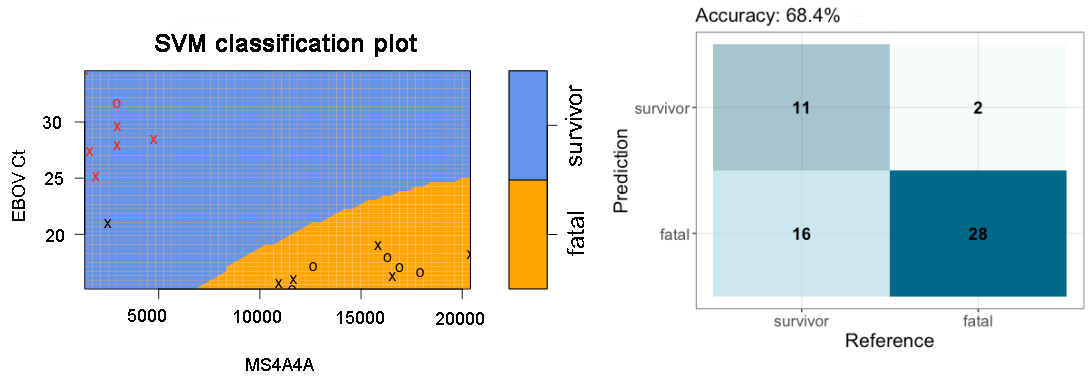
Overall, these results indicate that the best models for predicting the outcome of the 57 samples from the "blinded" cohort of patients were the one trained with variables EBOV Ct, ISG15, MS4A4A and TUBG1 (68.4% accuracy) and the model that combined MS4A4A, ISG15 and TUBG1 (67% accuracy). The clinical outcomes predicted with these models and the one that only used EBOV Ct values are presented in Table 4.1. The correct predictions of these models are highlighted in yellow.

These results suggested that a combination of EBOV Ct, MS4A4A, ISG15 and TUBG1 may be a suitable model for predicting mortality amongst EVD patients but also discriminating more survivors than a model with only EBOV Ct (54.3% accuracy). From Table 4.1 we can see that those patients whose EBOV Ct values ranged from 19 to 21; the EBOV Ct model predicted a fatal outcome while the model that used EBOV Ct, MS4A4A, ISG15 and TUBG1 as predictor variables correctly classified them as survivors (RNA-01, RNA-06, RNA-10, RNA-13, RNA-25, and RNA-26). The only two fatal cases that were misclassified by the best model (EBOV Ct, MS4A4A, ISG15 and TUBG1) had higher Ct values (RNA-54: Ct 23; RNA-59: Ct 26.42) and were patients that received treatment with favipiravir and did not have malaria co-infection (Appendix Table 6). Unfortunately, there is no information about the supportive care that these cohort of patients received during the outbreak.

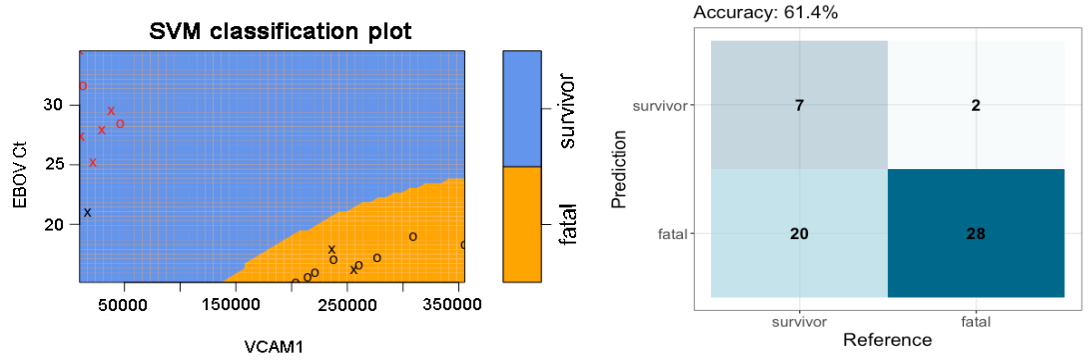
A) EBOV Ct + VCAM1 + ISG15 + MS4A4A + TUBG1



B) EBOV Ct + ISG15 +MS4A4A + TUBG1



C) EBOV Ct + VCAM1+ ISG15 + TUBG1



D) EBOV Ct

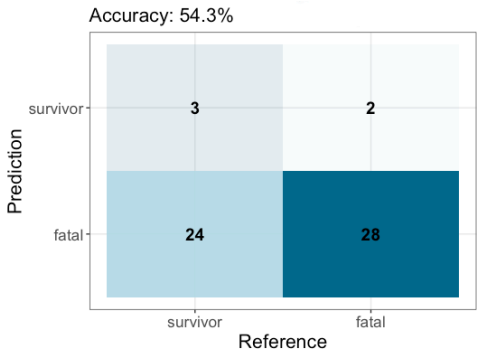


Figure 4.11 SVM models including EBOV Ct and confusion matrix showing models performance in the 57 samples of the "blinded" cohort of patients. *For models with a combination of more than two variables a classification plot was created to visualize the data of the trained model. The classification plots show only 2 dimensions. The decision boundaries are created for both outcomes and are coloured in orange (fatal) and light blue (survivor). The support vectors are represented by crosses and the data points are the circles (in both cases red is for survival outcome and black for fatal outcome). Confusion matrix plots are presented for better assessment and interpretation of the classifier performance. The confusion matrix reports the predicted and actual outcome. The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

*Table 4.1 Clinical outcomes predicted by SVM models for 57 samples from the "blinded" cohort of patients **

Sample_ID	EBOV Ct values	Actual Outcome	Predicted outcomes by SVM models**		
			EBOV Ct	MS4A4A +ISG15 +TUBG1	EBOV Ct + MS4A4A + ISG15 + TUBG1
RNA-01	19.1	survivor	fatal	survivor	survivor
RNA-02	17.25	survivor	fatal	survivor	fatal
RNA-03	17.09	survivor	fatal	survivor	fatal
RNA-04	19.08	survivor	fatal	survivor	fatal
RNA-05	17.73	survivor	fatal	survivor	fatal
RNA-06	19.76	survivor	fatal	survivor	survivor
RNA-07	19.04	survivor	fatal	survivor	fatal
RNA-08	14.42	survivor	fatal	survivor	fatal
RNA-09	23.19	survivor	survivor	survivor	survivor
RNA-10	19.54	survivor	fatal	survivor	survivor
RNA-11	16.08	survivor	fatal	survivor	fatal
RNA-13	21.29	survivor	fatal	survivor	survivor
RNA-14	15.99	survivor	fatal	survivor	fatal
RNA-15	19.35	survivor	fatal	survivor	fatal
RNA-17	17.71	survivor	fatal	survivor	fatal
RNA-18	18.5	survivor	fatal	fatal	fatal
RNA-19	16.52	survivor	fatal	survivor	fatal
RNA-20	17.44	survivor	fatal	survivor	fatal
RNA-21	15.24	survivor	fatal	survivor	fatal

RNA-22	17.24	survivor	fatal	survivor	survivor
RNA-23	19.55	survivor	fatal	fatal	fatal
RNA-24	19.95	survivor	fatal	survivor	fatal
RNA-25	19.31	survivor	fatal	survivor	survivor
RNA-26	19.81	survivor	fatal	survivor	survivor
RNA-28	14.66	fatal	fatal	fatal	fatal
RNA-29	14.61	fatal	fatal	survivor	fatal
RNA-30	15.3	fatal	fatal	survivor	fatal
RNA-31	15.51	fatal	fatal	survivor	fatal
RNA-32	13.5	fatal	fatal	fatal	fatal
RNA-33	14.91	fatal	fatal	survivor	fatal
RNA-34	19.95	fatal	fatal	fatal	fatal
RNA-35	16.69	fatal	fatal	fatal	fatal
RNA-36	14.99	fatal	fatal	survivor	fatal
RNA-37	14.6	fatal	fatal	survivor	fatal
RNA-38	15.34	fatal	fatal	survivor	fatal
RNA-39	15.07	fatal	fatal	survivor	fatal
RNA-40	14.03	fatal	fatal	fatal	fatal
RNA-41	15.33	fatal	fatal	fatal	fatal
RNA-42	15.82	fatal	fatal	survivor	fatal
RNA-43	12.87	fatal	fatal	survivor	fatal
RNA-44	16.33	fatal	fatal	survivor	fatal
RNA-45	15.4	fatal	fatal	fatal	fatal
RNA-47	17.44	fatal	fatal	fatal	fatal

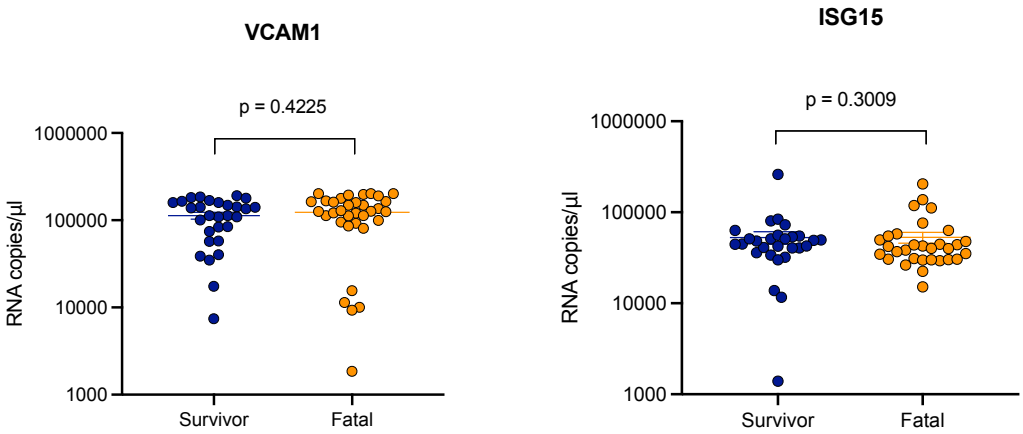
RNA-48	19.91	fatal	fatal	survivor	fatal
RNA-50	16.26	fatal	fatal	fatal	fatal
RNA-51	15.2	fatal	fatal	survivor	fatal
RNA-52	16.22	fatal	fatal	fatal	fatal
RNA-53	19.1	fatal	fatal	fatal	fatal
RNA-54	23	fatal	fatal	survivor	survivor
RNA-55	23.64	fatal	survivor	fatal	fatal
RNA-56	17.41	fatal	fatal	survivor	fatal
RNA-57	17.53	fatal	fatal	survivor	fatal
RNA-59	26.42	fatal	survivor	survivor	survivor
RNA-60	20.42	fatal	fatal	fatal	fatal
RNA-61	22.17	survivor	fatal	survivor	survivor
RNA-62	27.15	survivor	survivor	survivor	survivor
RNA-64	29.15	survivor	survivor	survivor	survivor

*This table shows 57 of the 64 initial cohort of patients since they did not have missing values for any of the predictor variables.
**Only the outcomes predicted with a model built with EBOV Ct and the models that had the best performance are shown.

4.2.3 Data analysis of the cohort of EVD patients with known outcome joined with the "blinded" cohort of patients.

The previous section showed that the best SVM model including EBOV Ct values was highly predictive of EVD mortality. However, the overall accuracy of this model was 68.4%. To examine the possibility of bias that influenced these results because of the small dataset size (training set of 18 subjects including 7 survivors and 11 fatal cases) used for building the models, it was important to compare both cohorts of patients (39 subjects from the cohort of patients with known outcome and 64 patients from the blinded cohort of samples).

At this point of the study, the outcome of the blinded cohort of patients was unblinded, so it was possible to compare the transcript abundance of VCAM1, ISG15, MS4A4A and TUBG1 between survivors and fatal cases. For this analysis, the Mann-Whitney *U* test was applied. As seen in Figure 4.12, there was a significant difference in the transcript abundance of MS4A4A ($p = 0.0013$) and TUBG1 ($p = 0.0443$) between survivors and fatal cases. No significant difference was found in the transcript abundance of VCAM1 ($p = 0.4225$) and ISG15 ($p = 0.3009$), which contradicts the previous results obtained for these genes when were tested in the cohort of patients with known outcome ($p < 0.0001$, $p = 0.0036$ respectively) shown previously in Figure 4.2.



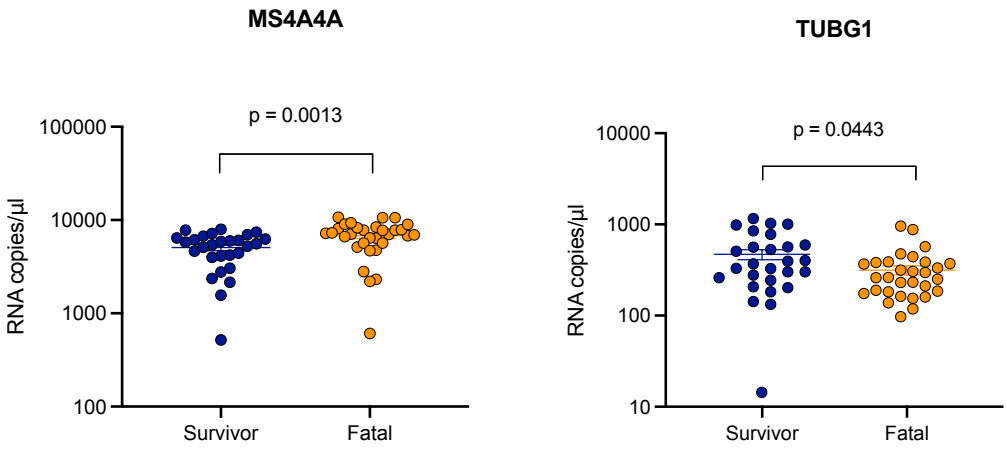


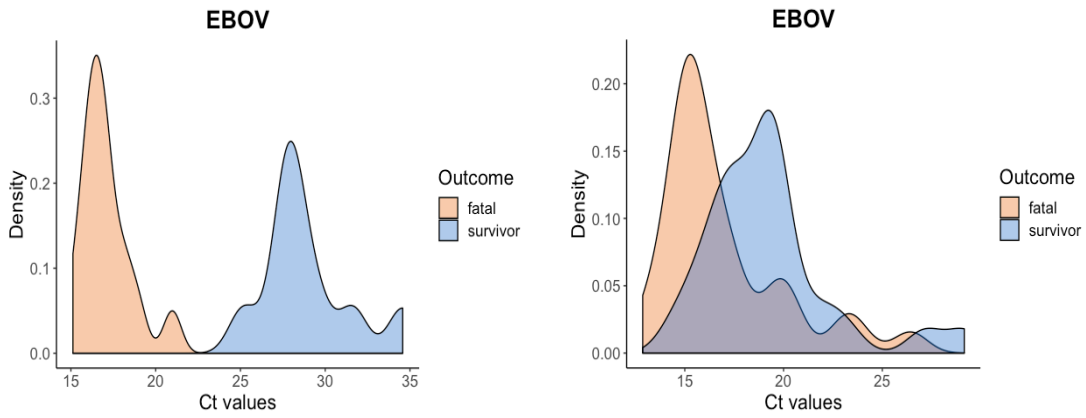
Figure 4.12 Transcript abundance of the four candidate predictive genes between survivors and non-survivors from the "blinded" cohort of patients. *The RNA copy number of the gene transcripts in the clinical samples were quantitated by RT-qPCRs using the standard curve method. The clinical samples were tested in duplicate and the mean RNA copy number per individual was obtained by plotting the mean Ct value against the standard curve created with in-vitro RNA dilutions of known concentration. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences in the transcript abundance between survivors and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

To visualize and compare the data between the cohort of patients with known outcome and the "blinded" cohort of patients, density plots were created to understand the distribution of each variable broken down by survivors and fatal cases. Figure 4.13 compares the density plots between both cohorts. The most interesting aspect of this graph was the perfect separation of survivors and fatal cases for the independent variable EBOV Ct in the cohort of patients with known outcome. Individuals with EBOV Ct values lower than 20 definitely belonged to fatal group, and higher Ct values were indicative of survival. This distribution was different from the one observed in the "blinded" cohort of patients, where the Ct values for survivors and fatal cases overlapped. Indeed, individuals that had Ct values lower than 20 survived.

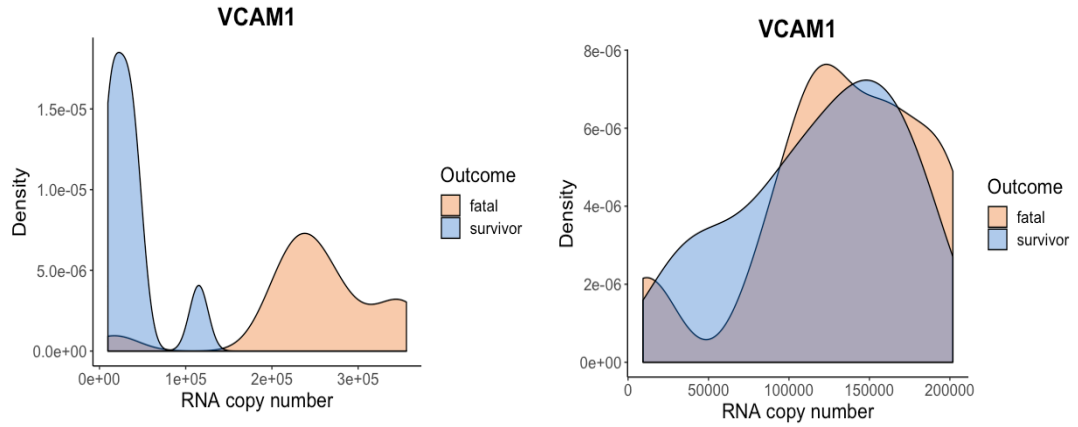
The distributions for VCAM1 and ISG15 were also different between both cohorts of patients showing overlapping of survivors and fatal cases in the "blinded" cohort of

patients. This corroborated the non-significant difference between survivors and fatal cases presented in Figure 4.12. TUBG1 showed a similar distribution in both cohorts but not a clear separation between survivors and fatal cases. MSA4A4 was the only gene transcript whose distribution was consistent in both cohort of patients showing a better separation between fatal cases and survivors, and the statistical results corroborated this difference. In addition, there were no observed differences in survivors and fatal cases according to age.

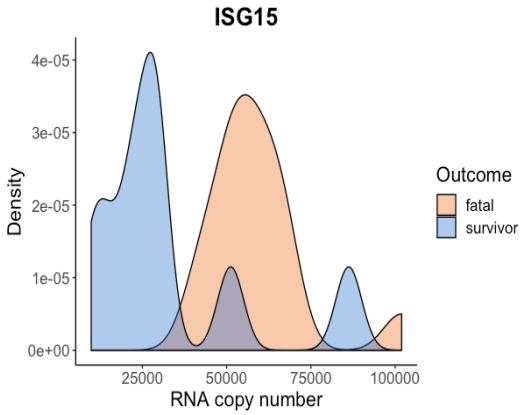
Cohort with known outcome **"blinded" cohort**



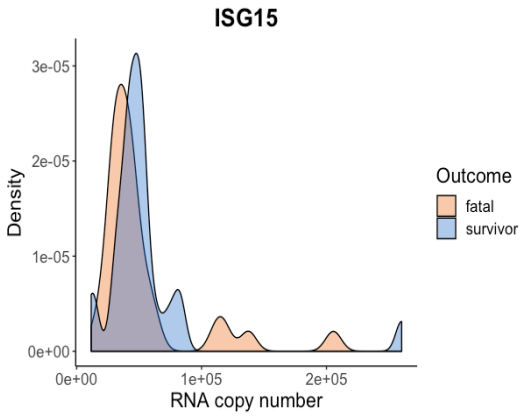
Cohort with known outcome **"blinded" cohort**



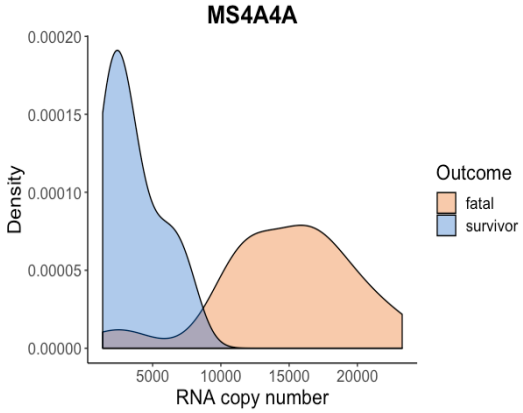
Cohort with known outcome



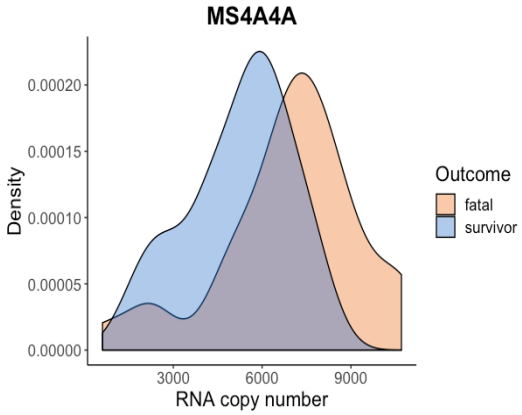
“blinded” cohort



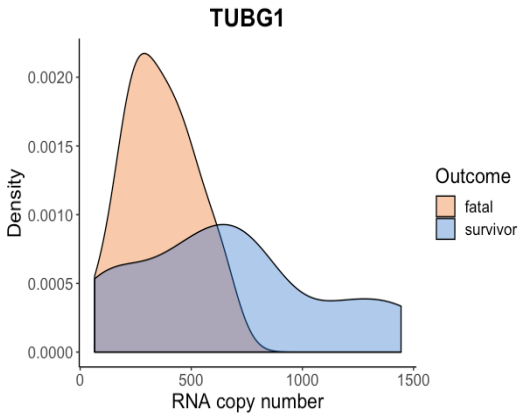
Cohort with known outcome



“blinded” cohort



Cohort with known outcome



“blinded” cohort

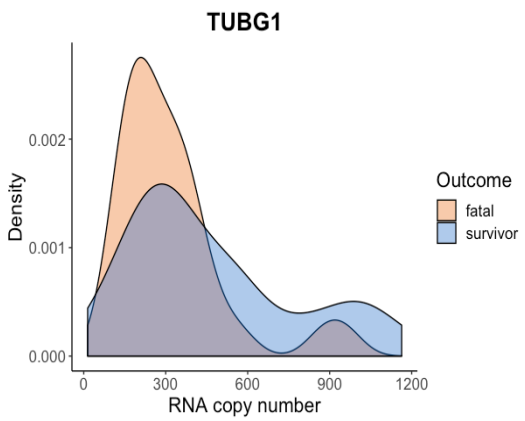
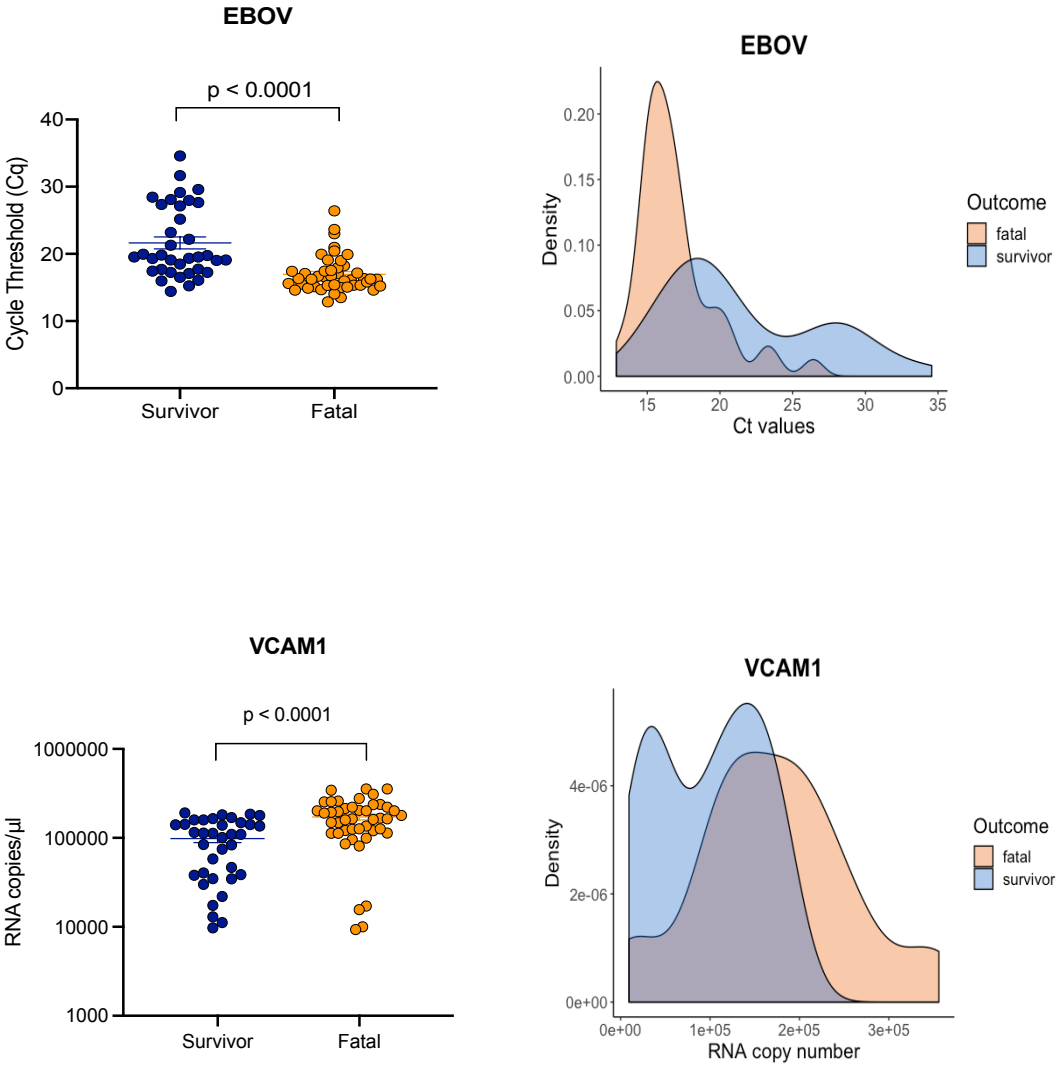


Figure 4.13 Density plots comparing the distribution of the four candidate predictive gene transcripts, EBOV Ct and age by outcome from both cohorts of EVD patients. *Kernel density estimation was used to visualize the distribution shape of the numeric variables (VCAM1, ISG15, MS4A4A TUBG1, EBOV Ct and age). The peaks of the density plot identify where values are concentrated over the interval of the continuous variable.*

The distributions of the predictor variables suggested that the small data size used for building the model may not represent the population. Therefore, to increase the reliability of measures, the samples that had complete values for all the predictor variables from both cohorts of patients (24 individuals from the cohort with known outcome and 57 individuals from the "Blinded" cohort) were joined in a single dataset for statistical analysis. It was assumed that both cohorts of patients came from the same population since all samples were collected during the EBOV outbreak in Guinea in 2014-2015 and followed the same collection methods and same RNA extraction. Thus, the whole dataset summed up to 81 subjects for analysis.

Figure 4.14 shows the transcript abundance of each predictor variable from the whole dataset (81 patients consisting of 36 survivors and 45 fatal cases) and the distribution of the values as density plots. Looking at Figure 4.14, we can see that there was a clear difference ($p < 0.0001$) in the EBOV Ct values reported from EVD survivors and fatal cases. Fatal cases had an average Ct value of 17 (median 16, range 13- 26) compared to Ct 22 (median 20, range 14-35) from survivors. This figure also shows the trend regarding the transcript abundance of the selected candidate

predictive genes. Interestingly, fatal cases had two times more MS4A4A and VCAM1 transcript copies compared to survivors ($p < 0.0001$). In contrast, the transcript abundance of TUBG1 was two times higher in survivors compared to fatal cases ($p = 0.0181$). Finally, there was no significant difference in the transcript abundance of ISG15 between survivors and fatal cases ($p = 0.2824$).



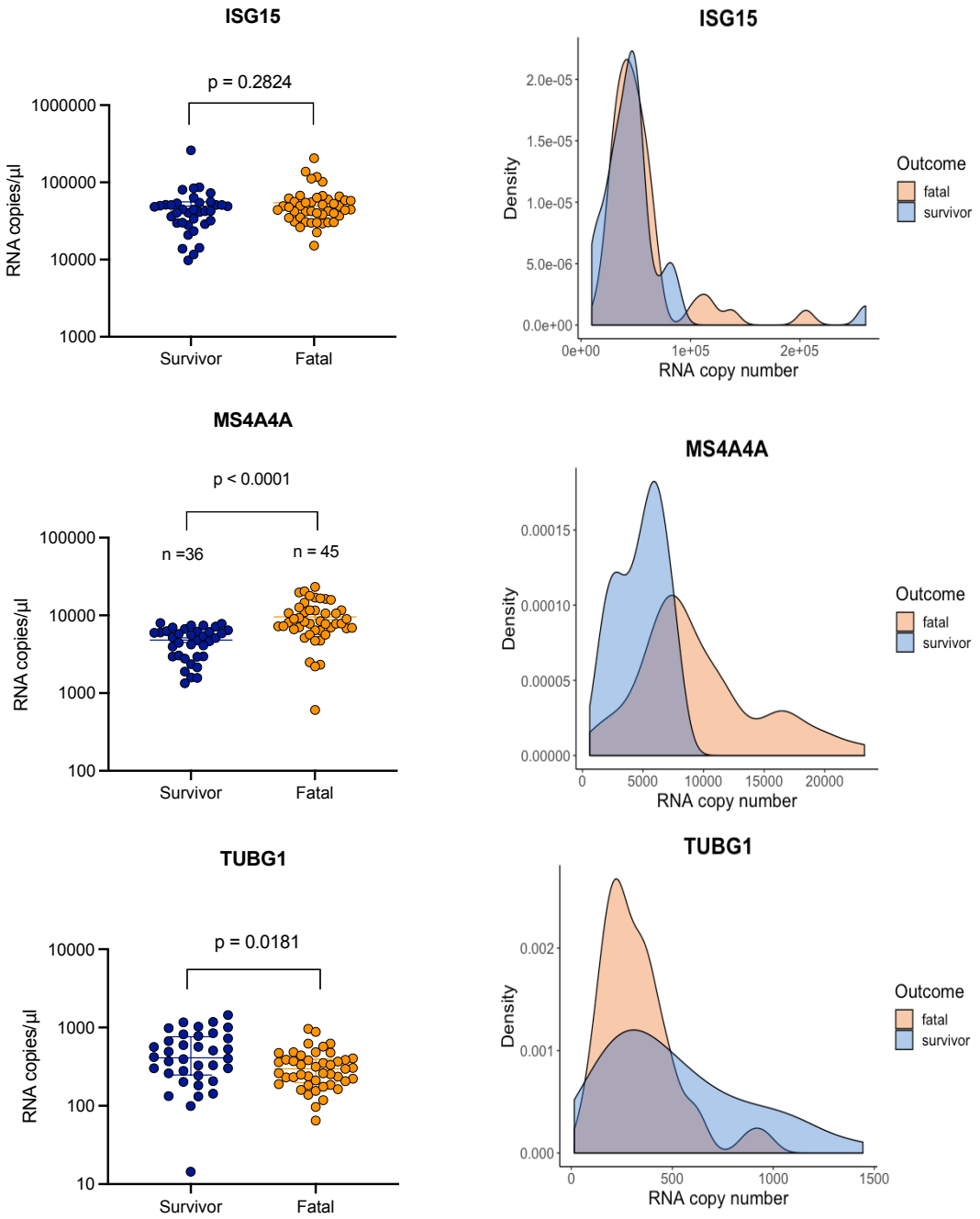


Figure 4.14 Distribution of the transcript abundance of the four candidate predictive genes and EBOV Ct values in the whole dataset of patients. *The graphs show the RNA copy number of VCAM1, ISG15, MS4A4A and TUBG1 in the clinical samples that were quantitated by RT-qPCRs using the standard curve method and the EBOV Ct values recorded from the same patients. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences between survivors and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

4.2.4 Machine learning modelling for the whole dataset of patients

4.2.4.1 Pre-processing of the whole dataset

To improve the accuracy of predictive models only the important variables for building the models were selected using the Boruta algorithm described before. The results obtained from this analysis are presented in Figure 4.15. The attributes that were relevant for the prediction of the outcome were MS4A4A, EBOV Ct, VCAM1, TUBG1 and ISG15. Surprisingly, the MS4A4A transcript had higher Z score values than EBOV Ct values, suggesting that may have better performance as a predictor variable than EBOV Ct. Age was a non-relevant attribute, thereby it was not considered for model building.

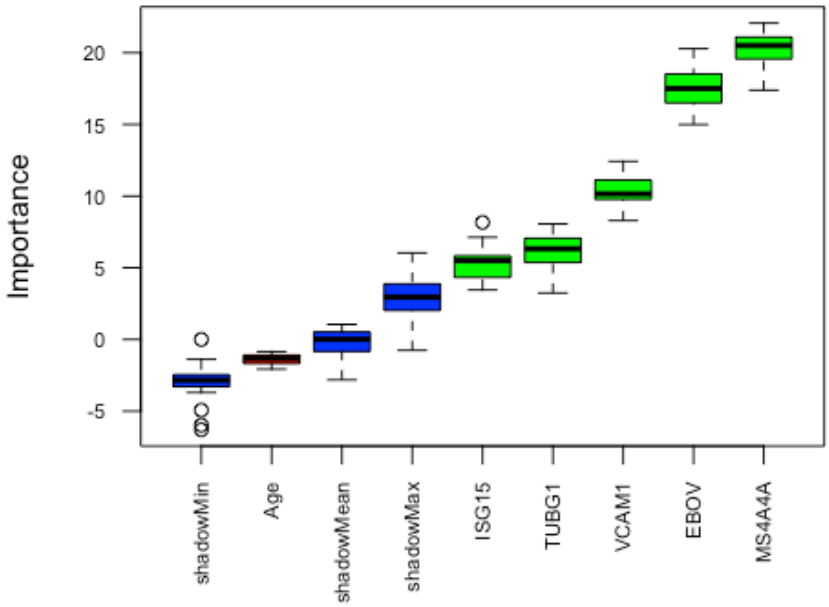


Figure 4.15 Boruta result plot for the predictor variables from the whole dataset. *The Boruta algorithm adds randomness to the data by creating shuffled copies ("shadows attributes") of all the attribute and runs a random forest classifier on the extended data to compute Z scores. It finds the maximum Z score among shadow attributes (MZSA) and assign a hit to every attribute that scored better than MZSA. At every iteration, checks whether an attribute has a higher Z score than the MZSA and consider them as "important" or as "unimportant" if the Z score is lower than MZSA, and removes them permanently. The algorithm stops after a limit of 500 random forest runs. The y-axis shows importance in terms of Z-scores. Blue boxplots correspond to minimal, average, and maximum Z score of a shadow attribute. Green boxplots represent Z scores of the confirmed important attributes. Red boxplot represented a rejected attribute. Box plot features are as follows: circles are outliers, dashed lines "whiskers" are 1.5 times the interquartile range, rectangle is the first and third quartiles, and horizontal bar is the median.*

The relationship between the attributes in the dataset was evaluated by Pearson correlation analysis. From Figure 4.16 we can see that EBOV Ct and VCAM1/MS4A4A had a moderate negative correlation (- 0.58 and - 0.45 respectively). VCAM1 and MS4A4A had a strong positive correlation (0.92) as seen before in section 4.2.2.1. A better separation between survivors and fatal cases was observed when EBOV Ct values were combined with VCAM1 and MS4A4A. To evaluate if the combination of these variables has a good prediction performance, new machine learning models were built.

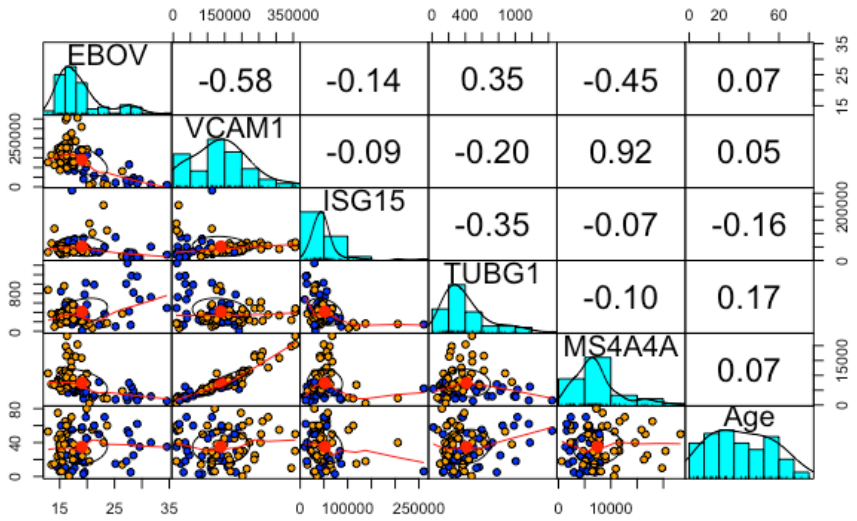


Figure 4.16 Scatterplot matrix of predictor variables for the whole dataset of EVD patients*. The plot shows the data from 81 patients (36 survivors represented as blue dots and 45 fatal cases represented as orange dots) from the whole dataset of EVD patients. Histograms are shown for each predictor variable. Below the diagonal, pairwise scatterplots show each combination of the host biomarkers, EBOV Ct and age. The upper panel above the diagonal shows pairwise Pearson's correlation coefficients obtained from the combination of the predictor variables.*

4.2.4.2 Evaluating different machine learning algorithms in the whole dataset

In order to choose the best machine learning algorithms to use for building models in the whole dataset, the strategy was to spot-check different algorithms as a starting point. The selection of models for spot-check was based on diversifying the choice of model complexity such as linear algorithms (Linear discriminant analysis, Logistic Regression), non-linear algorithms (Naive Bayes, k- Nearest Neighbours, Support

Vector Machine), trees (CART) and ensemble algorithms (Bagged CART, Random Forest, C5.0, Generalized Boosted Modelling).

Prior to model evaluation, the whole dataset was pre-processed as before following the flowchart presented in Figure 4.4. The whole dataset (81 patients) was split into 70% for training set (58 individuals consisting of 26 survivors and 32 fatal cases) and 30% for testing set (23 individuals consisting of 10 survivors and 13 fatal cases) that is hold back to use it during the model evaluation. Figure 4.17 provides the results obtained from the spot-check evaluation. From this graph we can see that three models were the most accurate classifiers: Support Vector Machine (86% accuracy, kappa = 0.71), k- Nearest Neighbors (85% accuracy, kappa = 0.71) and Random Forest (84% accuracy, kappa = 0.66). Only the two first algorithms were selected for further model tuning.

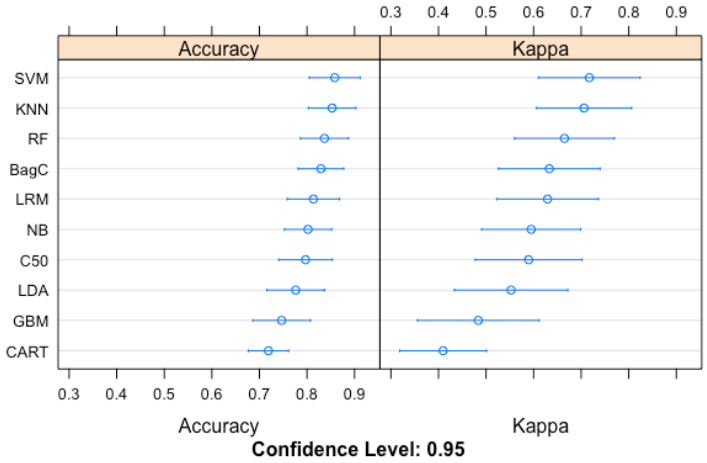


Figure 4.17 Evaluation of different machine learning algorithms in the whole dataset of patients*. The metrics used to evaluate all the machine learning algorithm were Accuracy and Kappa. Accuracy is the percentage of correctly classified patients out of the whole dataset. Kappa is based on the accuracy but corrected for a chance agreement. Kappa values of cero or less indicates a bad performance of the classifier and values near to 1 indicates a good classifier. SVM = Support Vector Machine, KNN = k- Nearest Neighbor model, RF = Random Forest, BagC = Bagged CART, LRM = Logistic Regression Model, NB = Naive Bayes, C50 = C5.0, LDA = Linear Discriminant Analysis, GBM = Generalized Boosted Modeling, CART*

4.2.4.3 Model parameter tuning and performance in the validation dataset

Each model was tuned according to each algorithm parameters. To evaluate the final model performance, the prediction was performed on the testing set (23 individuals consisting of 10 survivors and 13 fatal cases) which were the patients that were hold-out and not used to train the model. The performance of each model is presented as follow:

- Support Vector Machine models

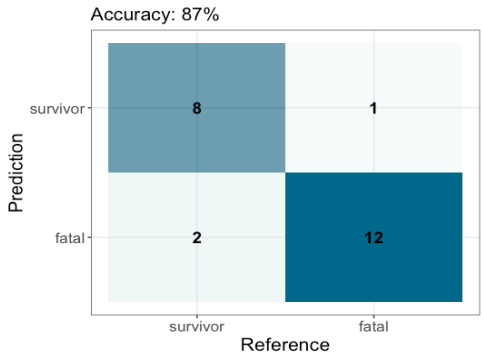
A SVM model that only used EBOV Ct values as the predictor variable correctly classified 87% (95% CI, 0.6641 - 0.9722) of patients from the testing set (Figure 4.18). Only one fatal case was misclassified with this model, 12/13 (92.3%) fatal cases and 8/10 (80%) survivors were correctly classified. Among all the models built with a single gene transcript, the MS4A4A model had the best performance and correctly classified 82.6% (95% CI,0.6122 - 0.9505) of patients from the testing set. This model correctly classified 12/13 (92.3%) fatal cases, and 7/10 (70%) survivors. VCAM1, ISG15 and TUBG1 models had lower predictive accuracy compared to EBOV or MS4A4A (70%, 65.2%, 74% respectively).

The combination of EBOV Ct and MS4A4A in a model improved the performance to 91.3% of accuracy (95% CI, 0.7196 - 0.9893) and correctly classified all the fatal cases and 8/10 (80%) survivors. Another model that had a high predictive accuracy 91.3% (95% CI, 0.7196 -0.9893) was the one that combined VCAM1, MS4A4A and ISG15. This model misclassified only one fatal case and one survivor indicating a 92.3% sensitivity and 90% specificity.

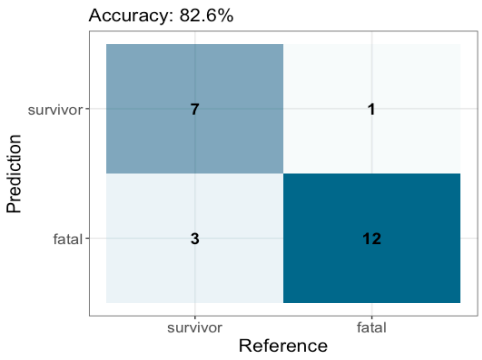
Between all the possible combinations of predictor variables, the model combining EBOV, ISG15 and MS4A4A variables had the best performance and was 100% (95% CI, 0.8518- 1) accurate classifying all the survivors and fatal cases from the testing set. Replacing MS4A4A for VCAM1 in the model also showed a good performance 96% (95% CI, 0.7805-0.9989). This model classified all fatal cases but

only one survivor was misclassified in 9/10 (90%) cases. The same high accuracy 96% (95% CI,0.7805-0.9989) was obtained in a model that used all the predictor variables (EBOV Ct, VCAM1, ISG15, MS4A4A and TUBG1) classifying all survivors but only one fatal case was misclassified in 12/13 (92.3%) cases.

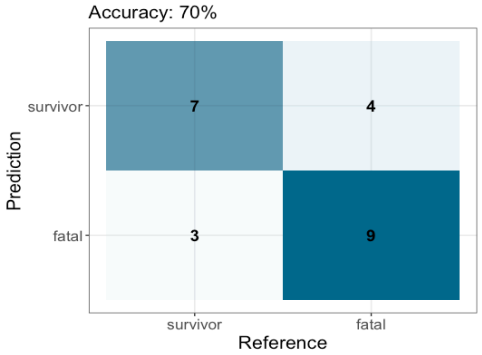
A) EBOV Ct



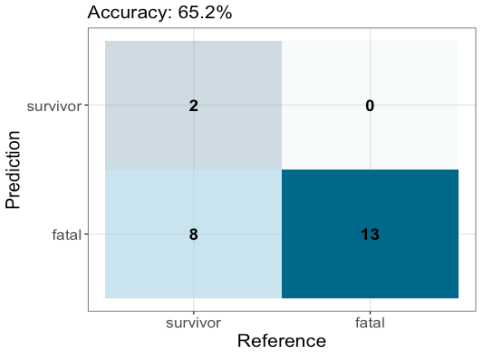
B) MS4A4A



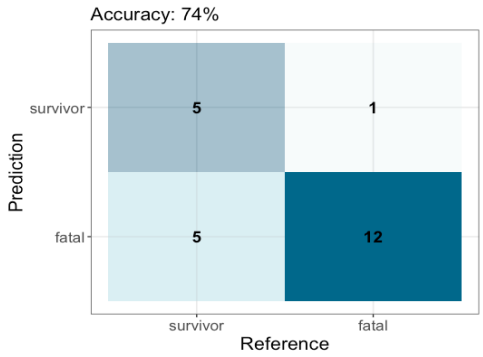
C) VCAM1



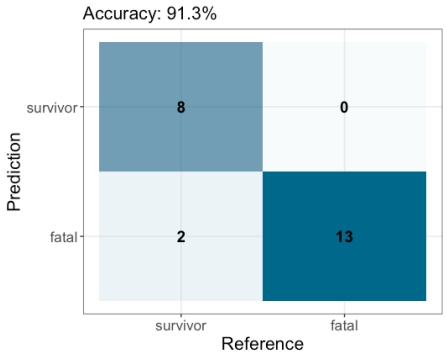
D) ISG15



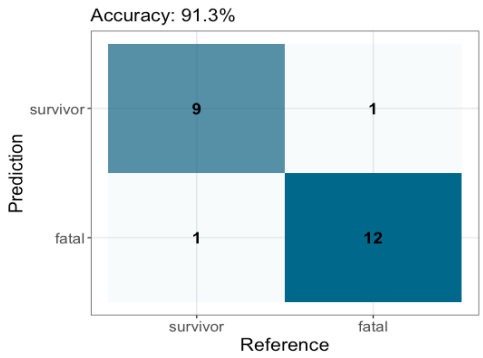
E) TUBG1



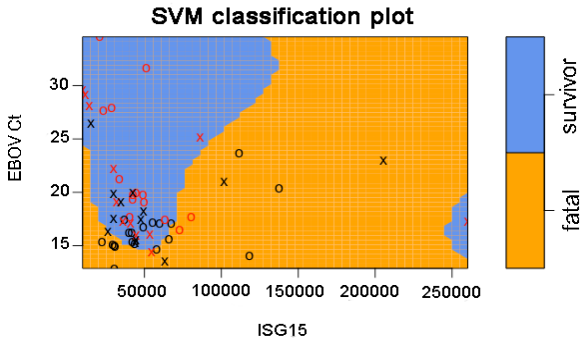
F) EBOV Ct + MSA4A4A

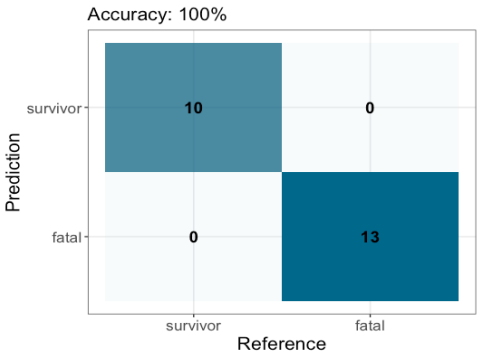


G) VCAM1+ MS4A4A + ISG15

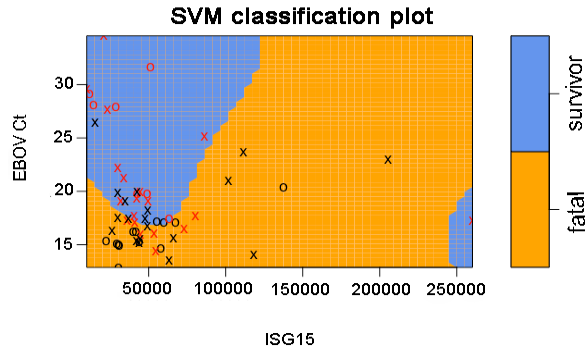


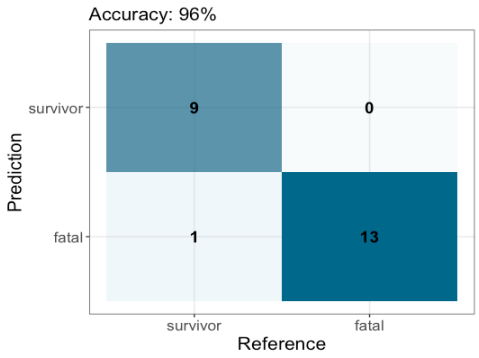
H) EBOV Ct + ISG15+ MS4A4A





I) EBOV Ct + ISG15 +VCAM1





J) EBOV Ct + VCAM + MS4A4A + ISG15 + TUBG1

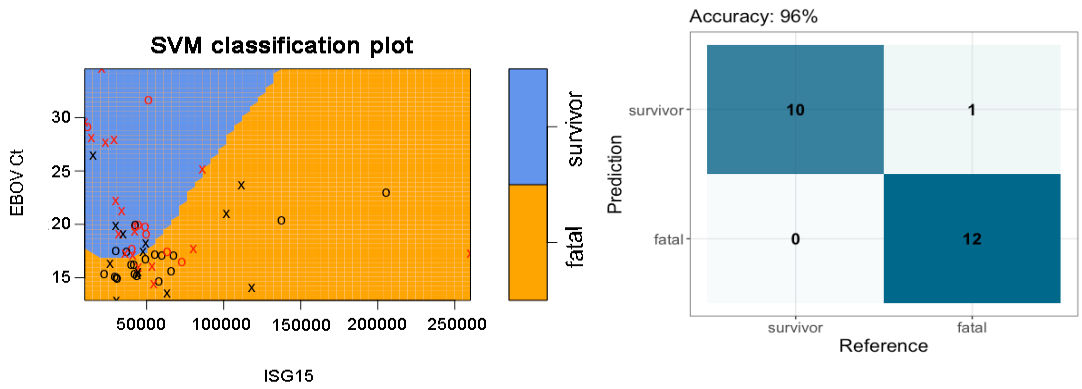


Figure 4.18 SVM models with confusion matrix showing models performance in the testing set. *For models with a combination of more than two variables a classification plot was created to visualize the data of the trained model. The classification plots show only 2 dimensions. The decision boundaries are created for both outcomes and are coloured in orange (fatal) and light blue (survivor). The support vectors are represented by crosses and the data points are the circles (in both cases red is for survival outcome and black for fatal outcome). The confusion matrix reports the predicted and actual outcome of the testing set (23 patients). The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

- k- Nearest Neighbour models

As seen in section 4.2.4.2, the K-nearest neighbour algorithm also showed a high accuracy when different algorithms were evaluated in the whole dataset of patients. This algorithm is described as instance-based learning, which does not learn a discriminative function from the training set, but memorize it instead (Figure 4.19) (Raschka & Mirjalili, 2017).

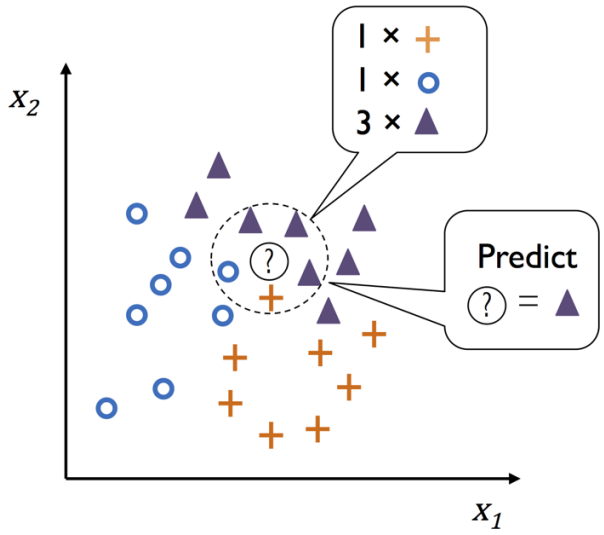
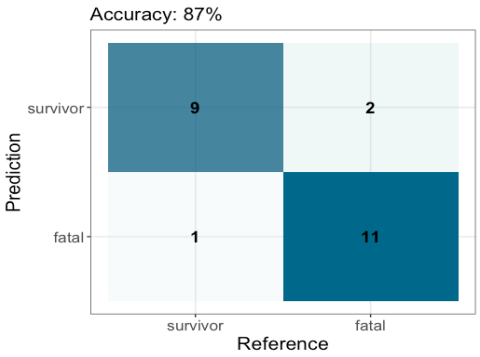


Figure 4.19 Representation of the KNN algorithm function. *After choosing the number of K and a distance metric, this algorithm finds the K nearest or similar samples in the training set to the sample we want to classify and assign the label by majority vote. For example, a new sample (?) is assigned the triangle label since it is the most voted (3) among its five nearest neighbours. Graph taken from Raschka and Mirjalili (2017).*

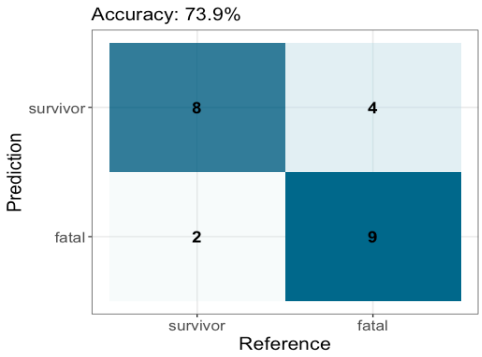
After tuning the models using K = 20 nearest neighbours, the model trained with EBOV Ct values and MS4A4A showed the highest accuracy 87% (95% CI, 0.6641 - 0.9722) in predicting the outcome of patients from the testing set (23 individuals consisting of 10 survivors and 13 fatal cases) (Figure 4.20). This model correctly classified almost all the fatal cases 11/13 (84.6%) and misclassified only one survivor 9/10 (90%).

The model built with only EBOV Ct values correctly predicted the outcome of 73.9% (95% CI, 0.5159-0.8977) patients from the testing set. This model correctly classified 9/13 (69.2%) fatal cases and 8/10 (80%) survivors. The use of all predictor variables together did not improve the prediction performance, and the accuracy decreased to 78.3% (95% CI, 0.563- 0.9254). In this model 10/13 (76.9%) fatal cases and 8/10 (80%) survivors were correctly classified.

A) EBOV Ct + MS4A4A



B) EBOV Ct



C) EBOV Ct + VCAM1 + ISG15 + MS4A4A + TUBG1

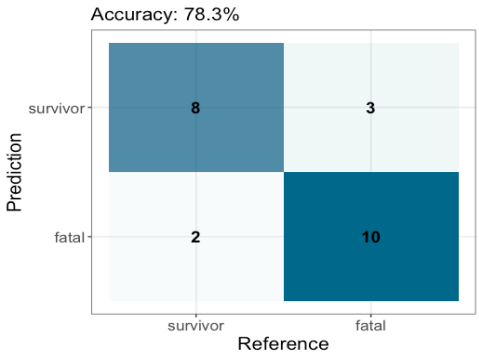


Figure 4.20 Confusion matrix showing the KNN models performance. *The confusion matrix reports the predicted and actual outcome of the testing set (23 patients). The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

4.2.5 Machine learning modelling for the DNase-treated samples

Due to the possibility that our results may have been influenced by the co-amplification with remaining genomic DNA, leftover samples were re-tested. Forty-eight leftover samples (23 survivors and 25 fatal cases) were first treated with DNase for a complete removal of genomic DNA. Since our previous results showed that MS4A4A is strong marker of EVD mortality, to confirm these findings the 48 samples were re-tested by SYBR green-based RT-qPCR (Appendix Table 7). The density analysis of the 48 samples shows the distribution of the EBOV Ct values and MS4A4A transcript abundance between survivors and fatal cases (Figure 4.21). No clear separation is observed between both groups.

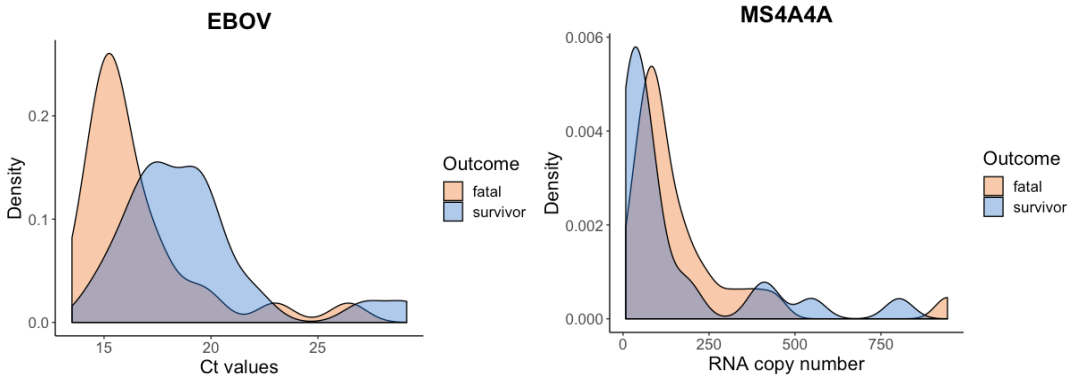


Figure 4.21 ***Density plots comparing the distribution of EBOV Ct values and MS4A4A transcript by outcome in the 48 DNase-treated clinical samples.*** *Kernel density estimation was used to visualize the distribution shape of the numeric variables (EBOV Ct values and MS4A4A). The peaks of the density plot identify where values are concentrated over the interval of the continuous variable.*

The Mann–Whitney U test analysis showed a significant difference (p = 0.0403) in the transcript abundance of MS4A4A between survivors and fatal cases (Figure 4.22).

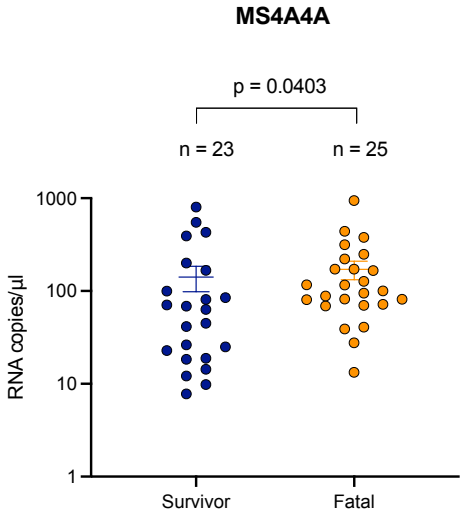


Figure 4.22 ***Transcript abundance of MS4A4A between survivors and fatal cases.*** *The RNA copy number of MS4A4A from 48 DNase-treated clinical samples were quantitated by RT-qPCR using the standard curve method. The clinical samples were tested in duplicate and the mean RNA copy number per individual was obtained by plotting the mean Ct value against the standard curve created with in-vitro RNA dilutions of known concentration. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences in the transcript abundance between survivors and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

Different SVM models were developed using EBOV Ct values and MS4A4A transcript as predictor variables. The data was randomly split in 70% (35 individuals) for the training set, and 30% (13 individuals) for the testing set. The models were built with the training set and the model performance was evaluated in the testing set. The model built with MS4A4A transcript correctly predicted the outcome of 77% (95% CI, 0.4619 - 0.9496) patients from the testing set (Figure 4.23). This model correctly classified all fatal cases (100%) and 3/6 (50%) survivors. The model built with EBOV Ct values correctly classified the outcome of 70% (95% CI, 0.3857 - 0.9091) patients including 5/7 (71.4%) fatal cases and 4/6 (67%) survivors. Combining both variables did not improve the prediction performance, and the accuracy decreased to 61.5% (95% CI, 0.3158 -0.8614).

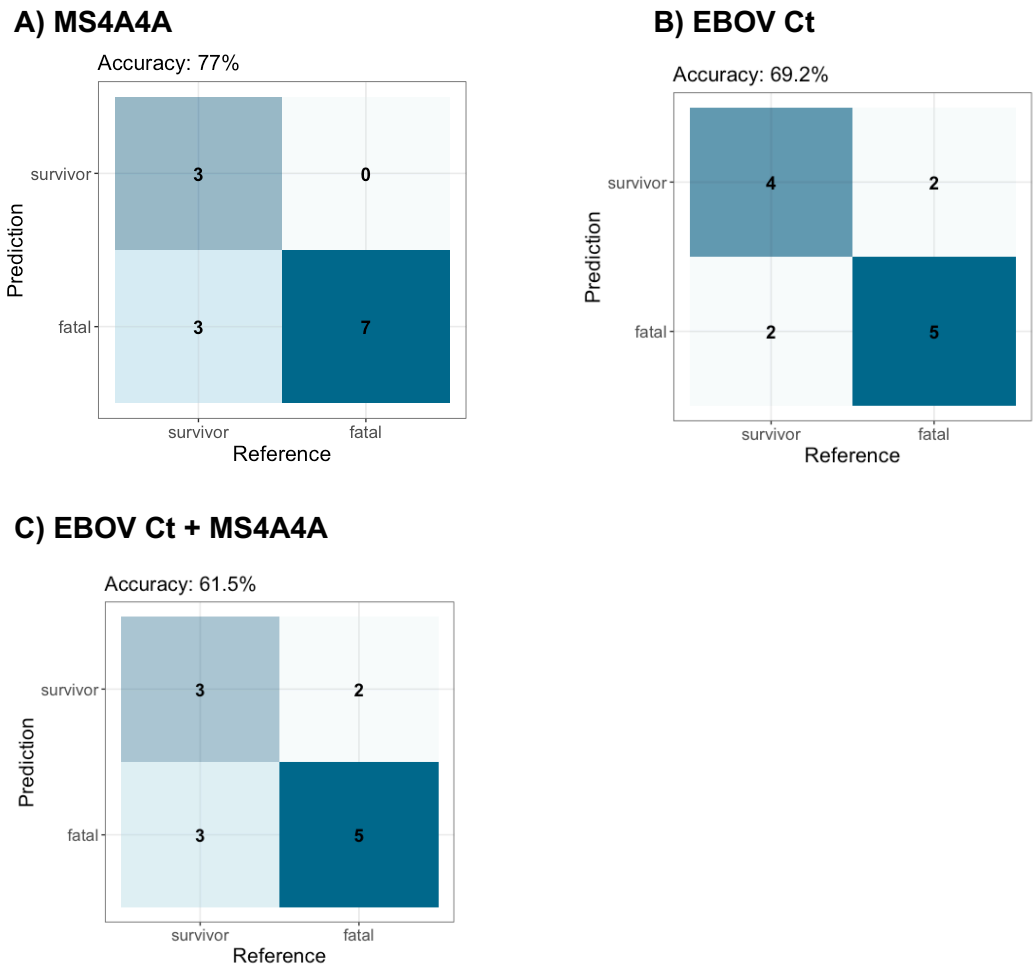


Figure 4.23 ***Confusion matrix showing the SVM models performance.*** *The confusion matrix reports the predicted and actual outcome of the testing set (13 patients). The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

Table 4.2 shows the predicted outcomes for the testing set using the EBOV Ct model and MS4A4A transcript model. The predictions from the model that combines both variables are not included because it had lower prediction accuracy than testing EBOV and MS4A4A alone. Correct predictions are highlighted in yellow.

Table 4.2 Predicted outcomes for the testing set

	Actual outcome	Days since symptom onset	EBOV Ct value	MS4A4A copy number	Predicted outcome	
					MS4A4A model	EBOV Ct model
RNA-04	survivor	0	19.08	167.0776	fatal	survivor
RNA-06	survivor	10	19.76	803.0531	survivor	survivor
RNA-08	survivor	8	14.42	18.84758	survivor	fatal
RNA-21	survivor	4	15.24	392.6698	fatal	fatal
RNA-23	survivor	3	19.55	44.85136	survivor	survivor
RNA-28	fatal	NA*	14.66	248.3702	fatal	fatal
RNA-36	fatal	7	14.99	69.08459	fatal	fatal
RNA-41	fatal	5	15.33	127.2507	fatal	fatal
RNA-50	fatal	10	16.26	100.0631	fatal	fatal
RNA-52	fatal	2	16.22	315.3989	fatal	fatal
RNA-54	fatal	2	23	221.0892	fatal	survivor
RNA-59	fatal	1	26.42	116.4044	fatal	survivor

* Data not available

To ensure that these results can replicate we randomly split the whole dataset more times into different training and testing sets. The results were similar, the model trained with only MS4A4A transcript as a predictor variable showed higher overall accuracy 69.2% (95% CI, 0.3857 - 0.9091) than the model that used EBOV Ct values 54% (95% CI, 0.2513 - 0.8078) (Figure 4.24). MS4A4A transcript showed to be a strong predictor of fatality since it correctly classified more fatal cases than EBOV Ct variable (85.7% versus 42.9%). Similar to previous results, combining both variables did not improve the prediction performance, and the accuracy dampened to 46.2% (95% CI, 0.1922 - 0.7487). Table 4.3 shows the predicted outcomes for this new random testing set by the EBOV Ct model and MS4A4A transcript model. Correct predictions are highlighted in yellow.

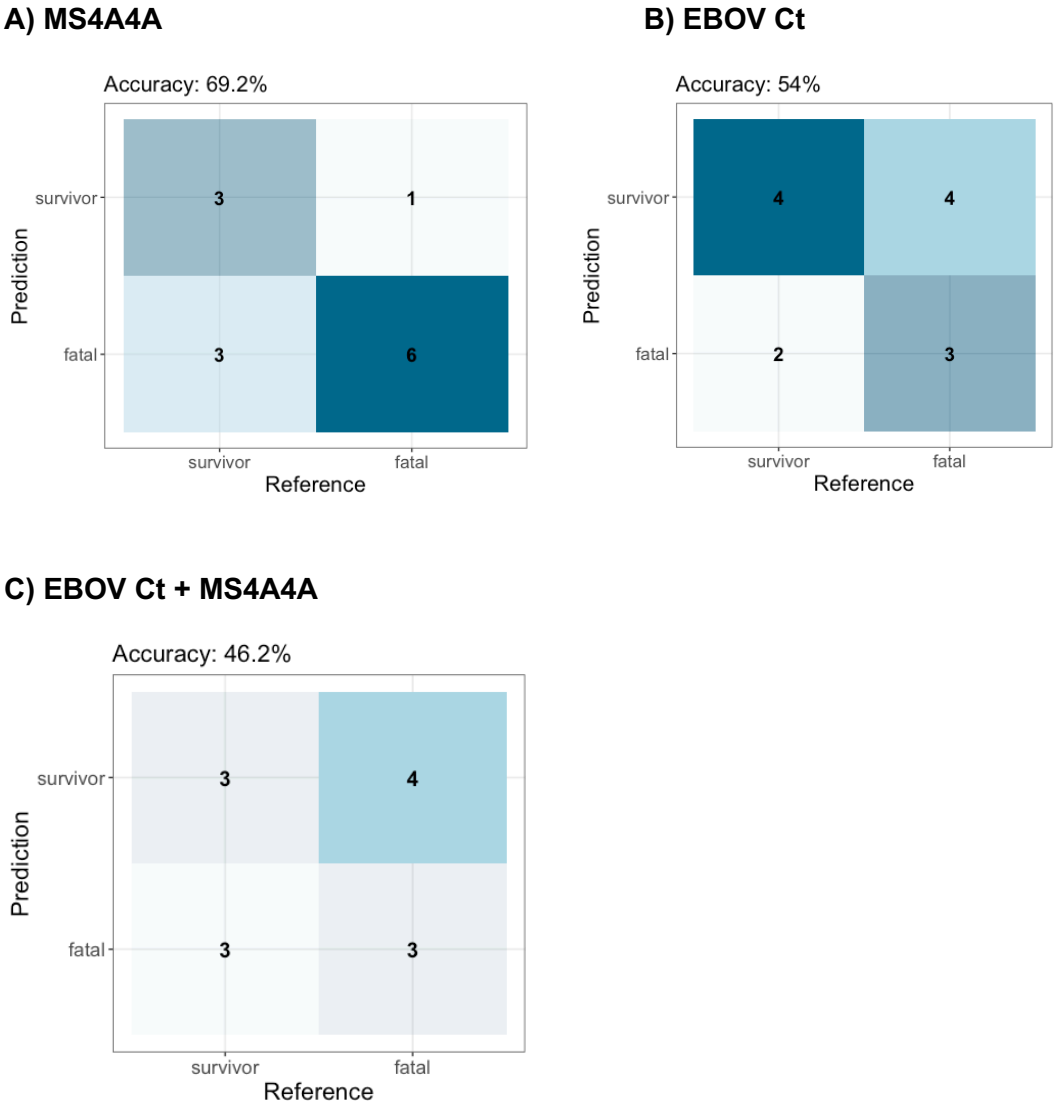


Figure 4.24 ***Confusion matrix showing the SVM models performance in a new random testing set.*** *The confusion matrix reports the predicted and actual outcome of the new testing set (13 patients). The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

Table 4.3 Predicted outcomes for the new random testing set

	Actual outcome	Days since symptom onset	EBOV Ct value	MS4A4A copy number	Predicted outcome	
					MS4A4A model	EBOV Ct model
RNA-02	survivor	11	17.25	200.395815	fatal	survivor
RNA-05	survivor	4	17.73	7.783173	survivor	survivor
RNA-07	survivor	7	19.04	84.784204	fatal	survivor
RNA-11	survivor	3	16.08	99.895662	fatal	fatal
RNA-14	survivor	6	15.99	41.39615	survivor	fatal
RNA-22	survivor	3	17.24	14.332849	survivor	survivor
RNA-34	fatal	10	19.95	81.761475	fatal	survivor
RNA-37	fatal	5	14.6	94.660444	fatal	fatal
RNA-40	fatal	5	14.03	13.291016	survivor	fatal
RNA-45	fatal	3	15.4	171.843689	fatal	fatal
RNA-56	fatal	3	17.41	88.432773	fatal	survivor
RNA-57	fatal	5	17.53	71.757409	fatal	survivor
RNA-59	fatal	1	26.42	116.40438	fatal	survivor

To analyse the potential effect of the time between symptom onset and sample collection/diagnosis on our models, we first analysed the distribution of this variable. Only one sample (RNA-28) was excluded from the analysis since data was not available. The density analysis of the variable “Days since symptom onset” did not show a clear association with the clinical outcome and its inclusion in a prediction model with EBOV Ct and MS4A4A transcripts did not improve the model performance, showing only 62% accuracy (95% CI, 0.3158- 0.8614) (Figure 4.25). MS4A4A transcript and EBOV Ct models had better performance when were used alone with 77% accuracy (95% CI, 0.4619 - 0.9496) compared to 69.2% accuracy (95% CI, 0.3857- 0.9091) of variable “Days since symptom onset”. These results suggest that both MS4A4A transcripts and EBOV Ct are strong predictors of EVD mortality and can be used independently.

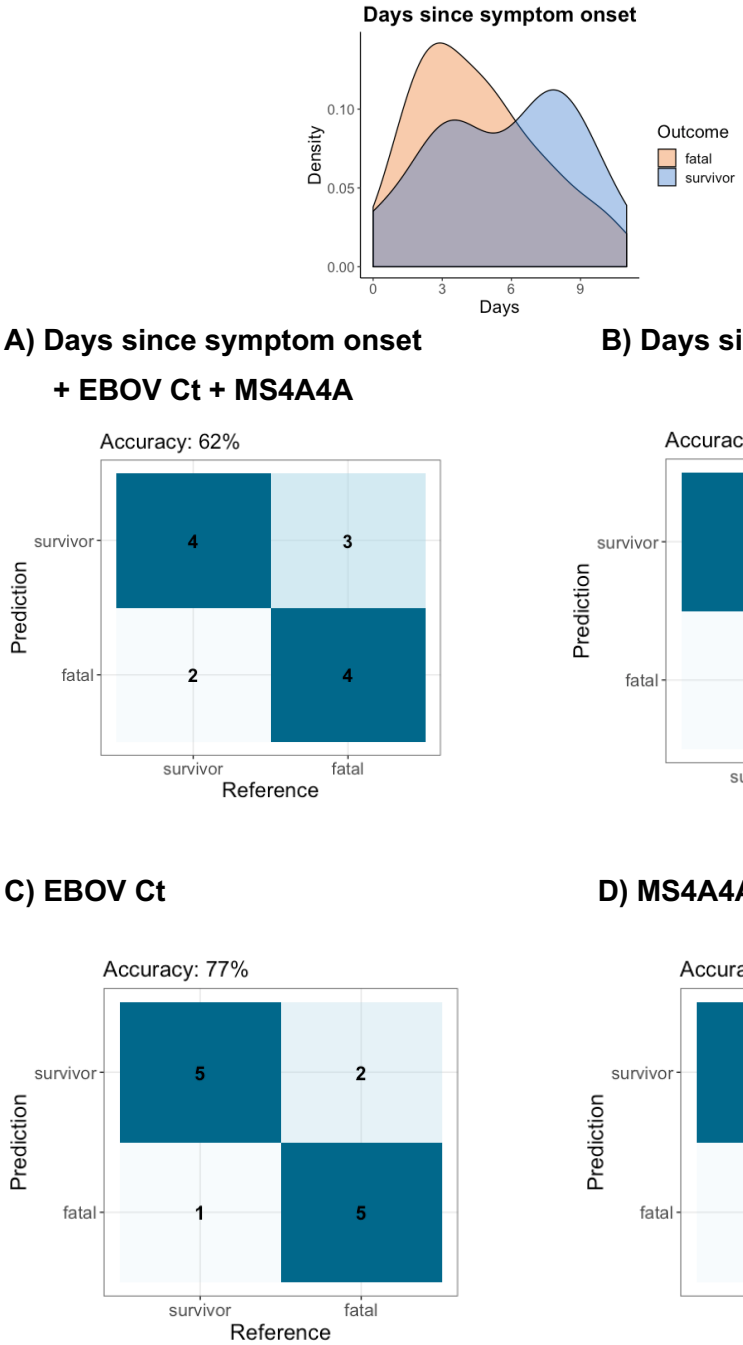


Figure 4.25 Density plot of the variable "Days since symptom onset" and confusion matrix showing the SVM models performance. *Kernel density estimation was used to visualize the distribution shape of the numeric variable "Days since symptom onset". The peaks of the density plot identify where values are concentrated over the interval of the continuous variable. The confusion matrix reports the predicted and actual outcome of the testing set (13 patients).*

4.3 Discussion

The first section of this chapter describes work that determined and analysed the transcript abundance of 11 potential candidate predictive genes (VCAM1, ISG15, CTSL, TUBG1, MS4A4A, TTC28, TGFBI, PLPP3, NF3L1, SLC25A5 and HOPX) by SYBR Green-based RT-qPCRs in 39 acute EVD patients with known outcome collected from the outbreak in Guinea in 2014-2015. From all the candidate predictive genes, the transcript abundance of VCAM1 ($p < 0.0001$), ISG15 ($p = 0.0036$), MS4A4A ($p < 0.0001$), TTC28 ($p = 0.0077$) and TGFBI ($p = 0.0303$) was significantly different between survivors and fatal cases. The fatal cases had higher transcript levels of MS4A4A (4 times more) and VCAM1 (3 times more) compared to survivors. These results are consistent with the transcriptomic data obtained by testing blood from a cohort of acutely infected patients in a previous study by our group (Liu et al., 2017). The transcriptomic data showed a higher VCAM1 (FDR = 7.25E-13) and MS4A4A (FDR = 6.19E-05) expression in the acute-fatal compared to the acute-survivor group (Liu et al., 2017). Our RT-qPCRs results showed a higher ISG15 transcript abundance (1 time more) in fatal cases compared to survivors. This result differed from the transcriptomic data obtained by Liu et al. (2017), where no significant difference (FDR = 0.02) was found in the ISG15 expression between survivors and fatal cases. Our analysis also showed that the transcript abundance of TUBG1 was not significantly different between survivors and fatal cases ($p = 0.1610$), corroborating the transcriptomic data (FDR = 0.56). The transcript abundance of CTSL1 in this cohort of samples was not different between survivors and fatal cases ($p = 0.5019$), in contrast to the transcriptomic data analysis (4.05E-09) by Liu et al. (2017).

Although the RT-qPCRs results showed that fatal cases had lower transcript levels of TTC28 (2 times less) and TGFBI (16 times less) than survivors in the cohort of samples with known outcome, the statistical results for these gene transcripts should be interpreted with caution because very few survivors were analysed. Likewise, the statistical analysis for PLPP3, NF3L1, SLC25A5 and HOPX was not possible due to the limited number of analysed samples.

The second section of this chapter was focused on building Support Vector Machine models using the data from the patients with known outcome to predict the clinical outcome of an independent set of samples consisting of 64 EVD patients (blinded samples). Although our attempt to correctly predict the outcome of the blinded cohort of samples was not completely successful, these results showed that the best models for predicting the disease outcome were models trained with EBOV Ct, ISG15, MS4A4A and TUBG1 (68.4% accuracy) and the model that combined MS4A4A, ISG15 and TUBG1 (67% accuracy). The former was very accurate predicting fatal cases and the latter survivors. Despite the low accuracy obtained from these models, the most interesting finding was that they had better performance than the model that only included EBOV Ct values (54.3%) as a predictor variable. These results suggested that EBOV Ct value is a strong predictor of EVD mortality; however, the overall prediction improves when other gene transcripts such as MS4A4A, VCAM1, ISG15 or TUBG1 are included in the prediction model.

The low accuracy (< 70%) obtained in the SVM models may have been influenced by two factors that induce to bias: 1) the extreme Ct values from the dataset as described in section 4.2.1 (survivors Ct values ranged from 25.18 to 34.5, and for fatal cases Ct values from 15.12 to 20.98), and 2) the small sample size. It is known that a small sample size could lead to low accuracy and an increase in the classification error. Therefore, using a higher sample size could enhance the prediction performance of machine learning models (Raudys & Jain, 1991). Hence, it was necessary to build more reliable models in a higher sample size with a wider range of Ct values joining the cohort of patients with known outcome and the “blinded” cohort of patients. This also helped to have a dataset that could represent better the population.

It is certain that prediction models could lead to overly optimistic results due to the possibility of bias during the model building. However, the results obtained using different machine learning algorithms such as SVM and KNN for predicting the outcome of patients from the testing set (unseen data taken from the whole dataset) were quite similar and the accuracy of the best models was greater than 90%.

Since our results were based on clinical samples that were not DNase treated to avoid loss of RNA due to an extra purification, we could not rule out the possibility of co-amplification of genomic DNA in some RT-qPCRs for primer pairs which were located within a single exon (VCAM1, HOPX, SLC25A5, MS4A4A and ISG15). Although is not clear whether genomic DNA contamination could have biased our results, we decided to re-test leftover samples after DNase treatment to validate our findings for the MS4A4A transcript. The results of 48 DNase-treated clinical samples corroborated our previous findings that showed a significant difference in the MS4A4A transcript abundance between survivors and fatal cases ($p = 0.0403$). MS4A4A transcript and EBOV Ct values demonstrated to be good predictors of EVD mortality; however, MS4A4A transcript outperformed EBOV Ct model (77% versus 69.2% accuracy).

Since the Ct values vary during the course of infection, it was important to understand the potential effect of the time between symptom onset and the sample collection/EBOV diagnosis in our predictive models. For this, we developed SVM models that included this variable. Our findings showed that MS4A4A transcript and EBOV Ct value greatly outperformed this variable in predicting mortality. Besides, MS4A4A transcript and EBOV Ct have the ability to predict the outcome in an independent manner. This is important since the time data on symptom onset is not very reliable because depends on the patient's state of health (whether is responsive and with a clear mind or not).

The analysis of the 48 DNase treated samples also showed that MS4A4A transcript is a marker that predict the outcome independent of Ct value. This dataset had a wider range of Ct values; Ct values ranged from 13.5 to 26.42 in fatal cases, and from 14.42 to 29.15 in survivors. Since Ct values vary during the course of the disease, an independent marker like MS4A4A transcript may be useful for predicting the outcome of patients where the Ct values at the time of diagnosis/hospital admission is uninformative (Ct values 19-22). Further investigations should examine a larger cohort of patients to validate our findings.

Chapter 5

Development of a Multiplex RT-qPCR based on four selected gene transcripts and application to clinical samples

5.1 Introduction

The development of a multiplex RT-qPCR assay has been conducting in parallel with the SYBR green-based RT-qPCR analysis of clinical samples shown in Chapter 4. Thus, the selected gene transcripts: VCAM1, ISG15, MS4A4A and TUBG1, were the targets used for the development of a multiplex RT-qPCR. This chapter sought to develop a multiplex RT-qPCR assay because compared to other PCR-based techniques it gives a faster detection since it can amplify multiple targets at the same time in a single reaction. This was very important in our study since we had several gene transcripts to test and the multiplex RT-qPCR can be used in situations when the sample input is limited, as occurred with the leftover clinical samples we received from the team in Germany. This technique also reduces the use and cost of reagents.

So far, multiplex RT-qPCR has been mainly used in the diagnosis of pathogens e.g. for the early diagnosis of highly pathogenic emerging viruses such as Influenza viruses (Cui et al., 2016; Ma et al., 2015), haemorrhagic fever viruses (Pang et al., 2014), Zika and Chikungunya (Broeders et al., 2020). Only few studies have used this technique for testing prognostic markers of a disease. Most of these studies were related to cancer research and have tested prognostic markers by multiplex RT-qPCR to predict the clinical outcome of patients receiving different treatments like immunotherapy (Gupta et al., 2019). In Ebola studies, PCR-based tests have been mainly used for the detection of the virus in clinical samples but not for testing host genes as prognostic markers. In this context it is necessary the development of field-deployable prognostic tools for the management of patients that could eventually improve the overall control of the epidemic.

The multiplex RT-qPCR developed in this chapter used a probe-based chemistry. This involved the design of specific primers and probes for each of the four gene transcripts (VCAM1, ISG15, MS4A4A and TUBG1) which are combined in a single reaction vessel. Currently, the multiplex RT-qPCR technique is able to detect four to six nucleic acid targets in a single sample (Law et al., 2015).

The aim of this chapter was to develop a multiplex RT-qPCR for the four gene transcripts: VCAM1, ISG15, MS4A4A and TUBG1 that could be used to predict the clinical outcome of EVD patients. The performance of the optimized assay was then evaluated in samples from the "blinded" cohort of EVD patients previously used in Chapter 4.

5.2 Results

5.2.1 Optimization of singleplex (individual) assays before multiplexing

While chapter 3 showed the development of one-step RT-qPCRs for each gene transcript based on SYBR-green chemistry, this chapter developed a multiplex RT-qPCR assay based on probe chemistry. To develop a multiplex RT-qPCR assay that could measure the transcript abundance of the four selected genes transcripts (VCAM1, ISG15, MS4A4A and TUBG1) in a single reaction tube, it was first necessary to optimize a singleplex RT-qPCR assay for each gene. For this, we used the same primers sets designed for the one-step SYBR green-based RT-qPCRs described in Chapter 3 and only probes were designed. Figure 5.1 shows the alignments of the transcript sequences from VCAM1, ISG15, MS4A4A and TUBG1 and highlights the location of the designed probes. The specificity of the probes was verified through BLAST search in GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The probes were labelled to distinguish each PCR amplicon: the fluorescent reporters on the 5' ends and the compatible quenchers on the 3' ends. A modification of VCAM1 and MS4A4A probes was done by adding Locked Nucleic Acids (LNA) which are nucleic acid analogues. The introduction of LNAs increases the thermal stability of the probe, which is important for making a strong bond with the complementary sequence. It also increases the annealing temperature which was necessary for a short length probe like VCAM1 (15 bases). Thus, all probes had a higher annealing temperature than the primers between 63.1° to 71.1°C. The sequences of the primers and probes used in this chapter are listed in Table 5.1

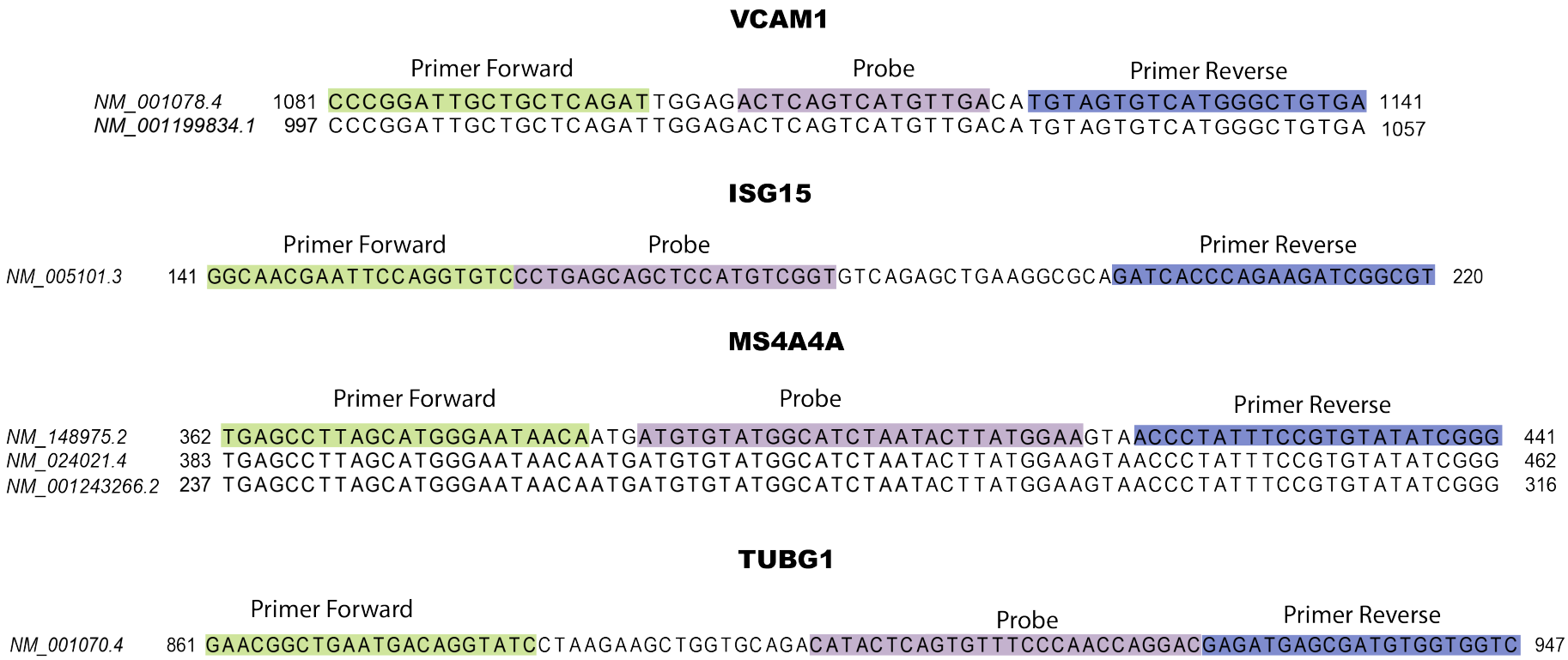


Figure 5.1 Location of the probes designed for the four selected gene transcripts. *The RefSeq-annotated transcript of each gene are identified by the accession number in the left side. The sequences were aligned using MAFFT multiple alignment program and visualized using Jalview (www.jalview.org).The green shadow shows the position of the primer forward, the purple shadow shows the position of the probes and the blue shadow shows the position of the primer reverse.*

Table 5.1 Sequences of primer and probe sets designed for the development of a multiplex RT-qPCR assay

Gene	NCBI Accession number	Sequences 5'- 3'	Amplicon length (bp)	Nucleotide position of the amplicon
VCAM1	NM_001199834.1	For: CCCGGATTGCTGCTCAGAT Rev: TCACAGCCCATGACACTACA Probe: *HEX*-A+CT+CA+GT+CAT+GTT+GA-*BHQ1*	61	997-1057
ISG15	NM_005101.3	For: GGCAACGAATTCCAGGTGTC Rev: ACGCCGATCTTCTGGGTGATC Probe: *TxRd*-CCTGAGCAGCTCCATGTCGGT-*BHQ2*	80	141-220
MS4A4A	NM_148975.2	For: TGAGCCTTAGCATGGGAATAACA Rev: CCCGATATACACGGAAATAGGGT Probe: *FAM*-+ATGTGTATGGCATCTAATACTTATGGAA-*BHQ1*	80	362-441
TUBG1	NM_001070.4	For: GAACGGCTGAATGACAGGTATC Rev: GACCACCACATCGCTCATCTC Probe: *CY5*-CATACTCAGTGTTTCCCAACCAGGAC-*BHQ3*	87	861-947

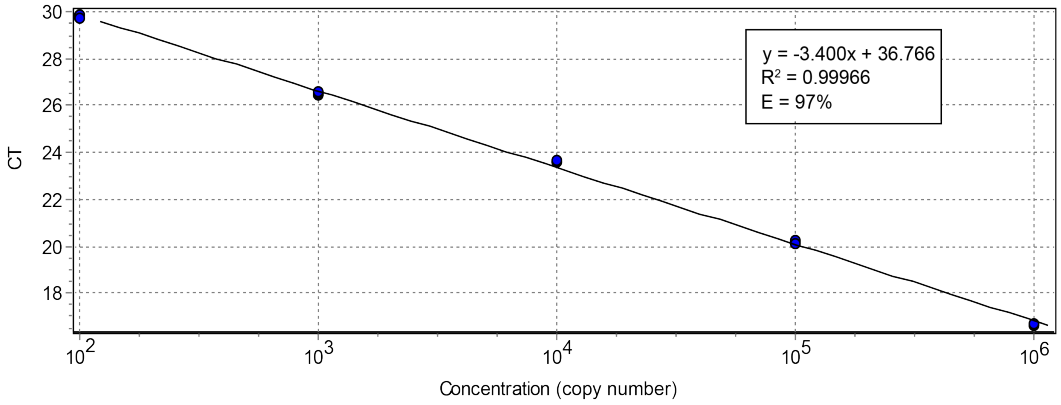
"+" indicates the LNA modification of the following base

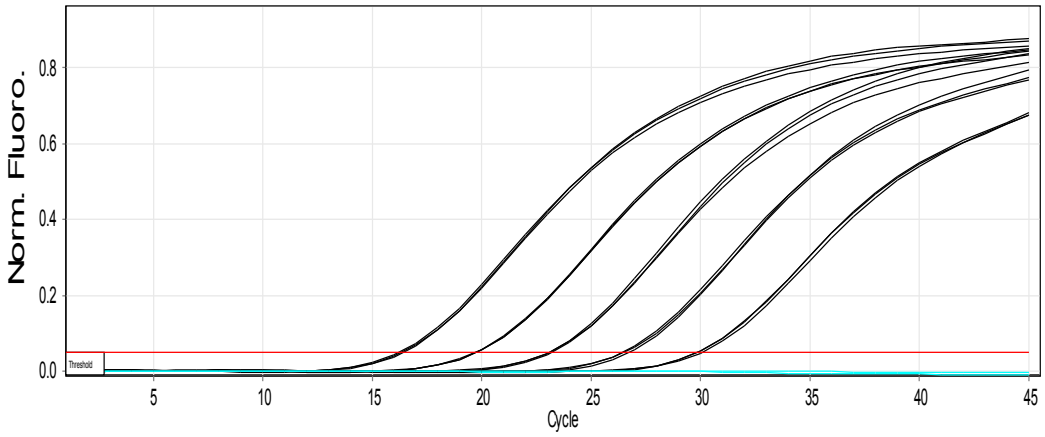
5.2.1.1 PCR efficiency and linearity of the singleplex RT-qPCR assays

The linearity and PCR efficiency of the individual assays was determined using the *in-vitro* transcript dilutions (3 replicates per dilution) ranging from 10^2 to 10^6 RNA copies/µl containing the target sequences for VCAM1, ISG15, MS4A4A and TUBG1 gene transcripts (Figure 5.2). The PCR efficiency under singleplex conditions for the four gene transcripts VCAM1, ISG15, MS4A4A and TUBG1 was greater than 90%. The slopes were also within the acceptable range of -3.2 to -3.5. All singleplex assays showed a good correlation coefficient ($R^2 > 0.990$). These results indicate a good confidence correlating the Ct value and the logarithm of the standard RNA copy number.

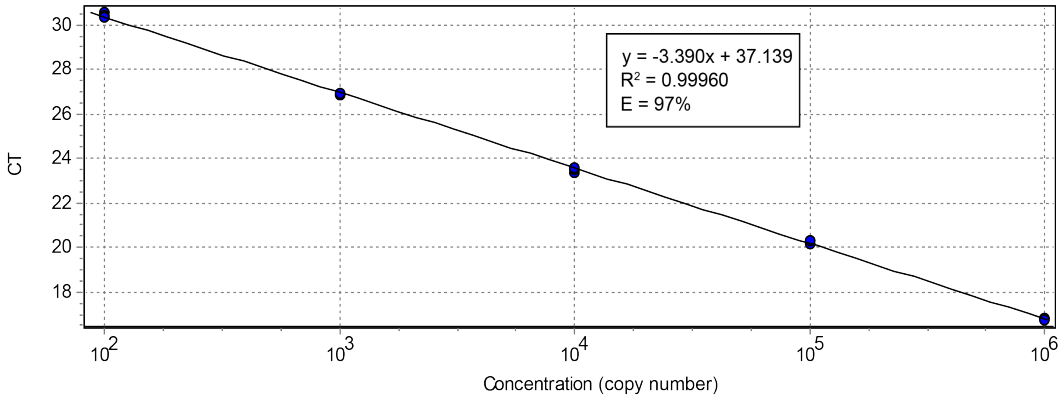
The amplification charts of each gene transcript are also shown in Figure 5.2. Tight replicate curves are observed for each of the five dilutions that creates the standard curve in the RT-qPCRs. The negative controls (no RNA template added) for VCAM1, ISG15, and MS4A4A did not amplify except TUBG1. In the TUBG1 assay, the negative control amplified above the threshold. Although, the assay was re-tested three times, and new reagents were prepared, the amplification in the negative control was still observed, suggesting the formation of primer or probe dimers that was not observed before in the SYBR green assay using the same primers. Despite TUBG1 assay had high PCR efficiency and good correlation coefficient, there was a probability that dimers may give an overestimation of the target. Thus, TUBG1 was not included in the group of gene transcripts for multiplexing.

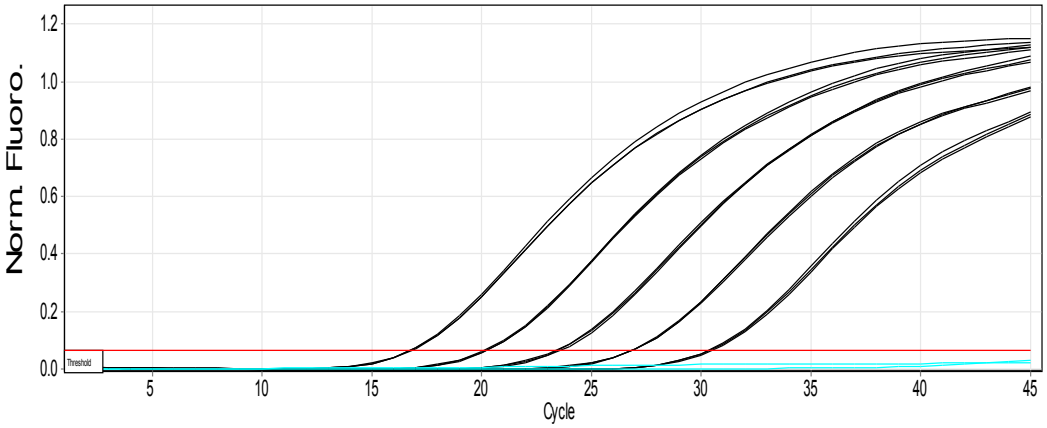
VCAM1


y = -3.400x + 36.766
R² = 0.99966
E = 97%
CT
Concentration (copy number)

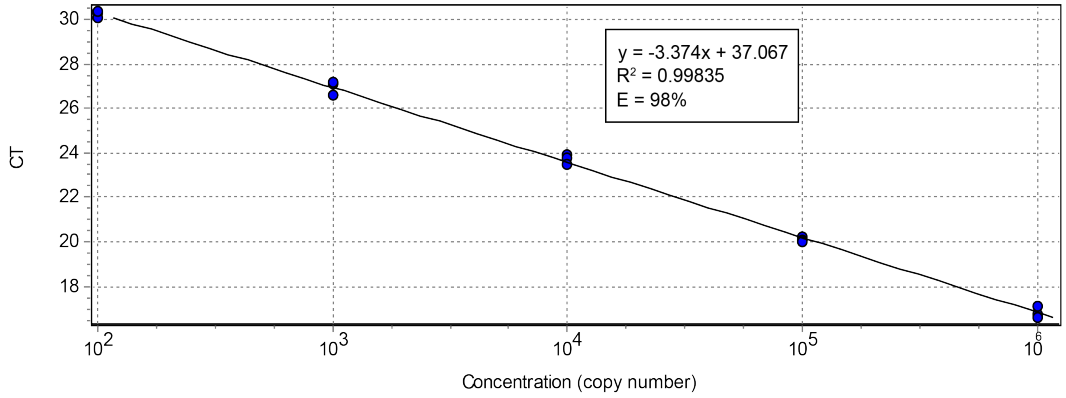

Norm. Fluoro.
Threshold
Cycle

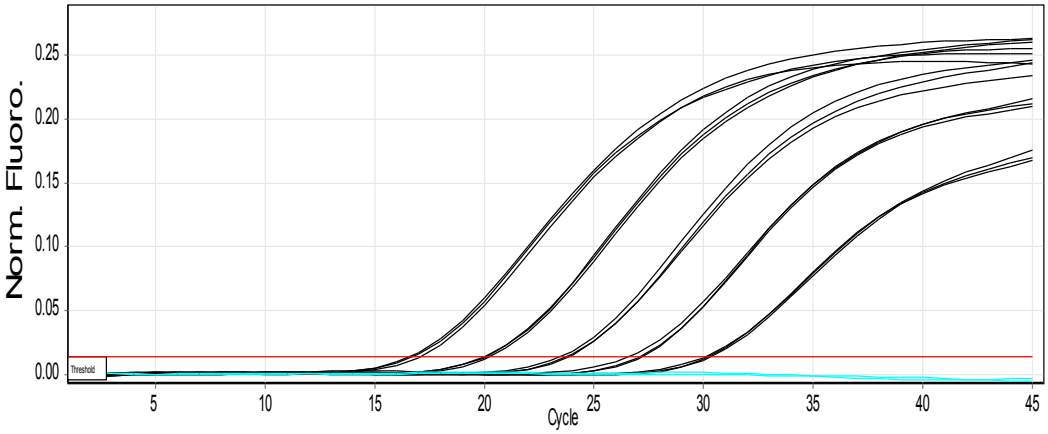
ISG15


y = -3.390x + 37.139
R² = 0.99960
E = 97%
CT
Concentration (copy number)


Norm. Fluoro.
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Threshold
5
10
15
20
25
30
35
40
45
Cycle

MS4A4A


CT
30
28
26
24
22
20
18
y = -3.374x + 37.067
R² = 0.99835
E = 98%
10^2
10^3
10^4
10^5
10^6
Concentration (copy number)


Norm. Fluoro.
0.25
0.20
0.15
0.10
0.05
0.00
Threshold
5
10
15
20
25
30
35
40
45
Cycle

TUBG1

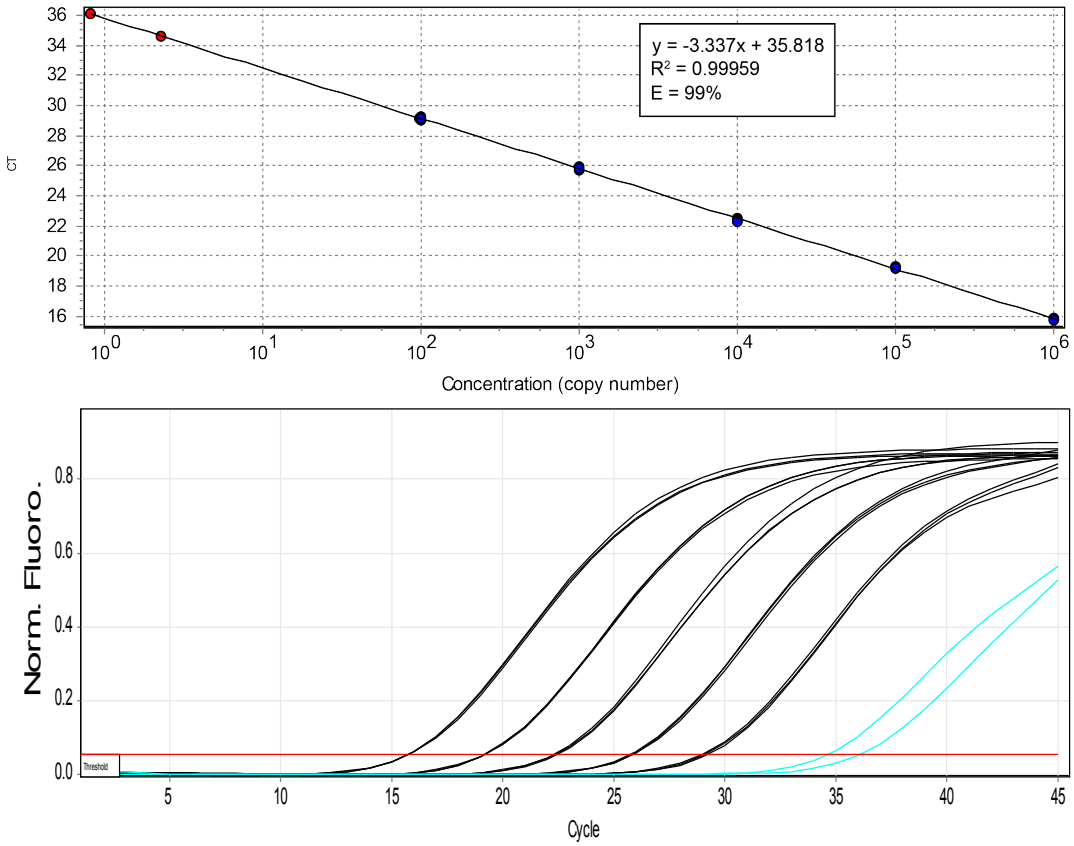


Figure 5.2 Linearity and PCR efficiency of the singleplex assays for each of the four selected gene transcripts. *The upper panels show the standard curves obtained created with the in-vitro RNA dilutions ranging from 10^2 to 10^6 RNA copies/µl containing the target sequences for VCAM1, ISG15, MS4A4A, and TUBG1 transcripts. The mean Ct value of the triplicates is plotted against the log of concentration of the target gene copies/reaction. The PCR efficiency for each target was calculated using the slope of the calibration curve with the formula $E = 10^{-1/slope}-1$. E = PCR efficiency, R^2 = correlation coefficient. Red dots in the TUBG1 standard curve are the detected negative controls. The lower panels show the amplification curves of the in-vitro RNA dilutions (black) tested in triplicate. The light blue lines represent the negative control (no template added) tested in duplicate.*

5.2.2 Multiplexing

Once the singleplex assays were optimized and showed a high efficiency, the primers, and probes for VCAM1, ISG15 and MS4A4A were combined in a multiplex assay. The primer and probe concentrations for each gene transcript used in the singleplex assays were the same used in the multiplex assay and there was no need to change other reaction components.

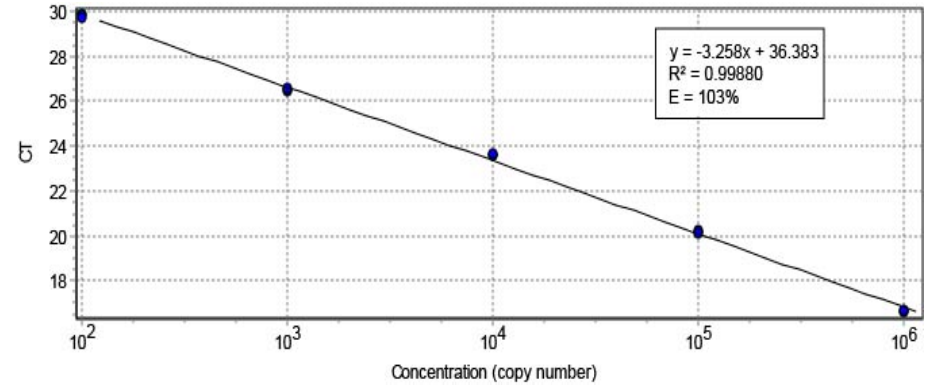
5.2.2.1 PCR efficiency and linearity of the optimized multiplex RT-qPCR assay

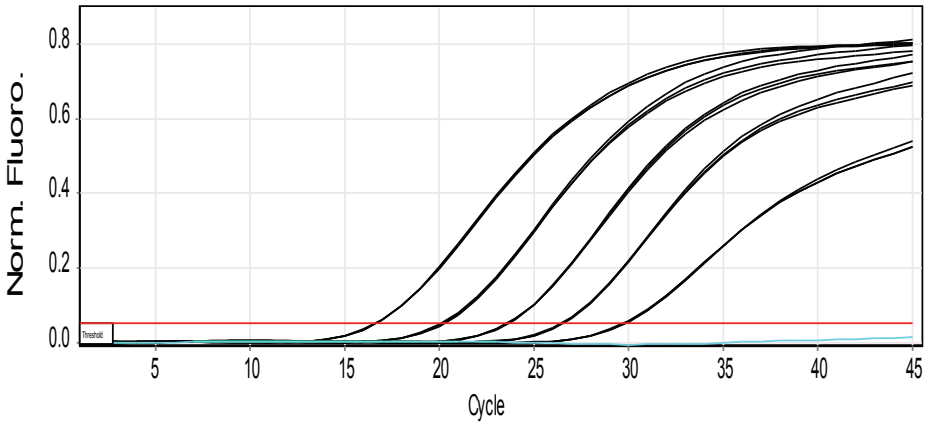
To evaluate the performance of the optimized multiplex RT-qPCR assay, the PCR efficiency and linearity was first checked. The PCR efficiency was estimated from the slope of the standard curve of each gene transcript in the multiplex assay and was within the acceptable values (90%-110%) for the three gene transcripts (103% for VCAM1, 106% for MS4A4A and 107% for ISG15) (Figure 5.3). The standard curves of each gene transcript also showed a good correlation coefficient greater than 0.99 indicating a good linearity. The R^2 for VCAM1 was 0.998, for MS4A4A was 0.997, and for ISG15 was 0.997. These results suggests that the multiplex RT-qPCR is a robust test that precisely correlate the Ct value and the logarithm of the standard RNA copy number.

5.2.2.2 Precision evaluation of the optimized multiplex RT-qPCR assay

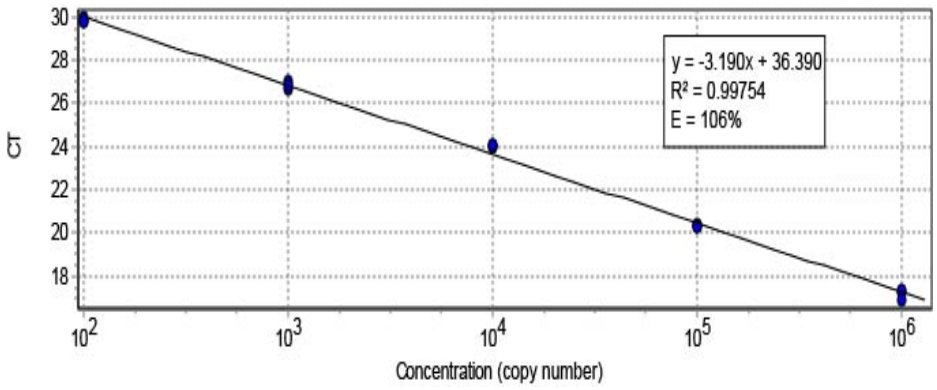
To evaluate the precision of the multiplex RT-qPCR, the intra and inter-assay variability was determined. For this, the multiplex RT-qPCR assay was run three times in order to measure the coefficient of variation (CV) within a single run and between the three runs. For the three gene transcripts, the CV of the intra- and inter-assays was less than 2%. The intra-assay CV for the standard dilutions were between 0.18 to 0.34% for VCAM1, 0.26 to 1.53% for MS4A4A, and 0.26 to 0.67% for ISG15 (Table 5.2). The inter-assay CV for the standard dilutions were 0.12 to 1.14% for VCAM1, 0.21 to 1.11% for MS4A4A, and 0.27 to 1.07% for ISG15 (Table 5.3). These results indicates that the multiplex assay has a high repeatability within a run and between runs.

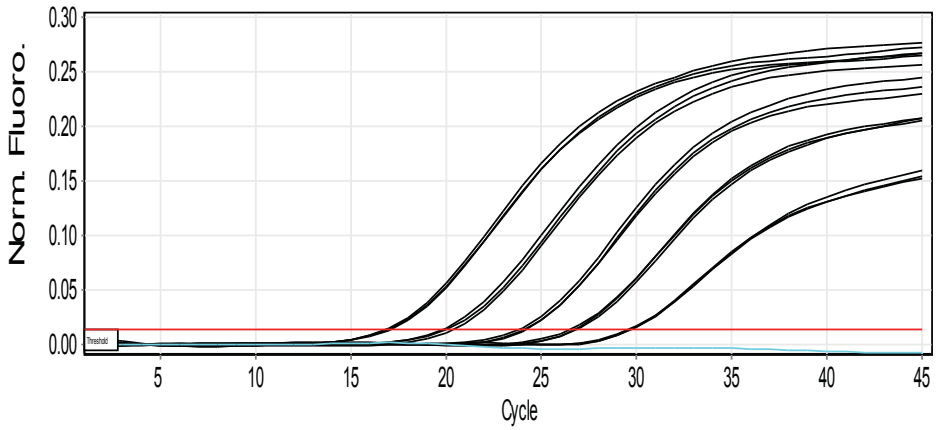
VCAM1





MS4A4A





ISG15

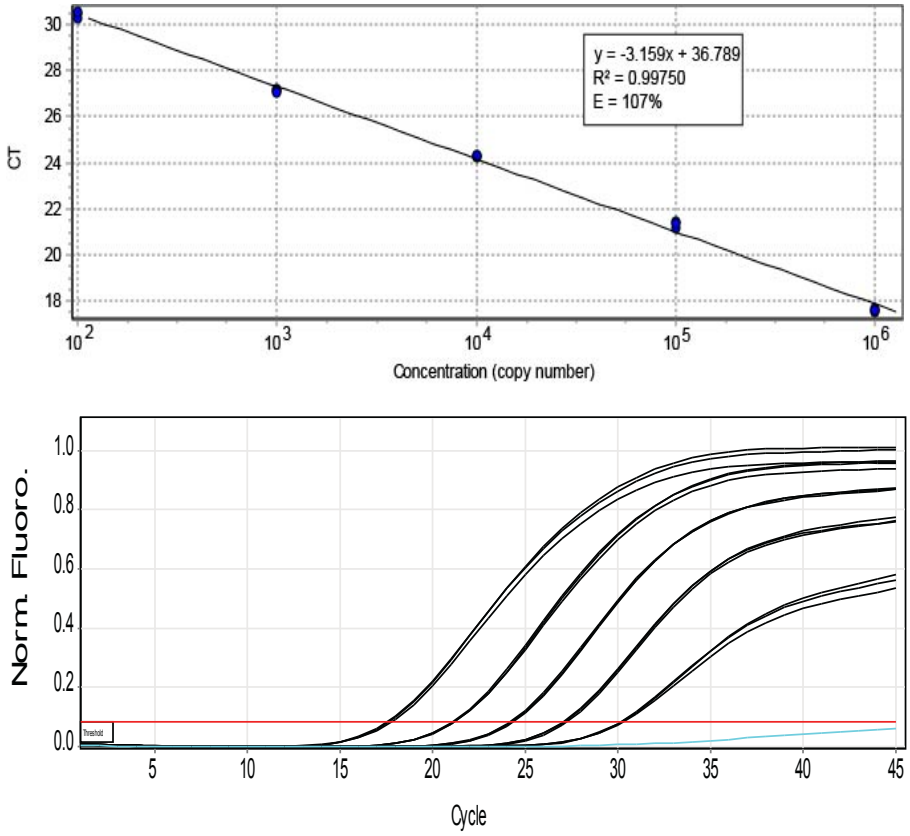


Figure 5.3 Linearity and PCR efficiency of the multiplex RT-qPCR assay for the detection of VCAM1, MS4A4A and ISG15 transcripts. *The upper panels show the standard curves obtained in the multiplex assay. The standard curves were created using in-vitro RNA dilutions ranging from 10^2 to 10^6 RNA copies/µl containing the target sequences for VCAM1, MS4A4A, and ISG15 transcripts and were included in every run. The mean Ct values are plotted against the log of concentration of the target gene copies/reaction. The PCR efficiency for each target was calculated using the slope of the calibration curve with the formula $E = 10^{-1/slope}-1$. E = PCR efficiency, R^2 = correlation coefficient. The lower panels show the amplification curves of the in-vitro RNA dilutions (black) tested in triplicate.*

Table 5.2 Intra-assay variability of the multiplex RT-qPCR assay

Multiplex assay		Copy number of cRNA standards				
		1.00E+06	**1.00E+05**	**1.00E+04**	**1.00E+03**	**1.00E+02**
VCAM1	CT values (mean ± SD)	16.65 ± 0.05	19.88 ± 0.04	23.41 ± 0.07	26.42 ± 0.09	29.78 ± 0.10
	CV	0.30%	0.18%	0.310%	0.34%	0.32%
MS4A4A	CT values (mean ± SD)	16.95 ± 0.10	20.06 ± 0.31	24.04 ± 0.21	26.70 ± 0.18	29.72 ± 0.08
	CV	0.61%	1.53%	0.87%	0.69%	0.26%
ISG15	CT values (mean ± SD)	17.68 ± 0.12	21.17 ± 0.06	24.28 ± 0.08	27.15 ± 0.12	30.21 ± 0.08
	CV	0.67%	0.29%	0.33%	0.43%	0.26%

SD = Standard deviation; CV = Coefficient of variation

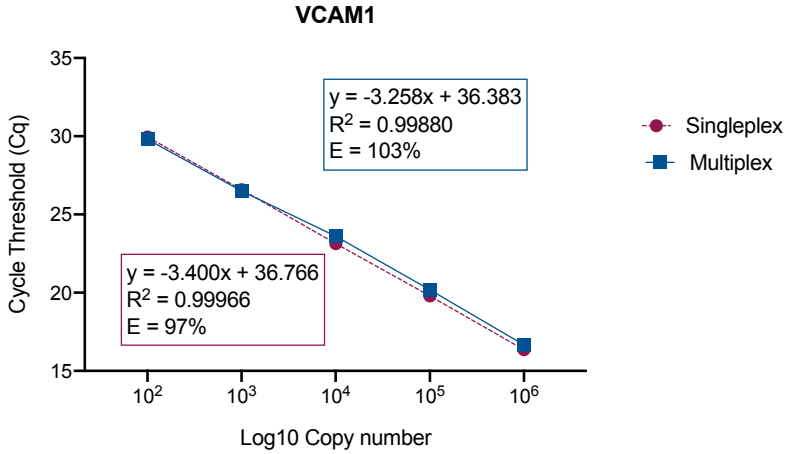
Table 5.3 Inter-assay variability of the multiplex RT-qPCR assay

Multiplex assay		Copy number of cRNA standards				
		1.00E+06	**1.00E+05**	**1.00E+04**	**1.00E+03**	**1.00E+02**
VCAM1	CT values (mean ± SD)	16.67 ± 0.02	20.13 ± 0.23	23.57 ± 0.15	26.52 ± 0.11	29.89 ± 0.17
	CV	0.12%	1.14%	0.63%	0.42%	0.56%
MS4A4A	CT values (mean ± SD)	17.09 ± 0.13	20.30 ± 0.23	24.07 ± 0.09	26.76 ± 0.06	29.81 ± 0.08
	CV	0.73%	1.11%	0.38%	0.21%	0.26%
ISG15	CT values (mean ± SD)	17.68 ± 0.10	21.37 ± 0.23	24.36 ± 0.15	27.19 ± 0.07	30.44 ± 0.22
	CV	0.54%	1.07%	0.63%	0.27%	0.71%

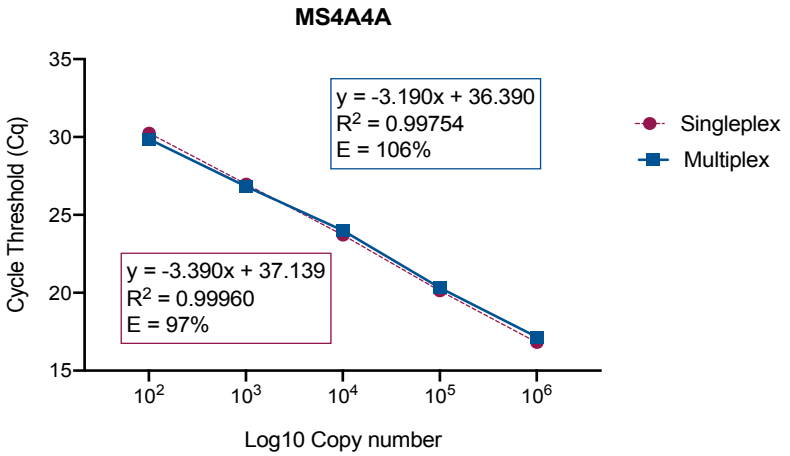
SD = Standard deviation; CV = Coefficient of variation

5.2.2.3 Comparison of the singleplex and multiplex assay for each gene transcript

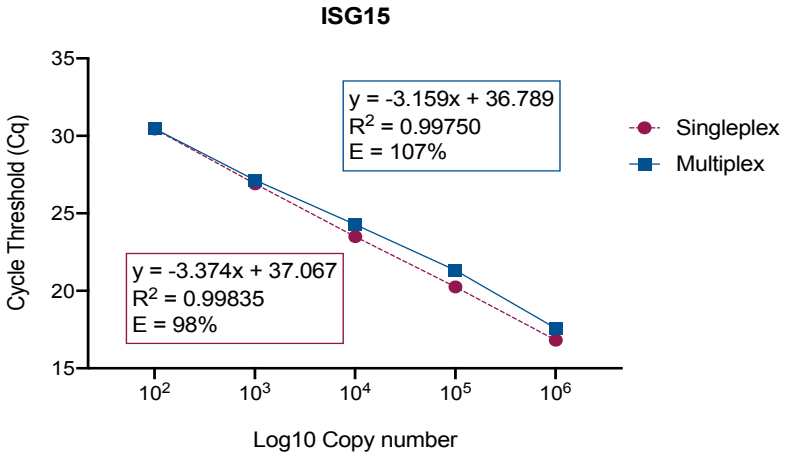
The Ct values obtained in the standard curves for each gene in the singleplex and multiplex assay were compared (Figure 5.4). There was a slight or no variation between the Ct values obtained for each gene when tested in a singleplex assay and when tested in the multiplex assay. The standard curves in VCAM1 and MS4A4A completely overlap, while the standard curve in ISG15 slightly varies in the dilutions that have high copy number (10^4 to 10^6) by less than or 1 cycle which was still acceptable. The PCR efficiency was higher in the multiplex assay compared to the singleplex but within the acceptable range from 90% to 110%. Likewise, the correlation coefficient was high for both assays ($R^2 > 0.99$). These results indicate consistency of the data between the singleplex and multiplex assays.



	Ct values				
	10^2	10^3	10^4	10^5	10^6
Singleplex	29.96 ± 0.12	26.59 ± 0.16	23.13 ± 0.07	19.80 ± 0.03	16.36 ± 0.09
Multiplex	29.80 ± 0.06	26.50 ± 0.06	23.61 ± 0.04	20.18 ± 0.06	16.67 ± 0.03



	Ct values				
	10^2	10^3	10^4	10^5	10^6
Singleplex	30.25 ± 0.14	26.97 ± 0.31	23.71 ± 0.21	20.12 ± 0.12	16.81 ± 0.25
Multiplex	29.86 ± 0.11	26.82 ± 0.18	24.00 ± 0.07	20.32 ± 0.03	17.15 ± 0.23



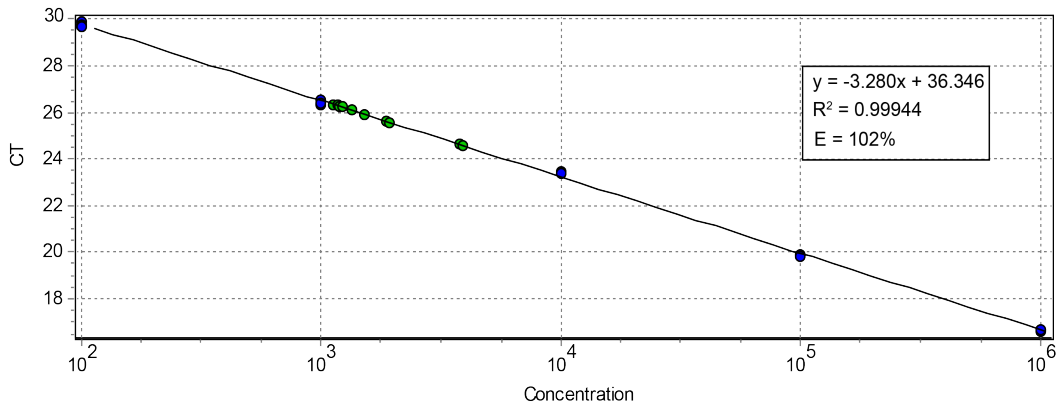
	Ct values				
	10^2	10^3	10^4	10^5	10^6
Singleplex	30.44 ± 0.11	26.90 ± 0.03	23.50 ± 0.11	20.26 ± 0.08	16.81 ± 0.04
Multiplex	30.46 ± 0.13	27.14 ± 0.09	24.27 ± 0.04	21.32 ± 0.13	17.58 ± 0.07

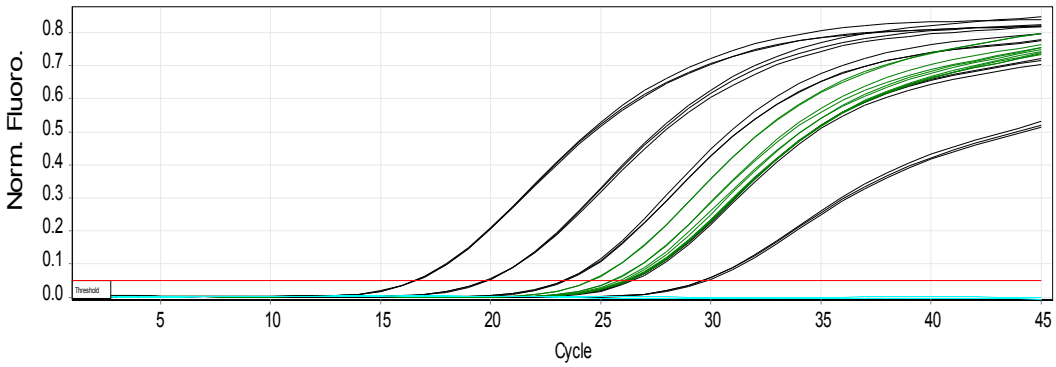
Figure 5.4 Comparison of the singleplex and multiplex RT-qPCR assay for the detection of VCAM1, MS4A4A and ISG15 transcripts. *The plots show the standard curves obtained from the singleplex (light red) and multiplex assays (light blue) for each gene transcript. The standard curves represent five dilutions of the in-vitro RNA standards (10^2 to 10^6 RNA copies/µl) containing the target sequence of VCAM1, MS4A4A and ISG15 transcripts. The plots only show the mean Ct value obtained from three replicates of each dilution, which is also listed in the table below.*

5.2.2.4 Application of the multiplex RT-qPCR assay in control samples

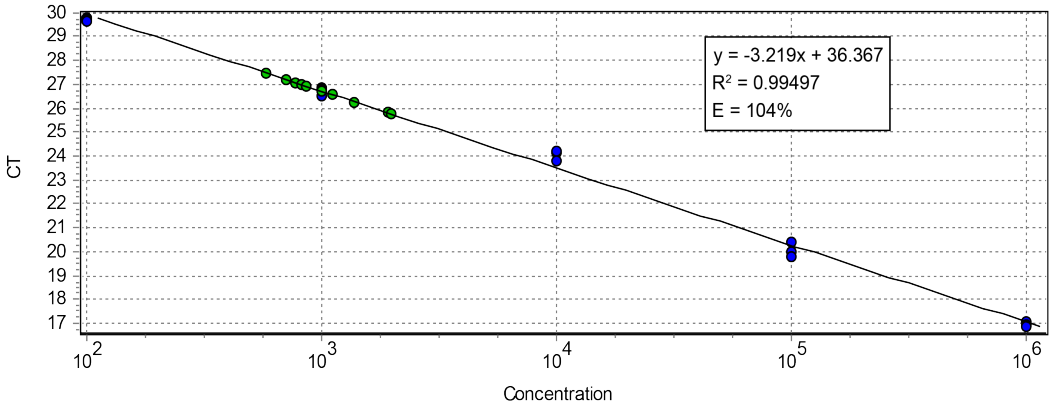
Control samples (corresponding to 5 healthy donors) were used to evaluate the performance of the multiplex RT-qPCR assay. The samples were tested in duplicate and the RNA copy number for each gene was obtained from the extrapolation of the Ct into the standard curve. For all the genes, the standard curves had PCR efficiencies within the acceptable range (90%-110%) (Figure 5.5). For VCAM1 was 102%, for MS4A4A was 104% and for ISG15 was 110%. The standard curves of the three gene transcripts also showed a good correlation coefficient ($R^2 > 0.99$) indicating a good linearity. The amplification curves of the replicates of the *in-vitro* transcript dilutions and control samples were tight, indicating a minimum variation in the Ct values. For VCAM1, the Ct values for the donors ranged from 24.6 to 26.29 and the transcript copy number ranged from 1179 to 3812 copies/µl. For MS4A4A, the Ct values ranged from 25.77 to 27.27 and the transcript copy number from 670 to 1955 copies/µl. For ISG15, the Ct values ranged from 23.66 to 25.21 and the transcript copy number from 4375 to 13883 copies/µl.

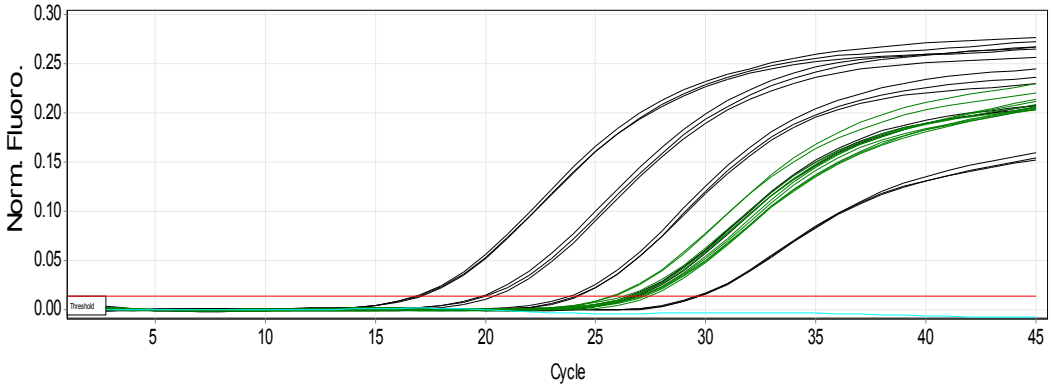
VCAM1



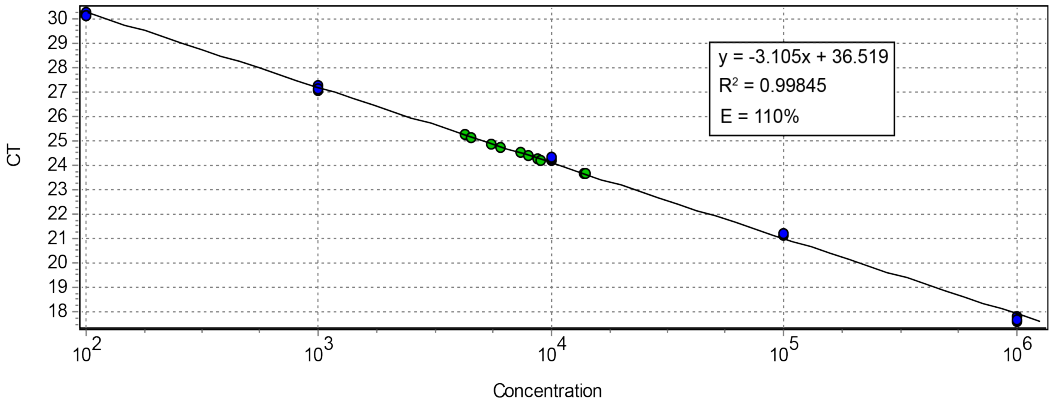


MS4A4A





ISG15



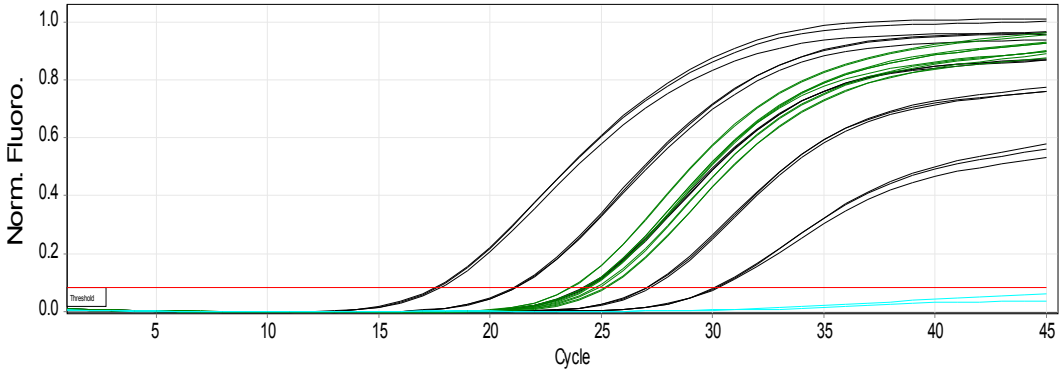


Figure 5.5 Application of the multiplex RT-qPCR for the simultaneous detection of VCAM1, MS4A4A and ISG15 transcripts in five control samples. *The upper panels show the standard curves that were created using in-vitro RNA dilutions ranging from 10^2 to 10^6 RNA copies/µl containing the target sequences for VCAM1, MS4A4A, and ISG15 transcripts and were included in every run. The green dots represent the control samples. The PCR efficiency for each target was calculated using the slope of the calibration curve with the formula $E = 10^{-1/slope}-1$. E = PCR efficiency, R^2 = correlation coefficient. The lower panels show the amplification curves representing the 5 serial dilutions (in-vitro RNA standards) tested in triplicate. green amplification curves represent the control samples tested in duplicate. The light blue lines represent the negative control (no template added) tested in duplicate.*

5.2.3 Application of the multiplex RT-qPCR in EVD samples

The multiplex RT-qPCR was then applied in the leftover of clinical samples from the "blinded" cohort of patients used in the analysis of Chapter IV. Forty-seven samples from the 64 clinical samples of the "blinded" cohort of patients had RNA leftover and were used for this analysis. These samples corresponded to 21 survivors and 26 fatal cases. The transcript abundance of VCAM1, MS4A4A and ISG15 measured by multiplex RT-qPCR in these samples is shown in Appendix Table 8. The difference between survivors and fatal cases was analysed by Mann-Whitney *U.* No significant difference in the transcript abundance of VCAM1 ($p = 0.2460$), MS4A4A ($p = 0.4904$) and ISG15 ($p = 0.5597$) was found between the survivors and fatal cases. Figure 5.6 compares the transcript abundance obtained from the multiplex and the SYBR green assays in the same group of 47 patients. Although the distribution of the transcript abundance of each gene looks similar between the multiplex and the SYBR green assay, the absolute RNA copy number differ between both assays.

For VCAM1, the statistical analysis shows a no significant difference between survivors and fatal cases detected by multiplex assay (p = 0.2460) and SYBR green assay (p = 0.3443). The transcript copy number for VCAM1 detected by the multiplex assay was higher in survivors (22539 RNA copies/µl) than fatal cases (19541 RNA copies/µl), while by SYBR green assay was the opposite, the fatal cases have higher transcript copy number (139961 RNA copies/µl) than survivors (126043 RNA copies/µl). For ISG15, a non-significant difference was found between survivors and fatal cases in both, by multiplex (p =0.5597) and SYBR green assay (p =0.3230). Both assays showed a higher transcript copy number in fatal cases compared to survivors, 93233 RNA copies/µl in fatal cases and 56577 RNA copies/µl in survivors by the multiplex assay, while the SYBR green assay detected 51872 RNA copies/µl in fatal cases and 45054 RNA copies/µl in survivors. For MS4A4A, the statistical analysis of the transcript abundance in these samples differ between the tests. The analysis for the multiplex assay found no difference between survivors and fatal cases (p =0.4904) while for SYBR green assay there was a significant difference (p = 0.0017). The mean transcript copy number obtained by the multiplex assay seems similar between survivors (9034 RNA copies/µl) and fatal cases (8263 RNA copies/µl). In contrast, the mean transcript copy number obtained by SYBR green RT-qPCR was higher in fatal cases (7222 RNA copies/µl) than survivors (5551 RNA copies/µl). Although a comparison between both assays may not be adequate since each assay has a different chemistry system and were run in different qPCR machines, a similar transcript copy number in the same individuals was expected since 10ng of RNA from each individual was the input sample in both assays. A possible explanation for this might be that remaining genomic DNA is present in the samples and are being co-amplified with the cDNA. Despite this possibility machine learning models were built for assessing the prediction performance of VCAM1, ISG15 and MS4A4A transcripts using the multiplex RT-qPCR data.

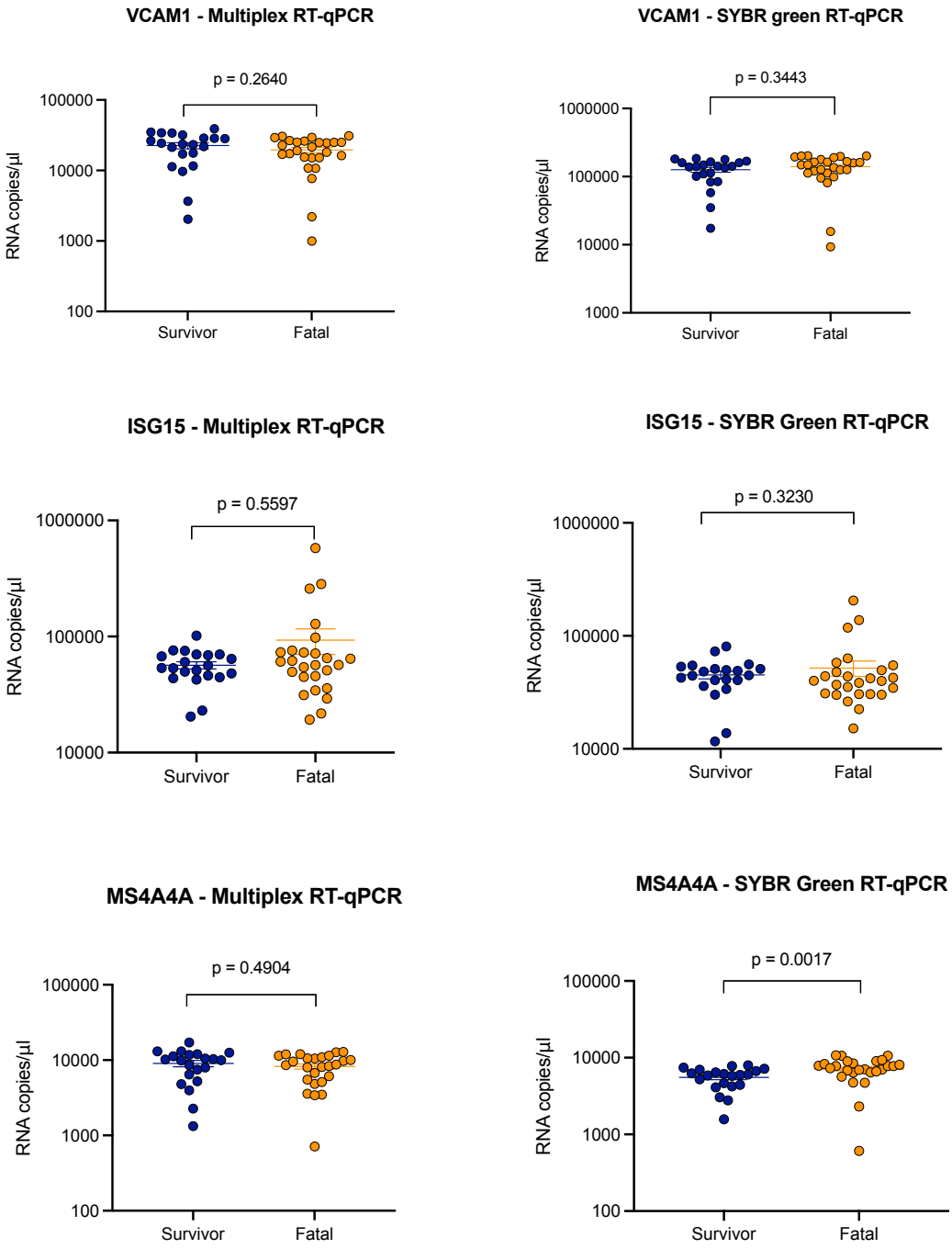


Figure 5.6 Comparison of the singleplex SYBR green-based RT-qPCR and the multiplex probe-based RT-qPCR assay. *The plots show the RNA copy number of VCAM1, ISG15 and MS4A4A measured by SYBR green RT-qPCR and multiplex RT-qPCR from 47 clinical samples (21 survivors and 26 fatal cases). The clinical samples were tested in duplicate in both assays and the mean RNA copy number per individual was obtained by plotting the mean Ct value against the standard curve created with in-vitro RNA dilutions of known concentration. Each dot represents an individual (blue for survivor, orange for fatal). The bars indicate the mean and the standard error of the mean (SEM). The differences in the transcript abundance between survivors and fatal cases were evaluated via the non-parametric Mann-Whitney U test.*

5.2.4 Classification models for the multiplex RT-qPCR data

Support vector machine models were trained to evaluate the prediction performance of VCAM1, ISG15 and MS4A4A transcripts using the multiplex results from the 47 EVD patients (21 survivors and 26 fatal cases). The same workflow followed in Chapter IV for training models was used in this chapter. The dataset was split in 70% for training set (34 individuals corresponding to 15 survivors and 19 fatal cases) and 30% for testing set (13 individuals corresponding to 6 survivors and 7 fatal cases). Different SVM models were trained and tuned in the training dataset and then used to evaluate its performance in the testing set. These models consisted in different combinations of the genes as predictors of the clinical outcome.

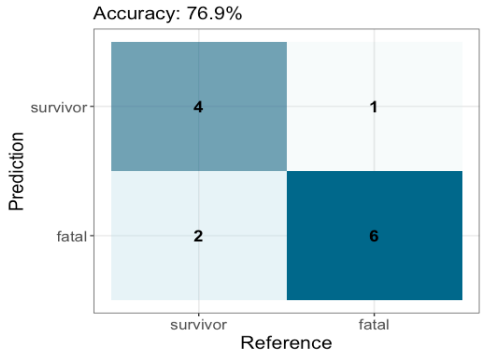
First, a model for each predictor variable was built in order to determine its individual performance. The SVM model that included only EBOV Ct as predictor was 76.9% accurate (95% CI, 0.4619 - 0.9496) classifying the EVD patients from the testing set (13 individuals) (Figure 5.7). This model correctly classified 85.7% (6/7) of fatal cases and 66.7% (4/6) of survivors. From the three selected gene transcripts, the SVM model for VCAM1 had the lowest accuracy performance 53.8% (95% CI, 0.2513 - 0.8078). The model correctly classified 85.7% (6/7) of fatal cases but only 16.7% (1/6) of survivors. The overall accuracy of the SVM model for either MS4A4A or ISG15 was 61.5% (95% CI, 0.3158 - 0.8614); however, the two models differed in the classification of fatal cases and survivors. The model that used MS4A4A as a single variable was the best predicting fatality compared to the other genes and EBOV Ct since correctly classified all fatal cases 100% (7/7), although only 16.7% (1/6) of survivors were correctly classified. In contrast, the SVM model for ISG15 was the best predicting survival 100% (6/6) but had a poor performance classifying fatal cases 28.6% (2/7).

To evaluate whether the prediction accuracy improves with the combination of different variables or not, several SVM models were trained. The model combining the three genes (VCAM1, ISG15 and MS4A4A) correctly classified 61.5% (95% CI,

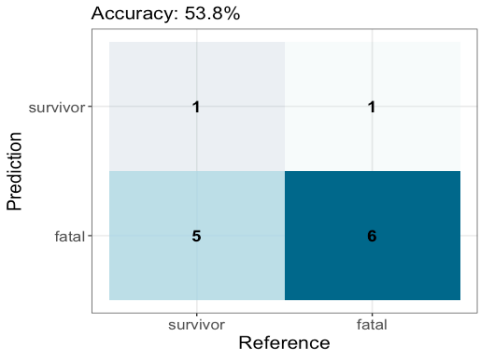
0.3158 - 0.8614) of patients from the testing set. This model accurately predicted all fatal cases 100% (7/7), but only 16.7% (1/6) of survivors. Then, different models combining EBOV Ct with each gene were built. The model combining EBOV Ct values and MS4A4A correctly classified 69.2% (95% CI, 0.3857 - 0.9091) of patients from the testing set. This model correctly classified almost all fatal cases 85.7% (6/7), and half of the survivors 50% (3/6). The models combining EBOV Ct and ISG15 had a higher overall accuracy 76.9% (95% CI, 0.4619 -0.9496) and correctly classified 85.7% (6/7) of fatal cases and 66.7% (4/6) of survivors. The model combining EBOV Ct and VCAM1 model had an accuracy of 84.6% (95% CI, 0.5455 - 0.9808) and was able to predict all the fatal cases 100% (7/7) and 66.7% (4/6) of survivors.

From all the possible combinations, a model combining EBOV Ct values and the three gene transcripts, or a model combining EBOV Ct values and only VCAM1 and MS4A4A showed a high accuracy 84.6% (95% CI, 0.5455 - 0.9808). These two models only misclassified one fatal case indicating 85.7% of sensitivity (6/7) and one survivor indicating 83.3% (5/6) of specificity. These findings indicates that the prediction accuracy of the clinical outcome increases when a model includes our candidate predictive gene transcripts rather than using only EBOV Ct values as a predictor variable.

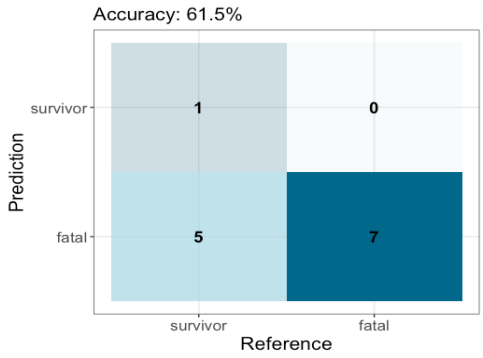
A) EBOV Ct



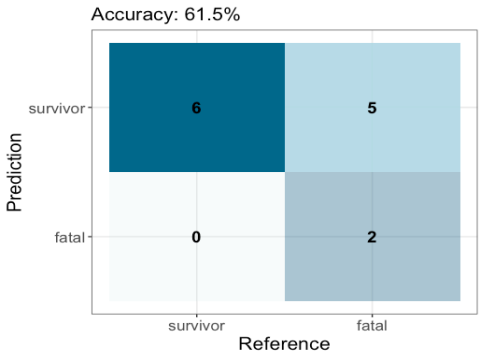
B) VCAM1



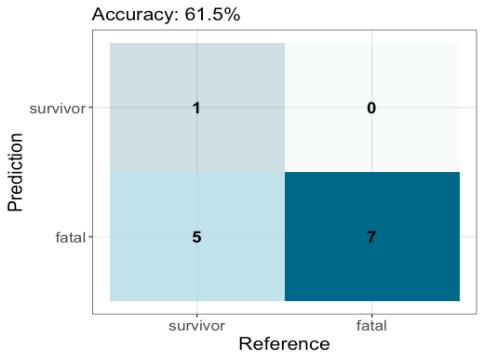
C) MS4A4A



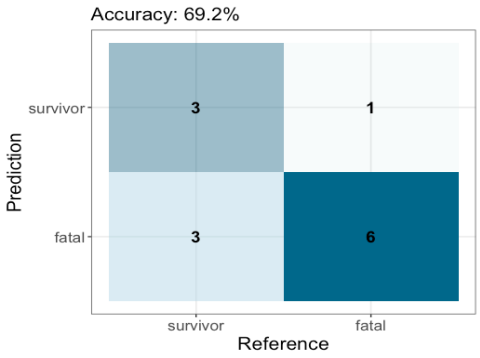
D) ISG15



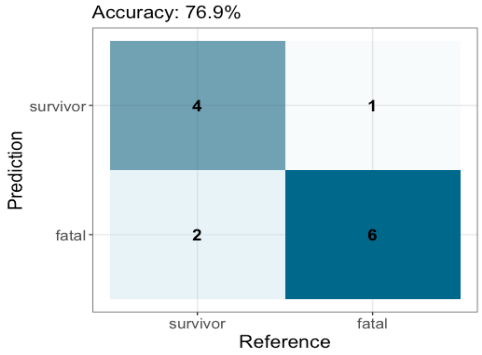
E) VCAM1 + MS4A4A + ISG15



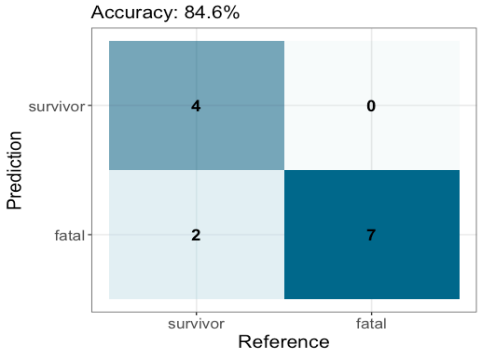
F) EBOV Ct + MS4A4A



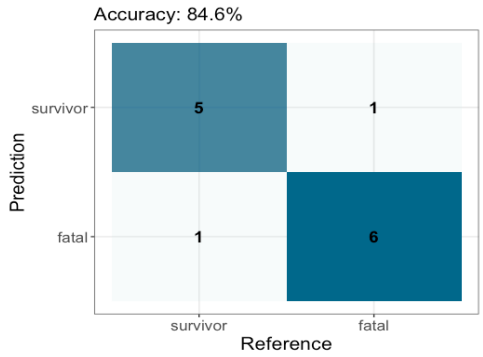
G) EBOV Ct + ISG15



H) EBOV Ct + VCAM1



I) EBOV Ct + VCAM1 + MS4A4A + ISG15



J) EBOV Ct + VCAM1 + MS4A4A

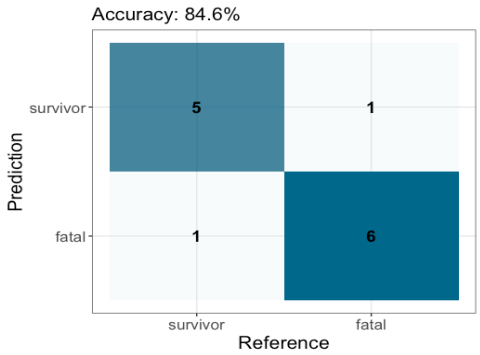


Figure 5.7 Confusion matrix showing SVM models performance in the testing set. *The confusion matrix reports the predicted and actual outcome of the testing set (13 individuals consisting of 6 survivors and 7 fatal cases). The intensity of the background colour is according to the frequency of patients. The accuracy metric it is defined as the fraction of patients that are correctly classified.*

5.3 Discussion

The purpose of this chapter was to develop a multiplex RT-qPCR that could detect and quantify simultaneously the transcript abundance of the selected gene transcripts that were found in Chapter IV with high prognostic value.

The optimized test performed efficiently without any cross-amplification. The multiplex RT-qPCR efficiency was 103% for VCAM1, 106% for MS4A4A and 107% for ISG15 whereas the singleplex efficiency was 97% for VCAM1, 97% for MS4A4A and 98% for ISG15. The correlation coefficient was greater than 0.99 for all the gene transcripts in the singleplex and multiplex assay.

The absolute transcript copy number obtained from the 47 clinical patients by multiplex RT-qPCR are inconsistent with the transcript abundance measured by the SYBR green assay. This discrepancy suggest that these results may not represent the exact absolute transcript copy number of VCAM1, MS4A4A and ISG15 since the primers for these genes are located within the same exon and if the clinical samples were not completely free of genomic DNA this could have served as a template and resulted in a co-amplification with the cDNA in both: SYBR green and multiplex assay. Unfortunately, DNase treatment is not included in the RNA extraction protocol followed by the EMlab using the QIAamp Viral RNA Mini Kit (Qiagen) and the RNA was not DNase treated after extraction to avoid losing RNA with an extra purification.

Despite this possibility, the SVM models trained with the multiplex RT-qPCR data support our hypothesis that the prognostic value improves when EBOV Ct is used along with our gene transcripts such as VCAM1, MS4A4A and ISG15. The SVM models showed that EBOV Ct alone is a strong predictor of fatality, but it does not correctly discriminate all survivors. VCAM1 and MS4A4A transcripts are also strong predictors of fatality when are used independently but did not correctly classify all survivors. Between the three gene transcripts, ISG15 is the strongest predictor of survival but a poor predictor of fatality.

The best classification model in this chapter was the combination of EBOV Ct and VCAM1, MS4A4A and ISG15 transcripts, which accurately classified 84.6% of the testing set. This model only misclassified one survivor and one fatal case. This is an encouraging result that may lead to an external validation on a larger cohort of clinical samples.

A multiplex RT-qPCR that includes both our candidate predictive gene transcripts and EBOV detection would be ideal. However, it was impossible to include the primers and probes of the two commercial kits (the RealStar Filovirus Screen and Zaire Ebolavirus RT-PCR kits) used for the diagnosis of EBOV during the West African EVD outbreak by the EMLab since the assay composition of these kits is confidential intellectual property of Altona Diagnostics (Hamburg, Germany). In further research, the generation of *in-vitro* transcripts of the target sequences of EBOV may be considered to have standards to include in the multiplex RT-qPCR assay for the simultaneous quantification of EBOV and these gene transcripts following the absolute quantification method applied in this test.

Chapter 6

General discussion

6.1 Use of predictive markers that determine the clinical outcome

EVD is one of the deadliest viral infections of humans, with case fatality rates recorded between 50 to 70%. Despite the ongoing efforts to control disease transmission using the recently licensed rVSV-ZEVOV-GP vaccine, a substantial gap remains in improving the outcome of acute EVD patients and survivors. There are still no licensed treatments and no prognostic tests for EVD. A prognostic test that predicts EVD severity could guide the clinical triage of patients and contribute to the improvement of a targeted therapy and the clinical outcome. During a pandemic, prognostic markers could relieve the clinical burden, support the decision making, logistical planning, and potentially reduce the mortality rate in limited-resource settings (Yan et al., 2020). In the absence of a prognostic test for EVD, the present study aimed to evaluate and validate host markers that were suggested as potential predictors of EVD outcome in a previous study from our group (Liu et al., 2017). The transcript abundance of a set of 11 genes (VCAM1, HOPX, TUBG1, PLPP3, MS4A4A, TGFBI, TTC28, NIF3L1, SLC25A5, CTSL and ISG15) was measured by RT-qPCR assays in cohorts of EVD patient samples from the West Africa outbreak in 2013-2016 and machine learning methods were used to build predictive models that could classify patients by their clinical outcome.

6.2 Prediction accuracy of the candidate predictive genes versus EBOV Ct value

RT-qPCR assays using SYBR green chemistry were first developed and optimized for quantifying the transcript abundance of the set of 11 candidate predictive genes. These RT-qPCRs were then used to measure the transcript abundance of the 11 genes in a cohort of 39 acute EVD patients (20 survivors and 19 fatal cases) with known outcome collected from the outbreak in Guinea in 2014-2015. Since it was not possible to measure all gene transcripts from the 39 individuals, only the data from four gene transcripts (VCAM1, ISG15, MS4A4A and TUBG1) was used to build SVM models to predict the clinical outcome of an independent set of samples consisting of 64 EVD patients (blinded samples). Although our attempt to predict correctly the outcome using a small cohort of blinded samples was not completely successful, our

results showed that models that included the predictive gene transcripts such as ISG15, MS4A4A and TUBG1 with or without EBOV Ct values (68.4% and 67%) were more accurate than a model that only used EBOV Ct values as a predictor variable (54.3%). These findings suggest a better prediction accuracy when models are built with a combination of our predictive markers and EBOV Ct values.

The low accuracy of these SVM models could be explained by the use of a small sample size to train the models (18 patients consisting of 7 survivors and 11 fatal cases) that resulted from the removal of missing values of the initial cohort of 39 patients and partitioning of the dataset in training and testing set. Therefore, to enhance the prediction performance of the machine learning models, a larger sample size was evaluated (a whole dataset of 81 EVD patients consisting of 36 survivors and 45 fatal cases). The analysis of this data showed that the transcript abundance of VCAM1 and MS4A4A was significantly higher in fatal cases than survivors and the transcript abundance of TUBG1 was significantly higher in survivors than fatal cases. There was no significant difference in the transcript abundance of ISG15 between survivors and fatal cases. SVM models were built with data from 58 individuals from this analysis (consisting of 26 survivors and 32 fatal cases) and the model's performance was evaluated by predicting the clinical outcome of 23 individuals (consisting of 10 survivors and 13 fatal cases). The performance evaluation of the SVM models built with a single variable showed that the EBOV Ct model had the highest accuracy (87%) followed by MS4A4A (82.6%), VCAM1 (70%), TUBG1 (74%) and ISG15 model (65.2%). EBOV Ct, MS4A4A, ISG15 and TUBG1 transcripts were strong predictors of mortality and correctly classified all fatal cases or misclassified only one fatal case. However, ISG15 and TUBG1 transcripts misclassified most survivors. From all the predictor variables evaluated independently, the EBOV Ct values and MS4A4A transcript were the most accurate predicting both fatal cases and survivors.

To enhance the prediction performance, SVM models were built with different combinations of predictor variables. Thus, the combinations of EBOV Ct values and MS4A4A or adding more gene transcripts such as VCAM1, ISG15 or TUBG1 increased

the prediction accuracy to more than 90% or even 100%. However, in the field it would be easier to analyse just one predictor in addition to the viral load.

Our results in the 48 DNase-treated samples corroborated the high capability of EBOV Ct values and MS4A4A to predict the risk of EVD mortality. It supports the idea that MS4A4A is a marker independent of EBOV Ct values, therefore useful for the prediction of the clinical outcome of EVD patients in situations where viral load alone is a poor predictor of the outcome.

6.3 MS4AA4 is a strong predictor of EVD fatality

MS4A4A (tetraspan molecule membrane spanning four domains A4A) belongs to the MS4A gene family that encodes multi-spanning membrane proteins, and its function is not yet fully understood. Recent studies have proposed that MS4A4A is a key modulator of the soluble triggering receptor expressed on myeloid cells 2 (sTREM2), can enhance the dectin-1 dependent activation of NK cell-mediated resistance to metastasis and could contribute to the development and pathophysiology of allergy since it is involved in the IgE-mediated mast cell activation (Arthur et al., 2020; Deming et al., 2019; Mattiola et al., 2019). It has been found expressed in peripheral blood monocytes and plasma cells but not lymphocyte B, T, NK cells and granulocytes from healthy human individuals (Sanyal et al., 2017). It is also seen transiently expressed in immature dendritic cells during *in-vitro* differentiation from monocytes and lost in mature dendritic cells (Sanyal et al., 2017). Its expression in human macrophages has also been corroborated by transcriptomic and proteomic data (Mattiola et al., 2019).

This study and previous transcriptomic data obtained in our group shows that MS4A4A is upregulated in EVD patients compared to healthy individuals (Liu et al., 2017). Our findings showed a higher MS4A4A transcript abundance in fatal cases compared to survivors and a high accuracy power for classifying fatal cases alone or in combination with EBOV Ct values. These results suggest that MS4A4A is a strong predictor of EVD fatality.

Several factors could explain this finding. Firstly, MS4A4A is expressed in macrophages which are primary target cells for EBOV infection (Sanyal et al., 2017; Wahl-Jensen et al., 2011). Secondly, MS4A4A is selectively expressed in M2 but not M1 macrophages (Sanyal et al., 2017). Macrophages have a functional plasticity that allows them to differentiate into distinct phenotypes in response to different stimuli (Mantovani et al., 2004). This ability is referred as "macrophage polarization" and can be mimicked *in-vitro* by treating macrophages with cytokines or other stimuli. M1 or "classically activated" macrophages are generated in the presence of IFNγ and by the exposure to lipopolysaccharide (LPS) or other microbial products while exposure to IL-4 and IL-13 induce a differentiation into M2 or "alternatively activated macrophages" (Mantovani et al., 2004; Mattiola et al., 2019). Both phenotypes have different functions, M1 are inflammatory and highly phagocytic macrophages, whereas M2 have been related to wound healing, repair, tumor progression, and downregulation of inflammatory responses (Sanyal et al., 2017). *In-vitro* monocytic differentiation experiments show that MS4A4A is maintained or is upregulated in M2 but not M1 macrophages inferring that MS4A4A can be used as a surface marker for alternative activated macrophages (Czimmerer et al., 2012; Sanyal et al., 2017).

Thirdly, M2 polarization is associated with an increased EBOV infection and fatality (Rogers et al., 2019). An *in-vitro* study of murine and human macrophages infected with a recombinant EBOV model (a vesicular stomatitis virus encoding EBOV GP) found that M2 macrophages induced by IL-4 and IL-13 were more susceptible to infection, indicating that M2 polarization of macrophages enhances EBOV infection. *In-vivo* and *ex-vivo* M2 polarization in mice has also been demonstrated to enhance EBOV infection and mortality (Rogers et al., 2019). The association between M2 polarization and severity or mortality in EVD was also suggested by McElroy et al. (2019) who found higher levels of the soluble macrophage activation marker sCD163 in fatal cases compared to survivors, and increased CD163-immunostained macrophages colocalizing with EBOV antigen in tissues of fatal cases (McElroy et al., 2019). However, unlike MS4A4A, CD163 is not a definitive marker of M2

macrophages since inflammatory macrophages can also express CD163 (Barros et al., 2013).

The increase in infection suggests that M2 polarized macrophages are more susceptible to EBOV entry possibly by increasing the expression of cell surface receptors necessary for EBOV. Rogers et al. (2019) found that the expression of mannose binding receptors such as DC-SIGN and the murine ortholog SIGNR3, which can act as receptors for EBOV, was enhanced in IL-4/IL-13 induced macrophages. It is unknown whether MS4A4A interacts with DC-SIGN to increase viral infection or not, but there is evidence of interaction with Dectin-1, which is also a C-type lectin receptor (a β-glucan receptor on macrophages) (Mattiola et al., 2019). Further studies are required to understand the interaction of MS4A4A, and cell surface receptors involved in EBOV infection.

6.4 Proposed tests to be used in the field for the quantification of the candidate predictive genes

One of the limitations of the RT-qPCR for the absolute quantification of nucleic acids is the use of external standards in each run and variations in PCR efficiency. To avoid this, it would be ideal to implement digital PCR technology for measuring the transcript abundance of the proposed prognostic markers. In contrast to RT-qPCR, digital PCR is capable to obtain absolute quantification without the need for running standard curves and has more robustness to variation in PCR efficiency (Hindson et al., 2013). A new method like the nanoliter-size droplet technology paired with digital PCR (ddPCR) also showed a precise absolute nucleic acid quantification, greater precision and reproducibility over real time PCR even in low abundant targets (Taylor et al., 2017). However, the implementation of digital PCR would not be feasible to use in the field since it has a higher cost compared to RT-qPCR since requires a droplet generator, droplet reader, a compatible thermal cycler, software, consumables, and appropriate expertise. As RT-qPCR platforms are widely distributed in the outbreak settings and were used for the diagnosis of EVD, we would suggest the implementation of the optimized SYBR green-based RT-qPCR assay developed in this

study for measuring the transcript abundance of MS4A4A as an additional laboratory test. The simultaneous quantification of EBOV and MS4A4A in a multiplex RT-qPCR would be ideal. However, we could not include the EBOV primers and probes from the RealStar Filovirus Screen and Zaire Ebolavirus RT-PCR kits in our multiplex RT-qPCR since is confidential intellectual property from Altona diagnostics. In the future, a point-of-care test that could predict EVD outcome would be ideal for remote locations without the need of diagnostic facilities.

6.5 Usefulness of MS4A4A as a prognostic marker in the current EVD situation

To date, the ongoing outbreak in Équateur Province in DRC shows signs of slowing with a few days with no new confirmed cases or deaths which is relatively encouraging. However, this EVD outbreak continues to be a serious concern and the case fatality rate as of 10 October 2020 continues to be around 40.3% (53/128) (World Health Organization, 2020e). Even with a successful vaccine and the implementation of a vaccination strategy, EVD in Africa is not completely defeated. Some of the remaining challenges include community resistance to response activities, the risk of safe and dignified burials, contacts lost to follow up, confirmed cases in the community, and lack of funding. Thus, EVD remains a public threat and for any new outbreaks it is crucial to contain epidemic spread and decrease mortality. One aspect of this is the need to improve the clinical outcome of EVD patients. Our study identified MS4A4A as a potential prognosis marker that could be used in the field to predict the clinical outcome of EVD patients. The potential usefulness of this host marker are detail as follows:

6.5.1 Usefulness of MS4A4A in the clinical decision

Data of EVD patients from the Ebola treatment centres in Africa, and from those evacuated to Europe or the USA, showed that approximately 50% of patients in either setting were critically ill or required advanced supportive care interventions (Kiiza et al., 2020; Uyeki et al., 2016). This highlights the need for early classification of patients who may develop severe illness, since supportive care interventions should

be tailored to patient-specific critical care. It is known that an appropriate supportive care intervention contribute to the improvement of the clinical outcome and lower the case fatality rate (18.5%) (Uyeki et al., 2016). However, it is still a challenge for resource-limited settings like Africa. Even though the case fatality rate dropped from 70% before 2014 to 39.5% during the West Africa outbreak, in more recent EVD outbreaks in DRC the mortality remains around 50% (50% in 2017, 61% in 2018, 66% in 2018-2020 outbreak and 40.3% in the ongoing outbreak). In an attempt to reduce mortality after admission to the Ebola treatment centres, the host marker MS4A4A may be included in the mortality prediction score for EVD to guide clinicians in the triage of patients. Thus, clinicians could direct the scarce resources towards those predicted as "fatal". As the model could help in the identification of severely ill patients, these patients may urgently receive mechanical ventilation, renal replacement therapy, antibiotics, and administration of investigational therapies for recovery (Uyeki et al., 2016).

Recently, WHO prepared a guidance for the supportive management of EVD patients in Ebola treatment centres. This guidance recommends a systematic clinical assessment on admission followed by regular re-assessments of vital signs, physical examination, fluid status, and laboratory monitoring (World Health Organization, 2019a). To support the evaluation of the clinical condition of the patient, the host marker MS4A4A could be included in laboratory monitoring that consists of serum biochemistry and haematology analysis of Na, K, HCO_3, BUN, creatinine, AST, ALT, Mg, glucose, Hg, PT/INR, creatine kinase, CL^-, Ca^{2+}, albumin and lactate. Indeed, most of these laboratory markers have been associated with fatal outcome as described in Chapter 1. Although our study showed that MS4A4A alone is a strong predictor of mortality, a further study should analyse the predictive power of prognostic models combining laboratory markers and MS4A4A as predictor variables. Unfortunately, we could not develop these models since we did not have this information from the cohorts of patients we studied.

One approach that could be accessible to clinicians in outbreak settings and to speed up the triage of patients would be the use of mobile apps based on prognostic models

as proposed by Colubri et al. (2019). Although Colubri et al. (2019) developed a model with a combination of viral load, age, time to presentation, clinical symptoms, and body temperature from EVD patients, it would be relevant to develop and validate a prognostic model that includes MS4A4A, EBOV Ct values and other laboratory markers to be deployed as mobile apps.

6.5.2 Usefulness of MS4A4A for the evaluation of therapeutics agents

The PALM trial conducted during the 2018-2020 outbreak in DRC identified a statistically significant survival benefit in patients treated with a combination of the standard care plus either mAb114 (a single human monoclonal antibody agent) or REGN-EB3 (a mixture of three human IgG1 monoclonal antibodies). However, the case fatality rate was still high in patients with high viral load at the time of admission (Ct values ≤ 22) despite the use of mAb114 (CFR = 70%) or REGN-EB3 (CFR = 64%)(Mulangu et al., 2019). This relative lack of efficacy in severe disease patients is not surprising since the efficacy of the four experimental therapeutics evaluated in the PALM trial (mAb114, REGN-EB3, remdesivir and ZMapp) was not assessed before in severe EVD during the preclinical evaluation in NHP models. Since NHPs recapitulate EVD at accelerated pace, therapeutic interventions were initiated at the time of virus exposure or shortly after disease signs were observed, but not at later time points like in patients in the outbreak setting (Iversen et al., 2020; Warren et al., 2020). The clinical interventions in patients were at around 11-12 days after viral exposure, since people often seek medical help around 5.5 days after the onset of symptoms when some patients may already have multi-organ dysfunction (Iversen et al., 2020; Mulangu et al., 2019). Since many patients with a severe disease arrive to the Ebola treatment centres during an outbreak it is necessary to develop effective therapeutic strategies that could improve the outcome of these patients. In line with this, we propose to study the increase of MS4A4A transcripts levels during the course of EBOV infection. This may be useful for assessing the efficacy of the experimental therapeutics since it could serve as a monitor of the improvement of the systemic inflammatory response after treatment interventions in either EVD patients or NHP models. This is relevant since the efficacy of experimental therapeutics currently

relies on assessing the viral replication but not the host response. Since our findings showed that the transcript abundance of MS4A4A is higher in fatal cases than survivors, we hypothesize that the improvement of the outcome after treatment interventions would mean a decrease in the transcript abundance of MSA4A4. Thus, it may be possible that MS4A4A could be used during the evaluation of a combined approach of therapeutics with an optimized supportive care because both cannot be uncoupled in the field. Supportive care remains the cornerstone of patient's treatment since it is often seen EVD patients with multiple organ dysfunction at the time of admission (Iversen et al., 2020; Mulangu et al., 2019).

6.5.3 Usefulness of MS4A4A for host-directed therapy

Host-directed therapies are any product that can modulate the host response in a beneficial way to counteract the viral infection by strengthening the host defence mechanism (e.g immune stimulation), by ameliorating the disease (e.g., modulating excessive inflammation) or both (Zumla et al., 2016). Some investigational host-directed therapies for EVD are repurposed drugs (Irbesartan, Atorvastatin), cytokine therapy (Pegylated interferon α and β), monoclonal antibody (ZMAb), and recombinant protein (rNAPc2)(Zumla et al., 2016). The potential of MS4A4A as an immunotherapeutic target has been suggested in cancer studies (Mattiola et al., 2019; Sanyal et al., 2017). Similar to recent findings in EBOV studies that indicates an association of M2 macrophage polarization with mortality, in cancer studies M2 tumour-enhancing macrophages have been associated with poor prognosis (Rogers et al., 2019; Sanyal et al., 2017; Zhang et al., 2011). Since MS4A4A expression was found restricted to M2 but not M1 macrophages, this suggested a way to target MS4A4A with monoclonal antibodies to deplete M2 macrophages without affecting the M1 pro-inflammatory macrophages (Mattiola et al., 2019; Sanyal et al., 2017). Indeed, targeting CD20 (known as MS4A1, and same protein family of MS4A4A) has shown to be effective in the treatment of autoimmune disorders (Franks et al., 2016). Thus, the potential use of MS4A4 as an immunotherapeutic target in EVD would be interesting to be investigated in future studies. This is encouraged by the results of other strategies that attempted to modulate the host response, mainly macrophages,

during EBOV infection. For example, treatment with Eritoran, a TLR-4 antagonist increased the survival of EBOV and SUDV infected mice (Younan et al., 2017). Likewise, the treatment of murine macrophages with IFNγ conferred resistance to EBOV infection, and the transfer of M1 macrophages to mice protected them against a lethal dose of recombinant EBOV model (Rhein et al., 2015; Rogers et al., 2020).

6.6 Potential utility of MS4A4A in other filovirus infections and other high consequence infectious diseases

Situations like the 2013-2016 EVD pandemic and the current COVID-19 pandemic and highlight the high pressure on healthcare facilities due to the high number of severely ill cases, and the need of a prognostic marker that distinguish those patients in need of urgent medical attention. It is also important to note that despite the advanced investigation of therapeutics for EVD, less attention has been paid to the investigation of other filovirus diseases such as Sudan virus disease (SVD), Bundibugyo virus disease (BVD), and Marburg virus disease (MVD). Indeed, some of the current therapeutics used in EVD may be ineffective in other filovirus diseases, e.g. the mAb treatments tested in the PALM trial (mAb114, REGN-EB3, and ZMapp) are specific to EBOV, therefore are ineffective against other ebolaviruses or marburgviruses (Iversen et al., 2020). One exception is remdesivir that showed an *in-vitro* activity against SUDV and MARV (Lo et al., 2017). Thus, to be prepared for future filovirus outbreaks, it would be relevant to investigate the expression levels of MS4A4A *in-vitro* or in animal models infected with these viruses. It is likely that MS4A4A is also associated with severity in other filovirus diseases since these viruses share a similar mechanism of entry. Besides, a previous study of SVD patients showed that a high level of the macrophage activation marker sCD163 was associated with fatal outcome, so it is very likely that MS4A4A would also be an important marker of fatality in SVD (McElroy et al., 2019). Likewise, the RT-qPCR assay developed in this thesis may be useful for the quantification of MS4A4A transcripts to compare survivors and fatal cases from other infectious diseases that elicit an M2 polarization like EVD. For the current pandemic, it would be important to investigate MS4A4A since severe lung disease has been associated to the increased induction of M2

macrophages in the infection caused by SARS-CoV and SARS-CoV-2 (Liu et al., 2020; Page et al., 2012).

6.7 Limitations of the study

The main limitation in our study was the RNA yield of the clinical samples that was insufficient for testing appropriately the 11 candidate predictive gene transcripts. Since these samples were not DNase treated to avoid loss of RNA due to an extra purification, we could not rule out the possibility of co-amplification of genomic DNA in some RT-qPCRs for primer pairs which were located within a single exon (VCAM1, HOPX, SLC25A5, MS4A4A and ISG15). However, 48 leftover clinical samples were DNase treated and re-tested by RT-qPCR for MS4A4A to validate our findings.

Other limitation in our study was the low sample size used to train the machine learning models (less than 100 individuals). However, to avoid an optimistically biased performance we separated the training and testing data as recommended by Vabalas et al. (2019) for validating machine learning algorithms in limited sample size and used a minimal number of predictive variables to avoid overfitting. This study has room for further improvement. In future work, training a predictive model using MS4A4A with or without EBOV Ct values in a larger cohort of patients from different Ebola treatment centres would be more accurate.

It is known that an advanced supportive care may improve the clinical outcome, as was observed in EVD patients that were evacuated to the USA or Europe during the 2013-2016 West African outbreak (McElroy et al., 2016). Whether a supportive care intervention in our cohorts of patients affected the measured transcript levels of our candidate predictive genes is unknown. We do not have information about the supportive care that these patients received.

Other factor that may have influenced the clinical outcome of these patients is the co-infection with parasitic diseases. Some studies have suggested that infection with *Plasmodium falciparum* protects from a fatal EBOV infection. Although this is

contradicting by other reports (Rosenke et al., 2016; Waxman et al., 2017). Interestingly, this protection has been suggested to be due to the induction of M1 polarization of tissue macrophages through the production of IFNγ (Rogers et al., 2020). In our study only one sample that had co-infection with *Plasmodium falciparum* was analysed, so it was not possible to evaluate differences in the MS4A4A transcript abundance between patients diagnosed with EVD alone and those diagnosed with both EVD and malaria. In addition, helminth infections may also impact the clinical outcome by inducing M2 polarization (Rolot & Dewals, 2018). Intestinal helminth infections are very prevalent in the West African countries and considered as a neglected tropical disease in this region (Hotez, 2015). Although we did not have information about co-infection with helminths in our cohorts of patients, we cannot rule out the possibility that a co-infection may have influenced the clinical outcome of these individuals. Further studies would be necessary to understand the impact of helminth infections in the EVD outcome and whether the transcript abundance of MS4A4A is altered or not.

6.8 Several other gene transcripts may also be important predictors of the clinical outcome

Several gene transcripts from the original set were not included in the machine learning approaches, as there was insufficient RNA to perform analysis on all patient samples. Nevertheless, some of them showed a significant difference in the transcript abundance between survivors and fatal cases e.g., TTC28 ($p = 0.0077$) and TGFBI ($p = 0.0303$). Although only 16 samples (4 survivors and 12 fatal cases) were analysed for TTC28, and 12 samples (2 survivors and 10 fatal cases) for TGFBI, transcript levels were quite different for TTC28 (2 times less) and TGFBI (16 times less) in fatal compared to survivors. It would have been more appropriate to select these genes as well for further investigation. However, as we initially aimed to predict the outcome of the “blinded” cohort of patients, priority was given to the genes that we had more data for building a SVM model such as VCAM1 (39 samples), ISG15 (32 samples), TUBG1 (31 samples) and MS4A4A (24 samples).

Based on our findings we recommend further analysis focused on three genes: VCAM1, ISG15 and TGFBI. This study showed that the transcript abundance of VCAM1 and ISG15 was significantly higher in fatal cases compared to survivors. In contrast, the transcript abundance of TGFBI was significantly lower in fatal cases than survivors.

VCAM1 is a glycoprotein predominantly expressed in the surface of endothelial cells and is upregulated during EBOV infection like other cell adhesion molecules (CAMs) such as ICAM-1 and shed soluble forms into plasma (Wahl-Jensen et al., 2005). Our study is consistent with that of Kerber et al. (2018) who found a significant difference in the kinetics expression of soluble VCAM1 (sVCAM1) between survivors and fatal cases from the 2013-2016 outbreak in Guinea, with increased levels of sVCAM1 in fatal cases and/or decreased levels in survivors. Our results showed an increased transcript copy number in both but higher levels in fatal cases. In contrast, other study only found increased levels of sICAM but normal levels of sVCAM1 in a small cohort of USA patients with EVD (McElroy et al., 2016). Finally, these results corroborate the importance of VCAM1 as a predictor of the EVD outcome which was consistently selected by three different machine learning classifiers in a previous work by our group (Liu et al., 2017).

ISG15 is a member of the ubiquitin family and it is strongly induced by type I interferons and pathogenic stimuli. It has also been proposed to have a function as a cytokine regulating the immune response and an antiviral role in diverse viral infections (Perng & Lenschow, 2018). ISG15 inhibits the viral replication by conjugation (ISGylation) of the viral proteins affecting the viral replication machinery, and by inhibiting the virus release as seen in EBOV infection (Durfee et al., 2010; Okumura et al., 2008). The increased transcript abundance of ISG15 in EVD patients in this study corroborates earlier findings that found an early and strong upregulation of ISG15 and other ISGs over the course of the disease in NHP models infected with EBOV (Caballero et al., 2016; Greenberg et al., 2020). Likewise, corroborates the higher ISG15 expression observed in non-survivors versus survivors in EBOV-infected NHP (Garamszegi et al., 2014).

TGFBI is an extracellular matrix protein. Although few samples were analysed for TGFBI, this study showed a lower transcript abundance in fatal cases compared to survivors which corroborates previous transcriptomic analysis of EVD patients (Liu et al., 2017). These results suggest a potential association between TGFBI downregulation and fatality in EVD. This may be somewhat associated to the role that TGFB has in promoting M2 polarization which has been related to EVD fatality (Gong et al., 2012). Contrary to this result, TGFBI is upregulated in Lassa virus infection (Malhotra et al., 2013). Similar to VCAM1, TGFBI was also consistently selected by three different machine learning models as a predictive marker to differentiate EVD survivors and fatal cases (Liu et al., 2017).

Further research is recommended to confirm the ability of VCAM1, ISG15 and TGFBI as potential markers of EVD outcome in a larger cohort.

6.9 Prognostic tests currently used in other infectious diseases

Although there are no commercially available prognostic tests for infectious diseases, there are several studies that have proposed severity prediction models to be used by clinicians to improve the triage and management of patients with Dengue virus, Influenza virus, SARS-CoV2 and Respiratory syncytial virus infection (Chen et al., 2020; Nguyen et al., 2017; Vos et al., 2019; Yan et al., 2020). These models are based on laboratory markers, clinical signs, viremia, or other risk factors of mortality. None of these models have included the gene transcripts evaluated in this thesis as prognostic markers. Further work is required to validate these severity prediction models and to assess their influence in the clinical treatment decision.

6.10 Conclusions

Taken together, this study has demonstrated that the transcript abundance of VCAM1, ISG15, TUBG1 but mainly MS4A4A may be useful as prognostic markers of EVD. Our findings suggest that MS4A4A is a strong predictor of EVD fatality and the

best predictor of EVD outcome within the genes set. One explanation for this finding is the expression of MS4A4A in M2 polarized macrophages which have been related to a severe form of EVD and fatality. In machine learning models, MS4A4A has shown a good performance as a predictor of the outcome similar and independent of EBOV Ct value. In future investigations, a larger sample size and the prediction of an independent dataset would be relevant to confirm these findings.

References

Agnandji, S. T., Huttner, A., Zinser, M. E., Njuguna, P., Dahlke, C., Fernandes, J. F., Yerly, S., Dayer, J. A., Kraehling, V., Kasonta, R., Adegnika, A. A., Altfeld, M., Auderset, F., Bache, E. B., Biedenkopf, N., Borregaard, S., Brosnahan, J. S., Burrow, R., Combescure, C., Desmeules, J., Eickmann, M., Fehling, S. K., Finckh, A., Goncalves, A. R., Grobusch, M. P., Hooper, J., Jambrecina, A., Kabwende, A. L., Kaya, G., Kimani, D., Lell, B., Lemaitre, B., Lohse, A. W., Massinga-Loembe, M., Matthey, A., Mordmuller, B., Nolting, A., Ogwang, C., Ramharter, M., Schmidt-Chanasit, J., Schmiedel, S., Silvera, P., Stahl, F. R., Staines, H. M., Strecker, T., Stubbe, H. C., Tsofa, B., Zaki, S., Fast, P., Moorthy, V., Kaiser, L., Krishna, S., Becker, S., Kieny, M. P., Bejon, P., Kremsner, P. G., Addo, M. M., & Siegrist, C. A. (2016). Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe. *N Engl J Med*, *374*(17), 1647-1660. https://doi.org/10.1056/NEJMoa1502924

Albert, H., Nathavitharana, R. R., Isaacs, C., Pai, M., Denkinger, C. M., & Boehme, C. C. (2016). Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *European Respiratory Journal*, *48*(2), 516. https://doi.org/10.1183/13993003.00543-2016

Aleksandrowicz, P., Marzi, A., Biedenkopf, N., Beimforde, N., Becker, S., Hoenen, T., Feldmann, H., & Schnittler, H. J. (2011). Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. *The Journal of infectious diseases*, *204 Suppl 3*, S957-967. https://doi.org/10.1093/infdis/jir326

Alvarez, C. P., Lasala, F., Carrillo, J., Muñiz, O., Corbí, A. L., & Delgado, R. (2002). C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *Journal of Virology*, *76*(13), 6841-6844. https://doi.org/10.1128/jvi.76.13.6841-6844.2002

Amman, B. R., Bird, B. H., Bakarr, I. A., Bangura, J., Schuh, A. J., Johnny, J., Sealy, T. K., Conteh, I., Koroma, A. H., Foday, I., Amara, E., Bangura, A. A., Gbakima, A. A., Tremeau-Bravard, A., Belaganahalli, M., Dhanota, J., Chow, A., Ontiveros, V., Gibson, A., Turay, J., Patel, K., Graziano, J., Bangura, C., Kamanda, E. S., Osborne, A., Saidu, E., Musa, J., Bangura, D., Williams, S. M. T., Wadsworth, R., Turay, M., Edwin, L., Mereweather-Thompson, V., Kargbo, D., Bairoh, F. V., Kanu, M., Robert, W., Lungai, V., Guetiya Wadoum, R. E., Coomber, M., Kanu, O., Jambai, A., Kamara, S. M., Taboy, C. H., Singh, T., Mazet, J. A. K., Nichol, S. T., Goldstein, T., Towner, J. S., & Lebbie, A. (2020). Isolation of Angola-like Marburg virus from Egyptian rousette bats from West Africa. *Nat Commun*, *11*(1), 510. https://doi.org/10.1038/s41467-020-14327-8

Aquila, G., Fortini, C., Pannuti, A., Delbue, S., Pannella, M., Morelli, M. B., Caliceti, C., Castriota, F., de Mattei, M., Ongaro, A., Pellati, A., Ferrante, P., Miele, L., Tavazzi, L., Ferrari, R., Rizzo, P., & Cremonesi, A. (2017). Distinct gene expression profiles associated with Notch ligands Delta-like 4 and Jagged1 in plaque material from peripheral artery disease patients: a pilot study. *Journal of Translational Medicine*, *15*(1), 98. https://doi.org/10.1186/s12967-017-1199-3

Arthur, G. K., Ehrhardt-Humbert, L. C., Snider, D. B., Jania, C., Tilley, S. L., Metcalfe, D. D., & Cruse, G. (2020). The FcεRIβ homologue, MS4A4A, promotes FcεRI signal transduction and store-operated Ca2+ entry in human mast cells. *Cellular Signalling*, *71*, 109617. https://doi.org/https://doi.org/10.1016/j.cellsig.2020.109617

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, *19*(8), 1165-1195. https://doi.org/10.1007/s00521-010-0362-z

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M. S., Keita, S., De Clerck, H., Tiffany, A., Dominguez, G., Loua, M., Traore, A., Kolie, M., Malano, E. R., Heleze, E., Bocquin, A., Mely, S., Raoul, H., Caro, V., Cadar, D., Gabriel, M., Pahlmann, M., Tappe, D., Schmidt-Chanasit, J., Impouma, B., Diallo, A. K., Formenty, P., Van Herp, M., & Gunther, S. (2014). Emergence of Zaire Ebola virus disease in Guinea. *N Engl J Med*, *371*(15), 1418-1425. https://doi.org/10.1056/NEJMoa1404505

Banerjee, A., Rapin, N., Bollinger, T., & Misra, V. (2017). Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. *Sci Rep*, *7*(1), 2232. https://doi.org/10.1038/s41598-017-01513-w

Barrette, R. W., Metwally, S. A., Rowland, J. M., Xu, L., Zaki, S. R., Nichol, S. T., Rollin, P. E., Towner, J. S., Shieh, W.-J., Batten, B., Sealy, T. K., Carrillo, C., Moran, K. E., Bracht, A. J., Mayr, G. A., Sirios-Cruz, M., Catbagan, D. P., Lautner, E. A., Ksiazek, T. G., White, W. R., & McIntosh, M. T. (2009). Discovery of Swine as a Host for the Reston ebolavirus. *Science*, *325*(5937), 204. https://doi.org/10.1126/science.1172705

Barros, M. H., Hauck, F., Dreyer, J. H., Kempkes, B., & Niedobitek, G. (2013). Macrophage polarisation: an immunohistochemical approach for identifying M1 and M2 macrophages. *PLoS One*, *8*(11), e80908. https://doi.org/10.1371/journal.pone.0080908

Bauer, A., Bostrom, A. G., Ball, J., Applegate, C., Cheng, T., Laycock, S., Rojas, S. M., Kirwan, J., & Zhou, J. (2019). Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production. *Hortic Res*, *6*, 70. https://doi.org/10.1038/s41438-019-0151-5

Bausch, D. G., Towner, J. S., Dowell, S. F., Kaducu, F., Lukwiya, M., Sanchez, A., Nichol, S. T., Ksiazek, T. G., & Rollin, P. E. (2007). Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. *The Journal of infectious diseases*, *196 Suppl 2*, S142-147. https://doi.org/10.1086/520545

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, *319*(13), 1317-1318.

Beckert, B., & Masquida, B. (2011). Synthesis of RNA by In Vitro Transcription. In H. Nielsen (Ed.), *RNA: Methods and Protocols* (pp. 29-41). Humana Press. https://doi.org/10.1007/978-1-59745-248-9_3

Berkley, S. (2018). Health security's blind spot. *Science*, *359*(6380), 1075. https://doi.org/10.1126/science.aat4714

Bosio, C. M., Aman, M. J., Grogan, C., Hogan, R., Ruthel, G., Negley, D., Mohamadzadeh, M., Bavari, S., & Schmaljohn, A. (2003). Ebola and Marburg Viruses Replicate in Monocyte-Derived Dendritic Cells without Inducing the Production of Cytokines and Full Maturation. *The Journal of infectious diseases*, *188*(11), 1630-1638. https://doi.org/10.1086/379199

Bower, H., & Glynn, J. R. (2017). A systematic review and meta-analysis of seroprevalence surveys of ebolavirus infection. *Sci Data*, *4*, 160133. https://doi.org/10.1038/sdata.2016.133

Bradley, J. H., Harrison, A., Corey, A., Gentry, N., & Gregg, R. K. (2018). Ebola virus secreted glycoprotein decreases the anti-viral immunity of macrophages in early inflammatory responses. *Cell Immunol*, *324*, 24-32. https://doi.org/10.1016/j.cellimm.2017.11.009

Bray, M., & Geisbert, T. W. (2005). Ebola virus: the role of macrophages and dendritic cells in the pathogenesis of Ebola hemorrhagic fever. *Int J Biochem Cell Biol*, *37*(8), 1560-1566. https://doi.org/10.1016/j.biocel.2005.02.018

Broadhurst, M. J., Brooks, T. J. G., & Pollock, N. R. (2016). Diagnosis of Ebola Virus Disease: Past, Present, and Future. *Clinical microbiology reviews*, *29*(4), 773-793. https://doi.org/10.1128/CMR.00003-16

Broeders, S., Garlant, L., Fraiture, M. A., Vandermassen, E., Suin, V., Vanhomwegen, J., Dupont-Rouzeyrol, M., Rousset, D., Van Gucht, S., & Roosens, N. (2020). A new multiplex RT-qPCR method for the simultaneous detection and discrimination of Zika and chikungunya viruses. *Int J Infect Dis*, *92*, 160-170. https://doi.org/10.1016/j.ijid.2019.12.028

Brudner, M., Karpel, M., Lear, C., Chen, L., Yantosca, L. M., Scully, C., Sarraju, A., Sokolovska, A., Zariffard, M. R., Eisen, D. P., Mungall, B. A., Kotton, D. N., Omari, A., Huang, I. C., Farzan, M., Takahashi, K., Stuart, L., Stahl, G. L., Ezekowitz, A. B., Spear, G. T., Olinger, G. G., Schmidt, E. V., & Michelow, I. C. (2013). Lectin-dependent enhancement of Ebola virus infection via soluble and transmembrane C-type lectin receptors. *PLoS One*, *8*(4), e60838. https://doi.org/10.1371/journal.pone.0060838

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., & Wittwer, C. T. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, *55*(4), 611. https://doi.org/10.1373/clinchem.2008.112797

Butler, D. (2018). Speedy Ebola tests help contain Africa's latest outbreak. *Nature*, *558*(7709), 172. https://doi.org/10.1038/d41586-018-05389-2

Caballero, I. S., Honko, A. N., Gire, S. K., Winnicki, S. M., Mele, M., Gerhardinger, C., Lin, A. E., Rinn, J. L., Sabeti, P. C., Hensley, L. E., & Connor, J. H. (2016). In vivo Ebola virus infection leads to a strong innate response in circulating immune cells. *BMC Genomics*, *17*, 707. https://doi.org/10.1186/s12864-016-3060-0

Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., Kuehne, A. I., Kranzusch, P. J., Griffin, A. M., Ruthel, G., Dal Cin, P., Dye, J. M., Whelan, S. P., Chandran, K., & Brummelkamp, T. R. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*, *477*(7364), 340-343. https://doi.org/10.1038/nature10348

Carroll, M. W. (2019). Retrospective versus real-time Ebola virus sequencing. *Lancet Infect Dis*, *19*(6), 567-568. https://doi.org/10.1016/S1473-3099(19)30207-5

Carroll, M. W., Matthews, D. A., Hiscox, J. A., Elmore, M. J., Pollakis, G., Rambaut, A., Hewson, R., Garcia-Dorival, I., Bore, J. A., Koundouno, R., Abdellati, S., Afrough, B., Aiyepada, J., Akhilomen, P., Asogun, D., Atkinson, B., Badusche, M., Bah, A., Bate, S., Baumann, J., Becker, D., Becker-Ziaja, B., Bocquin, A., Borremans, B., Bosworth, A., Boettcher, J. P., Cannas, A., Carletti, F., Castilletti, C., Clark, S., Colavita, F., Diederich, S., Donatus, A.,

Duraffour, S., Ehichioya, D., Ellerbrok, H., Fernandez-Garcia, M. D., Fizet, A., Fleischmann, E., Gryseels, S., Hermelink, A., Hinzmann, J., Hopf-Guevara, U., Ighodalo, Y., Jameson, L., Kelterbaum, A., Kis, Z., Kloth, S., Kohl, C., Korva, M., Kraus, A., Kuisma, E., Kurth, A., Liedigk, B., Logue, C. H., Ludtke, A., Maes, P., McCowen, J., Mely, S., Mertens, M., Meschi, S., Meyer, B., Michel, J., Molkenthin, P., Munoz-Fontela, C., Muth, D., Newman, E. N., Ngabo, D., Oestereich, L., Okosun, J., Olokor, T., Omiunu, R., Omomoh, E., Pallasch, E., Palyi, B., Portmann, J., Pottage, T., Pratt, C., Priesnitz, S., Quartu, S., Rappe, J., Repits, J., Richter, M., Rudolf, M., Sachse, A., Schmidt, K. M., Schudt, G., Strecker, T., Thom, R., Thomas, S., Tobin, E., Tolley, H., Trautner, J., Vermoesen, T., Vitoriano, I., Wagner, M., Wolff, S., Yue, C., Capobianchi, M. R., Kretschmer, B., Hall, Y., Kenny, J. G., Rickett, N. Y., Dudas, G., Coltart, C. E., Kerber, R., Steer, D., Wright, C., Senyah, F., Keita, S., Drury, P., Diallo, B., de Clerck, H., Van Herp, M., Sprecher, A., Traore, A., Diakite, M., Konde, M. K., Koivogui, L., Magassouba, N., Avsic-Zupanc, T., Nitsche, A., Strasser, M., Ippolito, G., Becker, S., Stoecker, K., Gabriel, M., Raoul, H., Di Caro, A., Wolfel, R., Formenty, P., & Gunther, S. (2015). Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature*, *524*(7563), 97-101. https://doi.org/10.1038/nature14594

CDC. (2020). *Years of Ebola Virus Disease Outbreaks*. https://www.cdc.gov/vhf/ebola/history/chronology.html

Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P., & Cunningham, J. M. (2005). Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science (New York, N.Y.)*, *308*(5728), 1643-1645. https://doi.org/10.1126/science.1110656

Chen, L., Han, X., Li, Y. L., Zhang, C., & Xing, X. (2020). FluA-p score: a novel prediction rule for mortality in influenza A-related pneumonia patients. *Respir Res*, *21*(1), 109. https://doi.org/10.1186/s12931-020-01379-z

Chen, R. H., Du, Y., Han, P., Wang, H. B., Liang, F. Y., Feng, G. K., Zhou, A. J., Cai, M. Y., Zhong, Q., Zeng, M. S., & Huang, X. M. (2016). ISG15 predicts poor prognosis and promotes cancer stem cell phenotype in nasopharyngeal carcinoma. *Oncotarget*, *7*(13), 16910-16922. https://doi.org/10.18632/oncotarget.7626

Chertow, D. S., Kleine, C., Edwards, J. K., Scaini, R., Giuliani, R., & Sprecher, A. (2014). Ebola Virus Disease in West Africa — Clinical Manifestations and Management. *New England Journal of Medicine*, *371*(22), 2054-2057. https://doi.org/10.1056/NEJMp1413084

Cnops, L., De Smet, B., Mbala-Kingebeni, P., van Griensven, J., Ahuka-Mundeke, S., & Arien, K. K. (2019). Where are the Ebola diagnostics from last time? *Nature*, *565*(7740), 419-421. https://doi.org/10.1038/d41586-019-00212-y

Colavita, F., Biava, M., Castilletti, C., Lanini, S., Miccio, R., Portella, G., Vairo, F., Ippolito, G., Capobianchi, M. R., Di Caro, A., & Lalle, E. (2019). Inflammatory and Humoral Immune Response during Ebola Virus Infection in Survivor and Fatal Cases Occurred in Sierra Leone during the 2014–2016 Outbreak in West Africa. *Viruses*, *11*(4), 373. https://doi.org/10.3390/v11040373

Colubri, A., Hartley, M. A., Siakor, M., Wolfman, V., Felix, A., Sesay, T., Shaffer, J. G., Garry, R. F., Grant, D. S., Levine, A. C., & Sabeti, P. C. (2019). Machine-learning Prognostic Models from the 2014-16 Ebola Outbreak: Data-

harmonization Challenges, Validation Strategies, and mHealth Applications. *EClinicalMedicine*, *11*, 54-64. https://doi.org/10.1016/j.eclinm.2019.06.003

Crowe, S. J., Maenner, M. J., Kuah, S., Erickson, B. R., Coffee, M., Knust, B., Klena, J., Foday, J., Hertz, D., Hermans, V., Achar, J., Caleo, G. M., Van Herp, M., Albarino, C. G., Amman, B., Basile, A. J., Bearden, S., Belser, J. A., Bergeron, E., Blau, D., Brault, A. C., Campbell, S., Flint, M., Gibbons, A., Goodman, C., McMullan, L., Paddock, C., Russell, B., Salzer, J. S., Sanchez, A., Sealy, T., Wang, D., Saffa, G., Turay, A., Nichol, S. T., & Towner, J. S. (2016). Prognostic Indicators for Ebola Patient Survival. *Emerg Infect Dis*, *22*(2), 217-223. https://doi.org/10.3201/eid2202.151250

Cui, D., Zhao, D., Xie, G., Yang, X., Huo, Z., Zheng, S., Yu, F., & Chen, Y. (2016). Simultaneous detection of influenza A subtypes of H3N2 virus, pandemic (H1N1) 2009 virus and reassortant avian H7N9 virus in humans by multiplex one-step real-time RT-PCR assay. *Springerplus*, *5*(1), 2054. https://doi.org/10.1186/s40064-016-3733-9

Czimmerer, Z., Varga, T., Poliska, S., Nemet, I., Szanto, A., & Nagy, L. (2012). Identification of novel markers of alternative activation and potential endogenous PPARgamma ligand production mechanisms in human IL-4 stimulated differentiating macrophages. *Immunobiology*, *217*(12), 1301-1314. https://doi.org/10.1016/j.imbio.2012.08.270

Dahlmann, F., Biedenkopf, N., Babler, A., Jahnen-Dechent, W., Karsten, C. B., Gnirß, K., Schneider, H., Wrensch, F., O'Callaghan, C. A., Bertram, S., Herrler, G., Becker, S., Pöhlmann, S., & Hofmann-Winkler, H. (2015). Analysis of Ebola Virus Entry Into Macrophages. *The Journal of infectious diseases*, *212 Suppl 2*(Suppl 2), S247-S257. https://doi.org/10.1093/infdis/jiv140

Dang, T. N., Naka, I., Sa-Ngasang, A., Anantapreecha, S., Chanama, S., Wichukchinda, N., Sawanpanyalert, P., Patarapotikul, J., Tsuchiya, N., & Ohashi, J. (2014). A replication study confirms the association of GWAS-identified SNPs at MICB and PLCE1in Thai patients with dengue shock syndrome. *BMC Medical Genetics*, *15*(1), 58. https://doi.org/10.1186/1471-2350-15-58

Davi, C., Pastor, A., Oliveira, T., Neto, F. B. d. L., Braga-Neto, U., Bigham, A. W., Bamshad, M., Marques, E. T. A., & Acioli-Santos, B. (2019). Severe Dengue Prognosis Using Human Genome Data and Machine Learning. *IEEE Transactions on Biomedical Engineering*, *66*(10), 2861-2868. https://doi.org/10.1109/TBME.2019.2897285

De Nys, H. M., Kingebeni, P. M., Keita, A. K., Butel, C., Thaurignac, G., Villabona-Arenas, C. J., Lemarcis, T., Geraerts, M., Vidal, N., Esteban, A., Bourgarel, M., Roger, F., Leendertz, F., Diallo, R., Ndimbo-Kumugo, S. P., Nsio-Mbeta, J., Tagg, N., Koivogui, L., Toure, A., Delaporte, E., Ahuka-Mundeke, S., Tamfum, J. M., Mpoudi-Ngole, E., Ayouba, A., & Peeters, M. (2018). Survey of Ebola Viruses in Frugivorous and Insectivorous Bats in Guinea, Cameroon, and the Democratic Republic of the Congo, 2015-2017. *Emerg Infect Dis*, *24*(12), 2228-2240. https://doi.org/10.3201/eid2412.180740

Dean, N. E., Halloran, M. E., Yang, Y., & Longini, I. M. (2016). Transmissibility and Pathogenicity of Ebola Virus: A Systematic Review and Meta-analysis of Household Secondary Attack Rate and Asymptomatic Infection. *Clin Infect Dis*, *62*(10), 1277-1286. https://doi.org/10.1093/cid/ciw114

Deming, Y., Filipello, F., Cignarella, F., Cantoni, C., Hsu, S., Mikesell, R., Li, Z., Del-Aguila, J. L., Dube, U., Farias, F. G., Bradley, J., Budde, J., Ibanez, L.,

Fernandez, M. V., Blennow, K., Zetterberg, H., Heslegrave, A., Johansson, P. M., Svensson, J., Nellgård, B., Lleo, A., Alcolea, D., Clarimon, J., Rami, L., Molinuevo, J. L., Suárez-Calvet, M., Morenas-Rodríguez, E., Kleinberger, G., Ewers, M., Harari, O., Haass, C., Brett, T. J., Benitez, B. A., Karch, C. M., Piccio, L., & Cruchaga, C. (2019). The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Science translational medicine*, *11*(505), eaau2291. https://doi.org/10.1126/scitranslmed.aau2291

Den Boon, S., Marston, B. J., Nyenswah, T. G., Jambai, A., Barry, M., Keita, S., Durski, K., Senesie, S. S., Perkins, D., Shah, A., Green, H. H., Hamblion, E. L., Lamunu, M., Gasasira, A., Mahmoud, N. O., Djingarey, M. H., Morgan, O., Crozier, I., & Dye, C. (2019). Ebola Virus Infection Associated with Transmission from Survivors. *Emerg Infect Dis*, *25*(2), 249-255. https://doi.org/10.3201/eid2502.181011

Deng, I., Duku, O., Gillo, A., Idris, A., Lolik, P., el Tahir, B., Tembura, W., Warille, N., Knobloch, J., & Cornet, M. (1978). Ebola Hemorrhagic-Fever In Sudan, 1976-Report Of A Who International Study Team. *Bulletin of the World Health Organization*, *56*(2), 247-270.

Diallo, B., Sissoko, D., Loman, N. J., Bah, H. A., Bah, H., Worrell, M. C., Conde, L. S., Sacko, R., Mesfin, S., Loua, A., Kalonda, J. K., Erondu, N. A., Dahl, B. A., Handrick, S., Goodfellow, I., Meredith, L. W., Cotten, M., Jah, U., Guetiya Wadoum, R. E., Rollin, P., Magassouba, N., Malvy, D., Anglaret, X., Carroll, M. W., Aylward, R. B., Djingarey, M. H., Diarra, A., Formenty, P., Keita, S., Gunther, S., Rambaut, A., & Duraffour, S. (2016). 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*, *63*(10), 1353-1356. https://doi.org/10.1093/cid/ciw601

Diallo, M. S. K., Rabilloud, M., Ayouba, A., Touré, A., Thaurignac, G., Keita, A. K., Butel, C., Kpamou, C., Barry, T. A., Sall, M. D., Camara, I., Leroy, S., Msellati, P., Ecochard, R., Peeters, M., Sow, M. S., Delaporte, E., Etard, J.-F., Aboubacar, D., Balde, A., Balde, I., Bamba, A., Camara, A., Conte, A. M., Delfraissy, J.-F., Diallo, A. B., Doumbouya, S., Kamano, E. S., Koivogui, J. B., Laniece-Delaunay, C., Levy, Y., Monemou, J. L., Povogui, M., Sakouvogui, M., Soumah, A. K., Subtil, F., Sylla, A. H., Taverne, B., & Yazdanpanah, Y. (2019). Prevalence of infection among asymptomatic and paucisymptomatic contact persons exposed to Ebola virus in Guinea: a retrospective, cross-sectional observational study. *The Lancet Infectious Diseases*, *19*(3), 308-316. https://doi.org/10.1016/S1473-3099(18)30649-2

Do, L. A. H., Pellet, J., van Doorn, H. R., Tran, A. T., Nguyen, B. H., Tran, T. T. L., Tran, Q. H., Vo, Q. B., Tran Dac, N. A., Trinh, H. N., Nguyen, T. T. H., Le Binh, B. T., Nguyen, H. M. K., Nguyen, M. T., Thai, Q. T., Vo, T. V., Ngo, N. Q. M., Dang, T. K. H., Cao, N. H., Tran, T. V., Ho, L. V., De Meulder, B., Auffray, C., Hofstra, J. J., Farrar, J., Bryant, J. E., de Jong, M., & Hibberd, M. L. (2017). Host Transcription Profile in Nasal Epithelium and Whole Blood of Hospitalized Children Under 2 Years of Age With Respiratory Syncytial Virus Infection. *The Journal of infectious diseases*, *217*(1), 134-146. https://doi.org/10.1093/infdis/jix519

Dovih, P., Laing, E. D., Chen, Y., Low, D. H. W., Ansil, B. R., Yang, X., Shi, Z., Broder, C. C., Smith, G. J. D., Linster, M., Ramakrishnan, U., & Mendenhall, I. H. (2019). Filovirus-reactive antibodies in humans and bats in Northeast

India imply zoonotic spillover. *PLoS Negl Trop Dis*, *13*(10), e0007733. https://doi.org/10.1371/journal.pntd.0007733

Dowell, S. F., Mukunu, R., Ksiazek, T. G., Khan, A. S., Rollin, P. E., & Peters, C. J. (1999). Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidemies a Kikwit. *The Journal of infectious diseases*, *179 Suppl 1*, S87-91. https://doi.org/10.1086/514284

Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., Park, D. J., Ladner, J. T., Arias, A., Asogun, D., Bielejec, F., Caddy, S. L., Cotten, M., D'Ambrozio, J., Dellicour, S., Di Caro, A., Diclaro, J. W., Duraffour, S., Elmore, M. J., Fakoli, L. S., Faye, O., Gilbert, M. L., Gevao, S. M., Gire, S., Gladden-Young, A., Gnirke, A., Goba, A., Grant, D. S., Haagmans, B. L., Hiscox, J. A., Jah, U., Kugelman, J. R., Liu, D., Lu, J., Malboeuf, C. M., Mate, S., Matthews, D. A., Matranga, C. B., Meredith, L. W., Qu, J., Quick, J., Pas, S. D., Phan, M. V. T., Pollakis, G., Reusken, C. B., Sanchez-Lockhart, M., Schaffner, S. F., Schieffelin, J. S., Sealfon, R. S., Simon-Loriere, E., Smits, S. L., Stoecker, K., Thorne, L., Tobin, E. A., Vandi, M. A., Watson, S. J., West, K., Whitmer, S., Wiley, M. R., Winnicki, S. M., Wohl, S., Wölfel, R., Yozwiak, N. L., Andersen, K. G., Blyden, S. O., Bolay, F., Carroll, M. W., Dahn, B., Diallo, B., Formenty, P., Fraser, C., Gao, G. F., Garry, R. F., Goodfellow, I., Günther, S., Happi, C. T., Holmes, E. C., Kargbo, B., Keïta, S., Kellam, P., Koopmans, M. P. G., Kuhn, J. H., Loman, N. J., Magassouba, N. F., Naidoo, D., Nichol, S. T., Nyenswah, T., Palacios, G., Pybus, O. G., Sabeti, P. C., Sall, A., Ströher, U., Wurie, I., Suchard, M. A., Lemey, P., & Rambaut, A. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, *544*(7650), 309-315. https://doi.org/10.1038/nature22040

Dunning, J., Sahr, F., Rojek, A., Gannon, F., Carson, G., Idriss, B., Massaquoi, T., Gandi, R., Joseph, S., Osman, H. K., Brooks, T. J., Simpson, A. J., Goodfellow, I., Thorne, L., Arias, A., Merson, L., Castle, L., Howell-Jones, R., Pardinaz-Solis, R., Hope-Gill, B., Ferri, M., Grove, J., Kowalski, M., Stepniewska, K., Lang, T., Whitehead, J., Olliaro, P., Samai, M., Horby, P. W., & team, R.-T. t. (2016). Experimental Treatment of Ebola Virus Disease with TKM-130803: A Single-Arm Phase 2 Clinical Trial. *PLoS Med*, *13*(4), e1001997. https://doi.org/10.1371/journal.pmed.1001997

Durfee, L. A., Lyon, N., Seo, K., & Huibregtse, J. M. (2010). The ISG15 conjugation system broadly targets newly synthesized proteins: implications for the antiviral function of ISG15. *Mol Cell*, *38*(5), 722-732. https://doi.org/10.1016/j.molcel.2010.05.002

Farrell, A., Wang, G., Rush, S. A., Martin, J. A., Belant, J. L., Butler, A. B., & Godwin, D. (2019). Machine learning of large-scale spatial distributions of wild turkeys with high-dimensional environmental data. *Ecol Evol*, *9*(10), 5938-5949. https://doi.org/10.1002/ece3.5177

Faye, O., Andronico, A., Faye, O., Salje, H., Boëlle, P.-Y., Magassouba, N. F., Bah, E. I., Koivogui, L., Diallo, B., Diallo, A. A., Keita, S., Konde, M. K., Fowler, R., Fall, G., Cauchemez, S., & Sall, A. A. (2015). Use of Viremia to Evaluate the Baseline Case Fatality Ratio of Ebola Virus Disease and Inform Treatment Studies: A Retrospective Cohort Study. *PLOS Medicine*, *12*(12), e1001908. https://doi.org/10.1371/journal.pmed.1001908

Feldmann, H., Nichol, S. T., Klenk, H. D., Peters, C. J., & Sanchez, A. (1994). Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein. *Virology*, *199*(2), 469-473. https://doi.org/10.1006/viro.1994.1147

Fitzpatrick, G., Vogt, F., Moi Gbabai, O. B., Decroo, T., Keane, M., De Clerck, H., Grolla, A., Brechard, R., Stinson, K., & Van Herp, M. (2015). 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*, *212*(11), 1752-1758. https://doi.org/10.1093/infdis/jiv304

Fjaerli, H. O., Bukholm, G., Krog, A., Skjaeret, C., Holden, M., & Nakstad, B. (2006). Whole blood gene expression in infants with respiratory syncytial virus bronchiolitis. *BMC Infect Dis*, *6*, 175. https://doi.org/10.1186/1471-2334-6-175

Forbes, K. M., Webala, P. W., Jaaskelainen, A. J., Abdurahman, S., Ogola, J., Masika, M. M., Kivisto, I., Alburkat, H., Plyusnin, I., Levanov, L., Korhonen, E. M., Huhtamo, E., Mwaengo, D., Smura, T., Mirazimi, A., Anzala, O., Vapalahti, O., & Sironen, T. (2019). Bombali Virus in Mops condylurus Bat, Kenya. *Emerg Infect Dis*, *25*(5). https://doi.org/10.3201/eid2505.181666

Formenty, P., Leroy, E. M., Epelboin, A., Libama, F., Lenzi, M., Sudeck, H., Yaba, P., Allarangar, Y., Boumandouki, P., Nkounkou, V. B., Drosten, C., Grolla, A., Feldmann, H., & Roth, C. (2006). Detection of Ebola virus in oral fluid specimens during outbreaks of Ebola virus hemorrhagic fever in the Republic of Congo. *Clin Infect Dis*, *42*(11), 1521-1526. https://doi.org/10.1086/503836

Franks, S. E., Getahun, A., Hogarth, P. M., & Cambier, J. C. (2016). Targeting B cells in treatment of autoimmunity. *Current opinion in immunology*, *43*, 39-45. https://doi.org/10.1016/j.coi.2016.09.003

Fronhoffs, S., Totzke, G., Stier, S., Wernert, N., Rothe, M., Brüning, T., Koch, B., Sachinidis, A., Vetter, H., & Ko, Y. (2002). A method for the rapid construction of cRNA standard curves in quantitative real-time reverse transcription polymerase chain reaction. *Molecular and Cellular Probes*, *16*(2), 99-110. https://doi.org/https://doi.org/10.1006/mcpr.2002.0405

Furuse, Y., Fallah, M., Oshitani, H., Kituyi, L., Mahmoud, N., Musa, E., Gasasira, A., Nyenswah, T., Dahn, B., & Bawo, L. (2017). Analysis of patient data from laboratories during the Ebola virus disease outbreak in Liberia, April 2014 to March 2015. *PLOS Neglected Tropical Diseases*, *11*(7), e0005804. https://doi.org/10.1371/journal.pntd.0005804

Garamszegi, S., Yen, J. Y., Honko, A. N., Geisbert, J. B., Rubins, K. H., Geisbert, T. W., Xia, Y., Hensley, L. E., & Connor, J. H. (2014). Transcriptional correlates of disease outcome in anticoagulant-treated non-human primates infected with ebolavirus. *PLoS Negl Trop Dis*, *8*(7), e3061. https://doi.org/10.1371/journal.pntd.0003061

Geisbert, T. W., Hensley, L. E., Larsen, T., Young, H. A., Reed, D. S., Geisbert, J. B., Scott, D. P., Kagan, E., Jahrling, P. B., & Davis, K. J. (2003). Pathogenesis of Ebola hemorrhagic fever in cynomolgus macaques: evidence that dendritic cells are early and sustained targets of infection. *Am J Pathol*, *163*(6), 2347-2370. https://doi.org/10.1016/S0002-9440(10)63591-2

Georges, A. J., Leroy, E. M., Renaut, A. A., Benissan, C. T., Nabias, R. J., Ngoc, M. T., Obiang, P. I., Lepage, J. P., Bertherat, E. J., Benoni, D. D., Wickings, E. J.,

Amblard, J. P., Lansoud-Soukate, J. M., Milleliri, J. M., Baize, S., & Georges-Courbot, M. C. (1999). Ebola hemorrhagic fever outbreaks in Gabon, 1994-1997: epidemiologic and health control issues. *The Journal of infectious diseases*, *179 Suppl 1*, S65-75. https://doi.org/10.1086/514290

Ghani, A. C., Burgess, D. H., Reynolds, A., & Rousseau, C. (2015). Expanding the role of diagnostic and prognostic tools for infectious diseases in resource-poor settings. *Nature*, *528*(7580), S50-52. https://doi.org/10.1038/nature16038

Giesecke, J. (2014). Primary and index cases. *The Lancet*, *384*(9959), 2024. https://doi.org/10.1016/S0140-6736(14)62331-X

Glynn, J. R., Bower, H., Johnson, S., Houlihan, C. F., Montesano, C., Scott, J. T., Semple, M. G., Bangura, M. S., Kamara, A. J., Kamara, O., Mansaray, S. H., Sesay, D., Turay, C., Dicks, S., Wadoum, R. E. G., Colizzi, V., Checchi, F., Samuel, D., & Tedder, R. S. (2017). 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. *The Lancet Infectious Diseases*, *17*(6), 645-653. https://doi.org/10.1016/S1473-3099(17)30111-1

Goldstein, T., Anthony, S. J., Gbakima, A., Bird, B. H., Bangura, J., Tremeau-Bravard, A., Belaganahalli, M. N., Wells, H. L., Dhanota, J. K., Liang, E., Grodus, M., Jangra, R. K., DeJesus, V. A., Lasso, G., Smith, B. R., Jambai, A., Kamara, B. O., Kamara, S., Bangura, W., Monagin, C., Shapira, S., Johnson, C. K., Saylors, K., Rubin, E. M., Chandran, K., Lipkin, W. I., & Mazet, J. A. K. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nat Microbiol*, *3*(10), 1084-1089. https://doi.org/10.1038/s41564-018-0227-2

Gong, D., Shi, W., Yi, S. J., Chen, H., Groffen, J., & Heisterkamp, N. (2012). TGFbeta signaling plays a critical role in promoting alternative macrophage activation. *BMC Immunol*, *13*, 31. https://doi.org/10.1186/1471-2172-13-31

Greenberg, A., Huber, B. R., Liu, D. X., Logue, J. P., Hischak, A. M. W., Hart, R. J., Abbott, M., Isic, N., Hisada, Y. M., Mackman, N., Bennett, R. S., Hensley, L. E., Connor, J. H., & Crossland, N. A. (2020). Quantification of Viral and Host Biomarkers in the Liver of Rhesus Macaques: A Longitudinal Study of Zaire Ebolavirus Strain Kikwit (EBOV/Kik). *Am J Pathol*, *190*(7), 1449-1460. https://doi.org/10.1016/j.ajpath.2020.03.003

Grolla, A., Jones, S., Kobinger, G., Sprecher, A., Girard, G., Yao, M., Roth, C., Artsob, H., Feldmann, H., & Strong, J. E. (2012). Flexibility of mobile laboratory unit in support of patient management during the 2007 Ebola-Zaire outbreak in the Democratic Republic of Congo. *Zoonoses Public Health*, *59 Suppl 2*, 151-157. https://doi.org/10.1111/j.1863-2378.2012.01477.x

Grolla, A., Lucht, A., Dick, D., Strong, J. E., & Feldmann, H. (2005). Laboratory diagnosis of Ebola and Marburg hemorrhagic fever. *Bull Soc Pathol Exot*, *98*(3), 205-209. https://www.ncbi.nlm.nih.gov/pubmed/16267962

Gupta, S., McCann, L., Chan, Y. G. Y., Lai, E. W., Wei, W., Wong, P. F., Smithy, J. W., Weidler, J., Rhees, B., Bates, M., Kluger, H. M., & Rimm, D. L. (2019). Closed system RT-qPCR as a potential companion diagnostic test for immunotherapy outcome in metastatic melanoma. *Journal for ImmunoTherapy of Cancer*, *7*(1), 254. https://doi.org/10.1186/s40425-019-0731-9

Haaskjold, Y. L., Bolkan, H. A., Krogh, K. O., Jongopi, J., Lundeby, K. M., Mellesmo, S., Garces, P. S., Josendal, O., Opstad, A., Svensen, E., Fuentes, L. M., Kamara, A. S., Riera, M., Arranz, J., Roberts, D. P., Stamper, P. D., Austin, P.,

Moosa, A. J., Marke, D., Hassan, S., Eide, G. E., Berg, A., & Blomberg, B. (2016). Clinical Features of and Risk Factors for Fatal Ebola Virus Disease, Moyamba District, Sierra Leone, December 2014-February 2015. *Emerg Infect Dis*, *22*(9), 1537-1544. https://doi.org/10.3201/eid2209.151621

Halpin, K., Hyatt, A. D., Fogarty, R., Middleton, D., Bingham, J., Epstein, J. H., Rahman, S. A., Hughes, T., Smith, C., Field, H. E., Daszak, P., & Henipavirus Ecology Research, G. (2011). Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. *The American Journal of Tropical Medicine and Hygiene*, *85*(5), 946-951. https://doi.org/10.4269/ajtmh.2011.10-0567

Hamblion, E. L., Raftery, P., Wendland, A., Dweh, E., Williams, G. S., George, R. N. C., Soro, L., Katawera, V., Clement, P., Gasasira, A. N., Musa, E., & Nagbe, T. K. (2018). The challenges of detecting and responding to a Lassa fever outbreak in an Ebola-affected setting. *Int J Infect Dis*, *66*, 65-73. https://doi.org/10.1016/j.ijid.2017.11.007

Hartley, M.-A., Young, A., Tran, A.-M., Okoni-Williams, H. H., Suma, M., Mancuso, B., Al-Dikhari, A., & Faouzi, M. (2017). Predicting Ebola Severity: A Clinical Prioritization Score for Ebola Virus Disease. *PLOS Neglected Tropical Diseases*, *11*(2), e0005265. https://doi.org/10.1371/journal.pntd.0005265

Hayes, C. G., Burans, J. P., Ksiazek, T. G., Del Rosario, R. A., Miranda, M. E. G., Manaloto, C. R., Barrientos, A. B., Robles, C. G., Dayrit, M. M., & Peters, C. J. (1992). Outbreak of Fatal Illness among Captive Macaques in the Philippines Caused by an Ebola-Related Filovirus. *The American Journal of Tropical Medicine and Hygiene*, *46*(6), 664-671. https://doi.org/https://doi.org/10.4269/ajtmh.1992.46.664

Henao-Restrepo, A. M., Camacho, A., Longini, I. M., Watson, C. H., Edmunds, W. J., Egger, M., Carroll, M. W., Dean, N. E., Diatta, I., Doumbia, M., Draguez, B., Duraffour, S., Enwere, G., Grais, R., Gunther, S., Gsell, P. S., Hossmann, S., Watle, S. V., Konde, M. K., Keita, S., Kone, S., Kuisma, E., Levine, M. M., Mandal, S., Mauget, T., Norheim, G., Riveros, X., Soumah, A., Trelle, S., Vicari, A. S., Rottingen, J. A., & Kieny, M. P. (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 Ca Suffit!). *Lancet*, *389*(10068), 505-518. https://doi.org/10.1016/S0140-6736(16)32621-6

Hensley, L. E., Young, H. A., Jahrling, P. B., & Geisbert, T. W. (2002). Proinflammatory response during Ebola virus infection of primate models: possible involvement of the tumor necrosis factor receptor superfamily. *Immunol Lett*, *80*(3), 169-179. https://doi.org/10.1016/s0165-2478(01)00327-3

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L., & Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods*, *10*(10), 1003-1005. https://doi.org/10.1038/nmeth.2633

Hoenen, T., Groseth, A., de Kok-Mercado, F., Kuhn, J. H., & Wahl-Jensen, V. (2011). Minigenomes, transcription and replication competent virus-like particles and beyond: reverse genetics systems for filoviruses and other negative stranded hemorrhagic fever viruses. *Antiviral Res*, *91*(2), 195-208. https://doi.org/10.1016/j.antiviral.2011.06.003

Hoenen, T., Groseth, A., & Feldmann, H. (2019). Therapeutic strategies to target the Ebola virus life cycle. *Nat Rev Microbiol*, *17*(10), 593-606. https://doi.org/10.1038/s41579-019-0233-2

Hotez, P. J. (2015). Neglected Tropical Diseases in the Ebola-Affected Countries of West Africa. *PLoS Negl Trop Dis*, *9*(6), e0003671. https://doi.org/10.1371/journal.pntd.0003671

Hsu, C. H., Champaloux, S. W., Keïta, S., Martel, L., Bilivogui, P., Knust, B., & McCollum, A. M. (2018). Sensitivity and Specificity of Suspected Case Definition Used during West Africa Ebola Epidemic. *Emerging infectious diseases*, *24*(1), 9-14. https://doi.org/10.3201/eid2401.161678

Hu, C.-A., Chen, C.-M., Fang, Y.-C., Liang, S.-J., Wang, H.-C., Fang, W.-F., Sheu, C.-C., Perng, W.-C., Yang, K.-Y., Kao, K.-C., Wu, C.-L., Tsai, C.-S., Lin, M.-Y., & Chao, W.-C. (2020). Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open*, *10*(2), e033898. https://doi.org/10.1136/bmjopen-2019-033898

Huang, J., Zhang, J., Li, H., Lu, Z., Shan, W., Mercado-Uribe, I., & Liu, J. (2013). VCAM1 expression correlated with tumorigenesis and poor prognosis in high grade serous ovarian cancer. *Am J Transl Res*, *5*(3), 336-346. https://www.ncbi.nlm.nih.gov/pubmed/23634244

Huizenga, E., van der Ende, J., Zwinkels, N., Jimissa, A., van der Ende-Bouwman, C., van Rooijen, R., Kargbo, B., Agnandji, S. T., Hanscheid, T., Goorhuis, A., & Grobusch, M. P. (2019). A Modified Case Definition to Facilitate Essential Hospital Care During Ebola Outbreaks. *Clin Infect Dis*, *68*(10), 1763-1768. https://doi.org/10.1093/cid/ciy798

Hume, A. J., & Mühlberger, E. (2019). Distinct Genome Replication and Transcription Strategies within the Growing Filovirus Family. *Journal of Molecular Biology*, *431*(21), 4290-4320. https://doi.org/https://doi.org/10.1016/j.jmb.2019.06.029

Hunt, L., Gupta-Wright, A., Simms, V., Tamba, F., Knott, V., Tamba, K., Heisenberg-Mansaray, S., Tamba, E., Sheriff, A., Conteh, S., Smith, T., Tobin, S., Brooks, T., Houlihan, C., Cummings, R., & Fletcher, T. (2015). Clinical presentation, biochemical, and haematological parameters and their association with outcome in patients with Ebola virus disease: an observational cohort study. *The Lancet Infectious Diseases*, *15*(11), 1292-1299. https://doi.org/10.1016/S1473-3099(15)00144-9

Ilunga Kalenga, O., Moeti, M., Sparrow, A., Nguyen, V. K., Lucey, D., & Ghebreyesus, T. A. (2019). The Ongoing Ebola Epidemic in the Democratic Republic of Congo, 2018-2019. *N Engl J Med*, *381*(4), 373-383. https://doi.org/10.1056/NEJMsr1904253

Iversen, P. L., Kane, C. D., Zeng, X., Panchal, R. G., Warren, T. K., Radoshitzky, S. R., Kuhn, J. H., Mudhasani, R. R., Cooper, C. L., Shurtleff, A. C., Nasar, F., Sunay, M. M. E., Duplantier, A. J., Eaton, B. P., Zumbrun, E. E., Bixler, S. L., Martin, S., Meinig, J. M., Chiang, C.-Y., Sanchez-Lockhart, M., Palacios, G. F., Kugelman, J. R., Martins, K. A., Pitt, M. L., Crozier, I., & Saunders, D. L. (2020). Recent successes in therapeutics for Ebola virus disease: no time for complacency. *The Lancet Infectious Diseases*, *20*(9), e231-e237. https://doi.org/10.1016/S1473-3099(20)30282-6

Jacob, S. T., Crozier, I., Fischer, W. A., 2nd, Hewlett, A., Kraft, C. S., Vega, M. A., Soka, M. J., Wahl, V., Griffiths, A., Bollinger, L., & Kuhn, J. H. (2020). Ebola

virus disease. *Nat Rev Dis Primers*, *6*(1), 13. https://doi.org/10.1038/s41572-020-0147-3

Jahrling, P. B., Geisbert, T. W., Johnson, E. D., Peters, C. J., Dalgard, D. W., & Hall, W. C. (1990). Preliminary report: isolation of Ebola virus from monkeys imported to USA. *The Lancet*, *335*(8688), 502-505. https://doi.org/https://doi.org/10.1016/0140-6736(90)90737-P

Jankeel, A., Menicucci, A. R., Woolsey, C., Fenton, K. A., Mendoza, N., Versteeg, K., Cross, R. W., Geisbert, T. W., & Messaoudi, I. (2020). Early transcriptional changes within liver, adrenal gland and lymphoid tissues significantly contribute to Ebola pathogenesis in cynomolgus macaques. *Journal of Virology*, JVI.00250-00220. https://doi.org/10.1128/JVI.00250-20

Jasenosky, L. D., Neumann, G., Lukashevich, I., & Kawaoka, Y. (2001). Ebola virus VP40-induced particle formation and association with the lipid bilayer. *J Virol*, *75*(11), 5205-5214. https://doi.org/10.1128/JVI.75.11.5205-5214.2001

Jayme, S. I., Field, H. E., de Jong, C., Olival, K. J., Marsh, G., Tagtag, A. M., Hughes, T., Bucad, A. C., Barr, J., Azul, R. R., Retes, L. M., Foord, A., Yu, M., Cruz, M. S., Santos, I. J., Lim, T. M., Benigno, C. C., Epstein, J. H., Wang, L. F., Daszak, P., & Newman, S. H. (2015). Molecular evidence of Ebola Reston virus infection in Philippine bats. *Virol J*, *12*, 107. https://doi.org/10.1186/s12985-015-0331-3

Johnson, G., Nolan, T., & Bustin, S. A. (2013). Real-Time Quantitative PCR, Pathogen Detection and MIQE. In M. Wilks (Ed.), *PCR Detection of Microbial Pathogens* (pp. 1-16). Humana Press. https://doi.org/10.1007/978-1-60327-353-4_1

Jong, V. L., Ahout, I. M., van den Ham, H. J., Jans, J., Zaaraoui-Boutahar, F., Zomer, A., Simonetti, E., Bijl, M. A., Brand, H. K., van, I. W. F., de Jonge, M. I., Fraaij, P. L., de Groot, R., Osterhaus, A. D., Eijkemans, M. J., Ferwerda, G., & Andeweg, A. C. (2016). Transcriptome assists prognosis of disease severity in respiratory syncytial virus infected infants. *Sci Rep*, *6*, 36603. https://doi.org/10.1038/srep36603

Kajihara, M., Hang'ombe, B. M., Changula, K., Harima, H., Isono, M., Okuya, K., Yoshida, R., Mori-Kajihara, A., Eto, Y., Orba, Y., Ogawa, H., Qiu, Y., Sawa, H., Simulundu, E., Mwizabi, D., Munyeme, M., Squarre, D., Mukonka, V., Mweene, A., & Takada, A. (2019). Marburgvirus in Egyptian Fruit Bats, Zambia. *Emerg Infect Dis*, *25*(8), 1577-1580. https://doi.org/10.3201/eid2508.190268

Kaletsky, R. L., Francica, J. R., Agrawal-Gamse, C., & Bates, P. (2009). Tetherin-mediated restriction of filovirus budding is antagonized by the Ebola glycoprotein. *Proceedings of the National Academy of Sciences*, *106*, 2886-2891.

Keita, A. K., Vidal, N., Toure, A., Diallo, M. S. K., Magassouba, N., Baize, S., Mateo, M., Raoul, H., Mely, S., Subtil, F., Kpamou, C., Koivogui, L., Traore, F., Sow, M. S., Ayouba, A., Etard, J. F., Delaporte, E., Peeters, M., & PostEbogui Study, G. (2019). A 40-Month Follow-Up of Ebola Virus Disease Survivors in Guinea (PostEbogui) Reveals Long-Term Detection of Ebola Viral Ribonucleic Acid in Semen and Breast Milk. *Open Forum Infect Dis*, *6*(12), ofz482. https://doi.org/10.1093/ofid/ofz482

Kelly-Cirino, C. D., Nkengasong, J., Kettler, H., Tongio, I., Gay-Andrieu, F., Escadafal, C., Piot, P., Peeling, R. W., Gadde, R., & Boehme, C. (2019).

Importance of diagnostics in epidemic and pandemic preparedness. *BMJ Glob Health*, *4*(Suppl 2), e001179. https://doi.org/10.1136/bmjgh-2018-001179

Kemenesi, G., Kurucz, K., Dallos, B., Zana, B., Foldes, F., Boldogh, S., Gorfol, T., Carroll, M. W., & Jakab, F. (2018). Re-emergence of Lloviu virus in Miniopterus schreibersii bats, Hungary, 2016. *Emerg Microbes Infect*, *7*(1), 66. https://doi.org/10.1038/s41426-018-0067-4

Kennedy, S. B., Bolay, F., Kieh, M., Grandits, G., Badio, M., Ballou, R., Eckes, R., Feinberg, M., Follmann, D., Grund, B., Gupta, S., Hensley, L., Higgs, E., Janosko, K., Johnson, M., Kateh, F., Logue, J., Marchand, J., Monath, T., Nason, M., Nyenswah, T., Roman, F., Stavale, E., Wolfson, J., Neaton, J. D., Lane, H. C., & Group, P. I. S. (2017). Phase 2 Placebo-Controlled Trial of Two Vaccines to Prevent Ebola in Liberia. *N Engl J Med*, *377*(15), 1438-1447. https://doi.org/10.1056/NEJMoa1614067

Kerber, R., Krumkamp, R., Diallo, B., Jaeger, A., Rudolf, M., Lanini, S., Bore, J. A., Koundouno, F. R., Becker-Ziaja, B., Fleischmann, E., Stoecker, K., Meschi, S., Mély, S., Newman, E. N. C., Carletti, F., Portmann, J., Korva, M., Wolff, S., Molkenthin, P., Kis, Z., Kelterbaum, A., Bocquin, A., Strecker, T., Fizet, A., Castilletti, C., Schudt, G., Ottowell, L., Kurth, A., Atkinson, B., Badusche, M., Cannas, A., Pallasch, E., Bosworth, A., Yue, C., Pályi, B., Ellerbrok, H., Kohl, C., Oestereich, L., Logue, C. H., Lüdtke, A., Richter, M., Ngabo, D., Borremans, B., Becker, D., Gryseels, S., Abdellati, S., Vermoesen, T., Kuisma, E., Kraus, A., Liedigk, B., Maes, P., Thom, R., Duraffour, S., Diederich, S., Hinzmann, J., Afrough, B., Repits, J., Mertens, M., Vitoriano, I., Bah, A., Sachse, A., Boettcher, J. P., Wurr, S., Bockholt, S., Nitsche, A., Županc, T. A., Strasser, M., Ippolito, G., Becker, S., Raoul, H., Carroll, M. W., De Clerck, H., Van Herp, M., Sprecher, A., Koivogui, L., Magassouba, N. F., Keïta, S., Drury, P., Gurry, C., Formenty, P., May, J., Gabriel, M., Wölfel, R., Günther, S., & Di Caro, A. (2016). Analysis of Diagnostic Findings From the European Mobile Laboratory in Guéckédou, Guinea, March 2014 Through March 2015. *The Journal of infectious diseases*, *214*(suppl 3), S250-S257. https://doi.org/10.1093/infdis/jiw269

Kerber, R., Krumkamp, R., Korva, M., Rieger, T., Wurr, S., Duraffour, S., Oestereich, L., Gabriel, M., Sissoko, D., Anglaret, X., Malvy, D., May, J., Zupanc, T. A., Munoz-Fontela, C., & Gunther, S. (2018). Kinetics of Soluble Mediators of the Host Response in Ebola Virus Disease. *The Journal of infectious diseases*, *218*(suppl_5), S496-S503. https://doi.org/10.1093/infdis/jiy429

Kerber, R., Lorenz, E., Duraffour, S., Sissoko, D., Rudolf, M., Jaeger, A., Cisse, S. D., Camara, A. M., Miranda, O., Castro, C. M., Akoi Bore, J., Raymond Koundouno, F., Repits, J., Afrough, B., Becker-Ziaja, B., Hinzmann, J., Mertens, M., Vitoriano, I., Hugh Logue, C., Bottcher, J. P., Pallasch, E., Sachse, A., Bah, A., Cabeza-Cabrerizo, M., Nitzsche, K., Kuisma, E., Michel, J., Holm, T., Zekeng, E. G., Cowley, L. A., Garcia-Dorival, I., Hetzelt, N., Baum, J. H. J., Portmann, J., Carter, L., Yenamaberhan, R. L., Camino, A., Enkirch, T., Singethan, K., Meisel, S., Mazzarelli, A., Kosgei, A., Kafetzopoulou, L., Rickett, N. Y., Patrono, L. V., Ghebreghiorghis, L., Arnold, U., Colin, G., Juchet, S., Marchal, C. L., Kolie, J. S., Beavogui, A. H., Wurr, S., Bockholt, S., Krumkamp, R., May, J., Stoecker, K., Fleischmann, E., Ippolito, G., Carroll, M. W., Koivogui, L., Magassouba, N., Keita, S., Gurry, C., Drury, P., Diallo, B., Formenty, P., Wolfel, R., Di Caro, A., Gabriel, M., Anglaret, X., Malvy, D., & Gunther, S. (2019). Laboratory Findings,

Compassionate Use of Favipiravir, and Outcome in Patients With Ebola Virus Disease, Guinea, 2015-A Retrospective Observational Study. *The Journal of infectious diseases*, *220*(2), 195-202. https://doi.org/10.1093/infdis/jiz078

Khan, A. S., Tshioko, F. K., Heymann, D. L., Le Guenno, B., Nabeth, P., Kerstiens, B., Fleerackers, Y., Kilmarx, P. H., Rodier, G. R., Nkuku, O., Rollin, P. E., Sanchez, A., Zaki, S. R., Swanepoel, R., Tomori, O., Nichol, S. T., Peters, C. J., Muyembe-Tamfum, J. J., & Ksiazek, T. G. (1999). The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidemies a Kikwit. *The Journal of infectious diseases*, *179 Suppl 1*, S76-86. https://doi.org/10.1086/514306

Kiiza, P., Mullin, S., Teo, K., Adhikari, N. K. J., & Fowler, R. A. (2020). Treatment of Ebola-related critical illness. *Intensive care medicine*, *46*(2), 285-297. https://doi.org/10.1007/s00134-020-05949-z

Kondratowicz, A. S., Lennemann, N. J., Sinn, P. L., Davey, R. A., Hunt, C. L., Moller-Tank, S., Meyerholz, D. K., Rennert, P., Mullins, R. F., Brindley, M., Sandersfeld, L. M., Quinn, K., Weller, M., McCray, P. B., Chiorini, J., & Maury, W. (2011). T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire Ebolavirus and <em>Lake Victoria Marburgvirus</em&gt. *Proceedings of the National Academy of Sciences*, *108*(20), 8426. https://doi.org/10.1073/pnas.1019030108

Koraka, P., Murgue, B., Deparis, X., Van Gorp, E. C., Setiati, T. E., Osterhaus, A. D., & Groen, J. (2004). Elevation of soluble VCAM-1 plasma levels in children with acute dengue virus infection of varying severity. *J Med Virol*, *72*(3), 445-450. https://doi.org/10.1002/jmv.20007

Kovats, S. (2015). Estrogen receptors regulate innate immune cells and signaling pathways. *Cell Immunol*, *294*(2), 63-69. https://doi.org/10.1016/j.cellimm.2015.01.018

Kralik, P., & Ricchi, M. (2017). A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything [Review]. *8*(108). https://doi.org/10.3389/fmicb.2017.00108

Kuhn, J. H., Adachi, T., Adhikari, N. K. J., Arribas, J. R., Bah, I. E., Bausch, D. G., Bhadelia, N., Borchert, M., Brantsæter, A. B., Brett-Major, D. M., Burgess, T. H., Chertow, D. S., Chute, C. G., Cieslak, T. J., Colebunders, R., Crozier, I., Davey, R. T., de Clerck, H., Delgado, R., Evans, L., Fallah, M., Fischer, W. A., Fletcher, T. E., Fowler, R. A., Grünewald, T., Hall, A., Hewlett, A., Hoepelman, A. I. M., Houlihan, C. F., Ippolito, G., Jacob, S. T., Jacobs, M., Jakob, R., Jacquerioz, F. A., Kaiser, L., Kalil, A. C., Kamara, R. F., Kapetshi, J., Klenk, H.-D., Kobinger, G., Kortepeter, M. G., Kraft, C. S., Kratz, T., Bosa, H. S. K., Lado, M., Lamontagne, F., Lane, H. C., Lobel, L., Lutwama, J., Lyon, G. M., Massaquoi, M. B. F., Massaquoi, T. A., Mehta, A. K., Makuma, V. M., Murthy, S., Musoke, T. S., Muyembe-Tamfum, J.-J., Nakyeyune, P., Nanclares, C., Nanyunja, M., Nsio-Mbeta, J., O'Dempsey, T., Pawęska, J. T., Peters, C. J., Piot, P., Rapp, C., Renaud, B., Ribner, B., Sabeti, P. C., Schieffelin, J. S., Slenczka, W., Soka, M. J., Sprecher, A., Strong, J., Swanepoel, R., Uyeki, T. M., van Herp, M., Vetter, P., Wohl, D. A., Wolf, T., Wolz, A., Wurie, A. H., & Yoti, Z. (2019). New filovirus disease classification and nomenclature. *Nature Reviews Microbiology*, *17*(5), 261-263. https://doi.org/10.1038/s41579-019-0187-4

Kuhn, J. H., Adkins, S., Alioto, D., Alkhovsky, S. V., Amarasinghe, G. K., Anthony, S. J., Avšič-Županc, T., Ayllón, M. A., Bahl, J., Balkema-Buschmann, A.,

Ballinger, M. J., Bartonička, T., Basler, C., Bavari, S., Beer, M., Bente, D. A., Bergeron, É., Bird, B. H., Blair, C., Blasdell, K. R., Bradfute, S. B., Breyta, R., Briese, T., Brown, P. A., Buchholz, U. J., Buchmeier, M. J., Bukreyev, A., Burt, F., Buzkan, N., Calisher, C. H., Cao, M., Casas, I., Chamberlain, J., Chandran, K., Charrel, R. N., Chen, B., Chiumenti, M., Choi, I.-R., Clegg, J. C. S., Crozier, I., da Graça, J. V., Dal Bó, E., Dávila, A. M. R., de la Torre, J. C., de Lamballerie, X., de Swart, R. L., Di Bello, P. L., Di Paola, N., Di Serio, F., Dietzgen, R. G., Digiaro, M., Dolja, V. V., Dolnik, O., Drebot, M. A., Drexler, J. F., Dürrwald, R., Dufkova, L., Dundon, W. G., Duprex, W. P., Dye, J. M., Easton, A. J., Ebihara, H., Elbeaino, T., Ergünay, K., Fernandes, J., Fooks, A. R., Formenty, P. B. H., Forth, L. F., Fouchier, R. A. M., Freitas-Astúa, J., Gago-Zachert, S., Gāo, G. F., García, M. L., García-Sastre, A., Garrison, A. R., Gbakima, A., Goldstein, T., Gonzalez, J.-P. J., Griffiths, A., Groschup, M. H., Günther, S., Guterres, A., Hall, R. A., Hammond, J., Hassan, M., Hepojoki, J., Hepojoki, S., Hetzel, U., Hewson, R., Hoffmann, B., Hongo, S., Höper, D., Horie, M., Hughes, H. R., Hyndman, T. H., Jambai, A., Jardim, R., Jiāng, D., Jin, Q., Jonson, G. B., Junglen, S., Karadağ, S., Keller, K. E., Klempa, B., Klingström, J., Kobinger, G., Kondō, H., Koonin, E. V., Krupovic, M., Kurath, G., Kuzmin, I. V., Laenen, L., Lamb, R. A., Lambert, A. J., Langevin, S. L., Lee, B., Lemos, E. R. S., Leroy, E. M., Li, D., Lǐ, J., Liang, M., Liú, W., Liú, Y., Lukashevich, I. S., Maes, P., Marciel de Souza, W., Marklewitz, M., Marshall, S. H., Martelli, G. P., Martin, R. R., Marzano, S.-Y. L., Massart, S., McCauley, J. W., Mielke-Ehret, N., Minafra, A., Minutolo, M., Mirazimi, A., Mühlbach, H.-P., Mühlberger, E., Naidu, R., Natsuaki, T., Navarro, B., Navarro, J. A., Netesov, S. V., Neumann, G., Nowotny, N., Nunes, M. R. T., Nylund, A., Økland, A. L., Oliveira, R. C., Palacios, G., Pallas, V., Pályi, B., Papa, A., Parrish, C. R., Pauvolid-Corrêa, A., Pawęska, J. T., Payne, S., Pérez, D. R., Pfaff, F., Radoshitzky, S. R., Rahman, A.-u., Ramos-González, P. L., Resende, R. O., Reyes, C. A., Rima, B. K., Romanowski, V., Robles Luna, G., Rota, P., Rubbenstroth, D., Runstadler, J. A., Ruzek, D., Sabanadzovic, S., Salát, J., Sall, A. A., Salvato, M. S., Sarpkaya, K., Sasaya, T., Schwemmle, M., Shabbir, M. Z., Shí, X., Shí, Z., Shirako, Y., Simmonds, P., Širmarová, J., Sironi, M., Smither, S., Smura, T., Song, J.-W., Spann, K. M., Spengler, J. R., Stenglein, M. D., Stone, D. M., Straková, P., Takada, A., Tesh, R. B., Thornburg, N. J., Tomonaga, K., Tordo, N., Towner, J. S., Turina, M., Tzanetakis, I., Ulrich, R. G., Vaira, A. M., van den Hoogen, B., Varsani, A., Vasilakis, N., Verbeek, M., Wahl, V., Walker, P. J., Wang, H., Wang, J., Wang, X., Wang, L.-F., Wèi, T., Wells, H., Whitfield, A. E., Williams, J. V., Wolf, Y. I., Wú, Z., Yang, X., Yáng, X., Yu, X., Yutin, N., Zerbini, F. M., Zhang, T., Zhang, Y.-Z., Zhou, G., & Zhou, X. (2020). 2020 taxonomic update for phylum Negarnaviricota (Riboviria: Orthornavirae), including the large orders Bunyavirales and Mononegavirales. *Archives of Virology*. https://doi.org/10.1007/s00705-020-04731-2

Kuhn, J. H., Becker, S., Ebihara, H., Geisbert, T. W., Johnson, K. M., Kawaoka, Y., Lipkin, W. I., Negredo, A. I., Netesov, S. V., Nichol, S. T., Palacios, G., Peters, C. J., Tenorio, A., Volchkov, V. E., & Jahrling, P. B. (2010). Proposal for a revised taxonomy of the family Filoviridae: classification, names of taxa and viruses, and virus abbreviations. *Archives of Virology*, *155*(12), 2083-2103. https://doi.org/10.1007/s00705-010-0814-x

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1-26. https://doi.org/http://dx.doi.org/10.18637/jss.v028.i05

Kukla, M., Zwirska-Korczala, K., Gabriel, A., Janczewska-Kazek, E., Berdowska, A., Wiczkowski, A., Rybus-Kalinowska, B., Kalinowski, M., Ziolkowski, A., Wozniak-Grygiel, E., Waluga, M., & Nowak, B. (2009). sPECAM-1 and sVCAM-1: Role in Pathogenesis and Diagnosis of Chronic Hepatitis C and Association with Response to Antiviral Therapy. *Therap Adv Gastroenterol*, *2*(2), 79-90. https://doi.org/10.1177/1756283X08100666

Kupferschmidt, K. (2019). This bat species may be the source of the Ebola epidemic that killed more than 11,000 people in West Africa. https://www.sciencemag.org/news/2019/01/bat-species-may-be-source-ebola-epidemic-killed-more-11000-people-west-africa

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *2010*, *36*(11), Journal of Statistical Software. https://doi.org/10.18637/jss.v036.i11

Laing, E. D., Mendenhall, I. H., Linster, M., Low, D. H. W., Chen, Y., Yan, L., Sterling, S. L., Borthwick, S., Neves, E. S., Lim, J. S. L., Skiles, M., Lee, B. P. Y., Wang, L. F., Broder, C. C., & Smith, G. J. D. (2018). Serologic Evidence of Fruit Bat Exposure to Filoviruses, Singapore, 2011-2016. *Emerg Infect Dis*, *24*(1), 114-117. https://doi.org/10.3201/eid2401.170401

Lanini, S., Portella, G., Vairo, F., Kobinger, G. P., Pesenti, A., Langer, M., Kabia, S., Brogiato, G., Amone, J., Castilletti, C., Miccio, R., Zumla, A., Capobianchi, M. R., Di Caro, A., Strada, G., Ippolito, G., & Group, I.-E. E. S. L. S. (2015). Blood kinetics of Ebola virus in survivors and nonsurvivors. *J Clin Invest*, *125*(12), 4692-4698. https://doi.org/10.1172/JCI83111

Law, J. W.-F., Ab Mutalib, N.-S., Chan, K.-G., & Lee, L.-H. (2015). Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations [Review]. *Frontiers in Microbiology*, *5*(770). https://doi.org/10.3389/fmicb.2014.00770

Le Guenno, B., Formenty, P., Wyers, M., Gounon, P., Walker, F., & Boesch, C. (1995). Isolation and partial characterisation of a new strain of Ebola virus. *Lancet*, *345*(8960), 1271-1274. https://doi.org/10.1016/s0140-6736(95)90925-7

Leroy, E. M., Baize, S., Volchkov, V. E., Fisher-Hoch, S. P., Georges-Courbot, M. C., Lansoud-Soukate, J., Capron, M., Debre, P., McCormick, J. B., & Georges, A. J. (2000). Human asymptomatic Ebola infection and strong inflammatory response. *Lancet*, *355*(9222), 2210-2215. https://doi.org/10.1016/s0140-6736(00)02405-3

Leroy, E. M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J. T., Gonzalez, J.-P., & Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. *Nature*, *438*(7068), 575-576. https://doi.org/10.1038/438575a

Li, J., Duan, H. J., Chen, H. Y., Ji, Y. J., Zhang, X., Rong, Y. H., Xu, Z., Sun, L. J., Zhang, J. Y., Liu, L. M., Jin, B., Zhang, J., Du, N., Su, H. B., Teng, G. J., Yuan, Y., Qin, E. Q., Jia, H. J., Wang, S., Guo, T. S., Wang, Y., Mu, J. S., Yan, T., Li, Z. W., Dong, Z., Nie, W. M., Jiang, T. J., Li, C., Gao, X. D., Ji, D., Zhuang, Y. J., Li, L., Wang, L. F., Li, W. G., Duan, X. Z., Lu, Y. Y., Sun, Z. Q., Kanu, A. B. J., Koroma, S. M., Zhao, M., Ji, J. S., & Wang, F. S. (2016). Age and Ebola viral load correlate with mortality and survival time in 288 Ebola virus

disease patients. *Int J Infect Dis*, *42*, 34-39. https://doi.org/10.1016/j.ijid.2015.10.021

Liang, W., Yao, J., Chen, A., Lv, Q., Zanin, M., Liu, J., Wong, S., Li, Y., Lu, J., Liang, H., Chen, G., Guo, H., Guo, J., Zhou, R., Ou, L., Zhou, N., Chen, H., Yang, F., Han, X., Huan, W., Tang, W., Guan, W., Chen, Z., Zhao, Y., Sang, L., Xu, Y., Wang, W., Li, S., Lu, L., Zhang, N., Zhong, N., Huang, J., & He, J. (2020). Early triage of critically ill COVID-19 patients using deep learning. *Nature Communications*, *11*(1), 3543. https://doi.org/10.1038/s41467-020-17280-8

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321-332. https://doi.org/10.1038/nrg3920

Lin, C.-C., Hsu, Y.-C., Li, Y.-H., Kuo, Y.-Y., Hou, H.-A., Lan, K.-H., Chen, T.-C., Tzeng, Y.-S., Kuo, Y.-Y., Kao, C.-J., Chuang, P.-H., Tseng, M.-H., Chiu, Y.-C., Chou, W.-C., & Tien, H.-F. (2017). Higher HOPX expression is associated with distinct clinical and biological features and predicts poor prognosis in de novo acute myeloid leukemia. *Haematologica*, *102*(6), 1044-1053. https://doi.org/10.3324/haematol.2016.161257

Liu, X., Speranza, E., Muñoz-Fontela, C., Haldenby, S., Rickett, N. Y., Garcia-Dorival, I., Fang, Y., Hall, Y., Zekeng, E. G., Lüdtke, A., Xia, D., Kerber, R., Krumkamp, R., Duraffour, S., Sissoko, D., Kenny, J., Rockliffe, N., Williamson, E. D., Laws, T. R., N'Faly, M., Matthews, D. A., Günther, S., Cossins, A. R., Sprecher, A., Connor, J. H., Carroll, M. W., & Hiscox, J. A. (2017). Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. *Genome biology*, *18*(1), 4-4. https://doi.org/10.1186/s13059-016-1137-3

Liu, X., Zhu, A., He, J., Chen, Z., Liu, L., Xu, Y., Ye, F., Feng, H., Luo, L., Cai, B., Mai, Y., Lin, L., Zhang, Z., Chen, S., Shi, J., Wen, L., Wei, Y., Zhuo, J., Zhao, Y., Li, F., Wei, X., Chen, D., Zhang, X., Zhong, N., Huang, Y., Liu, H., Wang, J., Xu, X., Wang, J., Chen, R., Chen, X., Zhong, N., Zhao, J., Li, Y., Zhao, J., & Chen, J. (2020). Single-Cell Analysis Reveals Macrophage-Driven T Cell Dysfunction in Severe COVID-19 Patients. *medRxiv*, 2020.2005.2023.20100024. https://doi.org/10.1101/2020.05.23.20100024

Lo, M. K., Jordan, R., Arvey, A., Sudhamsu, J., Shrivastava-Ranjan, P., Hotard, A. L., Flint, M., McMullan, L. K., Siegel, D., Clarke, M. O., Mackman, R. L., Hui, H. C., Perron, M., Ray, A. S., Cihlar, T., Nichol, S. T., & Spiropoulou, C. F. (2017). GS-5734 and its parent nucleoside analog inhibit Filo-, Pneumo-, and Paramyxoviruses. *Sci Rep*, *7*, 43395. https://doi.org/10.1038/srep43395

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 9. https://doi.org/10.1057/s41599-020-0501-9

Lowe, R., Barcellos, C., Brasil, P., Cruz, O. G., Honorio, N. A., Kuper, H., & Carvalho, M. S. (2018). The Zika Virus Epidemic in Brazil: From Discovery to Future Implications. *Int J Environ Res Public Health*, *15*(1). https://doi.org/10.3390/ijerph15010096

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. *R Journal*, *6*, 79-89. https://doi.org/10.32614/RJ-2014-008

Luthra, P., Ramanan, P., Mire, C. E., Weisend, C., Tsuda, Y., Yen, B., Liu, G., Leung, D. W., Geisbert, T. W., Ebihara, H., Amarasinghe, G. K., & Basler, C. F.

(2013). Mutual antagonism between the Ebola virus VP35 protein and the RIG-I activator PACT determines infection outcome. *Cell host & microbe*, *14*(1), 74-84. https://doi.org/10.1016/j.chom.2013.06.010

Ma, X., Xu, H., Shi, L., Yang, P., Zhang, L., Sun, X., Zhen, W., & Hu, K. (2015). A multiplex PCR assay for the detection of five influenza viruses using a dual priming oligonucleotide system. *BMC Infectious Diseases*, *15*(1), 93. https://doi.org/10.1186/s12879-015-0818-y

Macneil, A., Reed, Z., & Rollin, P. E. (2011). Serologic cross-reactivity of human IgM and IgG antibodies to five species of Ebola virus. *PLoS Negl Trop Dis*, *5*(6), e1175. https://doi.org/10.1371/journal.pntd.0001175

Maganga, G. D., Kapetshi, J., Berthet, N., Kebela Ilunga, B., Kabange, F., Mbala Kingebeni, P., Mondonge, V., Muyembe, J.-J. T., Bertherat, E., Briand, S., Cabore, J., Epelboin, A., Formenty, P., Kobinger, G., González-Angulo, L., Labouba, I., Manuguerra, J.-C., Okwo-Bele, J.-M., Dye, C., & Leroy, E. M. (2014). Ebola Virus Disease in the Democratic Republic of Congo. *New England Journal of Medicine*, *371*(22), 2083-2091. https://doi.org/10.1056/NEJMoa1411099

Malhotra, S., Yen, J. Y., Honko, A. N., Garamszegi, S., Caballero, I. S., Johnson, J. C., Mucker, E. M., Trefry, J. C., Hensley, L. E., & Connor, J. H. (2013). Transcriptional profiling of the circulating immune response to lassa virus in an aerosol model of exposure. *PLoS Negl Trop Dis*, *7*(4), e2171. https://doi.org/10.1371/journal.pntd.0002171

Malvy, D., McElroy, A. K., de Clerck, H., Gunther, S., & van Griensven, J. (2019). Ebola virus disease. *Lancet*, *393*(10174), 936-948. https://doi.org/10.1016/S0140-6736(18)33132-5

Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., & Locati, M. (2004). The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol*, *25*(12), 677-686. https://doi.org/10.1016/j.it.2004.09.015

Marsh, G. A., Haining, J., Robinson, R., Foord, A., Yamada, M., Barr, J. A., Payne, J., White, J., Yu, M., Bingham, J., Rollin, P. E., Nichol, S. T., Wang, L.-F., & Middleton, D. (2011). Ebola Reston Virus Infection of Pigs: Clinical Significance and Transmission Potential. *The Journal of infectious diseases*, *204*(suppl_3), S804-S809. https://doi.org/10.1093/infdis/jir300

Marzi, A., Halfmann, P., Hill-Batorski, L., Feldmann, F., Shupert, W. L., Neumann, G., Feldmann, H., & Kawaoka, Y. (2015). Vaccines. An Ebola whole-virus vaccine is protective in nonhuman primates. *Science (New York, N.Y.)*, *348*(6233), 439-442. https://doi.org/10.1126/science.aaa4919

Mattiola, I., Tomay, F., De Pizzol, M., Silva-Gomes, R., Savino, B., Gulic, T., Doni, A., Lonardi, S., Astrid Boutet, M., Nerviani, A., Carriero, R., Molgora, M., Stravalaci, M., Morone, D., Shalova, I. N., Lee, Y., Biswas, S. K., Mantovani, G., Sironi, M., Pitzalis, C., Vermi, W., Bottazzi, B., Mantovani, A., & Locati, M. (2019). The macrophage tetraspan MS4A4A enhances dectin-1-dependent NK cell-mediated resistance to metastasis. *Nat Immunol*, *20*(8), 1012-1022. https://doi.org/10.1038/s41590-019-0417-y

Mbala-Kingebeni, P., Aziza, A., Di Paola, N., Wiley, M. R., Makiala-Mandanda, S., Caviness, K., Pratt, C. B., Ladner, J. T., Kugelman, J. R., Prieto, K., Chitty, J. A., Larson, P. A., Beitzel, B., Ayouba, A., Vidal, N., Karhemere, S., Diop, M., Diagne, M. M., Faye, M., Faye, O., Aruna, A., Nsio, J., Mulangu, F., Mukadi, D., Mukadi, P., Kombe, J., Mulumba, A., Villabona-Arenas, C. J., Pukuta, E.,

Gonzalez, J., Bartlett, M. L., Sozhamannan, S., Gross, S. M., Schroth, G. P., Tim, R., Zhao, J. J., Kuhn, J. H., Diallo, B., Yao, M., Fall, I. S., Ndjoloko, B., Mossoko, M., Lacroix, A., Delaporte, E., Sanchez-Lockhart, M., Sall, A. A., Muyembe-Tamfum, J. J., Peeters, M., Palacios, G., & Ahuka-Mundeke, S. (2019). Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri Provinces of the Democratic Republic of the Congo: a rapid genomic assessment. *Lancet Infect Dis*, *19*(6), 648-657. https://doi.org/10.1016/S1473-3099(19)30118-5

McElroy, A. K., Akondy, R. S., Davis, C. W., Ellebedy, A. H., Mehta, A. K., Kraft, C. S., Lyon, G. M., Ribner, B. S., Varkey, J., Sidney, J., Sette, A., Campbell, S., Stroher, U., Damon, I., Nichol, S. T., Spiropoulou, C. F., & Ahmed, R. (2015). Human Ebola virus infection results in substantial immune activation. *Proc Natl Acad Sci U S A*, *112*(15), 4719-4724. https://doi.org/10.1073/pnas.1502619112

McElroy, A. K., Erickson, B. R., Flietstra, T. D., Rollin, P. E., Nichol, S. T., Towner, J. S., & Spiropoulou, C. F. (2014). Ebola hemorrhagic Fever: novel biomarker correlates of clinical outcome. *The Journal of infectious diseases*, *210*(4), 558-566. https://doi.org/10.1093/infdis/jiu088

McElroy, A. K., Harmon, J. R., Flietstra, T. D., Campbell, S., Mehta, A. K., Kraft, C. S., Lyon, M. G., Varkey, J. B., Ribner, B. S., Kratochvil, C. J., Iwen, P. C., Smith, P. W., Ahmed, R., Nichol, S. T., & Spiropoulou, C. F. (2016). Kinetic Analysis of Biomarkers in a Cohort of US Patients With Ebola Virus Disease. *Clin Infect Dis*, *63*(4), 460-467. https://doi.org/10.1093/cid/ciw334

McElroy, A. K., Shrivastava-Ranjan, P., Harmon, J. R., Martines, R. B., Silva-Flannery, L., Flietstra, T. D., Kraft, C. S., Mehta, A. K., Lyon, G. M., Varkey, J. B., Ribner, B. S., Nichol, S. T., Zaki, S. R., & Spiropoulou, C. F. (2019). Macrophage Activation Marker Soluble CD163 Associated with Fatal and Severe Ebola Virus Disease in Humans(1). *Emerg Infect Dis*, *25*(2), 290-298. https://doi.org/10.3201/eid2502.181326

Melendez, M. M., McNurlan, M. A., Mynarcik, D. C., Khan, S., & Gelato, M. C. (2008). Endothelial adhesion molecules are associated with inflammation in subjects with HIV disease. *Clin Infect Dis*, *46*(5), 775-780. https://doi.org/10.1086/527563

Milligan, I. D., Gibani, M. M., Sewell, R., Clutterbuck, E. A., Campbell, D., Plested, E., Nuthall, E., Voysey, M., Silva-Reyes, L., McElrath, M. J., De Rosa, S. C., Frahm, N., Cohen, K. W., Shukarev, G., Orzabal, N., van Duijnhoven, W., Truyers, C., Bachmayer, N., Splinter, D., Samy, N., Pau, M. G., Schuitemaker, H., Luhn, K., Callendret, B., Van Hoof, J., Douoguih, M., Ewer, K., Angus, B., Pollard, A. J., & Snape, M. D. (2016). Safety and Immunogenicity of Novel Adenovirus Type 26- and Modified Vaccinia Ankara-Vectored Ebola Vaccines: A Randomized Clinical Trial. *JAMA*, *315*(15), 1610-1623. https://doi.org/10.1001/jama.2016.4218

Mitchell, T. (1997). *Machine learning*. McGraw-Hill.

Mohan, G. S., Li, W., Ye, L., Compans, R. W., & Yang, C. (2012). Antigenic subversion: a novel mechanism of host immune evasion by Ebola virus. *PLoS Pathog*, *8*(12), e1003065. https://doi.org/10.1371/journal.ppat.1003065

Muhlberger, E. (2007). Filovirus replication and transcription. *Future Virol*, *2*(2), 205-215. https://doi.org/10.2217/17460794.2.2.205

Mühlberger, E., Weik, M., Volchkov, V. E., Klenk, H. D., & Becker, S. (1999). Comparison of the transcription and replication strategies of marburg virus and

Ebola virus by using artificial replication systems. *Journal of Virology*, *73*(3), 2333-2342. https://pubmed.ncbi.nlm.nih.gov/9971816
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC104478/
Mulangu, S., Dodd, L. E., Davey, R. T., Jr., Tshiani Mbaya, O., Proschan, M., Mukadi, D., Lusakibanza Manzo, M., Nzolo, D., Tshomba Oloma, A., Ibanda, A., Ali, R., Coulibaly, S., Levine, A. C., Grais, R., Diaz, J., Lane, H. C., Muyembe-Tamfum, J. J., Group, P. W., Sivahera, B., Camara, M., Kojan, R., Walker, R., Dighero-Kemp, B., Cao, H., Mukumbayi, P., Mbala-Kingebeni, P., Ahuka, S., Albert, S., Bonnett, T., Crozier, I., Duvenhage, M., Proffitt, C., Teitelbaum, M., Moench, T., Aboulhab, J., Barrett, K., Cahill, K., Cone, K., Eckes, R., Hensley, L., Herpin, B., Higgs, E., Ledgerwood, J., Pierson, J., Smolskis, M., Sow, Y., Tierney, J., Sivapalasingam, S., Holman, W., Gettinger, N., Vallee, D., Nordwall, J., & Team, P. C. S. (2019). A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N Engl J Med*, *381*(24), 2293-2303. https://doi.org/10.1056/NEJMoa1910993
Mutua, G., Anzala, O., Luhn, K., Robinson, C., Bockstal, V., Anumendem, D., & Douoguih, M. (2019). Safety and Immunogenicity of a 2-Dose Heterologous Vaccine Regimen With Ad26.ZEBOV and MVA-BN-Filo Ebola Vaccines: 12-Month Data From a Phase 1 Randomized Clinical Trial in Nairobi, Kenya. *The Journal of infectious diseases*, *220*(1), 57-67. https://doi.org/10.1093/infdis/jiz071
Negredo, A., Palacios, G., Vazquez-Moron, S., Gonzalez, F., Dopazo, H., Molero, F., Juste, J., Quetglas, J., Savji, N., de la Cruz Martinez, M., Herrera, J. E., Pizarro, M., Hutchison, S. K., Echevarria, J. E., Lipkin, W. I., & Tenorio, A. (2011). Discovery of an ebolavirus-like filovirus in europe. *PLoS Pathog*, *7*(10), e1002304. https://doi.org/10.1371/journal.ppat.1002304
Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models.
Ng, M., Ndungo, E., Kaczmarek, M. E., Herbert, A. S., Binger, T., Kuehne, A. I., Jangra, R. K., Hawkins, J. A., Gifford, R. J., Biswas, R., Demogines, A., James, R. M., Yu, M., Brummelkamp, T. R., Drosten, C., Wang, L. F., Kuhn, J. H., Muller, M. A., Dye, J. M., Sawyer, S. L., & Chandran, K. (2015). Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. *Elife*, *4*. https://doi.org/10.7554/eLife.11785
Nguyen, M. T., Ho, T. N., Nguyen, V. V., Nguyen, T. H., Ha, M. T., Ta, V. T., Nguyen, L. D., Phan, L., Han, K. Q., Duong, T. H., Tran, N. B., Wills, B., Wolbers, M., & Simmons, C. P. (2017). An Evidence-Based Algorithm for Early Prognosis of Severe Dengue in the Outpatient Setting. *Clin Infect Dis*, *64*(5), 656-663. https://doi.org/10.1093/cid/ciw863
Okumura, A., Pitha, P. M., & Harty, R. N. (2008). ISG15 inhibits Ebola VP40 VLP budding in an L-domain-dependent manner by blocking Nedd4 ligase activity. *Proceedings of the National Academy of Sciences*, *105*(10), 3974. https://doi.org/10.1073/pnas.0710629105
Olival, K. J., Islam, A., Yu, M., Anthony, S. J., Epstein, J. H., Khan, S. A., Khan, S. U., Crameri, G., Wang, L. F., Lipkin, W. I., Luby, S. P., & Daszak, P. (2013). Ebola virus antibodies in fruit bats, bangladesh. *Emerg Infect Dis*, *19*(2), 270-273. https://doi.org/10.3201/eid1902.120524
Omatsu, T., Bak, E. J., Ishii, Y., Kyuwa, S., Tohya, Y., Akashi, H., & Yoshikawa, Y. (2008). Induction and sequencing of Rousette bat interferon alpha and beta

genes. *Vet Immunol Immunopathol*, *124*(1-2), 169-176. https://doi.org/10.1016/j.vetimm.2008.03.004

Page, C., Goicochea, L., Matthews, K., Zhang, Y., Klover, P., Holtzman, M. J., Hennighausen, L., & Frieman, M. (2012). Induction of alternatively activated macrophages enhances pathogenesis during severe acute respiratory syndrome coronavirus infection. *J Virol*, *86*(24), 13334-13349. https://doi.org/10.1128/JVI.01689-12

Pallett, S. J. C., Rayment, M., Patel, A., Fitzgerald-Smith, S. A. M., Denny, S. J., Charani, E., Mai, A. L., Gilmour, K. C., Hatcher, J., Scott, C., Randell, P., Mughal, N., Jones, R., Moore, L. S. P., & Davies, G. W. (2020). Point-of-care serological assays for delayed SARS-CoV-2 case identification among health-care workers in the UK: a prospective multicentre cohort study. *The Lancet Respiratory Medicine*, *8*(9), 885-894. https://doi.org/10.1016/S2213-2600(20)30315-5

Pan, X., Chen, Y., & Gao, S. (2020). Four genes relevant to pathological grade and prognosis in ovarian cancer. *Cancer Biomark*, *29*(2), 169-178. https://doi.org/10.3233/CBM-191162

Pan, Y., Zhang, W., Cui, L., Hua, X., Wang, M., & Zeng, Q. (2014). Reston virus in domestic pigs in China. *Archives of Virology*, *159*(5), 1129-1132. https://doi.org/10.1007/s00705-012-1477-6

Pang, Z., Li, A., Li, J., Qu, J., He, C., Zhang, S., Li, C., Zhang, Q., Liang, M., & Li, D. (2014). Comprehensive multiplex one-step real-time TaqMan qRT-PCR assays for detection and quantification of hemorrhagic fever viruses. *PLoS One*, *9*(4), e95635. https://doi.org/10.1371/journal.pone.0095635

Perkins, M. D., Dye, C., Balasegaram, M., Brechot, C., Mombouli, J. V., Rottingen, J. A., Tanner, M., & Boehme, C. C. (2017). Diagnostic preparedness for infectious disease outbreaks. *Lancet*, *390*(10108), 2211-2214. https://doi.org/10.1016/S0140-6736(17)31224-2

Perng, Y. C., & Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*, *16*(7), 423-439. https://doi.org/10.1038/s41579-018-0020-5

Pleet, M. L., DeMarino, C., Lepene, B., Aman, M. J., & Kashanchi, F. (2017). The Role of Exosomal VP40 in Ebola Virus Disease. *DNA Cell Biol*, *36*(4), 243-248. https://doi.org/10.1089/dna.2017.3639

Prevail II Writing Group, Multi-National, P. I. I. S. T., Davey, R. T., Jr., Dodd, L., Proschan, M. A., Neaton, J., Neuhaus Nordwall, J., Koopmeiners, J. S., Beigel, J., Tierney, J., Lane, H. C., Fauci, A. S., Massaquoi, M. B. F., Sahr, F., & Malvy, D. (2016). A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. *N Engl J Med*, *375*(15), 1448-1456. https://doi.org/10.1056/NEJMoa1604330

Pringle, C., Alexander, D., Billeter, M., Collins, P., Kingsbury, D., Lipkind, M., Nagai, Y., Orvell, C., Rima B, Rott R, & ter Meulen, V. (1991). The order Mononegavirales. *Archives of Virology*, *117*(1-2), 137-140.

Prins, K. C., Cardenas, W. B., & Basler, C. F. (2009). Ebola virus protein VP35 impairs the function of interferon regulatory factor-activating kinases IKKepsilon and TBK-1. *J Virol*, *83*(7), 3069-3077. https://doi.org/10.1128/JVI.01875-08

Qin, E., Bi, J., Zhao, M., Wang, Y., Guo, T., Yan, T., Li, Z., Sun, J., Zhang, J., Chen, S., Wu, Y., Li, J., & Zhong, Y. (2015). Clinical Features of Patients With Ebola Virus Disease in Sierra Leone. *Clin Infect Dis*, *61*(4), 491-495. https://doi.org/10.1093/cid/civ319

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouedraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sanchez, A., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G. G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallash, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Trina, R., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., O'Shea, M. K., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Woelfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keita, S., Rambaut, A., Formenty, P., Gunther, S., & Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228-232. https://doi.org/10.1038/nature16996

R Core team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.* In https://www.R-project.org/.

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition*. Packt Publishing.

Rasmussen, A. L., Okumura, A., Ferris, M. T., Green, R., Feldmann, F., Kelly, S. M., Scott, D. P., Safronetz, D., Haddock, E., LaCasse, R., Thomas, M. J., Sova, P., Carter, V. S., Weiss, J. M., Miller, D. R., Shaw, G. D., Korth, M. J., Heise, M. T., Baric, R. S., de Villena, F. P., Feldmann, H., & Katze, M. G. (2014). Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science*, *346*(6212), 987-991. https://doi.org/10.1126/science.1259595

Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(3), 252-264. https://doi.org/10.1109/34.75512

Reaves, E. J., Mabande, L. G., Thoroughman, D. A., Arwady, M. A., & Montgomery, J. M. (2014). Control of Ebola virus disease - firestone district, liberia, 2014. *MMWR Morb Mortal Wkly Rep*, *63*(42), 959-965. https://www.ncbi.nlm.nih.gov/pubmed/25340914

Reid, S. P., Leung, L. W., Hartman, A. L., Martinez, O., Shaw, M. L., Carbonnelle, C., Volchkov, V. E., Nichol, S. T., & Basler, C. F. (2006). Ebola virus VP24 binds karyopherin alpha1 and blocks STAT1 nuclear accumulation. *J Virol*, *80*(11), 5156-5167. https://doi.org/10.1128/JVI.02349-05

Reisler, R. B., Yu, C., Donofrio, M. J., Warren, T. K., Wells, J. B., Stuthman, K. S., Garza, N. L., Vantongeren, S. A., Donnelly, G. C., Kane, C. D., Kortepeter, M. G., Bavari, S., & Cardile, A. P. (2017). Clinical Laboratory Values as Early Indicators of Ebola Virus Infection in Nonhuman Primates. *Emerging infectious diseases*, *23*(8), 1316-1324. https://doi.org/10.3201/eid2308.170029

Reynard, S., Journeaux, A., Gloaguen, E., Schaeffer, J., Varet, H., Pietrosemoli, N., Mateo, M., Baillet, N., Laouenan, C., Raoul, H., Mullaert, J., & Baize, S. (2019). Immune parameters and outcomes during Ebola virus disease. *JCI Insight*, *4*(1). https://doi.org/10.1172/jci.insight.125106

Rhein, B. A., Powers, L. S., Rogers, K., Anantpadma, M., Singh, B. K., Sakurai, Y., Bair, T., Miller-Hunt, C., Sinn, P., Davey, R. A., Monick, M. M., & Maury, W. (2015). Interferon-gamma Inhibits Ebola Virus Infection. *PLoS Pathog*, *11*(11), e1005263. https://doi.org/10.1371/journal.ppat.1005263

Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun*, *11*(1), 3923. https://doi.org/10.1038/s41467-020-17419-7

Rieger, T., Kerber, R., El Halas, H., Pallasch, E., Duraffour, S., Günther, S., & Ölschläger, S. (2016). Evaluation of RealStar Reverse Transcription-Polymerase Chain Reaction Kits for Filovirus Detection in the Laboratory and Field. *The Journal of infectious diseases*, *214*(suppl 3), S243-S249. https://doi.org/10.1093/infdis/jiw246

Rodríguez, A., Rodríguez, M., Córdoba, J. J., & Andrade, M. J. (2015). Design of Primers and Probes for Quantitative Real-Time PCR Methods. In C. Basu (Ed.), *PCR Primer Design* (pp. 31-56). Springer New York. https://doi.org/10.1007/978-1-4939-2365-6_3

Rogers, K. J., Brunton, B., Mallinger, L., Bohan, D., Sevcik, K. M., Chen, J., Ruggio, N., & Maury, W. (2019). IL-4/IL-13 polarization of macrophages enhances Ebola virus glycoprotein-dependent infection. *PLOS Neglected Tropical Diseases*, *13*(12), e0007819-e0007819. https://doi.org/10.1371/journal.pntd.0007819

Rogers, K. J., Shtanko, O., Vijay, R., Mallinger, L. N., Joyner, C. J., Galinski, M. R., Butler, N. S., & Maury, W. (2020). Acute Plasmodium Infection Promotes Interferon-Gamma-Dependent Resistance to Ebola Virus Infection. *Cell Rep*, *30*(12), 4041-4051 e4044. https://doi.org/10.1016/j.celrep.2020.02.104

Rollin, P. E., Williams, R. J., Bressler, D. S., Pearson, S., Cottingham, M., Pucak, G., Sanchez, A., Trappier, S. G., Peters, R. L., Greer, P. W., Zaki, S., Demarcus, T., Hendricks, K., Kelley, M., Simpson, D., Geisbert, T. W., Jahrling, P. B., Peters, C. J., & Ksiazek, T. G. (1999). Ebola (Subtype Reston) Virus among Quarantined Nonhuman Primates Recently Imported from the Philippines to the United States. *The Journal of infectious diseases*, *179*(Supplement_1), S108-S114. https://doi.org/10.1086/514303

Rolot, M., & Dewals, B. G. (2018). Macrophage Activation and Functions during Helminth Infection: Recent Advances from the Laboratory Mouse. *J Immunol Res*, *2018*, 2790627. https://doi.org/10.1155/2018/2790627

Rong, X. M., Yang, L., Chu, H. D., & Fan, M. (2020). Effect of delay in diagnosis on transmission of COVID-19. *Math Biosci Eng*, *17*(3), 2725-2740. https://doi.org/10.3934/mbe.2020149

Rosenke, K., Adjemian, J., Munster, V. J., Marzi, A., Falzarano, D., Onyango, C. O., Ochieng, M., Juma, B., Fischer, R. J., Prescott, J. B., Safronetz, D., Omballa, V., Owuor, C., Hoenen, T., Groseth, A., Martellaro, C., van Doremalen, N., Zemtsova, G., Self, J., Bushmaker, T., McNally, K., Rowe, T., Emery, S. L., Feldmann, F., Williamson, B. N., Best, S. M., Nyenswah, T. G., Grolla, A., Strong, J. E., Kobinger, G., Bolay, F. K., Zoon, K. C., Stassijns, J., Giuliani, R., de Smet, M., Nichol, S. T., Fields, B., Sprecher, A., Massaquoi, M., Feldmann, H., & de Wit, E. (2016). Plasmodium Parasitemia Associated With

Increased Survival in Ebola Virus-Infected Patients. *Clin Infect Dis*, *63*(8), 1026-1033. https://doi.org/10.1093/cid/ciw452

Rouquet, P., Froment, J. M., Bermejo, M., Kilbourn, A., Karesh, W., Reed, P., Kumulungui, B., Yaba, P., Delicat, A., Rollin, P. E., & Leroy, E. M. (2005). Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001-2003. *Emerg Infect Dis*, *11*(2), 283-290. https://doi.org/10.3201/eid1102.040533

Ruan, J., Zheng, H., Fu, W., Zhao, P., Su, N., & Luo, R. (2014). Increased expression of cathepsin L: a novel independent prognostic marker of worse outcome in hepatocellular carcinoma patients. *PLoS One*, *9*(11), e112136. https://doi.org/10.1371/journal.pone.0112136

Ryan, M. C., Zeeberg, B. R., Caplen, N. J., Cleland, J. A., Kahn, A. B., Liu, H., & Weinstein, J. N. J. B. B. (2008). SpliceCenter: A suite of web-based bioinformatic applications for evaluating the impact of alternative splicing on RT-PCR, RNAi, microarray, and peptide-based studies [journal article]. *9*(1), 313. https://doi.org/10.1186/1471-2105-9-313

Sadek, R. F., Khan, A. S., Stevens, G., Peters, C. J., & Ksiazek, T. G. (1999). Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995: determinants of survival. *The Journal of infectious diseases*, *179 Suppl 1*, S24-27. https://doi.org/10.1086/514311

Saeed, M. F., Kolokoltsov, A. A., Albrecht, T., & Davey, R. A. (2010). Cellular entry of ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLOS Pathogens*, *6*(9), e1001110-e1001110. https://doi.org/10.1371/journal.ppat.1001110

Sanchez, A., Ksiazek, T. G., Rollin, P. E., Miranda, M. E., Trappier, S. G., Khan, A. S., Peters, C. J., & Nichol, S. T. (1999). Detection and molecular characterization of Ebola viruses causing disease in human and nonhuman primates. *The Journal of infectious diseases*, *179 Suppl 1*, S164-169. https://doi.org/10.1086/514282

Sanchez, A., Lukwiya, M., Bausch, D., Mahanty, S., Sanchez, A. J., Wagoner, K. D., & Rollin, P. E. (2004). Analysis of human peripheral blood samples from fatal and nonfatal cases of Ebola (Sudan) hemorrhagic fever: cellular responses, virus load, and nitric oxide levels. *J Virol*, *78*(19), 10370-10377. https://doi.org/10.1128/JVI.78.19.10370-10377.2004

Sanders, R., Mason, D. J., Foy, C. A., & Huggett, J. F. (2014). Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Anal Bioanal Chem*, *406*(26), 6471-6483. https://doi.org/10.1007/s00216-014-7857-x

Sanyal, R., Polyak, M. J., Zuccolo, J., Puri, M., Deng, L., Roberts, L., Zuba, A., Storek, J., Luider, J. M., Sundberg, E. M., Mansoor, A., Baigorri, E., Chu, M. P., Belch, A. R., Pilarski, L. M., & Deans, J. P. (2017). MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. *Immunol Cell Biol*, *95*(7), 611-619. https://doi.org/10.1038/icb.2017.18

Schieffelin, J. S., Shaffer, J. G., Goba, A., Gbakie, M., Gire, S. K., Colubri, A., Sealfon, R. S. G., Kanneh, L., Moigboi, A., Momoh, M., Fullah, M., Moses, L. M., Brown, B. L., Andersen, K. G., Winnicki, S., Schaffner, S. F., Park, D. J., Yozwiak, N. L., Jiang, P.-P., Kargbo, D., Jalloh, S., Fonnie, M., Sinnah, V., French, I., Kovoma, A., Kamara, F. K., Tucker, V., Konuwa, E., Sellu, J., Mustapha, I., Foday, M., Yillah, M., Kanneh, F., Saffa, S., Massally, J. L. B., Boisen, M. L., Branco, L. M., Vandi, M. A., Grant, D. S., Happi, C., Gevao, S.

M., Fletcher, T. E., Fowler, R. A., Bausch, D. G., Sabeti, P. C., Khan, S. H., & Garry, R. F. (2014). Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone. *371*(22), 2092-2100. https://doi.org/10.1056/NEJMoa1411680

Schuh, A. J., Amman, B. R., Jones, M. E., Sealy, T. K., Uebelhoer, L. S., Spengler, J. R., Martin, B. E., Coleman-McCray, J. A., Nichol, S. T., & Towner, J. S. (2017). Modelling filovirus maintenance in nature by experimental transmission of Marburg virus between Egyptian rousette bats. *Nat Commun*, *8*, 14446. https://doi.org/10.1038/ncomms14446

Schuh, A. J., Amman, B. R., Sealy, T. K., Kainulainen, M. H., Chakrabarti, A. K., Guerrero, L. W., Nichol, S. T., Albarino, C. G., & Towner, J. S. (2019). Antibody-Mediated Virus Neutralization Is Not a Universal Mechanism of Marburg, Ebola, or Sosuga Virus Clearance in Egyptian Rousette Bats. *The Journal of infectious diseases*, *219*(11), 1716-1721. https://doi.org/10.1093/infdis/jiy733

Semper, A. E., Broadhurst, M. J., Richards, J., Foster, G. M., Simpson, A. J., Logue, C. H., Kelly, J. D., Miller, A., Brooks, T. J., Murray, M., & Pollock, N. R. (2016). Performance of the GeneXpert Ebola Assay for Diagnosis of Ebola Virus Disease in Sierra Leone: A Field Evaluation Study. *PLoS Med*, *13*(3), e1001980. https://doi.org/10.1371/journal.pmed.1001980

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., Wang, W., Eden, J. S., Shen, J. J., Liu, L., Holmes, E. C., & Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. *Nature*, *556*(7700), 197-202. https://doi.org/10.1038/s41586-018-0012-7

Siegert, R., Shu, H. L., Slenczka, W., Peters, D., & Muller, G. (1967). On the etiology of an unknown human infection originating from monkeys. *Dtsch Med Wochenschr*, *92*(51), 2341-2343. https://doi.org/10.1055/s-0028-1106144 (Zur Atiologie einer umbekannten, von Affen ausgegangenen menschlichen Infektionskrankheit.)

Sissoko, D., Laouenan, C., Folkesson, E., M'Lebing, A.-B., Beavogui, A.-H., Baize, S., Camara, A.-M., Maes, P., Shepherd, S., Danel, C., Carazo, S., Conde, M. N., Gala, J.-L., Colin, G., Savini, H., Bore, J. A., Le Marcis, F., Koundouno, F. R., Petitjean, F., Lamah, M.-C., Diederich, S., Tounkara, A., Poelart, G., Berbain, E., Dindart, J.-M., Duraffour, S., Lefevre, A., Leno, T., Peyrouset, O., Irenge, L., Bangoura, N. F., Palich, R., Hinzmann, J., Kraus, A., Barry, T. S., Berette, S., Bongono, A., Camara, M. S., Chanfreau Munoz, V., Doumbouya, L., Souley, H., Kighoma, P. M., Koundouno, F. R., Réné, L., Loua, C. M., Massala, V., Moumouni, K., Provost, C., Samake, N., Sekou, C., Soumah, A., Arnould, I., Komano, M. S., Gustin, L., Berutto, C., Camara, D., Camara, F. S., Colpaert, J., Delamou, L., Jansson, L., Kourouma, E., Loua, M., Malme, K., Manfrin, E., Maomou, A., Milinouno, A., Ombelet, S., Sidiboun, A. Y., Verreckt, I., Yombouno, P., Bocquin, A., Carbonnelle, C., Carmoi, T., Frange, P., Mely, S., Nguyen, V.-K., Pannetier, D., Taburet, A.-M., Treluyer, J.-M., Kolie, J., Moh, R., Gonzalez, M. C., Kuisma, E., Liedigk, B., Ngabo, D., Rudolf, M., Thom, R., Kerber, R., Gabriel, M., Di Caro, A., Wölfel, R., Badir, J., Bentahir, M., Deccache, Y., Dumont, C., Durant, J.-F., El Bakkouri, K., Gasasira Uwamahoro, M., Smits, B., Toufik, N., Van Cauwenberghe, S., Ezzedine, K., Dortenzio, E., Pizarro, L., Etienne, A., Guedj, J., Fizet, A., Barte de Sainte Fare, E., Murgue, B., Tran-Minh, T., Rapp, C., Piguet, P., Poncin, M., Draguez, B., Allaford Duverger, T., Barbe, S., Baret, G., Defourny, I., Carroll, M., Raoul, H., Augier, A., Eholie, S. P., Yazdanpanah, Y., Levy-

Marchal, C., Antierrens, A., Van Herp, M., Günther, S., de Lamballerie, X., Keïta, S., Mentre, F., Anglaret, X., Malvy, D., & Group, J. S. (2016). Experimental Treatment with Favipiravir for Ebola Virus Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea. *PLOS Medicine*, *13*(3), e1001967. https://doi.org/10.1371/journal.pmed.1001967

Slenczka, W., & Klenk, H. D. (2007). Forty years of marburg virus. *The Journal of infectious diseases*, *196 Suppl 2*, S131-135. https://doi.org/10.1086/520551

Stanley, D. A., Honko, A. N., Asiedu, C., Trefry, J. C., Lau-Kilby, A. W., Johnson, J. C., Hensley, L., Ammendola, V., Abbate, A., Grazioli, F., Foulds, K. E., Cheng, C., Wang, L., Donaldson, M. M., Colloca, S., Folgori, A., Roederer, M., Nabel, G. J., Mascola, J., Nicosia, A., Cortese, R., Koup, R. A., & Sullivan, N. J. (2014). Chimpanzee adenovirus vaccine generates acute and durable protective immunity against ebolavirus challenge. *Nat Med*, *20*(10), 1126-1129. https://doi.org/10.1038/nm.3702

Stantchev, T. S., Zack-Taylor, A., Mattson, N., Strebel, K., Broder, C. C., & Clouse, K. A. (2019). Cytokine Effects on the Entry of Filovirus Envelope Pseudotyped Virus-Like Particles into Primary Human Macrophages. *Viruses*, *11*(10), 889. https://doi.org/10.3390/v11100889

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W., & Kubista, M. (2015). How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*, *3*, 9-16. https://doi.org/https://doi.org/10.1016/j.bdq.2015.01.005

Swanepoel, R., Smit, S. B., Rollin, P. E., Formenty, P., Leman, P. A., Kemp, A., Burt, F. J., Grobbelaar, A. A., Croft, J., Bausch, D. G., Zeller, H., Leirs, H., Braack, L. E., Libande, M. L., Zaki, S., Nichol, S. T., Ksiazek, T. G., Paweska, J. T., International, S., & Technical Committee for Marburg Hemorrhagic Fever Control in the Democratic Republic of, C. (2007). Studies of reservoir hosts for Marburg virus. *Emerg Infect Dis*, *13*(12), 1847-1851. https://doi.org/10.3201/eid1312.071115

Takada, A., Fujioka, K., Tsuiji, M., Morikawa, A., Higashi, N., Ebihara, H., Kobasa, D., Feldmann, H., Irimura, T., & Kawaoka, Y. (2004). Human macrophage C-type lectin specific for galactose and N-acetylgalactosamine promotes filovirus entry. *J Virol*, *78*(6), 2943-2947. https://doi.org/10.1128/jvi.78.6.2943-2947.2004

Taniguchi, S., Watanabe, S., Masangkay, J. S., Omatsu, T., Ikegami, T., Alviola, P., Ueda, N., Iha, K., Fujii, H., Ishii, Y., Mizutani, T., Fukushi, S., Saijo, M., Kurane, I., Kyuwa, S., Akashi, H., Yoshikawa, Y., & Morikawa, S. (2011). Reston Ebolavirus antibodies in bats, the Philippines. *Emerg Infect Dis*, *17*(8), 1559-1560. https://doi.org/10.3201/eid1708.101693

Tapia, M. D., Sow, S. O., Lyke, K. E., Haidara, F. C., Diallo, F., Doumbia, M., Traore, A., Coulibaly, F., Kodio, M., Onwuchekwa, U., Sztein, M. B., Wahid, R., Campbell, J. D., Kieny, M.-P., Moorthy, V., Imoukhuede, E. B., Rampling, T., Roman, F., De Ryck, I., Bellamy, A. R., Dally, L., Mbaya, O. T., Ploquin, A., Zhou, Y., Stanley, D. A., Bailer, R., Koup, R. A., Roederer, M., Ledgerwood, J., Hill, A. V. S., Ballou, W. R., Sullivan, N., Graham, B., & Levine, M. M. (2016). Use of ChAd3-EBO-Z Ebola virus vaccine in Malian and US adults, and boosting of Malian adults with MVA-BN-Filo: a phase 1, single-blind, randomised trial, a phase 1b, open-label and double-blind, dose-escalation trial, and a nested, randomised, double-blind, placebo-controlled trial. *The*

Lancet. Infectious diseases, *16*(1), 31-42. https://doi.org/10.1016/S1473-3099(15)00362-X

Taylor, S. C., Laperriere, G., & Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci Rep*, *7*(1), 2409. https://doi.org/10.1038/s41598-017-02217-x

Tong, M., Jiang, Y., Xia, D., Xiong, Y., Zheng, Q., Chen, F., Zou, L., Xiao, W., & Zhu, Y. (2020). Elevated Expression of Serum Endothelial Cell Adhesion Molecules in COVID-19 Patients. *The Journal of infectious diseases*, *222*(6), 894-898. https://doi.org/10.1093/infdis/jiaa349

Tong, Y. G., Shi, W. F., Liu, D., Qian, J., Liang, L., Bo, X. C., Liu, J., Ren, H. G., Fan, H., Ni, M., Sun, Y., Jin, Y., Teng, Y., Li, Z., Kargbo, D., Dafae, F., Kanu, A., Chen, C. C., Lan, Z. H., Jiang, H., Luo, Y., Lu, H. J., Zhang, X. G., Yang, F., Hu, Y., Cao, Y. X., Deng, Y. Q., Su, H. X., Sun, Y., Liu, W. S., Wang, Z., Wang, C. Y., Bu, Z. Y., Guo, Z. D., Zhang, L. B., Nie, W. M., Bai, C. Q., Sun, C. H., An, X. P., Xu, P. S., Zhang, X. L., Huang, Y., Mi, Z. Q., Yu, D., Yao, H. W., Feng, Y., Xia, Z. P., Zheng, X. X., Yang, S. T., Lu, B., Jiang, J. F., Kargbo, B., He, F. C., Gao, G. F., Cao, W. C., & China Mobile Laboratory Testing Team in Sierra, L. (2015). Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature*, *524*(7563), 93-96. https://doi.org/10.1038/nature14490

Towner, J. S., Amman, B. R., Sealy, T. K., Carroll, S. A. R., Comer, J. A., Kemp, A., Swanepoel, R., Paddock, C. D., Balinandi, S., Khristova, M. L., Formenty, P. B. H., Albarino, C. G., Miller, D. M., Reed, Z. D., Kayiwa, J. T., Mills, J. N., Cannon, D. L., Greer, P. W., Byaruhanga, E., Farnon, E. C., Atimnedi, P., Okware, S., Katongole-Mbidde, E., Downing, R., Tappero, J. W., Zaki, S. R., Ksiazek, T. G., Nichol, S. T., & Rollin, P. E. (2009). Isolation of Genetically Diverse Marburg Viruses from Egyptian Fruit Bats. *PLOS Pathogens*, *5*(7), e1000536. https://doi.org/10.1371/journal.ppat.1000536

Towner, J. S., Rollin, P. E., Bausch, D. G., Sanchez, A., Crary, S. M., Vincent, M., Lee, W. F., Spiropoulou, C. F., Ksiazek, T. G., Lukwiya, M., Kaducu, F., Downing, R., & Nichol, S. T. (2004). Rapid diagnosis of Ebola hemorrhagic fever by reverse transcription-PCR in an outbreak setting and assessment of patient viral load as a predictor of outcome. *Journal of Virology*, *78*(8), 4330-4341. https://doi.org/10.1128/jvi.78.8.4330-4341.2004

Towner, J. S., Sealy, T. K., Khristova, M. L., Albarino, C. G., Conlan, S., Reeder, S. A., Quan, P. L., Lipkin, W. I., Downing, R., Tappero, J. W., Okware, S., Lutwama, J., Bakamutumaho, B., Kayiwa, J., Comer, J. A., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2008). Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog*, *4*(11), e1000212. https://doi.org/10.1371/journal.ppat.1000212

Trigunaite, A., Dimo, J., & Jorgensen, T. N. (2015). Suppressive effects of androgens on the immune system. *Cell Immunol*, *294*(2), 87-94. https://doi.org/10.1016/j.cellimm.2015.02.004

Turmelle, A. S., Jackson, F. R., Green, D., McCracken, G. F., & Rupprecht, C. E. (2010). Host immunity to repeated rabies virus infection in big brown bats. *The Journal of general virology*, *91*(Pt 9), 2360-2366. https://doi.org/10.1099/vir.0.020073-0

Uyeki, T. M., Mehta, A. K., Davey, R. T., Jr., Liddell, A. M., Wolf, T., Vetter, P., Schmiedel, S., Grunewald, T., Jacobs, M., Arribas, J. R., Evans, L., Hewlett,

A. L., Brantsaeter, A. B., Ippolito, G., Rapp, C., Hoepelman, A. I., Gutman, J., Working Group of the, U. S. E. C. N. o. C. M. o. E. V. D. P. i. t. U. S., & Europe. (2016). Clinical Management of Ebola Virus Disease in the United States and Europe. *N Engl J Med*, *374*(7), 636-646. https://doi.org/10.1056/NEJMoa1504874

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

van Griensven, J., Edwards, T., Baize, S., & Ebola-Tx, C. (2016). Efficacy of Convalescent Plasma in Relation to Dose of Ebola Virus Antibodies. *N Engl J Med*, *375*(23), 2307-2309. https://doi.org/10.1056/NEJMc1609116

van Manen, D., van 't Wout, A. B., & Schuitemaker, H. (2012). Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology*, *9*, 70. https://doi.org/10.1186/1742-4690-9-70

Velasquez, G. E., Aibana, O., Ling, E. J., Diakite, I., Mooring, E. Q., & Murray, M. B. (2015). Time From Infection to Disease and Infectiousness for Ebola Virus Disease, a Systematic Review. *Clin Infect Dis*, *61*(7), 1135-1140. https://doi.org/10.1093/cid/civ531

Vetter, P., Fischer, W. A., 2nd, Schibler, M., Jacobs, M., Bausch, D. G., & Kaiser, L. (2016). Ebola Virus Shedding and Transmission: Review of Current Evidence. *The Journal of infectious diseases*, *214*(suppl 3), S177-S184. https://doi.org/10.1093/infdis/jiw254

Vishwakarma, S., Agarwal, R., Goel, S. K., Panday, R. K., Singh, R., Sukumaran, R., Khare, S., & Kumar, A. (2017). Altered Expression of Sphingosine-1-Phosphate Metabolizing Enzymes in Oral Cancer Correlate With Clinicopathological Attributes. *Cancer Invest*, *35*(2), 139-141. https://doi.org/10.1080/07357907.2016.1272695

Volchkov, V. E., Feldmann, H., Volchkova, V. A., & Klenk, H. D. (1998). Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc Natl Acad Sci U S A*, *95*(10), 5762-5767. https://doi.org/10.1073/pnas.95.10.5762

Vos, L. M., Oosterheert, J. J., Hoepelman, A. I. M., Bont, L. J., Coenjaerts, F. E. J., & Naaktgeboren, C. A. (2019). External validation and update of a prognostic model to predict mortality in hospitalized adults with RSV: A retrospective Dutch cohort study. *J Med Virol*, *91*(12), 2117-2124. https://doi.org/10.1002/jmv.25568

Wahl-Jensen, V., Kurz, S., Feldmann, F., Buehler, L. K., Kindrachuk, J., DeFilippis, V., da Silva Correia, J., Früh, K., Kuhn, J. H., Burton, D. R., & Feldmann, H. (2011). Ebola Virion Attachment and Entry into Human Macrophages Profoundly Effects Early Cellular Gene Expression. *PLOS Neglected Tropical Diseases*, *5*(10), e1359. https://doi.org/10.1371/journal.pntd.0001359

Wahl-Jensen, V. M., Afanasieva, T. A., Seebach, J., Stroher, U., Feldmann, H., & Schnittler, H. J. (2005). Effects of Ebola virus glycoproteins on endothelial cell activation and barrier function. *J Virol*, *79*(16), 10442-10450. https://doi.org/10.1128/JVI.79.16.10442-10450.2005

Warren, T., Zumbrun, E., Weidner, J. M., Gomba, L., Rossi, F., Bannister, R., Tarrant, J., Reed, M., Lee, E., Raymond, J. L., Wells, J., Shamblin, J., Wetzel, K., Donnelly, G., Van Tongeren, S., Lackemeyer, N., Steffens, J., Kimmel, A., Garvey, C., Bloomfield, H., Blair, C., Singh, B., Bavari, S., Cihlar, T., & Porter, D. (2020). Characterization of Ebola Virus Disease (EVD) in Rhesus

Monkeys for Development of EVD Therapeutics. *Viruses*, *12*(1). https://doi.org/10.3390/v12010092

Wauquier, N., Becquart, P., Gasquet, C., & Leroy, E. M. (2009). Immunoglobulin G in Ebola outbreak survivors, Gabon. *Emerg Infect Dis*, *15*(7), 1136-1137. https://doi.org/10.3201/eid1507.090402

Wauquier, N., Becquart, P., Padilla, C., Baize, S., & Leroy, E. M. (2010). Human fatal zaire ebola virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis. *PLOS Neglected Tropical Diseases*, *4*(10), e837. https://doi.org/10.1371/journal.pntd.0000837

Wawina-Bokalanga, T., Vanmechelen, B., Lhermitte, V., Marti-Carreras, J., Vergote, V., Koundouno, F. R., Akoi-Bore, J., Thom, R., Tipton, T., Steeds, K., Moussa, K. B., Amento, A., Laenen, L., Duraffour, S., Gabriel, M., Ruibal, P., Hall, Y., Kader-Konde, M., Gunther, S., Baele, G., Munoz-Fontela, C., Van Weyenbergh, J., Carroll, M. W., & Maes, P. (2021). Human Diversity of Killer Cell Immunoglobulin-Like Receptors and Human Leukocyte Antigen Class I Alleles and Ebola Virus Disease Outcomes. *Emerg Infect Dis*, *27*(1), 76-84. https://doi.org/10.3201/eid2701.202177

Waxman, M., Aluisio, A. R., Rege, S., & Levine, A. C. (2017). Characteristics and survival of patients with Ebola virus infection, malaria, or both in Sierra Leone: a retrospective cohort study. *Lancet Infect Dis*, *17*(6), 654-660. https://doi.org/10.1016/S1473-3099(17)30112-3

WHO Ebola Response Team. (2014). Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine*, *371*(16), 1481-1495. https://doi.org/10.1056/NEJMoa1411100

WHO Ebola Response Team. (2016). Ebola Virus Disease among Male and Female Persons in West Africa. *New England Journal of Medicine*, *374*(1), 96-98. https://doi.org/10.1056/NEJMc1510305

WHO Ebola Response Team, Agua-Agum, J., Ariyarajah, A., Blake, I. M., Cori, A., Donnelly, C. A., Dorigatti, I., Dye, C., Eckmanns, T., Ferguson, N. M., Fowler, R. A., Fraser, C., Garske, T., Hinsley, W., Jombart, T., Mills, H. L., Murthy, S., Nedjati Gilani, G., Nouvellet, P., Pelletier, L., Riley, S., Schumacher, D., Shah, A., & Van Kerkhove, M. D. (2015). Ebola virus disease among children in West Africa. *N Engl J Med*, *372*(13), 1274-1277. https://doi.org/10.1056/NEJMc1415318

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wong, G., Audet, J., Fernando, L., Fausther-Bovendo, H., Alimonti, J. B., Kobinger, G. P., & Qiu, X. (2014). Immunization with vesicular stomatitis virus vaccine expressing the Ebola glycoprotein provides sustained long-term protection in rodents. *Vaccine*, *32*(43), 5722-5729. https://doi.org/10.1016/j.vaccine.2014.08.028

World Health Organization. (1978). Ebola haemorrhagic fever in Zaire, 1976. Report of an international commission. *Bull. World Health Organ*, *56*, 271-293.

[Record #4229 is using a reference type undefined in this output style.]

World Health Organization. (2003). *Outbreaks of Ebola haemorrhagic fever, Congo and Gabon, October 2001-July 2002* (Weekly Epidemiological Report, Issue.

World Health Organization. (2007). *Ebola virus haemorrhagic fever, Democratic Republic of the Congo-Update* (Weekly Epidemiological Record, Issue.

World Health Organization. (2014). *Laboratory guidance for the diagnosis of Ebola virus disease : interim recommendations: September 2014*. https://apps.who.int/iris/handle/10665/134009

World Health Organization. (2015a). *Emergency guidance: selection and use of Ebola in vitro diagnostic (IVD) assays*. https://apps.who.int/iris/handle/10665/175554

World Health Organization. (2015b). *Factors that contributed to undetected spread of the Ebola virus and impeded rapid containment*. https://www.who.int/csr/disease/ebola/one-year-report/factors/en/

World Health Organization. (2016a). *Clinical management of patients with viral haemorrhagic fever: a pocket guide for front-line health workers: interim emergency guidance for country adaptation*. World Health Organization. https://apps.who.int/iris/handle/10665/205570

World Health Organization. (2016b). *Ebola Outbreak 2014-2016*. https://www.who.int/csr/disease/ebola/en/

World Health Organization. (2018). *Ebola Virus Disease Democratic Republic of Congo: External Situation Report 17 / 2018* CC BY-NC-SA 3.0 IGO). https://apps.who.int/iris/handle/10665/273348

World Health Organization. (2019a). *Optimized supportive care for Ebola virus disease: clinical management standard operating procedures*. World Health Organization. https://apps.who.int/iris/handle/10665/325000

World Health Organization. (2019b). *Preliminary results on the efficay of rVSV-ZEBOV-GP Ebola vaccine using the ring vaccination strategy in the control of an Ebola outbreak in the Democratic Republic of the Congo: an example of integration of research into epidemic response.* https://www.who.int/csr/resources/publications/ebola/ebola-ring-vaccination-results-12-april-2019.pdf?ua=1

World Health Organization. (2020a). *Ebola in the Democratic Republic of the Congo 2020- Équateur Province*. Retrieved 24 October from https://www.who.int/emergencies/diseases/ebola/ebola-health-update---%C3%A9quateur-province-democratic-republic-of-the-congo-2020

World Health Organization. (2020b). Ebola virus disease – Democratic Republic of the Congo (Update, 26 June 2020) – Maladie à virus Ebola – République démocratique du Congo (Mise à jour, 26 juin 2020). *Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire*, *95*(27), 301-306. https://apps.who.int/iris/handle/10665/332951

World Health Organization. (2020c). *Ebola Virus Disease Democratic Republic of Congo: External situation report 98 / 2020*. https://apps.who.int/iris/handle/10665/332654

World Health Organization. (2020d). Weekly Bulletin on Outbreak and other Emergencies: Week 36: 31 August - 06 September 2020. https://apps.who.int/iris/handle/10665/334185

World Health Organization. (2020e). Weekly Bulletin on Outbreak and other Emergencies: Week 41: 05 - 11 October 2020. https://apps.who.int/iris/handle/10665/336026

Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., Wang, L. F., Shi, Z. L., & Zhou, P. (2018). Dampened STING-Dependent Interferon Activation in Bats. *Cell Host Microbe*, *23*(3), 297-301 e294. https://doi.org/10.1016/j.chom.2018.01.006

Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., Huang, X., Xiao, Y., Cao, H., Chen, Y., Ren, T., Wang,

F., Xiao, Y., Huang, S., Tan, X., Huang, N., Jiao, B., Cheng, C., Zhang, Y., Luo, A., Mombaerts, L., Jin, J., Cao, Z., Li, S., Xu, H., & Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, *2*(5), 283-288. https://doi.org/10.1038/s42256-020-0180-7

Yan, T., Mu, J., Qin, E., Wang, Y., Liu, L., Wu, D., Jia, H., Li, Z., Guo, T., Wang, X., Qin, Y., Li, Y., Chen, S., Zhang, Y., Zhang, J., Wu, Y., Wang, S., & Li, J. (2015). Clinical characteristics of 154 patients suspected of having Ebola virus disease in the Ebola holding center of Jui Government Hospital in Sierra Leone during the 2014 Ebola outbreak. *Eur J Clin Microbiol Infect Dis*, *34*(10), 2089-2095. https://doi.org/10.1007/s10096-015-2457-z

Yang, X. L., Tan, C. W., Anderson, D. E., Jiang, R. D., Li, B., Zhang, W., Zhu, Y., Lim, X. F., Zhou, P., Liu, X. L., Guan, W., Zhang, L., Li, S. Y., Zhang, Y. Z., Wang, L. F., & Shi, Z. L. (2019). Characterization of a filovirus (Mengla virus) from Rousettus bats in China. *Nat Microbiol*, *4*(3), 390-395. https://doi.org/10.1038/s41564-018-0328-y

Younan, P., Ramanathan, P., Graber, J., Gusovsky, F., & Bukreyev, A. (2017). The Toll-Like Receptor 4 Antagonist Eritoran Protects Mice from Lethal Filovirus Challenge. *MBio*, *8*(2). https://doi.org/10.1128/mBio.00226-17

Younan, P., Santos, R. I., Ramanathan, P., Iampietro, M., Nishida, A., Dutta, M., Ammosova, T., Meyer, M., Katze, M. G., Popov, V. L., Nekhai, S., & Bukreyev, A. (2019). Ebola virus-mediated T-lymphocyte depletion is the result of an abortive infection. *PLOS Pathogens*, *15*(10), e1008068-e1008068. https://doi.org/10.1371/journal.ppat.1008068

Yuan, J., Zhang, Y., Li, J., Zhang, Y., Wang, L. F., & Shi, Z. (2012). Serological evidence of ebolavirus infection in bats, China. *Virol J*, *9*, 236. https://doi.org/10.1186/1743-422X-9-236

Zhang, B., Yao, G., Zhang, Y., Gao, J., Yang, B., Rao, Z., & Gao, J. (2011). M2-polarized tumor-associated macrophages are associated with poor prognoses resulting from accelerated lymphangiogenesis in lung adenocarcinoma. *Clinics (Sao Paulo, Brazil)*, *66*(11), 1879-1886. https://doi.org/10.1590/s1807-59322011001100006

Zhang, X., Rong, Y., Sun, L., Liu, L., Su, H., Zhang, J., Teng, G., Du, N., Chen, H., Fang, Y., Zhan, W., Kanu, A. B., Koroma, S. M., Jin, B., Xu, Z., & Song, H. (2015). Prognostic Analysis of Patients with Ebola Virus Disease. *PLoS Negl Trop Dis*, *9*(9), e0004113. https://doi.org/10.1371/journal.pntd.0004113

Zhou, N., Pan, T., Zhang, J., Li, Q., Zhang, X., Bai, C., Huang, F., Peng, T., Zhang, J., Liu, C., Tao, L., & Zhang, H. (2016). Glycopeptide Antibiotics Potently Inhibit Cathepsin L in the Late Endosome/Lysosome and Block the Entry of Ebola Virus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). *The Journal of biological chemistry*, *291*(17), 9218-9232. https://doi.org/10.1074/jbc.M116.716100

Zhou, P., Cowled, C., Todd, S., Crameri, G., Virtue, E. R., Marsh, G. A., Klein, R., Shi, Z., Wang, L. F., & Baker, M. L. (2011). Type III IFNs in pteropid bats: differential expression patterns provide evidence for distinct roles in antiviral immunity. *J Immunol*, *186*(5), 3138-3147. https://doi.org/10.4049/jimmunol.1003115

Zhu, J., Chen, X., Liao, Z., He, C., & Hu, X. (2015). TGFBI protein high expression predicts poor prognosis in colorectal cancer patients. *International journal of*

clinical and experimental pathology, *8*(1), 702-710. https://pubmed.ncbi.nlm.nih.gov/25755764
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4348825/

Zhu, Y., Cherukuri, N. C., Jackel, J. N., Wu, Z., Crary, M., Buckley, K. J., Bisaro, D. M., & Parris, D. S. (2012). Characterization of the RNA silencing suppression activity of the Ebola virus VP35 protein in plants and mammalian cells. *J Virol*, *86*(6), 3038-3049. https://doi.org/10.1128/JVI.05741-11

Zignego, A. L., Wojcik, G. L., Cacoub, P., Visentini, M., Casato, M., Mangia, A., Latanich, R., Charles, E. D., Gragnani, L., Terrier, B., Piazzola, V., Dustin, L. B., Khakoo, S. I., Busch, M. P., Lauer, G. M., Kim, A. Y., Alric, L., Thomas, D. L., & Duggal, P. (2014). Genome-wide association study of hepatitis C virus- and cryoglobulin-related vasculitis. *Genes Immun*, *15*(7), 500-505. https://doi.org/10.1038/gene.2014.41

Zumla, A., Rao, M., Wallis, R. S., Kaufmann, S. H. E., Rustomjee, R., Mwaba, P., Vilaplana, C., Yeboah-Manu, D., Chakaya, J., Ippolito, G., Azhar, E., Hoelscher, M., Maeurer, M., & Host-Directed Therapies Network, c. (2016). Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. *The Lancet. Infectious diseases*, *16*(4), e47-e63. https://doi.org/10.1016/S1473-3099(16)00078-5

Zuniga, J., Buendia-Roldan, I., Zhao, Y., Jimenez, L., Torres, D., Romo, J., Ramirez, G., Cruz, A., Vargas-Alarcon, G., Sheu, C. C., Chen, F., Su, L., Tager, A. M., Pardo, A., Selman, M., & Christiani, D. C. (2012). Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur Respir J*, *39*(3), 604-610. https://doi.org/10.1183/09031936.00020611

Appendix

Appendix Table 1. RNA concentration of the cohort of 39 EVD patients with known outcome

Sample_ID	RNA concentration ng/ µl
EM_013676	5.4
EM_013796	Low out of range
EM_076846	6.96
EM_075933	5.56
EM_076035	12.9
EM_008015	7.04
EM_075155	5.86
EM_022044	4.8
EM_013261	8.28
EM_076814	Low out of range
EM_0895	5.98
EM_0367	4.6
EM_023321	Low out of range
EM_013283	6.02
EM_077285	7.64
EM_015781	5.34
EM_0017	6.34
EM_004032	Low out of range
EM_076295	Low out of range
EM_013860	8.12
EM_075355	7.84
EM_075935	Low out of range
EM_016460	10.1
EM_075314	7.64
EM_008007	12.6
EM_0727	10.5
EM_0061	6.56
EM_0763	14.7
EM_080293	Low out of range
EM_075159	7.22
EM_0894	9.12
EM_080093	5
EM_077046	7.6
EM_076569	15
EM_076866	17.4
EM_0783	12.8
EM_076545	10.3
EM_076882	13.3
EM_080108	5.5

Appendix Table 2. Transcript abundance of the 11 candidate predictive genes quantified by RT-qPCR from testing 5 control samples and 39 EVD patients with known outcome

Sample_ID	VCAM1	ISG15	CTSL	TUBG1	MS4A4A	TTC28	TGFBI	PLPP3	NF3L1	SLC25A5	HOPX
Donor 1	11686.13858	5242.27765	441.47646	135.47640	1039.36406	68.97138	6278.20041	167.44563	4982.02687	7595.17642	6224.37507
Donor 2	20074.81423	6883.60588	377.28792	88.37637	1506.46239	79.17396	4829.43874	12.99610	3863.83449	6506.04545	5503.04715
Donor 3	17535.92266	3904.66924	180.54552	173.58964	971.67212	91.25269	7884.95952	16.45213	3816.70421	8456.61642	7160.58279
Donor 4	30651.92075	5329.85262	226.98251	78.20144	1993.99197	70.49724	4262.24242	17.06604	4129.12163	9325.77264	10183.67044
Donor 5	24793.28728	4080.52575	258.54555	92.09127	1672.38429	86.28244	4552.97802	9.80021	3535.47046	8987.29253	10462.26782
EM_013676	33651.07565	17305.63671	1983.03272	934.20737	ND*	ND	ND	ND	ND	ND	ND
EM_013796	157072.97066	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_076846	46524.96830	27904.47722	9622.92316	820.55884	4758.07089	ND	ND	ND	ND	ND	ND
EM_075933	30328.74277	220892.73376	3415.90509	164.01675	ND	ND	ND	ND	ND	ND	ND
EM_076035	38093.39343	9784.35974	1120.62426	672.49861	2965.74361	94.42692	3568.25351	33.23519	4479.12441	18251.62518	ND
EM_008015	12950.96956	51193.23401	3001.02720	724.72959	2929.96690	ND	ND	ND	ND	ND	ND
EM_075155	17934.26722	11898.95074	2595.27814	477.53258	ND	ND	ND	ND	ND	ND	ND
EM_022044	135295.03942	14882.44944	371.63095	179.22659	ND	ND	ND	ND	ND	ND	ND
EM_013261	11168.81382	29790.47308	2940.63651	1443.86534	1587.45818	109.85650	3942.46390	ND	ND	ND	ND
EM_076814	65569.17978	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_0895	21922.35830	86219.84835	1167.08298	99.76071	1897.00691	ND	ND	ND	ND	ND	ND
EM_0367	57896.98703	18762.18891	2912.72963	ND	ND	ND	ND	ND	ND	ND	ND
EM_023321	321675.79783	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND

EM_013283	9743.03827	20878.86713	1966.03832	492.13530	1342.56261	ND	ND	ND	ND	ND	ND
EM_077285	34600.36384	14171.66985	5702.84096	1176.09098	6193.18619	98.07155	ND	ND	ND	ND	ND
EM_015781	168266.73552	44221.64881	1789.82002	134.19983	ND	ND	ND	ND	ND	ND	ND
EM_0017	115184.18060	23352.68245	2690.35591	420.61513	7419.78863	ND	ND	ND	ND	ND	ND
EM_004032	86716.89648	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_076295	178407.44678	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_013860	29973.98800	28807.98375	4762.56166	131.42202	2948.78459	68.00195	ND	ND	ND	ND	ND
EM_075355	344407.93077	61971.73925	2681.17556	235.94922	19752.84787	ND	ND	ND	ND	ND	ND
EM_075935	146039.54370	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_016460	237755.56857	59933.37476	682.80590	483.92448	16914.43442	29.16604	99.79415	44.27615	ND	ND	ND
EM_075314	355361.29674	49453.24447	9227.77356	402.68567	20409.16495	ND	ND	ND	ND	ND	ND
EM_008007	254607.90920	67465.42867	992.84758	483.02655	14609.20408	44.24546	223.14419	107.11175	2991.03532	50231.80894	ND
EM_0727	356053.62865	67193.53008	2816.72904	479.99178	23281.73892	30.46923	240.73290	149.17687	ND	ND	ND
EM_0061	17145.71404	101920.52341	1740.63431	64.81578	2491.72388	ND	ND	ND	ND	ND	ND
EM_0763	260395.24077	55827.71615	1247.92940	221.66486	17929.66423	19.15526	119.21772	97.86812	751.39969	58539.05294	70991.04443
EM_080293	143981.92155	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
EM_075159	214498.22060	66078.74121	2095.65547	362.43341	10943.34389	80.50910	ND	ND	ND	ND	ND
EM_0894	236264.12165	58918.06637	1901.78874	207.10724	16310.79821	43.12587	176.05320	ND	ND	ND	ND
EM_080093	181885.12651	34951.20373	1670.58390	110.48702	ND	ND	ND	ND	ND	ND	ND
EM_077046	203901.39120	53027.73441	2480.03037	253.19339	11606.18183	53.28530	ND	ND	ND	ND	ND
EM_076569	221858.66276	49233.76862	3069.07116	623.63133	11649.51735	63.47506	458.15880	79.25875	4123.28391	45428.35664	53796.97259

EM_076866	255992.02461	41762.89035	7384.01589	623.18320	16563.03512	92.57343	593.26189	188.37908	3671.42580	45667.01961	50802.59509
EM_0783	309091.94034	42245.03357	7239.75854	335.56032	15839.15371	70.83181	185.67578	211.62932	1950.35170	49253.68827	ND
EM_076545	221225.77026	51330.26437	1360.79225	238.70682	11669.60265	51.40996	110.66112	97.23544	ND	ND	ND
EM_076882	276796.70836	55320.95897	2976.70119	352.03694	12640.54114	67.94587	184.23730	118.51083	3485.00267	43156.69717	54703.67801
EM_080108	139519.58797	13596.30214	520.24635	21.68663	ND	ND	ND	ND	ND	ND	ND

*ND = not determined due to insufficient RNA sample

Appendix Table 3. Characteristics of the cohort of 39 EVD patients with known outcome

Sample_ID	EBOV Ct value	Age	Sex*	EMLab site	Year of sampling	Profession	Comorbidity	Treatment	Outcome
EM_013676	26.33	8	F	Coyah	2015	other	no malaria	No	survivor
EM_013796	27.08	18	F	Coyah	2015	housewife	no malaria	No	survivor
EM_076846	28.45	25	F	Guéckédou	2014	NA**	no malaria	No	survivor
EM_075933	27.09	30	F	Guéckédou	2014	NA	no malaria	No	survivor
EM_076035	29.58	35	F	Guéckédou	2014	NA	no malaria	No	survivor
EM_008015	31.64	40	F	Coyah	2015	housewife	no malaria	No	survivor
EM_075155	28.83	45	F	Guéckédou	2014	housewife	no malaria	No	survivor
EM_022044	29.24	50	F	Coyah	2015	housewife	no malaria	No	survivor
EM_013261	27.37	54	F	Coyah	2015	housewife	no malaria	No	survivor
EM_076814	28.97	65	F	Guéckédou	2014	NA	no malaria	No	survivor
EM_0895	25.18	15	M	Guéckédou	2014	NA	no malaria	No	survivor
EM_0367	27.35	25	M	Guéckédou	2014	hunter, student	no malaria	No	survivor
EM_023321	27.08	30	M	Coyah	2015	housewife	no malaria	No	survivor
EM_013283	34.57	35	M	Coyah	2015	farmer	no malaria	No	survivor
EM_077285	28.1	40	M	Guéckédou	2014	NA	no malaria	No	survivor
EM_015781	32.18	45	M	Coyah	2015	other	no malaria	Favipiravir	survivor
EM_0017	27.65	51	M	Guéckédou	2014	farmer	no malaria	No	survivor
EM_004032	27.4	60	M	Guéckédou	2014	other	no malaria	Favipiravir	survivor
EM_076295	28.58	71	M	Guéckédou	2014	other	no malaria	No	survivor
EM_013860	27.95	5	F	Coyah	2015	other	no malaria	No	survivor
EM_075355	16.12	10	F	Guéckédou	2014	NA	no malaria	No	fatal
EM_075935	20.45	20	F	Guéckédou	2014	NA	no malaria	No	fatal
EM_016460	17.07	25	F	Coyah	2015	NA	no malaria	No	fatal
EM_075314	18.25	30	F	Guéckédou	2014	housewife	no malaria	No	fatal
EM_008007	17.08	40	F	Coyah	2015	housewife	no malaria	No	fatal

EM_0727	16.34	50	F	Guéckédou	2014	NA	no malaria	No	fatal
EM_0061	20.98	55	F	Guéckédou	2014	housewife	no malaria	No	fatal
EM_0763	16.59	65	F	Guéckédou	2014	NA	no malaria	No	fatal
EM_080293	19.3	70	F	Guéckédou	2014	housewife	no malaria	No	fatal
EM_075159	15.59	9	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_0894	17.91	20	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_080093	17.33	25	M	Guéckédou	2014	farmer	no malaria	No	fatal
EM_077046	15.12	30	M	Guéckédou	2014	other	no malaria	No	fatal
EM_076569	16.72	40	M	Guéckédou	2014	farmer	no malaria	No	fatal
EM_076866	16.22	50	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_0783	18.99	55	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_076545	15.99	65	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_076882	17.16	70	M	Guéckédou	2014	NA	no malaria	No	fatal
EM_080108	18.01	80	M	Guéckédou	2014	NA	no malaria	No	fatal

*F = female; M = male; **NA = data not available

Appendix Table 4. RNA concentration of the 64 EVD patients from the "blinded" cohort of samples

Sample_ID	RNA concentration ng/ µl
RNA-01	5.34
RNA-02	15.6
RNA-03	15.7
RNA-04	16.5
RNA-05	6.84
RNA-06	19
RNA-07	12.3
RNA-08	13.1
RNA-09	5.86
RNA-10	18.1
RNA-11	17.6
RNA-12	Low out of range
RNA-13	8.12
RNA-14	12.8
RNA-15	6.14
RNA-16	Low out of range
RNA-17	21.2
RNA-18	17.1
RNA-19	7.26
RNA-20	7.2
RNA-21	16.8
RNA-22	6.28
RNA-23	9.02
RNA-24	14.8
RNA-25	6.08
RNA-26	8.08
RNA-27	Low out of range
RNA-28	10.4
RNA-29	9.36
RNA-30	13.2
RNA-31	7.24
RNA-32	25.4

RNA-33	20.8
RNA-34	11.4
RNA-35	30.2
RNA-36	8.82
RNA-37	17.8
RNA-38	12.3
RNA-39	7.94
RNA-40	7.84
RNA-41	18.3
RNA-42	14.5
RNA-43	22.6
RNA-44	8.84
RNA-45	12.1
RNA-46	4
RNA-47	16.3
RNA-48	4.2
RNA-49	Low out of range
RNA-50	9.16
RNA-51	8.32
RNA-52	10.2
RNA-53	12.9
RNA-54	7.28
RNA-55	4.6
RNA-56	10.7
RNA-57	11.9
RNA-58	Low out of range
RNA-59	7.6
RNA-60	5.42
RNA-61	10.6
RNA-62	7.76
RNA-63	4.2
RNA-64	6.38

Appendix Table 5. Transcript abundance of the 4 candidate predictive genes quantified by RT-qPCR from testing 5 control samples and 64 EVD patients from the "blinded" cohort of samples

Sample_ID	VCAM1	ISG15	MS4A4A	TUBG1
Donor 1	8727.42752	7647.40248	552.44773	148.26357
Donor 2	10171.63980	6979.39775	519.78395	76.63345
Donor 3	7985.83130	5166.11509	433.83158	161.12324
Donor 4	18119.19344	4583.78006	1009.72090	78.82248
Donor 5	14519.12060	3580.18236	930.29021	122.21942
RNA-01	191505.82976	31995.48060	5422.97335	14.37721
RNA-02	141523.40691	36012.58353	6043.55946	596.68779
RNA-03	142194.47161	41119.15293	5972.98089	182.20802
RNA-04	148528.34159	49763.01116	6260.36657	301.53413
RNA-05	83536.80180	80488.10555	4217.05054	206.34228
RNA-06	112876.75554	49103.06291	5236.35980	1027.02770
RNA-07	135617.65320	48329.88493	6143.10226	400.77303
RNA-08	159708.10817	54694.06341	6696.27085	202.18910
RNA-09	38651.38504	51087.71150	2363.78144	369.63809
RNA-10	181922.05216	40559.48239	6982.01957	778.84435
RNA-11	138647.00602	53496.80872	5781.86631	507.36030
RNA-12	57166.74090	ND*	ND	ND

RNA-13	57927.73540	33835.11718	3047.07598	984.99033
RNA-14	159723.31369	44741.09007	6436.72749	564.14681
RNA-15	40270.06140	83782.80418	2146.72670	278.80286
RNA-16	No amplification	No amplification	No amplification	No amplification
RNA-17	109514.64318	40530.81294	4656.30179	531.30612
RNA-18	185028.50246	50948.25120	7415.50741	330.19750
RNA-19	84307.07673	72954.41781	4109.99266	301.97073
RNA-20	109400.69426	63308.59169	5110.38362	260.63828
RNA-21	140288.36653	56010.21721	5869.70586	847.61888
RNA-22	113097.65187	260186.35146	5576.08405	142.93491
RNA-23	168801.62137	50988.52607	7965.36387	394.42508
RNA-24	164964.98138	44487.62455	7185.38030	566.33556
RNA-25	74498.30018	42375.19295	3972.84285	243.28418
RNA-26	100880.00170	42564.89568	4423.05421	327.19850
RNA-27	91981.11845	ND	ND	ND
RNA-28	197630.77588	57921.53986	8417.04157	387.17359
RNA-29	178514.59699	49710.67081	6391.73778	231.17918
RNA-30	201621.22005	43958.05562	7045.60441	231.86884
RNA-31	113233.94634	44500.09451	5135.43058	182.87150
RNA-32	163541.04848	63327.19968	7001.91104	367.79269
RNA-33	148227.54484	30946.98641	7791.17255	381.88785

RNA-34	194252.97821	42631.86562	10602.38312	138.22268
RNA-35	201864.02412	35256.93137	10617.42410	474.47922
RNA-36	120719.74055	30530.68669	6432.52493	260.25497
RNA-37	159239.42998	38324.70333	7763.13517	305.34760
RNA-38	95491.56975	22446.28982	4741.92252	211.52420
RNA-39	85767.60423	29307.55569	5656.40443	262.22113
RNA-40	112521.62892	118322.56817	6791.60136	96.97805
RNA-41	166672.41302	54848.84994	8091.59067	387.80988
RNA-42	136107.75041	30189.48956	7788.99375	959.61714
RNA-43	126855.05391	30446.36602	10706.59167	878.55790
RNA-44	99406.64730	26282.68474	5659.12417	297.58613
RNA-45	148893.18087	42154.29448	7857.72888	314.60170
RNA-46	11408.55441	76098.31902	2802.11322	ND
RNA-47	164024.64200	36909.92393	9008.61427	441.64695
RNA-48	121137.00319	29886.52512	7223.76408	118.29060
RNA-49	1855.54461	ND	ND	ND
RNA-50	161449.63004	39924.38916	8294.15576	333.72382
RNA-51	125333.14966	43743.43750	7279.17457	159.76693
RNA-52	201798.26263	40020.98006	9321.72258	185.51054
RNA-53	189912.28293	34670.08797	8978.17070	189.10925
RNA-54	15614.31154	205547.94032	2314.59587	154.92555

RNA-55	10034.07940	111601.46716	2202.88289	174.95566
RNA-56	112759.06300	47759.74868	6628.27250	569.26410
RNA-57	126330.50826	30079.54571	6922.29436	250.58458
RNA-58	128996.38712	ND	ND	ND
RNA-59	80895.27259	15178.91557	4703.89702	371.07360
RNA-60	9298.36357	137532.61728	607.84116	162.85828
RNA-61	178573.29717	30084.19548	7789.31235	132.83524
RNA-62	34890.25513	13797.35440	2777.64092	1006.85627
RNA-63	7415.23029	1394.60925	517.43595	ND
RNA-64	17457.89919	11630.51198	1569.12823	1163.45488

*ND = not determined due to insufficient RNA sample

Appendix Table 6. Characteristics of the 64 EVD patients from the "blinded" cohort of samples

Sample_D	EBOV_Ct	Age	Sex	EMLab site	Year of sampling	Comorbidity	Treatment	Outcome
RNA-01	19.1	58	F	Guèckèdou	2014	no malaria	No	survivor
RNA-02	17.25	55	M	Guèckèdou	2014	no malaria	No	survivor
RNA-03	17.09	40	F	Guèckèdou	2014	no malaria	No	survivor
RNA-04	19.08	18	M	Guèckèdou	2014	no malaria	No	survivor
RNA-05	17.73	24	M	Guèckèdou	2014	no malaria	No	survivor
RNA-06	19.76	70	F	Guèckèdou	2014	no malaria	No	survivor
RNA-07	19.04	21	M	Guèckèdou	2014	no malaria	No	survivor
RNA-08	14.42	12	M	Guèckèdou	2014	no malaria	No	survivor
RNA-09	23.19	70	M	Guèckèdou	2014	no malaria	No	survivor
RNA-10	19.54	5	F	Guèckèdou	2014	no malaria	No	survivor
RNA-11	16.08	25	F	Guèckèdou	2014	no malaria	No	survivor
RNA-12	22.05	50	M	Guèckèdou	2014	no malaria	No	survivor
RNA-13	21.29	60	M	Guèckèdou	2014	no malaria	No	survivor
RNA-14	15.99	11	F	Guèckèdou	2014	no malaria	No	survivor
RNA-15	19.35	30	F	Guèckèdou	2014	no malaria	No	survivor
RNA-16	15.77	38	F	Guèckèdou	2014	no malaria	No	survivor
RNA-17	17.71	10	M	Guèckèdou	2014	no malaria	No	survivor
RNA-18	18.5	30	M	Guèckèdou	2014	no malaria	No	survivor

RNA-19	16.52	3	F	Guèckèdou	2014	no malaria	No	survivor
RNA-20	17.44	19	F	Guèckèdou	2014	no malaria	No	survivor
RNA-21	15.24	60	F	Guèckèdou	2014	no malaria	No	survivor
RNA-22	17.24	6	M	Coyah	2015	no malaria	No	survivor
RNA-23	19.55	16	F	Guèckèdou	2014	no malaria	No	survivor
RNA-24	19.95	45	F	Guèckèdou	2014	no malaria	No	survivor
RNA-25	19.31	3	M	Guèckèdou	2014	no malaria	No	survivor
RNA-26	19.81	45	M	Guèckèdou	2014	no malaria	No	survivor
RNA-27	17.9	55	F	Guèckèdou	2014	no malaria	No	fatal
RNA-28	14.66	75	F	Guèckèdou	2014	no malaria	No	fatal
RNA-29	14.61	45	F	Guèckèdou	2014	no malaria	No	fatal
RNA-30	15.3	48	M	Guèckèdou	2014	no malaria	No	fatal
RNA-31	15.51	18	M	Guèckèdou	2014	no malaria	No	fatal
RNA-32	13.5	14	F	Guèckèdou	2014	malaria	No	fatal
RNA-33	14.91	38	F	Guèckèdou	2014	no malaria	No	fatal
RNA-34	19.95	30	F	Guèckèdou	2014	no malaria	No	fatal
RNA-35	16.69	0.6	M	Guèckèdou	2014	no malaria	No	fatal
RNA-36	14.99	6	F	Guèckèdou	2014	no malaria	No	fatal
RNA-37	14.6	24	M	Guèckèdou	2014	no malaria	No	fatal
RNA-38	15.34	30	M	Guèckèdou	2014	no malaria	No	fatal
RNA-39	15.07	50	F	Guèckèdou	2014	no malaria	No	fatal

RNA-40	14.03	15	M	Guèckèdou	2014	no malaria	No	fatal
RNA-41	15.33	6	M	Guèckèdou	2014	no malaria	No	fatal
RNA-42	15.82	38	M	Guèckèdou	2014	no malaria	No	fatal
RNA-43	12.87	53	M	Guèckèdou	2014	no malaria	No	fatal
RNA-44	16.33	60	F	Guèckèdou	2014	no malaria	No	fatal
RNA-45	15.4	45	M	Guèckèdou	2014	no malaria	No	fatal
RNA-46	22.1	0.5	F	Guèckèdou	2014	no malaria	No	fatal
RNA-47	17.44	20	F	Guèckèdou	2014	no malaria	No	fatal
RNA-48	19.91	80	F	Coyah	2015	no malaria	No	fatal
RNA-49	20.27	73	M	Coyah	2015	no malaria	No	fatal
RNA-50	16.26	59	M	Coyah	2015	no malaria	No	fatal
RNA-51	15.2	25	F	Guèckèdou	2014	no malaria	No	fatal
RNA-52	16.22	60	M	Coyah	2015	no malaria	No	fatal
RNA-53	19.1	24	F	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-54	23	40	F	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-55	23.64	60	F	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-56	17.41	29	M	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-57	17.53	38	M	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-58	18.9	50	M	Coyah	2015	no malaria	Favipiravir	fatal
RNA-59	26.42	14	F	Guèckèdou	2014	no malaria	Favipiravir	fatal
RNA-60	20.42	18	F	Guèckèdou	2014	no malaria	Favipiravir	fatal

RNA-61	22.17	50	F	Guèckèdou	2014	no malaria	Favipiravir	survivor
RNA-62	27.15	18	M	Guèckèdou	2014	no malaria	Favipiravir	survivor
RNA-63	20.06	19	M	Coyah	2015	no malaria	Favipiravir	survivor
RNA-64	29.15	62	M	Guèckèdou	2014	no malaria	Favipiravir	survivor

Appendix Table 7. Characteristics of the 48 DNase-treated samples and MS4A4A transcript abundance quantified by RT-qPCR

Sample_ID	Outcome	Days since symptom onset	Age	Sex	EMLab site	Year	Comorbidity	Treatment	EBOV Ct value	MS4A4A copy number
RNA-02	survivor	11	55	M	Guèckèdou	2014	no malaria	No	17.25	200.39582
RNA-03	survivor	7	40	F	Guèckèdou	2014	no malaria	No	17.09	68.81439
RNA-04	survivor	0	18	M	Guèckèdou	2014	no malaria	No	19.08	167.07763
RNA-05	survivor	4	24	M	Guèckèdou	2014	no malaria	No	17.73	7.78317
RNA-06	survivor	10	70	F	Guèckèdou	2014	no malaria	No	19.76	803.05306
RNA-07	survivor	7	21	M	Guèckèdou	2014	no malaria	No	19.04	84.78420
RNA-08	survivor	8	12	M	Guèckèdou	2014	no malaria	No	14.42	18.84758
RNA-10	survivor	8	5	F	Guèckèdou	2014	no malaria	No	19.54	548.45975
RNA-11	survivor	3	25	F	Guèckèdou	2014	no malaria	No	16.08	99.89566
RNA-13	survivor	9	60	M	Guèckèdou	2014	no malaria	No	21.29	9.82720
RNA-14	survivor	6	11	F	Guèckèdou	2014	no malaria	No	15.99	41.39615
RNA-17	survivor	8	10	M	Guèckèdou	2014	no malaria	No	17.71	63.33648
RNA-18	survivor	8	30	M	Guèckèdou	2014	no malaria	No	18.5	428.45637
RNA-19	survivor	2	3	F	Guèckèdou	2014	no malaria	No	16.52	25.02896
RNA-20	survivor	3	19	F	Guèckèdou	2014	no malaria	No	17.44	26.27417
RNA-21	survivor	4	60	F	Guèckèdou	2014	no malaria	No	15.24	392.66976
RNA-22	survivor	3	6	M	Coyah	2015	no malaria	No	17.24	14.33285
RNA-23	survivor	3	16	F	Guèckèdou	2014	no malaria	No	19.55	44.85136

RNA-24	survivor	8	45	F	Guèckèdou	2014	no malaria	No	19.95	81.07485
RNA-26	survivor	9	45	M	Guèckèdou	2014	no malaria	No	19.81	18.33814
RNA-28	fatal	NA*	75	F	Guèckèdou	2014	no malaria	No	14.66	248.37024
RNA-29	fatal	8	45	F	Guèckèdou	2014	no malaria	No	14.61	165.94605
RNA-30	fatal	2	48	M	Guèckèdou	2014	no malaria	No	15.3	27.67241
RNA-31	fatal	3	18	M	Guèckèdou	2014	no malaria	No	15.51	39.05938
RNA-32	fatal	4	14	F	Guèckèdou	2014	malaria	No	13.5	943.57356
RNA-33	fatal	5	38	F	Guèckèdou	2014	no malaria	No	14.91	377.96708
RNA-34	fatal	10	30	F	Guèckèdou	2014	no malaria	No	19.95	81.76147
RNA-36	fatal	7	6	F	Guèckèdou	2014	no malaria	No	14.99	69.08459
RNA-37	fatal	5	24	M	Guèckèdou	2014	no malaria	No	14.6	94.66044
RNA-38	fatal	6	30	M	Guèckèdou	2014	no malaria	No	15.34	116.05709
RNA-39	fatal	2	50	F	Guèckèdou	2014	no malaria	No	15.07	80.86238
RNA-40	fatal	5	15	M	Guèckèdou	2014	no malaria	No	14.03	13.29102
RNA-41	fatal	5	6	M	Guèckèdou	2014	no malaria	No	15.33	127.25070
RNA-42	fatal	1	38	M	Guèckèdou	2014	no malaria	No	15.82	171.92892
RNA-44	fatal	7	60	F	Guèckèdou	2014	no malaria	No	16.33	40.86581
RNA-45	fatal	3	45	M	Guèckèdou	2014	no malaria	No	15.4	171.84369
RNA-47	fatal	8	20	F	Guèckèdou	2014	no malaria	No	17.44	440.83137
RNA-50	fatal	10	59	M	Coyah	2015	no malaria	No	16.26	100.06306
RNA-51	fatal	3	25	F	Guèckèdou	2014	no malaria	No	15.2	70.04940
RNA-52	fatal	2	60	M	Coyah	2015	no malaria	No	16.22	315.39890
RNA-53	fatal	3	24	F	Guèckèdou	2014	no malaria	Favipiravir	19.1	81.70676

RNA-54	fatal	2	40	F	Guèckèdou	2014	no malaria	Favipiravir	23	221.08921
RNA-56	fatal	3	29	M	Guèckèdou	2014	no malaria	Favipiravir	17.41	88.43277
RNA-57	fatal	5	38	M	Guèckèdou	2014	no malaria	Favipiravir	17.53	71.75741
RNA-59	fatal	1	14	F	Guèckèdou	2014	no malaria	Favipiravir	26.42	116.40438
RNA-61	survivor	4	50	F	Guèckèdou	2014	no malaria	Favipiravir	22.17	22.83285
RNA-62	survivor	6	18	M	Guèckèdou	2014	no malaria	Favipiravir	27.15	70.65749
RNA-64	survivor	0	62	M	Guèckèdou	2014	no malaria	Favipiravir	29.15	12.14904

*Data not available

Appendix Table 8. Transcript abundance of VCAM1, MS4A4A and ISG15 obtained by multiplex RT-qPCR from testing 47 clinical samples from the "blinded" cohort of patients.

Sample_ID	VCAM1	MS4A4A	ISG15
RNA-02	31,866	11,972	42,626
RNA-03	34,991	13,102	53,476
RNA-04	33,913	12,537	64,295
RNA-05	17,086	6,455	60,413
RNA-06	23,277	10,179	70,236
RNA-07	28,236	10,347	69,070
RNA-08	34,226	13,064	67,472
RNA-10	39,061	17,081	51,146
RNA-11	28,549	11,282	75,545
RNA-13	11,592	3,959	43,807
RNA-14	28,701	11,739	48,129
RNA-17	21,757	7,986	49,809
RNA-18	26,390	10,502	56,446
RNA-19	9,712	4,798	101,877
RNA-21	17,623	7,465	70,147
RNA-23	21,338	8,588	75,999
RNA-24	24,355	9,872	53,531
RNA-26	11,290	5,220	46,066
RNA-28	30,528	11,473	97,782
RNA-29	24,688	9,722	64,362
RNA-30	26,582	10,484	44,988
RNA-32	24,970	10,913	65,214
RNA-33	21,553	8,119	56,782
RNA-34	29,278	10,565	75,607
RNA-35	30,957	12,676	61,660
RNA-36	18,148	8,587	49,619

RNA-37	25,060	10,005	29,192
RNA-38	10,731	3,571	19,248
RNA-40	15,103	6,098	258,775
RNA-41	26,060	11,419	128,715
RNA-42	16,296	8,365	35,777
RNA-43	17,344	11,988	45,639
RNA-44	10,818	4,834	31,353
RNA-45	19,132	8,034	60,780
RNA-47	22,718	12,830	51,252
RNA-50	25,142	8,719	72,981
RNA-51	16,961	6,729	57,019
RNA-52	29,386	11,902	71,417
RNA-53	25,043	9,566	54,496
RNA-54	2,206	3,468	579,028
RNA-56	15,198	5,113	73,216
RNA-57	15,462	5,523	34,352
RNA-59	7,694	3,426	21,727
RNA-60	999	716	283,074
RNA-61	23,654	9,969	44,515
RNA-62	3,658	2,270	23,043
RNA-64	2,035	1,332	20,475