Aus dem Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie - IBE
Lehrstuhl für Epidemiologie

Institut der Ludwig-Maximilians-Universität München

Characterisation of Herpesviruses in KORA Cohort and Association with Type 2 Diabetes

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

vorgelegt von

Tim Wölfle

aus

Freiburg im Breisgau

Jahr

2021

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

Erste Gutachterin:	Prof. Dr. Annette Peters
Zweiter Gutachter:	Prof. Dr. med. Jochen Seißler
Dritte Gutachterin:	Prof. Dr. Anne-Laure Boulesteix
ggf. weitere Gutachter:	_____
	_____
Mitbetreuung durch den promovierten Mitarbeiter:	_____
Dekan:	Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 25.10.2021

Contents

Zusammenfassung

Einleitung: Herpesviren sind eine Klasse hochansteckender DNA–Viren mit hoher Prävalenz und geringer Pathogenität. Diabetes mellitus Typ 2 ist eine der häufigsten Stoffwechselerkrankungen mit weltweit steigender Prävalenz und ist ein Hauptrisikofaktor für kardiovaskuläre Erkrankungen, darunter Myokardinfarkt und Schlaganfall. Diese Studie untersucht die Assoziation des Serostatus‘ von 7 humanen Herpesviren mit Diabetes mellitus Typ 2: Herpes–Simplex–Virus 1 & 2, Varizella–Zoster–Virus, Epstein–Barr–Virus, Cytomegalovirus und Humanes Herpesvirus 6 & 7.

Methoden: Zwei Zeitpunkte der bayrischen KORA Kohortenstudie werden verwendet: „F4“ mit 2‘950 TeilnehmerInnen, die zwischen 2006–2008 untersucht wurden, und „FF4“ mit 2‘129 TeilnehmerInnen, die 2013/2014 untersucht wurden. Beide haben eine Überschneidung von 1‘967 TeilnehmerInnen. Alle von ihnen haben zweimalige Testungen des oralen Glukose Toleranztests und einer viralen Multiplex–Serologie erhalten. Logistische Regressionsmodelle wurden verwendet, um den binären Serostatus mit Prävalenz und Inzidenz von (Prä)Diabetes zu vergleichen, generalisierte Schätzmodelle, um die beiden Zeitpunkte für die Prävalenzanalysen zu kombinieren, und lineare Regressionsmodelle für Vergleiche mit HbA1c, jeweils korrigiert für Alter und Geschlecht.

Ergebnisse: Alle humanen Herpesviren außer das Varizella–Zoster–Virus zeigten eine höhere Prävalenz in Frauen als in Männern. Serostatus und Antikörperreaktivität von Herpes–Simplex–Virus 1 & 2, Epstein–Barr–Virus und Cytomegalovirus waren signifikant und konsistent mit höherem Alter assoziiert. Schwache Evidenz für erhöhte Antikörperreaktivität in den kalten als in den warmen Monaten wurde für alle Herpesviren außer für das Herpes–Simplex–Virus 2 und das Humane Herpesvirus 6 gefunden.

Das Herpes–Simplex–Virus 2 und das Cytomegalovirus waren signifikant mit Inzidenz von (Prä)Diabetes assoziiert, korrigiert für Alter und Geschlecht. Das Herpes–Simplex–Virus 2 war ebenfalls signifikant mit HbA1c assoziiert, einem wichtigen Labormarker für Langzeitblutzucker, korrigiert sowohl für Alter und Geschlecht, als auch für Diabetes–Status. Keine Dosiseffekte von viraler Antikörperreaktivität auf Prävalenz oder Inzidenz von (Prä)Diabetes wurden gefunden.

Diskussion: Assoziationen des Cytomegalovirus mit Diabetes–Prävalenz wurden in den letzten 20 Jahren mehrfach berichtet. Eine neue Erkenntnis dieser Arbeit ist ein Zusammenhang des Cytomegalovirus–Serostatus’ mit der (Prä)Diabetes–Inzidenz. Dies bestätigt die Ergebnisse von Yoo und Kollegen, die 2019 eine Assoziation von durchgemachter Cytomegalovirus–

Erkrankung mit Diabetes–Inzidenz zeigten, jedoch nicht von dessen Serostatus. Für Herpes–Simplex–Virus 2 wurde bisher noch keine Assoziation mit der Inzidenz von (Prä)Diabetes gezeigt, wofür die vorliegende Arbeit erste Hinweise liefert. Diese Ergebnisse unterstreichen die Notwendigkeit weiterer Forschung und verstärkter Anstrengung bei viralen Präventionsmaßnahmen und bei der Entwicklung von Impfstoffen gegen Herpesviren.

Acknowledgements

I sincerely thank my first two supervisors, Prof. Dr. Annette Peters from LMU München and Helmholtz Zentrum München and Prof. Dr. Marc Chadeau–Hyam from Imperial College London for their invaluable discussions and suggestions. I also thank Prof. Dr. Jochen Seißler from LMU München for being the third supervisor of this thesis. Additionally, I would like to thank Dr. Birgit Linkohr for her help with the KORA application and Dr. Tim Waterboer for his help with understanding the technical details of the multiplex viral serology.

Abstract

Introduction: Herpesviruses are a class of highly prevalent and contagious DNA viruses with low pathogenicity. Type 2 diabetes is among the most important metabolic diseases with increasing worldwide prevalence and is one of the major risk factors for cardiovascular disease, including myocardial infarction and stroke. This study examines the association of serostatus of 7 human herpesviruses with type 2 diabetes, namely herpes simplex virus 1 & 2, varicella–zoster virus, Epstein–Barr virus, cytomegalovirus and human herpesvirus 6 & 7.

Methods: Two timepoints of the population cohort KORA were used: "F4" with 2,950 participants examined between 2006–2008 and "FF4" with 2,129 participants examined in 2013/2014, with an overlap of 1,967 participants. All of them have undergone both oral glucose tolerance tests and viral multiplex serology twice. Logistic regression models were used for comparison of binary viral serostatus with (pre)diabetes prevalence and incidence, generalised estimating equations were applied to combine both timepoints for analyses on (pre)diabetes prevalence, and linear regression was used for comparison with HbA1c, always adjusting for age and sex.

Results: All human herpesviruses except varicella–zoster virus were significantly more prevalent in women than in men. Serostatus and antibody reactivity of herpes simplex virus 1 & 2, Epstein–Barr virus and cytomegalovirus were significantly and consistently associated with older age. Weak evidence of increased antibody reactivity in the colder compared to the warmer months has been found for all herpesviruses except for herpes simplex virus 2 and human herpesvirus 6.

Herpes simplex virus 2 and cytomegalovirus were significantly associated with (pre)diabetes incidence, after adjustment for age and sex. Herpes simplex virus 2 was also significantly associated with HbA1c, an important laboratory marker of long–term blood sugar, after adjustment for both sex, age and even diabetes status. No dose effects of viral antibody reactivities with prevalence or incidence of (pre)diabetes were found.

Discussion: Associations of cytomegalovirus with diabetes prevalence have been reported multiple times for the past 20 years. This work's novel finding of an association of cytomegalovirus serostatus with (pre)diabetes incidence confirms a 2019 Korean study by Yoo and colleagues demonstrating an association of history of cytomegalovirus disease with diabetes incidence. Herpes simplex virus 2 has thus far not been associated with (pre)diabetes incidence, making this the first study to provide such evidence. These results highlight the need for further research, more efforts in viral prevention strategies and the development of effective vaccines against herpesviruses.

List of Figures

List of Tables

List of Abbreviations

CI	Confidence Interval
CMV	Cytomegalovirus (= HHV–5)
T2D	Type 2 Diabetes Mellitus
DNA	Deoxyribonucleic Acid
EBV	Epstein–Barr Virus (= HHV–4)
F4 / FF4	Follow–up timepoints of KORA S4 cohort (F4 ‘06–‘08, FF4 ‘13–‘14)
GEE	Generalised Estimating Equation
HHV–6, HHV–7	Human Herpesvirus 6 & 7
HIV	Human Immunodeficiency Virus
HSV–1, HSV–2	Herpes Simplex Virus 1 & 2 (= HHV–1 & 2)
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IR	Incidence Rate
IQR	Inter–Quartile–Range (25^{th} to 75^{th} percentile)
KORA	Kooperative Gesundheitsforschung in der Region Augsburg (Cooperative Health Research in the Region Augsburg)
KSHV	Kaposi’s Sarcoma–Associated Virus (= HHV–8)
MFI	Median Fluorescence Intensity (measure of antibody reactivity)

ns	not significant
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PY	Packyears (average number of packs smoked per day × years)
RR	Risk Ratio
SARS–CoV–2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	Standard Deviation
SE	Standard Error
VZV	Varicella–Zoster Virus (= HHV–3)

1 Introduction

The current SARS–CoV–2 pandemic puts the field of virology into the spotlight of international attention, utterly unanticipated at the beginning of this thesis in 2019. In June 2020, Rubino et al. have written a letter to the editor of the New England Journal of Medicine, stating there was a "bidirectional relationship between Covid–19 and diabetes" [1]. This interesting reflection points at the intersection of the two major branches of epidemiology, communicable and non–communicable disease, where virology has been known to play an important role at the latest since the discovery of the oncogenic Epstein–Barr–Virus (EBV) in 1964 by Epstein, Achong and Barr [2]. This thesis explores another such intersection: the association of herpesviruses, among the most prevalent viruses, with type 2 diabetes mellitus (T2D), one of the most prevalent metabolic diseases.

1.1 Overview of Human Herpesviruses

Herpesviridae are a taxonomical family in the order of Herpesvirales, consisting of 107 viral species according to the 2018b release of Virus Taxonomy, published and maintained by the International Committee on Taxonomy of Viruses [3]. They all share the same basic virion morphology defined by 4 layers: the core (containing linear, double–stranded DNA), the icosahedron capsid, the tegument and the lipid envelope [4].

Eight of these viral species are known to infect humans, named here in the order of their discovery: Herpes simplex viruses 1 & 2 (HSV–1 & 2), varicella–zoster virus (VZV), Epstein–Barr virus (EBV), cytomegalovirus (CMV), human herpesviruses 6, 7 and 8 (HHV–6, 7, 8). As all herpesviruses, they achieve lifelong latent infection in their hosts after systemic primary infection and depending on the cell type of latency they can be further subclassified into three subfamilies: Alphaherpesvirinae persist in neurons (HSV–1 & HSV–2, VZV), Betaherpesvirinae persist in monocytes / macrophages (CMV, HHV–6 & HHV–7) and Gammaherpesvirinae persist in lymphocytes (EBV, HHV–8) [4].

1.1.1 Herpes Simplex Virus 1 & 2

HSV–1 and HSV–2 are both transferred through intimate, close contact of mucosal areas like the mouth and genitalia or defective skin with infectious material stemming from mucocutaneous lesions. Whereas HSV–2 is mostly transmitted sexually, HSV–1 is more commonly transmitted non–sexually during childhood. Primary infection is usually asymptomatic or mild with mucocutaneous eruption (mucosal ulcers and skin vesicles) and potential systemic symptoms including fever, headache, myalgia and lymphadenopathy [5].

Upon primary infection of their host neurons, viruses are transported retrograde inside the axon towards the dorsal root ganglia, usually the trigeminal ganglion upon infection of the mouth or sacral ganglia upon infection of the genitalia. The root ganglia remain infected for life in a state of latency with potential sporadic reactivation ("recurrent infection") upon local or systemic stimuli, in which newly replicated viruses travel anterograde close to the initial infection site and cause mucocutaneous eruptions [5].

Clinical diseases are mainly orolabial herpes (more commonly HSV–1), genital herpes (traditionally more commonly HSV–2 but recently also more and more HSV–1 [6–8]) and keratoconjunctivitis (mostly HSV–1) potentially leading to vision loss. However, more rarely, severe life–threatening encephalitis (mostly HSV–1) and other syndromes of the central nervous system can be caused by primary as well as recurrent infection, especially in immunocompromised individuals [5]. Life–threatening neonatal infections through vertical transmission from mother to child occur rarely, usually during birth but sometimes also transplacental, leading to congenital infection with stigma like microcephaly and microphthalmia [5].

1.1.2 Varicella–Zoster Virus

The viral lifecycle of VZV is more complex than that of the other two human alphaherpesviruses discussed above. Its primary site of infection is respiratory mucosal epithelium after inhalation of viral particles from aerosol droplets or, more rarely, conjunctival mucosa after direct contact. Afterwards, VZV viruses infect T–cells and other leukocytes, leading to cell–associated viremia in the last days of the incubation period, which usually lasts 10–21 days. This allows the virus to travel to the skin, where the primary manifestation occurs [9].

Varicella (or "chickenpox") leads to mucosal ulcers and – most evident – erythematous papular skin lesions, crusting within about 2 days. In healthy children, about 100–300 such lesions typically occur [9]. These mucocutaneous lesions can involve sensory nerve axons, allowing the neurotropic VZV virus to retrogradely infect sensory ganglia, where the virus can achieve latency and stay dormant. Years or even decades later, dermatome–specific reactivation can occur, resulting in painful herpes zoster ("shingles") [9,10].

While HSV–1 is a neurotropic alphaherpesvirus as well, VZV latency is different from HSV–1 insofar as that reactivation becomes more likely with older age, it usually occurs only once (not multiple times) and no reactivation stimuli are known. In addition, zoster lesions are dermatome–specific and very painful whereas HSV–1 lesions are focal and usually not painful. Asymptomatic reactivation with virus shedding is frequent for HSV–1 but does not occur with VZV [10].

Due to the neurotropic nature of VZV, it can lead to more dangerous diseases than skin varicella and zoster, especially in immunocompromised individuals. Meningoencephalitis and cerebellar ataxia used to be common complications in children before vaccination campaigns started. Myelitis, optic neuritis and Guillain–Barré syndrome have also been described [9]. Severe life–threatening lung edema can occur in adults, who are more susceptible to varicella pneumonia than children. Varicella hepatitis and coagulopathy are especially problematic in the immunocompromised. Varicella embryopathy upon primary infection of the mother within the first half of pregnancy can lead to microcephaly and intrauterine encephalitis with seizures and mental retardation among other defects, whereas herpes zoster does not seem to increase the risk for varicella embryopathy [9].

1.1.3 Epstein–Barr Virus

EBV (also called human herpesvirus 4 by order of discovery) is one of the two human gammaherpesviruses – the other being Kaposi's sarcoma–associated virus (also called human herpesvirus 8), which was not part of the KORA study. The primary mode of transmission of EBV is salivary / oral contact whereas secondary and rarer modes of transmissions are sexual and parenteral (e.g. blood transfusion and organ transplantation) [11]. The main cellular reservoir are lymphocytes, mostly B–cells expressing complement receptor 2, in which EBV establishes lifelong latency. Intermittent lytic reactivation with discharge of infectious virions in the oral cavity is common [12].

In most cases, infection with EBV remains asymptomatic, especially in infancy and childhood, the first peak of infection due to oral contact of parents to infants. The second peak of infection usually occurs in teenagers due to oral contact of intimate partners [13] and is symptomatic in less than 50%. The main clinical presentation of EBV is infectious mononucleosis, also called glandular fever or "kissing disease", comprising pharyngitis, fever, cervical lymphadenopathy, fatigue and sometimes hepatosplenomegaly [11]. Lymphocytosis is common and consists of initial proliferation of EBV–infected B–cells followed by reactive proliferation of T–cells and NK–cells. While most symptoms usually last a few weeks, fatigue can carry on for months [11].

While the level of infection of EBV is more than 90% and thus nearly ubiquitous in human adult populations [11,13], some carriers develop malignancies, especially immunocompromised individuals with immunodeficiency, either congenital or acquired (e.g. through HIV or post organ–transplantation) [11]. EBV–associated malignancies include Hodgkin's and Non–Hodgkin's (e.g. Burkitt's) B–cell–lymphomas as well as nasopharyngeal and gastric carcinoma and sarcoma [11].

In addition to malignancies, EBV has also been reported to cause to a number of neurologic diseases including encephalitis, meningitis and myelitis [11]. Multiple sclerosis, a multifactorial inflammatory demyelinating disease of the central nervous system with unknown cause, is thought to be potentially associated with EBV infection [14–16].

1.1.4 Cytomegalovirus

CMV was the first betaherpesvirus to be identified. Its main methods of transmission are thought to be contact transmission (e.g. hand–to–mouth contact with infectious fluids like saliva among toddlers), sexual transmission, nosocomial transmission (e.g. organ / bone marrow transplantation or blood perfusion) and vertical transmission (e.g. intrauterine transplacental transmission or postnatally through breast feeding) [17].

Congenital CMV infection can lead to severe neurological birth defects including microcephaly leading to learning disability, chorioretinitis leading to vision loss and sensorineural hearing loss [16]. In immunocompetent adults however, infection is mostly subclinical, even though CMV is the second most common cause of infectious mononucleosis after EBV [18].

In immunocompromised individuals on the other hand, CMV can lead to devastating multi–organ failure. Before specific antiviral therapy was available, CMV was one of the number one death causes in post–transplant allograft recipients as well as in AIDS patients [17].

Lastly, the role of CMV in chronic non–communicable diseases like certain human cancers and atherosclerosis is highly debated [17].

1.1.5 Human Herpesvirus 6A, 6B & 7

HHV–6 & HHV–7 have been discovered in 1986 and 1990 respectively [19,20]. They are classified in the roseolovirus genus of the hetaherpesvirus family as all of them – but most commonly HHV–6B – can cause exanthema subitum at primary infection, also called roseola infantum, “3–day–fever” or “sixth disease” [21]. The main symptom is a pink rash usually lasting less than three days and often preceded by high fever. Complications include febrile seizures and encephalitis and are rare in healthy infants but more common in immunocompromised individuals [21].

As with most herpesviruses, the prevalence is very high and the main method of transmission appears to be through saliva in early childhood. HHV–6 seroconversion usually happens around the age of one year and HHV–7 seroconversion usually happens a little later around the age of two years [22]. Even though HHV–6A and HHV–6B share 90% of their genome, they are often considered as two biologically and immunologically different variants [23].

Interestingly, both HHV–6 variants are able to fully integrate their genomes into the chromosome telomeres of infected host cells, setting them apart from other herpesviruses, which keep their genomes in circular episomes [24]. When germline cells are infected, the virus can be passed on vertically to offspring, leading to inherited chromosomally integrated HHV–6 (iciHHV–6), which is present in roughly 1% of the worldwide population [24].

While HHV–6 and HHV–7 mostly infect monocytes and macrophages upon primary infection, they have also been found in brain samples and are considered to be neurotropic to a certain degree [22]. In addition, especially HHV–6 has been associated with a number of neurological diseases including Alzheimer's disease [24], epilepsy [25] and multiple sclerosis [26].

1.2 Overview of Diabetes Mellitus Type 2

T2D is one of the most widespread diseases in the developed world with a 2010 prevalence estimate of 7.1% in Germany. It is particularly a disease of the elderly, with the prevalence starting to raise at the age of 50 and peaking at age 80 at 25%. The highest estimated incidence rate (IR) is around the age of 85 at 2.4% per year [27]. This morbidity leads to a high burden of mortality with an estimated 16% of all deaths attributable to T2D in Germany in 2010 [28].

The diagnosis of diabetes can be made by either multiple measures of elevated fasting glucose (≥ 7mM / 126mg/dL), by an elevated HbA1c measure (≥ 6.5%) or by an abnormal oral glucose tolerance test (OGTT). For the OGTT, the patient usually consumes 75g of sugar and has their subsequent increase of blood sugar quantified. Plasma glucose levels ≥ 11.1mM / 200mg/dL after 2 hours confirm the diagnosis [29]. HbA1c is a laboratory measure of the glycated proportion of haemoglobin and is used as a long–term marker of hyperglycaemia [30].

Only slightly increased levels of glucose are considered prediabetic: For fasting glucose they are called *impaired fasting glucose* (IFG: between 5.5mM / 100mg/dL – 7.0mM / 126mg/dL) and for the OGTT they are called *impaired glucose tolerance* (IGT: 2 hour response between 7.8mM / 140mg/dL – 11.1mM / 200 mg/dL) according to the American Diabetes Association [29]. Lifestyle changes like moderate physical activity and improved diet are the best prophylactic strategies to prevent prediabetes from converting to full T2D and can even revert IFG & IGT to normal [30]. Kowall et al. have shown that the incidence rate of T2D among prediabetics is highest in combined IFG & IGT (up to 7.6% per year) [31]. In order to reflect this continuum, prediabetes and T2D have often been grouped together in this study.

Different ethnic susceptibilities have been demonstrated, with Asians being particularly predisposed to the development of prediabetes & T2D [32]. Many genetic risk factors have been identified, making T2D a polygenic disease with marked gene–environment interaction [33].

Many environmental risk factors for prediabetes and T2D are known, including unhealthy diet [34], too little exercise [35], obesity [36], air pollution [37] and inflammation [38,39].

1.3 Type 2 Diabetes and Viral Infections

Because of immunosuppression, increased blood sugar and microangiopathy, the morbidity and mortality of infectious diseases are generally increased in diabetic patients [40,41]. Links between T2D and several viral infections like hepatitis B [42] and hepatitis C [43,44] have been shown, generally with the conclusion that diabetes precedes and increases the risk of the viral infection. Until recently, viruses seemed to have played a proposed aetiological role only in type 1 diabetes, which is a much rarer autoimmune disease than T2D, affecting younger people and characterised by an absolute absence of insulin. In that light, enteroviruses and the hepatitis C virus have been suggested as risk factors for the development of type 1 diabetes [45,46].

However, recently a possible increase in the risk of T2D caused by infection with human herpesviruses has been shown, especially for CMV [47] and human herpesvirus 8 (HHV–8), also called Kaposi's sarcoma–associated herpesvirus. The former will be discussed in detail in section 4.3, but HHV–8 is briefly discussed here, as it has not been part of the KORA study. Its association with T2D has been proposed by Pompei et al. [48] and increased prevalences of HHV–8 among diabetics have been described in populations from Sardinia, Italy [49], sub–Saharan Africa [50] and western China [51]. However, Cui et al. rightly point out, that these cross–sectional and case–control studies can neither indicate causality nor chronology, leading to a classic "chicken or the egg" dilemma [51]. As reverse causality is a big issue, prospective study designs are needed to demonstrate an increased incidence of T2D in herpesvirus–positive compared to –negative patients.

1.4 Aims of this Thesis

This thesis examines the hypothesis that serotitres of human herpesviruses (HSV–1, HSV–2, VZV, EBV, CMV, HHV–6 and HHV–7) are associated with (pre)diabetes prevalence and incidence. Moreover, the hypothesis that herpesvirus seropositivity is associated with inadequate glycemic control as measured by HbA1c is tested. Both binary viral serostatus as well as continuous antibody reactivity are considered. In order to assess possible confounders, the herpesviruses are first characterised thoroughly with regards to sex, age and season. In addition, common co–occurrence patterns are evaluated and tested for associations with (pre)diabetes.

2 Methods

2.1 KORA Cohort

KORA is a German acronym meaning “Kooperative Gesundheitsforschung in der Region Augsburg”, translating to “Cooperative Health Research in the Region of Augsburg”. It is a collection of population–based cohorts from the region of the city Augsburg in south–eastern Germany (Bavaria). It was created in 1996 as an extension to the MONICA project, focussing on cardiovascular disease. However, since then the research areas have broadened and particularly the availability of several follow–up timepoints, biosamples and genetic information make it a very valuable research platform [52].

This thesis is using information from the F4 and FF4 follow–up timepoints of the original S4 population–based cohort, which was a random sample from residential registers from age 25 to 74 with 4,261 participants between 1999–2001. F4 was a follow–up performed between 2006–2008 with 3,080 participants and FF4 another follow–up from 2013/2014 with 2,279 participants, representing response rates of 72.3% and 53.5% respectively (Figure 1). The participants of both of these groups have undergone extensive phenotyping including viral–serology multiplex–assays for human herpesviruses [53] as well as oral glucose–tolerance tests (OGTT).

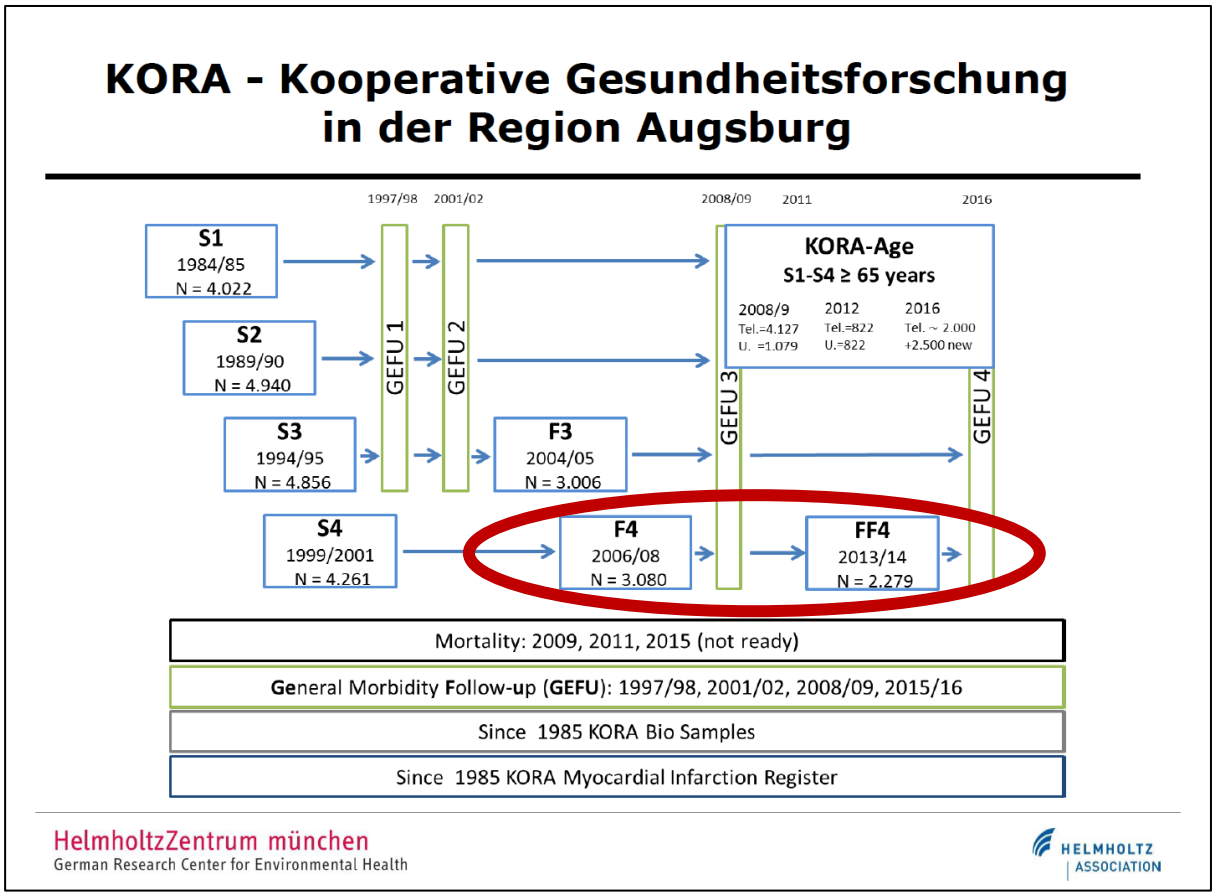


Figure 1: Overview of KORA subgroups with F4/FF4 highlighted in red ellipse (friendly permission by Dr. Birgit Linkohr)

After the exclusion of participants with inconclusive or missing OGTT results, type 1 diabetes, drug–induced diabetes, missing viral–serology (in both of the two timepoints) or withdrawn consent there were n=2,950 of 3,080 participants left in F4 and 2,129 of 2,279 in FF4, with an overlap of 1,967 probands allowing incidence–analysis. Figure 2 demonstrates the participant selection, Table 1 gives a demographical overview of the participants and Figure 14 in a later chapter demonstrates the sample–overlap as well as their diabetic prevalence and incidence.

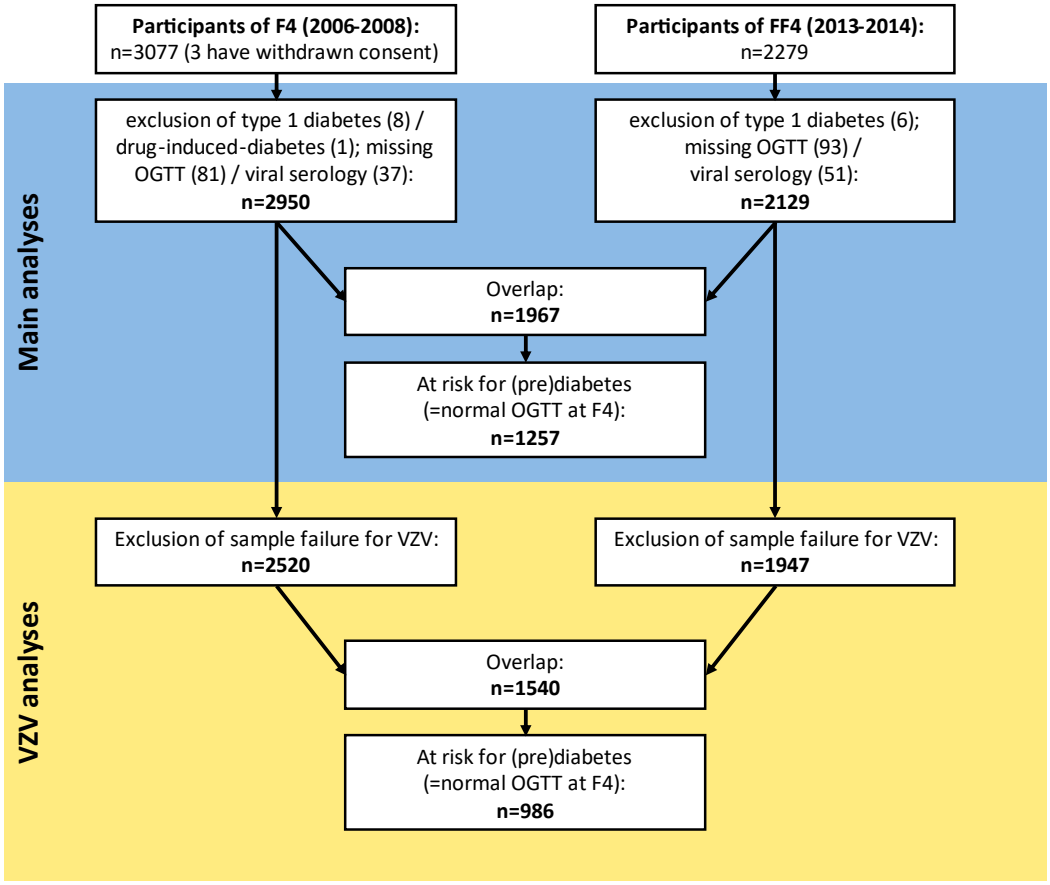


Figure 2: Flowchart demonstrating participant selection

Table 1: Demographical overview of cohort at F4 and FF4 timepoints as well as overlap

	F4 (n=2,950)	**FF4 (n=2,129)**	**Overlap (n=1,967)**
Percent male	48.5%	48.8%	49.2%
Median Age (Range)	56 (32–81)	60 (39–88)	54 (32–81) at F4 61 (39–88) at FF4
Median BMI (IQR)	27 (24–30)	27 (24–31)	27 (24–30) at F4 27 (24–31) at FF4
Percent ever smoker (median PY)	57.3% (15.3 PY)	56.9% (15.8 PY)	56.8% (15.0 PY) at F4 56.9% (15.4 PY) at FF4

Years of education (IQR)	11 (10–13 IQR)	11 (10–13 IQR)	11 (10–13 IQR)
Number IFG/IGT (percent)	845 (28.6%)	754 (35.4%)	542 (27.5%) at F4, 713 (36.2%) at FF4
Number T2D (percent)	337 (11.4%)	310 (14.6%)	168 (8.5%) at F4, 287 (14.6%) at FF4

2.2 Viral Multiplex Assays

In total, 7 human herpesviruses were measured with a multiplex assay: HSV–1, HSV–2, VZV, EBV, CMV, HHV–6 and HHV–7. Multiplex serology is a high–throughput method based on beads combining antibodies against different pathogens, allowing cost–efficient detection in low–volume samples. Antigen–binding of the primary antibodies is quantified through incubation with biotinylated goat–α–human IgM/IgG/IgA secondary antibodies and a reporter dye (streptavidin–R–phycoerythrin). Each bead set represents one antigen and consists of hundreds of beads, whose median fluorescence intensities (MFI) are reported as results [53]. These "antibody reactivities" correlate very well with traditional titre studies and indirectly relate to viral load [53].

Importantly, validation of the multiplex assay with traditional reference assays has already been performed successfully for herpesviruses 1–5 but is still ongoing for HHV–6 and HHV–7, which is why their results have to be treated with caution [15,53].

Three of the viruses (HSV–1, HSV–2 and HHV–7) are represented by one antigen each, whereas VZV is represented by two antigens and EBV, CMV and HHV–6 are represented by four antigens each. VZV is deemed seropositive when either one of the two antigens is above threshold and EBV, CMV and HHV–6 are deemed seropositive when 2 out of 4 antigens are above threshold. Table 2 gives an overview of these antigens, their respective biological description and MFI–thresholds.

Table 2: Antigens for 7 human herpesviruses

Virus	Antigen	Description	MFI–Threshold
HSV–1	gg	Glyoprotein G	100
HSV–2	mgg	Unique sequence of membrane–bound part of mature glycoprotein G	100
VZV **(1 / 2)**	giorf67 georf68	Glycoprotein I(ORF67) Glycoprotein E(ORF68)	100 80
EBV **(2 / 4)**	ead ebna1	Early antigen–diffuse (EA–D) Nuclear antigen 1 (EBNA–1)	150 150

	vcap18	Viral capsid antigen (VCA p18)	200
	zebra	Virus protein ZEBRA	150
CMV (2 / 4)	pp28	Protein pp 28	150
	pp52	Protein pp 52	150
	pp65	Protein pp 65	150
	pp150	Protein pp 150 (N–terminus)	150
HHV–6 (2 / 4)	ie1atr	6A Immediate–early 1 protein (truncated)	50
	p100tr	6A antigenic tegument protein (truncated)	50
	ie1btr	6B Immediate–early 1 protein variant B (trunc.)	50
	p101ktr	6B antigenic tegument protein (truncated)	50
HHV–7	u14	Protein U14, putative structural function	100

2.3 Statistics

2.3.1 T–Test, Mann–Whitney–Test, Chi–Squared–Test, Pearson–Correlation

Significance tests were performed with two–sample or paired t–tests when the data are interval and roughly follow a normal distribution. For data that do not follow a normal distribution, the non–parametric Wilcoxon–Mann–Whitney rank–sum–test was used [54].

To assess whether the observed frequencies in contingency tables differ significantly from expected frequencies under the null–hypothesis of even distribution, Pearson's chi–squared test (χ^2) was used [54].

Correlation coefficients *R* were calculated using Pearson's correlation method if not otherwise stated [54].

2.3.2 Normal Confidence Intervals, Poisson Rate Confidence Intervals

Confidence intervals (CI) are often calculated expecting an underlying normal distribution of the data, making use of the central limit theorem and the Standard Error (SE). For a two–sided 95% CI , the lower limit must lie at the 2.5th percentile and the upper limit at the 97.5th percentile of the distribution. As the former lies at -1.959964 ≈ -1.96 and the latter at 1.959964 ≈ 1.96 on the quantile function of the standard normal distribution and the SE of the mean is $SD/\sqrt{n}$, the formula for a 95% CI of a mean is thus [54]:

$$95\%\ CI\ of\ a\ mean_{lower,upper} = mean \pm 1.96 * \frac{SD}{\sqrt{n}}$$

As there is an asymptotic approximation of the SE of a log Odd's Ratio (OR) of the type $\frac{A}{B} / \frac{C}{D}$, the formula for the corresponding 95% CI is the following [54]:

$$95\%\ CI\ of\ an\ OR\ _{lower,upper} = e^{\log(OR) \pm 1.96 * \sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}}$$

Similarly, the formula for the 95% CI of a Risk Ratio (RR) of the type $\frac{A}{A+B} / \frac{C}{C+D}$ is [54]:

$$95\%\ CI\ of\ a\ RR\ _{lower,upper} = e^{\log(RR) \pm 1.96 * \sqrt{\frac{1}{A}-\frac{1}{A+B}+\frac{1}{C}-\frac{1}{C+D}}}$$

The incidence rate (IR) $\frac{k}{n}$ is better modelled by a Poisson distribution than a normal distribution and thus a Poisson CI is warranted, which has the following formula where *k* is the number of incident cases and *n* the total number persons (or person–time) [55]:

$$95\%\ Poisson\ CI\ for\ IR\ _{lower} = \frac{\chi^2_{(2*k),0.025}}{2*n}, CI_{upper} = \frac{\chi^2_{(2*(k+1)),0.975}}{2*n}$$

2.3.2 Linear Regression

Multivariate linear regression models take the following form where y_i is the continuous (i.e. normally distributed) dependent ("outcome", "response") variable and $x_{i1} - x_{ik}$ are the k independent ("input", "predictor") variables of participant *i* (ranging from 1 to *n*, the total number of participants). The coefficients $\beta_0 - \beta_k$ are estimated through the ordinary least squares approach to minimise the error term ε_i [54].

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

In matrix notation this equals the following [56]:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_k \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

The closed form analytic ordinary least square solution is [56]:

$$\boldsymbol{\beta} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T y}$$

It is important to note that the results of linear regression models can only be considered valid when the following assumptions are met: linearity of the dependent variable, independence of error terms (thus of observations), constant variance in the dependent variable (also called homoscedasticity) and lack of perfect correlation among independent variables (also called perfect

multicollinearity). The mathematical generalisation of linear regression models is called generalised linear models [56].

2.3.3 Logistic Regression

For binary independent outcome variables (e.g. healthy vs. (pre)diabetic or seronegative vs. seropositive), multivariate logistic regression can be used to model the probability *p* of the variable to be positive ("true"). The log odds (also called "logit") are modelled similarly to multivariate linear regression, however the error term ε_i is unobserved and thus often not included in the formula [54]:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

The probability *p* can be mathematically recovered by exponentiation and algebraic manipulation [56]:

$$p = \frac{1}{1 + e^{-\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}}$$

Unlike linear regression, no closed form solution for determining the optimal β–coefficients exists. However, iterative optimisation algorithms like gradient descent can be used to maximize the log likelihood of the model and thus approximate the optimal solution. Logistic regression can be regarded as a special form of generalised linear models, where the dependent variable follows a Bernoulli distribution (rather than a normal distribution as in linear regression) and is linked to the continuous results of the linear model through the logit function [56].

2.3.4 Generalised Estimating Equations

Generalised estimating equations (GEE) are an adaptation of generalised linear models that can be used when the assumption of independent observations is violated. Outcome variables are correlated for example in repeated–measurements and time–series data, as is the case in the longitudinal KORA cohort with the two timepoints F4 and FF4. Just like generalised linear models, GEE can model continuous outcome variables (normally distributed, see linear regression) and binary outcome variables (Bernoulli distributed) with a logit link function (see logistic regression) [57].

In addition to the distribution and the link function, the correlation structure of the observations must be specified in GEE as the working variance–covariance–matrix. Typical options are exchangeable, auto–regressive and unstructured [57]. As there are only two observations per participant in the KORA study, all of these options are the same with *p* representing the correlation between F4 and FF4:

$$VarCov = \begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix}$$

In opposition to generalised linear models, GEE are not solved through maximisation of likelihood but through an iterative so called "quasi–likelihood" estimation because the joint distribution is not fully specified and thus likelihood cannot be calculated. Consequently, methods for testing fit, comparing models and conducting inferences based on likelihood cannot be used (e.g. likelihood ratio test) [57].

Inference on coefficients is usually performed with an empirical estimator of the SE proposed by Eicker, Huber and White, which is also called the "robust" or "sandwich" estimator. It allows for heteroscedasticity (inconstant variance) and in large–enough samples it provides a good estimate of standard errors even when the covariance structure was misspecified initially [57].

2.3.5 Multiple Testing & Bonferroni Correction

When multiple statistical inference tests are performed, the risk of obtaining false–positive results increases, leading to more type I error. To control the type I error rate α (the significance level) for multiple testing, the so–called family–wise error rate can be used. It states the probability of making at least one type I error (false positive discoveries) among multiple tests. One classic way to control the family–wise error rate is the Bonferroni correction, which simply divides the significance level for each individual test α by the total number of tests n and thus rejects a null–hypothesis H_0 if the corresponding p–value is below $\frac{\alpha}{n}$ [54].

The ease of the Bonferroni correction comes at the cost of being overly strict or conservative in many cases, especially when many of the tests are correlated. For example in this study, the so–called effective number of tests is usually smaller than n, because the prevalences of the 7 viruses or 17 antigens examined are highly correlated. Nonetheless, the Bonferroni method is highly interpretable and results surviving its strict correction are quite unlikely to be false positive if the model assumptions hold true [54].

2.3.6 Statistical Software Used

All statistical analyses were performed with the open source statistical software language "R" version 3.6 [58]. Packages used are "corrplot" for correlation heatmaps [59], "ggalluvial" for Sankey plots [60], "ggpubr" for general plotting [61], "upsetr" for co–occurrence maps [62], "gee" for generalised estimating equations [63] and "rmeta" for forest plots [64].

3 Results

3.1 Characterisation of Herpesviruses in KORA

3.1.1 Correlation of Viral Antigens

As discussed in section 2.2, 17 antigens form the basis for the detection of the 7 herpesviruses. While HSV–1, HSV–2 and HHV–7 are only represented by one antigen each, the other 4 are represented by multiple antigens. Figure 3 shows pairwise correlations for all 17 antigens with those belonging to the same virus grouped together in black bordered boxes. On the left–hand side (A), pairwise correlations including all participants from F4 can be seen (n=2,950; except for VZV n=2,520 because of antibody failure in 430 samples) and on the right–hand side (B), pairwise correlations of antigen–seropositives only.

Overall, the correlations are larger when including seronegatives (Figure 3A): The two VZV antigens are moderately correlated with R=0.3, as are the four EBV–antigens with R ranging from 0.2–0.6. The four CMV antigens are most strongly correlated with R ranging from 0.8–0.9 and the four HHV–6 antigens are only weakly correlated with R ranging from 0.1 to 0.3.

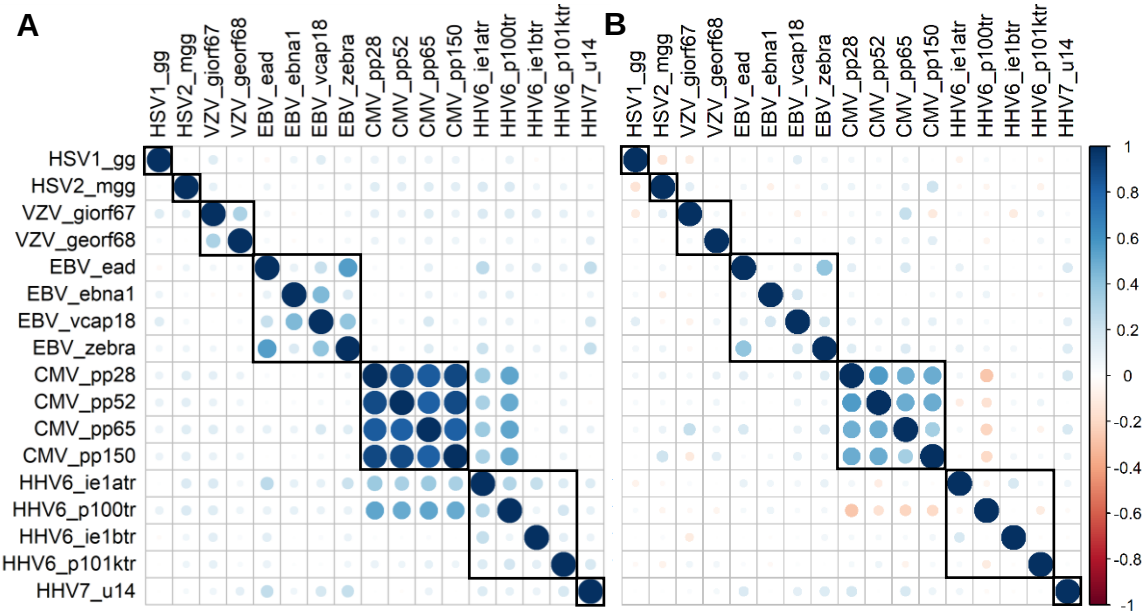


Figure 3: Pairwise correlation heatmap of viral antigens in all participants (A) and seropositive participants only (B)

There seems to be only little intercorrelation between viruses with the highest coming from the four CMV antigens: they intercorrelate moderately with HHV–6 *ie1atr* (R ranging from 0.3–0.4) and strongly with HHV–6 *p100tr* (R=0.5). Interestingly, the intercorrelation with HHV–6 *ie1atr* is lost when only considering seropositives (Figure 3B) and the direction of the intercorrelation with HHV–6 *p100tr* is reversed: now it is ranging from R = -0.3 – -0.2, with the biggest absolute correlation coefficient coming from CMV *pp28*. Figure 4 demonstrates this strong

intercorrelation. The reason remains unclear, it could partly be due to co–occurrence (later it can be seen that CMV and HHV–6 often occur together, see Figure 7) or to cross–reactivity of the antibodies. The observed negative correlation for seropositives for both antigens is most likely an artifact due to a suboptimal threshold for HHV–6 *p100tr*, highlighting that the thresholds are not optimised for correlation analyses. By inspecting the left–hand side of Figure 4 visually, it appears that a small part of the CMV *pp28*–positive–only–cloud in the bottom–right quarter is included in the top–right quarter. The choice of thresholds is further discussed in the Limitations section of the Discussion.

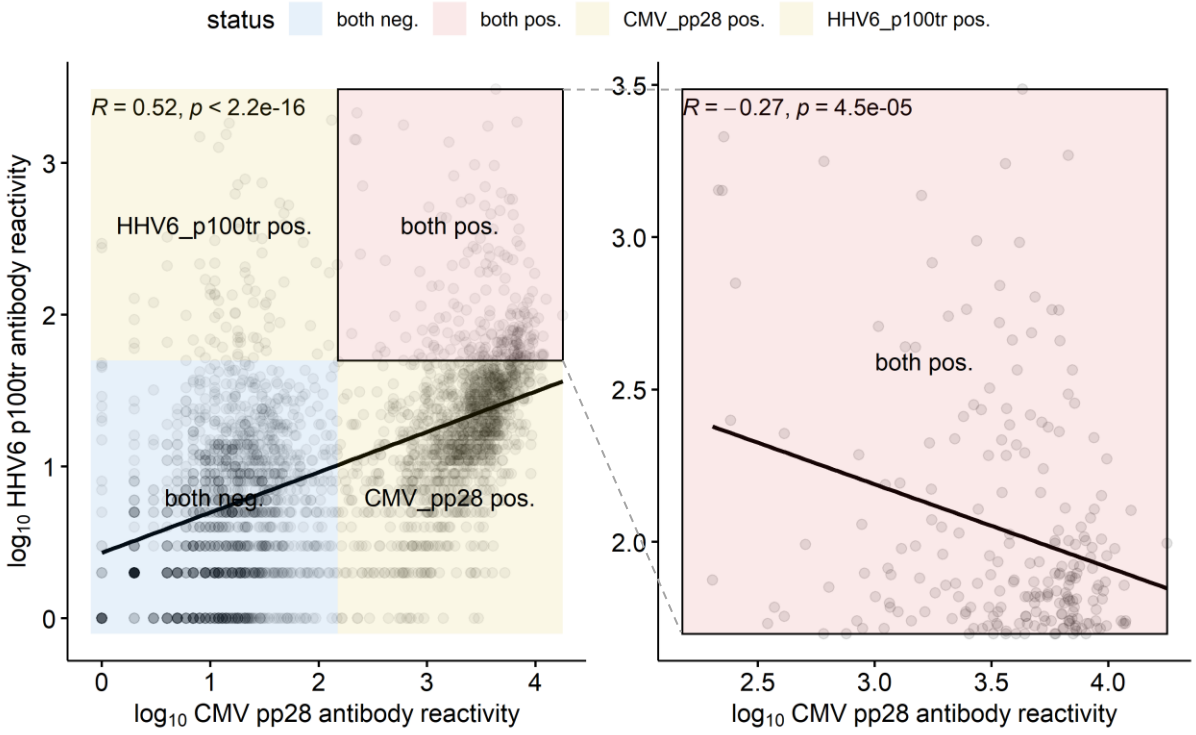


Figure 4: Correlation of CMV pp28 and HHV–6 p100tr antigens for all participants (left) and seropositives only (right)

3.1.2 Prevalence and Incidence of Viruses

Figure 5 demonstrates viral prevalence at both F4 and FF4 timepoints for all 7 human herpesviruses in a subgroup of 1,967 participants who have had viral serology performed at both timepoints. VZV–assays were lost for 427 participants due to technical failure of antibody analysis with 1,540 participants remaining. HSV–1, HSV–2 and HHV–7 are only represented by one antigen each. VZV is considered positive (=prevalent) when 1 out of 2 antigens are above threshold and EBV, CMV & HHV–6 are considered positive when 2 out of 4 antigens are above threshold (see 2.2). EBV is most prevalent (98% at F4), followed by HSV–1 (88% at F4), HHV–7 (85% at F4), VZV (79% at F4), CMV (46% at F4), HHV–6 (39%) and lastly, least–prevalently, HSV–2 (11% at F4).

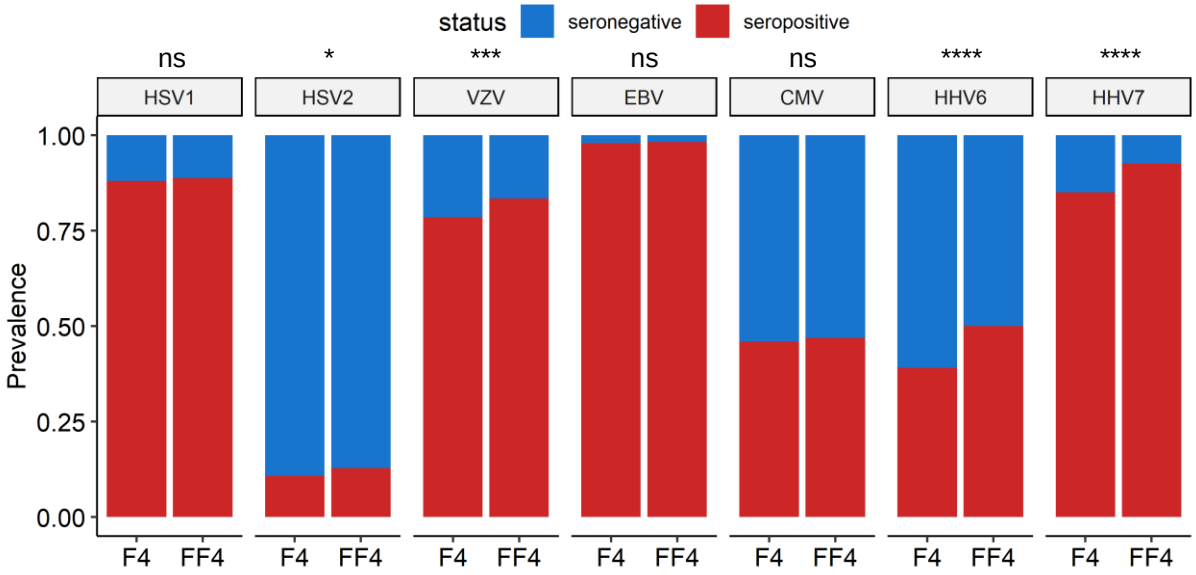


Figure 5: Viral prevalences at F4 and FF4 of 1,967 overlapping participants (except for VZV: 1,540 overlapping participants)

Figure 5 shows that the prevalences of HSV–2, VZV, HHV–6 and HHV–7 increase significantly between the two timepoints (significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). As the prevalence plot is only showing overlapping participants present in both timepoints, dropout cannot account for these differences. Thus, they have arisen due to so called seroconversions. Table 3 gives the numbers and rates of positive seroconversions for all 7 herpesviruses. Because the mean time difference between F4 and FF4 is 6.5 years, the crude rate is for that timeframe, but the table also provides a normalised rate for one year.

Table 3: Rates of positive seroconversion for human herpesviruses (95% Poisson rate confidence intervals in brackets)

Virus	Pos. seroconversions	Negative at F4	Rate of positive seroconversions per 6.5 years	Rate of positive seroconversions per year
HSV–1	17	236	7.20% (4.20% – 11.53%)	1.11% (0.65% – 1.77%)
HSV–2	61	1756	3.47% (2.66% – 4.46%)	0.53% (0.41% – 0.69%)
VZV	121	331	36.56% (30.33% – 43.68%)	5.62% (4.67% – 6.72%)
EBV	10	41	24.39% (11.70% – 44.85%)	3.75% (1.80% – 6.90%)
CMV	38	1064	3.57% (2.53% – 4.90%)	0.55% (0.39% – 0.75%)
HHV–6	383	1199	31.94% (28.82% – 35.31%)	4.91% (4.43% – 5.43%)
HHV–7	182	295	61.69% (53.06% – 71.34%)	9.49% (8.16% – 10.98%)

As described in the introduction, human herpesviruses are usually persistent in their hosts, meaning that they remain in the cell nuclei lifelong, mostly inactive and symptom–free in a latency state. However, that does not mean that they are always detectable by antibodies in blood: Their serostatus can change due to many factors including immune–system– and viral–activity. Incidence of most herpesviruses usually occurs in early childhood, much earlier than

at the median age of 56 years at F4, but infections at older age are possible. A positive seroconversion could thus represent an incident case but it could also be due to an increase of antibody reactivity (modelling titre and viral load) of a beforehand undetectable virus.

In the same light, a person losing seropositivity cannot be considered "healed" of the virus, it is much more likely to remain in an undetectable latency state. Table 4 demonstrates these rates of negative seroconversion for all 7 viruses and it becomes clear that HSV–2 and HHV–6 are the two viruses with the highest number thereof.

Table 4: Rates of negative seroconversions for human herpesviruses (95% Poisson rate confidence intervals in brackets)

Virus	Neg. seroconversions	Positive at F4	Rate of negative seroconversions per 6.5 years	Rate of negative seroconversions per year
HSV–1	2	1731	0.12% (0.01% – 0.42%)	0.02% (0.00% – 0.06%)
HSV–2	21	211	9.95% (6.16% – 15.21%)	1.53% (0.95% – 2.34%)
VZV	45	1209	3.72% (2.71% – 4.98%)	0.57% (0.42% – 0.77%)
EBV	4	1926	0.21% (0.06% – 0.53%)	0.03% (0.01% – 0.08%)
CMV	18	903	1.99% (1.18% – 3.15%)	0.31% (0.18% – 0.48%)
HHV–6	167	768	21.74% (18.57% – 25.30%)	3.35% (2.86% – 3.89%)
HHV–7	33	1672	1.97% (1.36% – 2.77%)	0.30% (0.21% – 0.43%)

As detailed in 2.2, viral status depends on the thresholds of 1 to 4 antigens per virus. When looking at one antigen at a time, it can be seen that participants who change their serostatus have – on average – antibody reactivities closer to the threshold line than participants who have a stable serostatus between the two timepoints F4 and FF4. I interpret this as an indicator that most seroconversions to positive are in fact not incident cases but rather participants with undetectable antibodies at baseline. This is shown exemplarily for the HHV–6 antigen *p101ktr* in Figure 6 but holds true for all antigens except for HSV–1, see Supplementary Figures 1–16.

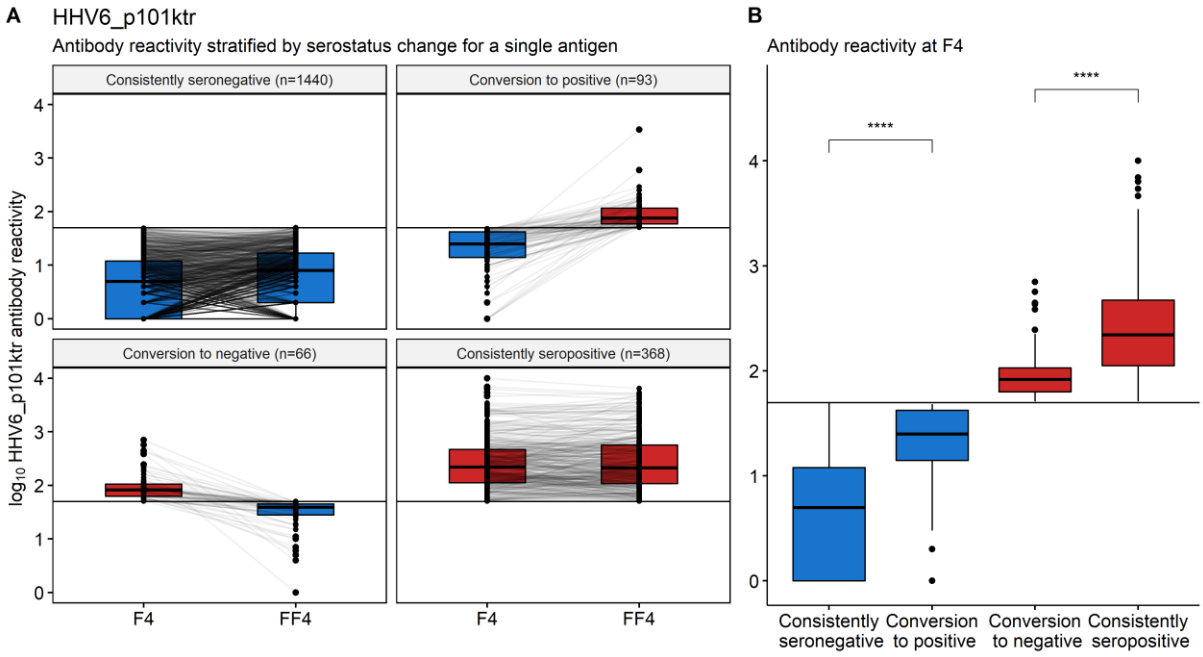


Figure 6: HHV–6 p101ktr antibody reactivity is closer to the threshold line for participants who change serostatus than for those who do not

3.1.3 Co–Occurrence Maps

Most people are infected with multiple of the 7 human herpesviruses studied, which is not surprising, considering the high prevalences discussed above. Figure 7 shows the distribution of these co–occurrences sorted by set sizes in an intersection–plot produced by the software UpsetR [62]. Out of 2,520 participants with complete viral–serologies at F4, the largest group with 442 participants is infected with EBV, HSV–1, HHV–7 and VZV, followed by a group of 332 participants with seropositivity for 6 out of 7 viruses (all but HSV–2).

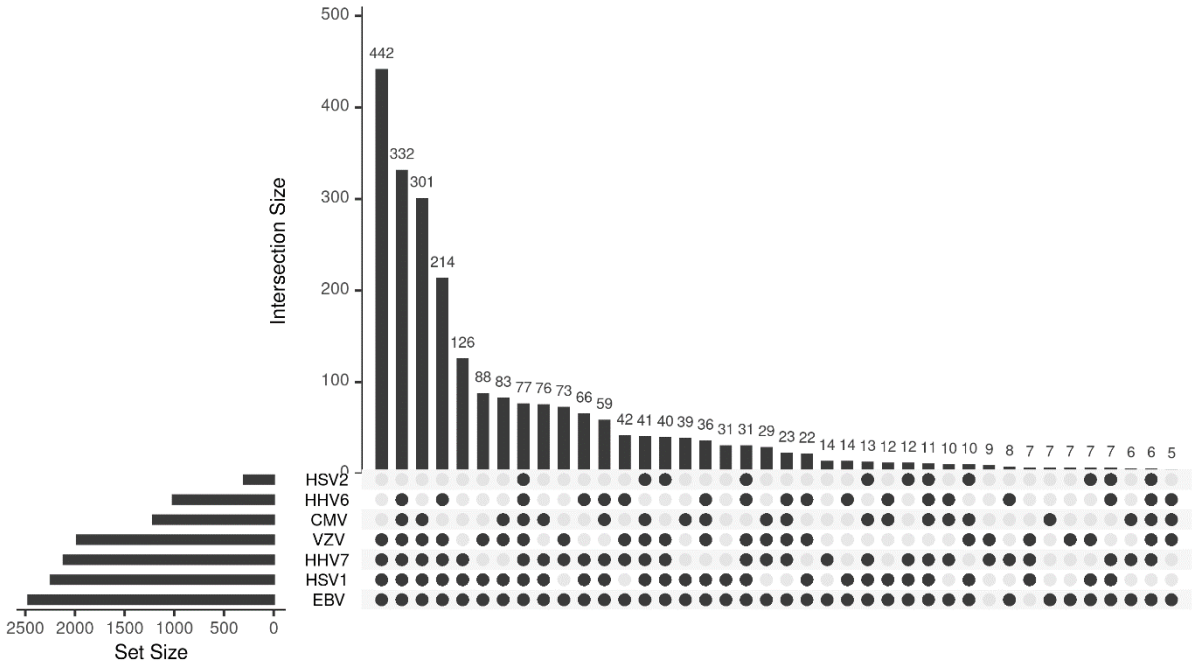


Figure 7: Co–occurrence of human herpesviruses at F4 (n=2,520, right tail cut off)

The number of seropositive viruses per person is fairly normally distributed with a mean of 4.4±1.1 SD at F4 and 4.7±1.1 SD at FF4 in the overlapping 1,540 probands (95% CI for difference: 0.21–0.37, p<1e-12). It has been shown in the section before (3.1.2) that the prevalences for HSV–2, VZV, HHV–6 and HHV–7 are significantly higher at FF4 compared to F4. Furthermore it was concluded that there are more conversions to seropositive for all viruses than conversions to seronegative (undetected). This is also reflected in the number of co–occurrences: Figure 8 shows in detail how all overlapping participants change the groups corresponding to their number of seropositive viruses. A third of all probands are positive for more viruses at FF4 than at F4 (34%), a little more than half are positive for the same number of viruses (54%) and only 12% are positive for less viruses at FF4 than at F4.

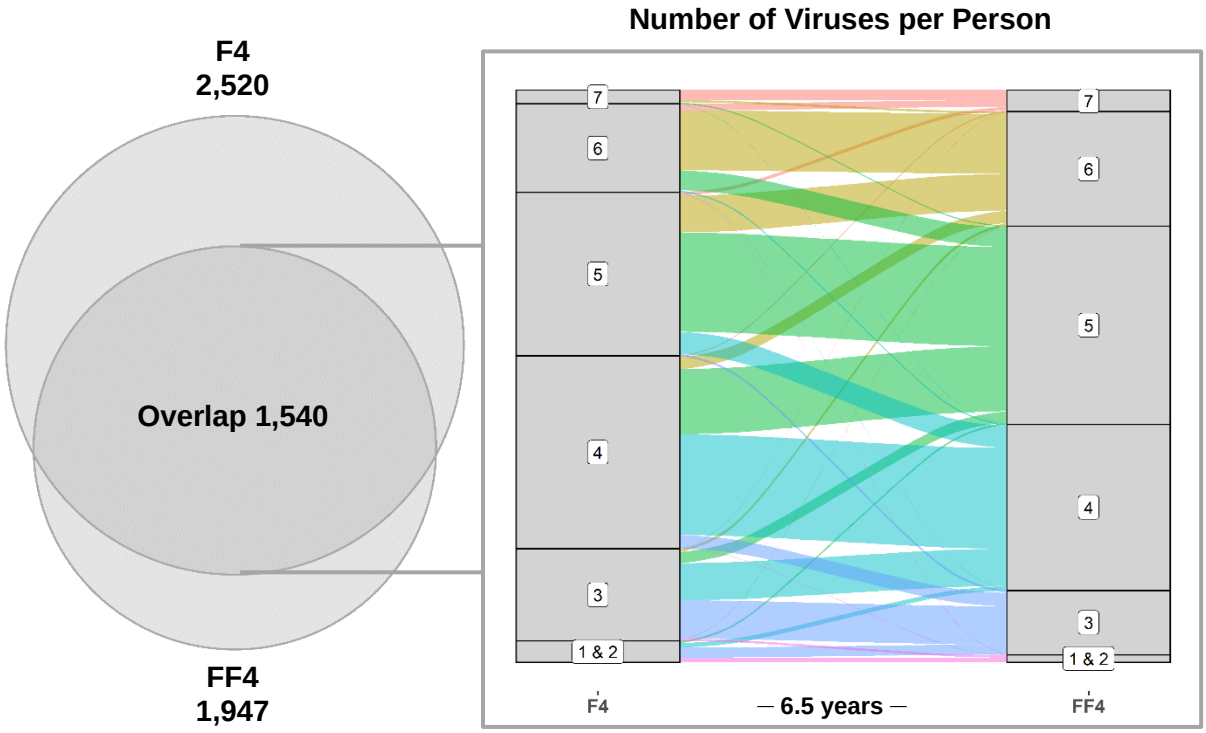


Figure 8: Changes of number of viruses per person from F4 to FF4 (subset of samples with complete serology)

3.1.4 Association of Viruses with Age

In order to robustly assess the association of human herpesviruses with type 2 diabetes later, first it has to be established how the viruses themselves are influenced by common confounders. Figure 9 shows the relationship of HSV–1, HSV–2 and HHV–7 status (left boxplots) and their viral antibody reactivities (right scatterplots) versus age in F4. These 3 viruses are represented by only one antigen each and are thus easier to interpret than the other 4 viruses examined.

One can appreciate that for HSV–1 there is a highly significant positive association of age and both viral status as well as antibody reactivity at the F4 timepoint. HSV–1 carriers – the vast majority of probands, as HSV–1 is highly prevalent – are on average 6.2 years older than non–

carriers (p < 0.0001). Among those carriers, the viral antibody reactivity correlates positively with age with a Pearson correlation coefficient of R = 0.12 (p = 2.5e-9). The picture looks similar for HSV–2, even though it is much less prevalent than HSV–1, and quite opposite for HHV–7: There is a significant negative association of both viral status and viral antibody reactivity with age. Most of these findings remain stable at the FF4 timepoint (see Table 5).

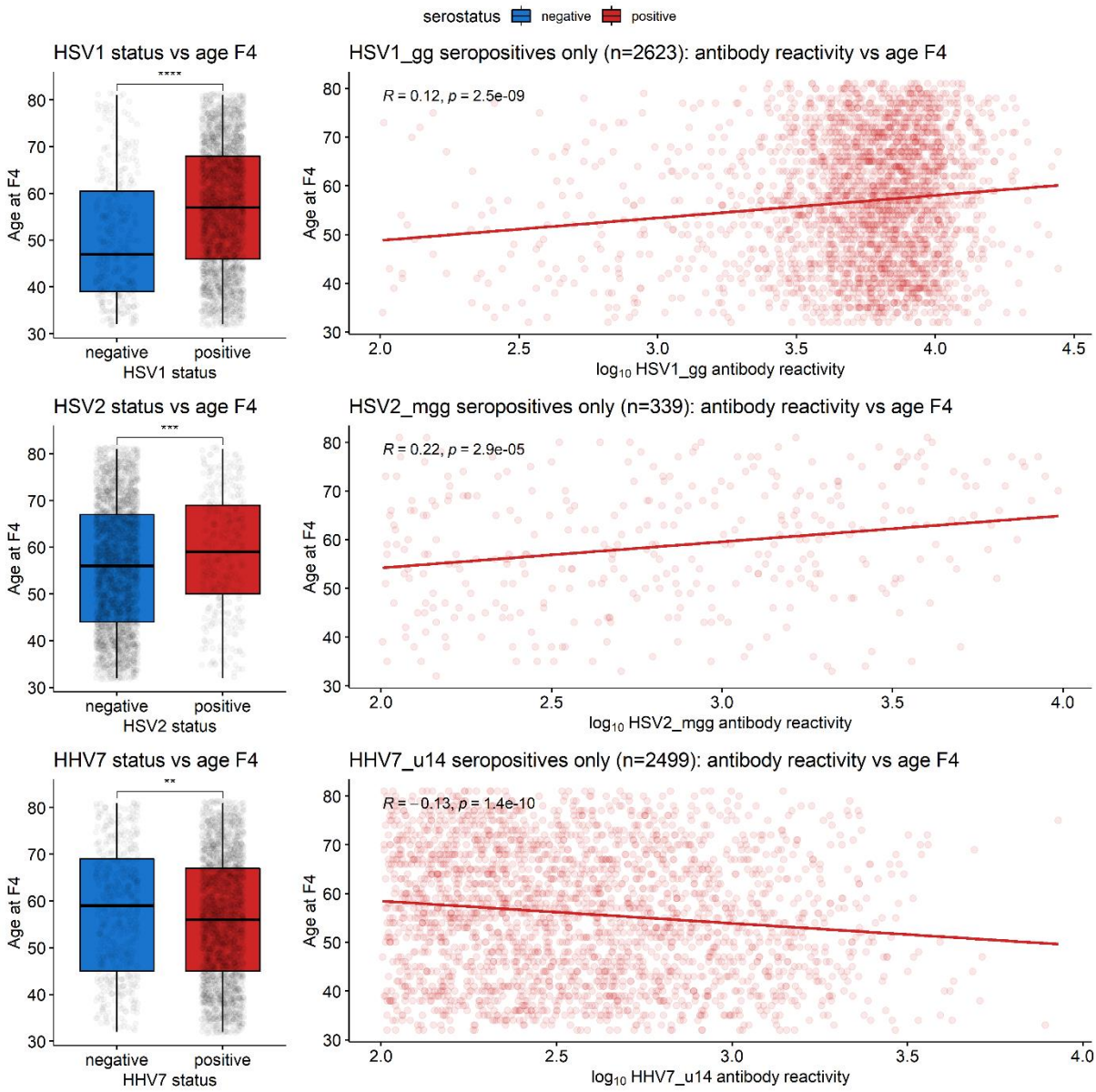


Figure 9: HSV–1, HSV–2 and HHV–7 viral status / antibody reactivity vs. age at F4

Table 5 summarises the age associations with viral status and antibody reactivity for all viruses and both timepoints. Observe how VZV, EBV, CMV and HHV–6 are represented by multiple antigens and how their behaviour is not always unidirectional. This underlines that biological systems are complex and may not give unambiguous results. (Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001)

Table 5: Viral status: Mean age difference for seropositives (95% CI in brackets), viral antibody reactivity: age correlation coefficient

	Viral status (F4)	Viral antibody reactivity (F4)	Viral status (FF4)	Viral antibody reactivity (FF4)
HSV–1	**6.2y (4.7–7.8)**	**gg: R= .12 ******	**5.8y (4.2–7.4)**	**gg: R= .12 ********
HSV–2	**3.0y (1.5–4.4)**	**mgg: R= .22 ********	**1.6y (0.1–3.1)**	**mgg: R= .24 *****
VZV	-0.5y (-1.7–0.7)	giorf67: R=-.02 **georf68: R= .05 ***	0.9y (-0.6–2.4)	girf67: R=-.05 **georf68: R= .06 ***
EBV	**5.4y (2.2–8.6)**	ead: R= .03 **ebna1: R=-.04 *** **vcap18: R= .06 **** **zebra: R= .05 ****	**4.9y (1.3–8.6)**	**ead: R= .06 **** ebna1: R=-.02 **vcap18: R= .08 ****** **zebra: R= .06 ****
CMV	**4.8y (3.9–5.7)**	pp28: R=-.01 **pp52: R= .09 ****** pp65: R= .05 pp150: R= .02	**3.7y (2.6–4.7)**	pp28: R=-.01 **pp52: R= .09 **** **pp65: R= .09 **** **pp150: R= .07 ***
HHV–6	0.2y (-0.8–1.1)	ie1atr: R=-.02 p100tr: R= .04 **ie1btr: R=-.06 *** p101ktr: R= .01	-0.3y (-1.3–0.8)	ie1atr: R= .04 p100tr: R= .05 ie1btr: R=-.05 p101ktr: R= .04
HHV–7	**-1.8y (-3.2– -0.5)**	**u14: R=-.13 ********	0.7y (-1.3–2.7)	**u14: R=-.12 ********

An example of such an equivocal result is the age–association of the viral antibody reactivity of EBV–antigen *ebna1*, whose correlation coefficient is significantly negative at F4 and thus opposing the direction of the other antigens and viral status (see Table 5 and Supplementary Figures 17–25). However, this antigen is only significant at the 0.05 level and none of the p–values in the above table have been corrected for multiple–testing, as this is an exploratory analysis. Consequently, this specific *ebna1* finding might be a false–positive, which is even more likely considering that it is not significant at the FF4 timepoint. In 3.1.7 confounders will be combined and Bonferroni correction will be applied, after which *ebna1* loses its significance.

To summarise, it seems that status and antibody reactivity of HSV–1, HSV–2, EBV and CMV are consistently positively associated with age in both F4 and FF4 whereas HHV–7 is more or less consistently negatively associated with age. No strong age associations seem to exist for VZV & HHV–6.

3.1.5 Association of Viruses with Sex

Differences in virus serotitres by sex were examined. CMV for example seems to be more common in women than in men (OR_f = 1.6, 95% CI 1.4–1.8) and the MFIs of seropositive women are significantly higher for its 4 antigens than those of seropositive men at F4 (see

Figure 10). The picture looks very similar at FF4, only CMV *pp65* loses its significance (see Table 6 and Supplementary Figures 26–34).

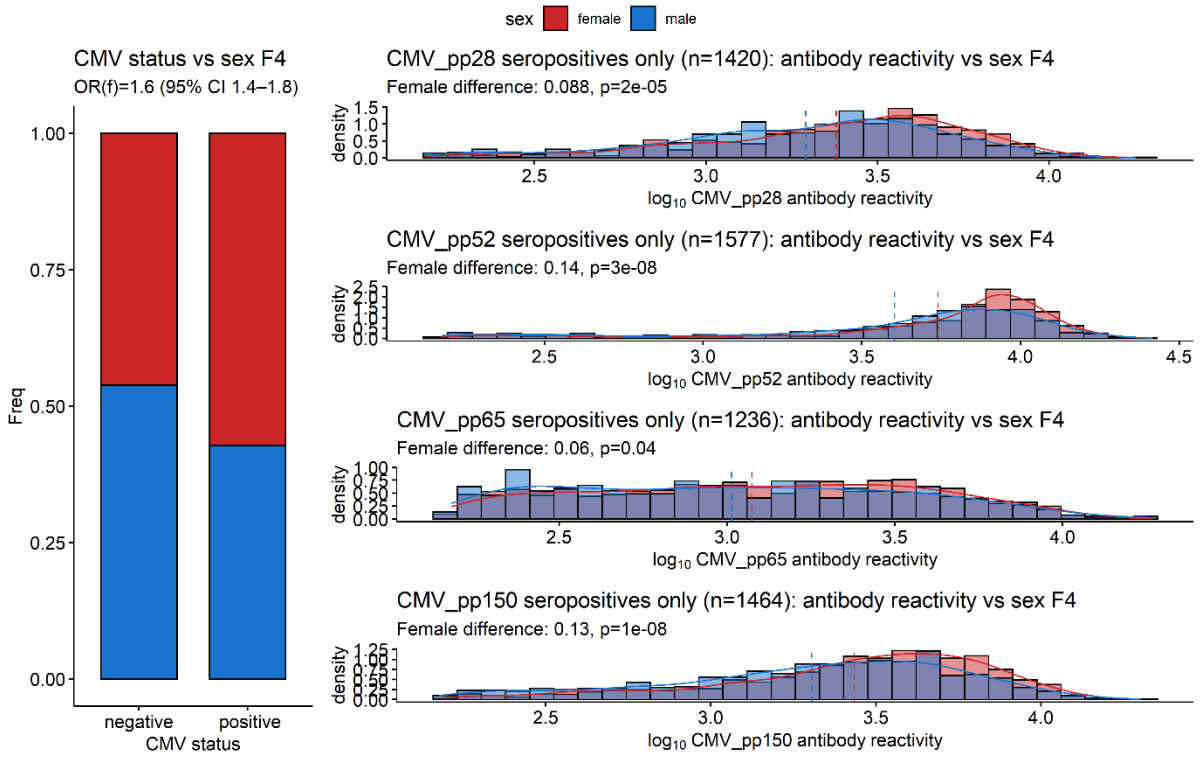


Figure 10: Distribution of CMV status / antibody reactivity stratified by sex at F4

Table 6 summarises these associations for all seven viruses at both F4 and FF4. It becomes clear that most viruses are significantly more common in women, except for VZV, which is more common in men (OR_f = 0.65, 95% CI 0.52–0.81 at F4). For some viruses, this trend is also reflected on the viral antibody reactivity level: Seropositive women with EBV, CMV and HHV–7 have consistently higher viral antibody reactivities than seropositive men, who in turn consistently have higher antibody reactivities of one of the two VZV antigens. For HSV–1, the effect direction of the association of sex with viral status does not match that of viral antibody reactivity: Whereas women have the higher odds to be seropositive on the one hand, men have higher antibody reactivities among the seropositives on the other. (Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001)

Table 6: Viral status: ORs for females to be seropositive (95% CI in brackets), Viral antibody reactivities: mean difference for females (Δ_f)

	Viral status (F4)	**Viral antibody reactivity (F4)**	**Viral status (FF4)**	**Viral antibody reactivity (FF4)**
HSV–1	**OR_f = 1.3 (1.0–1.7)**	**gg: Δ_f =-.05 *****	OR_f = 1.2 (0.9–1.7)	**gg: Δ_f =-.05 ****
HSV–2	OR_f = 1.2 (1.0–1.4)	**mgg: Δ_f = .24 ********	**OR_f = 1.2 (1.0–1.5)**	mgg: Δ_f = .08
VZV	**OR_f = 0.7 (0.5–0.8)**	giorf67: Δ_f = .03 **georf68: Δ_f =-.09 *****	**OR_f = 0.7 (0.5–0.9)**	girf67: Δ_f = .05 **georf68: Δ_f =-.10 ********

EBV	**OR_f = 2.3 (1.2–4.4)**	**ead:** **ebna1:** vcap18: **zebra:**	**Δ_f = .06 **** **Δ_f =-.05 **** Δ_f = .00 **Δ_f = .10 ******	**OR_f = 2.4 (1.1–5.2)**	ead: **ebna1:** vcap18: **zebra:**	Δ_f = .05 **Δ_f =-.04 *** Δ_f =-.01 **Δ_f = .10 ******
CMV	**OR_f = 1.6 (1.4–1.8)**	**pp28:** **pp52:** **pp65:** **pp150:**	**Δ_f = .09 ****** **Δ_f = .19 ****** **Δ_f = .06 *** **Δ_f = .10 ******	**OR_f = 1.4 (1.2–1.6)**	**pp28:** **pp52:** pp65: **pp150:**	**Δ_f = .06 *** **Δ_f = .20 ****** Δ_f = .00 **Δ_f = .10 ******
HHV–6	**OR_f = 1.2 (1.1–1.4)**	ie1atr: p100tr: **ie1btr:** p101ktr:	Δ_f =-.03 Δ_f =-.02 **Δ_f = .07 *** Δ_f = .06	OR_f = 1.0 (0.9–1.2)	ie1atr: p100tr: **ie1btr:** p101ktr:	Δ_f = .02 Δ_f =-.06 **Δ_f = .1 ***** Δ_f = .03
HHV–7	**OR_f = 1.6 (1.2–2.0)**	**u14:**	**Δ_f = .12 ******	**OR_f = 1.7 (1.2–2.5)**	**u14:**	**Δ_f = .12 ******

3.1.6 Influence of Season on Herpesviruses

Winter is flu–season and usually comes with a higher risk of viral infections like the common cold (caused for example by rhinoviruses, coronaviruses, adenoviruses) or influenza. In order to assess any possible associations of season with herpesviruses, the association of viral status and antibody reactivity with the months during which blood was drawn can be considered. In this thesis, November–April have been grouped into a "cold season" and May–October into a "warm season" in order to facilitate statistical testing, reflecting the normal seasonal variation of temperatures in Bavaria.

Figure 11 shows the analysis exemplarily for HSV–1 at both F4 and FF4: While the season does not seem to have any effect on viral status (seropositivity, see bar charts on the left), it does seem to influence viral antibody reactivity at F4: The antibody reactivities of seropositives who were measured in the cold season were significantly higher than of those who were measured in the warm season (Δ_{cold} = 0.05 ***). However, this result is not stable and cannot be seen at FF4.

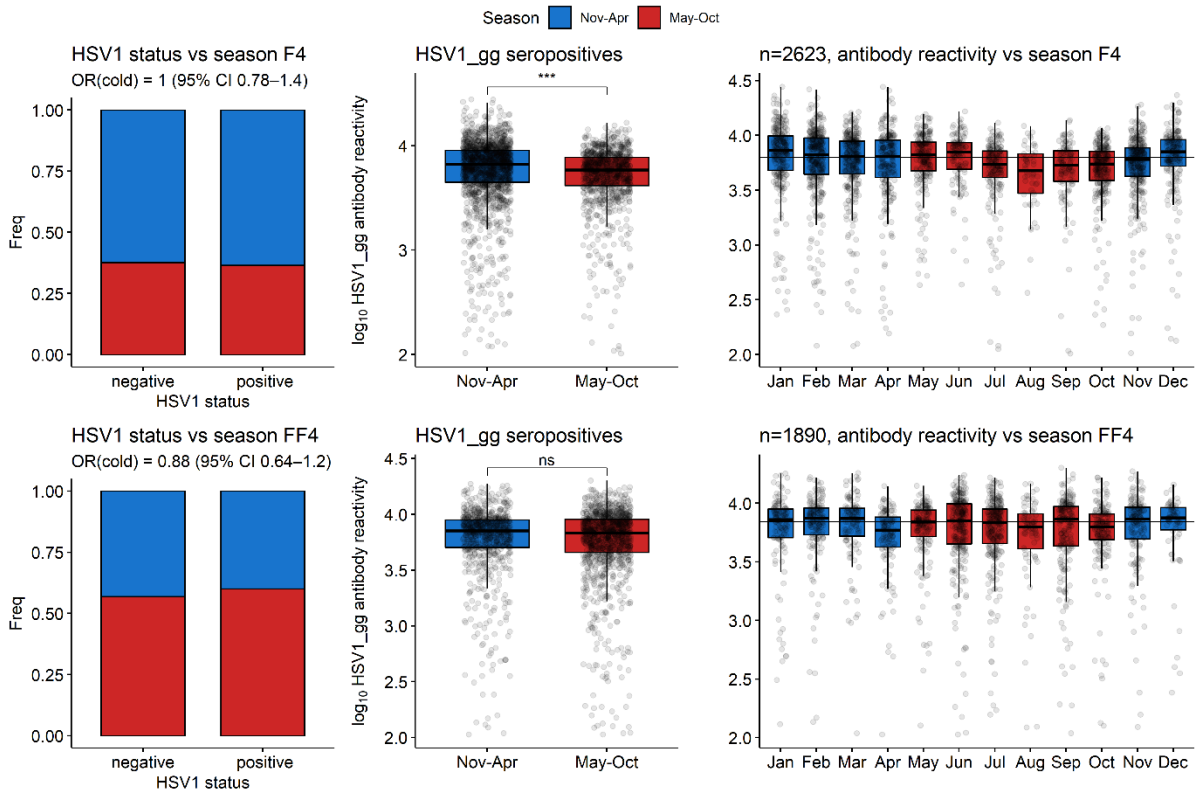


Figure 11: Influence of season on viral status / antibody reactivity of HSV–1 at F4 and FF4

Season is significantly associated with viral status (seropositivity) for VZV, HHV–6 and HHV–7 at F4 and only for HHV–6 at FF4. However, the HHV–6 associations have opposite directions at F4 and FF4, thus decreasing consistency. Moreover, season is correlated with viral antibody reactivity among seropositives for many of the viral antigens at F4 but only for very few at FF4 (see Table 7). (Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

Table 7: Association of season with viral status / antibody reactivity at F4 and FF4 (bold cells significant)

	Viral status (F4)	**Viral antibody reactivity (F4)**	**Viral status (FF4)**	**Viral antibody reactivity (FF4)**
HSV–1	OR_{cold} = 1.0 (0.8–1.4)	**gg: Δ_c = .05 *** **	OR_{cold} = 0.9 (0.6–1.2)	gg: Δ_c = .02
HSV–2	OR_{cold} = 0.9 (0.8–1.1)	mgg: Δ_c = .00	OR_{cold} = 1.0 (0.8–1.2)	mgg: Δ_c =-.04
VZV	**OR_{cold} = 1.3 (1.1–1.6)**	**giorf67: Δ_c =-.09 *** **georf68: Δ_c = .07 ****	OR_{cold} = 1.0 (0.8–1.4)	girf67: Δ_c = .03 georf68: Δ_c = .01
EBV	OR_{cold} = 1.2 (0.6–2.2)	ead: Δ_c = .02 **ebna1: Δ_c = .04 *** **vcap18: Δ_c = .06 ****** zebra: Δ_c = .03	OR_{cold} = 1.1 (0.5–2.3)	**ead: Δ_c = .07 *** **ebna1: Δ_c =-.04 **** vcap18: Δ_c = .01 **zebra: Δ_c = .04 ***
CMV	OR_{cold} = 1.0 (0.9–1.2)	**pp28: Δ_c = .09****** **pp52: Δ_c = .09***** pp65: Δ_c =-.01 **pp150: Δ_c = .07 ****	OR_{cold} = 1.1 (0.9–1.3)	pp28: Δ_c = .02 pp52: Δ_c =-.02 pp65: Δ_c = .06 pp150: Δ_c =-.01

HHV–6	**OR_{cold} = 0.8 (0.7–0.9)**	ie1atr: Δ_c =-.04 p100tr: Δ_c =-.03 **ie1btr: Δ_c = .06 *** p101ktr: Δ_c =-.06	**OR_{cold} = 1.5 (1.3–1.8)**	ie1atr: Δ_c = .03 p100tr: Δ_c = .01 ie1btr: Δ_c =-.03 p101ktr: Δ_c = .07
HHV–7	**OR_{cold} = 1.4 (1.1–1.8)**	**u14: Δ_c = .04 ***	OR_{cold} = 1.2 (0.8–1.8)	u14: Δ_c =-.01

How come there is no consistency between F4 and FF4 at all? The F4 timepoint is slightly more powerful with roughly 800 participants more than FF4, but there also seem to be other influences. Figure 12 demonstrates that age is more or less equally distributed among the seasons and does not seem to influence the results. However, it can be appreciated that whereas in F4 the majority of probands was examined in the cold season, this proportion swaps in FF4, where the majority was examined in the warm season.

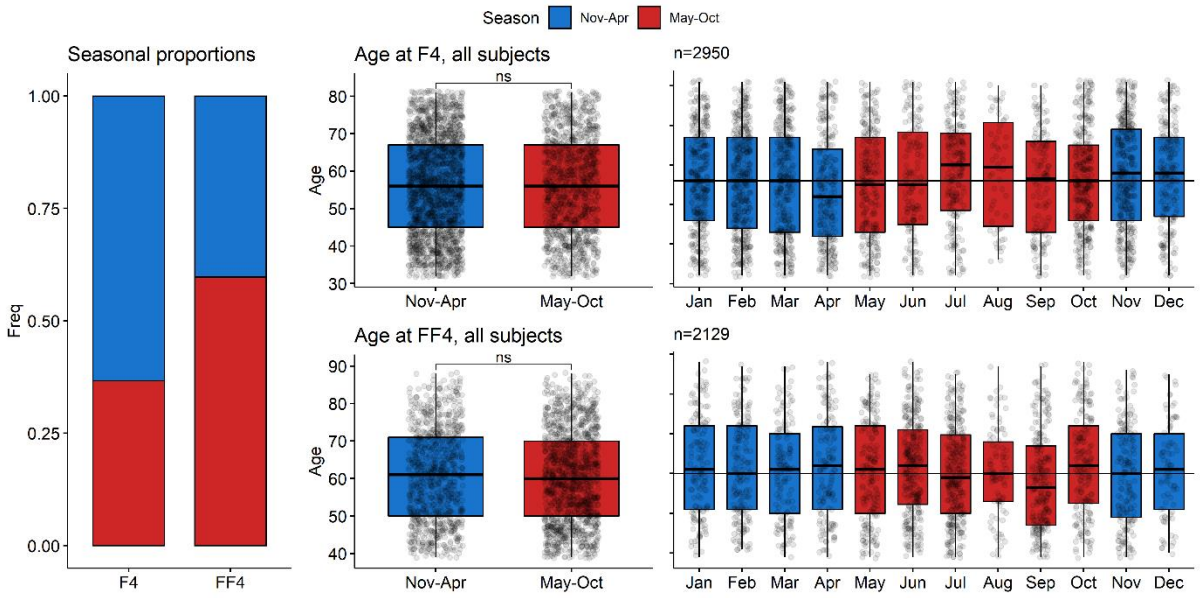


Figure 12: Seasonal proportions and age at F4 / FF4 (depending on month of examination)

Given that the season of examination is much less stable between the two timepoints than age and sex, a paired analysis seems warranted. After all, F4 and FF4 are not independent, they have a large overlap of up to 1,967 participants. Maybe the inconsistencies have to do with people "switching" season: What if probands with intrinsically higher antibody reactivities were examined in the cold season at F4 and again in the warm season at FF4? To examine these effects, paired analyses of samples seropositive at both timepoints have been performed. Figure 13 shows this exemplarily for the same virus as Figure 11, HSV–1.

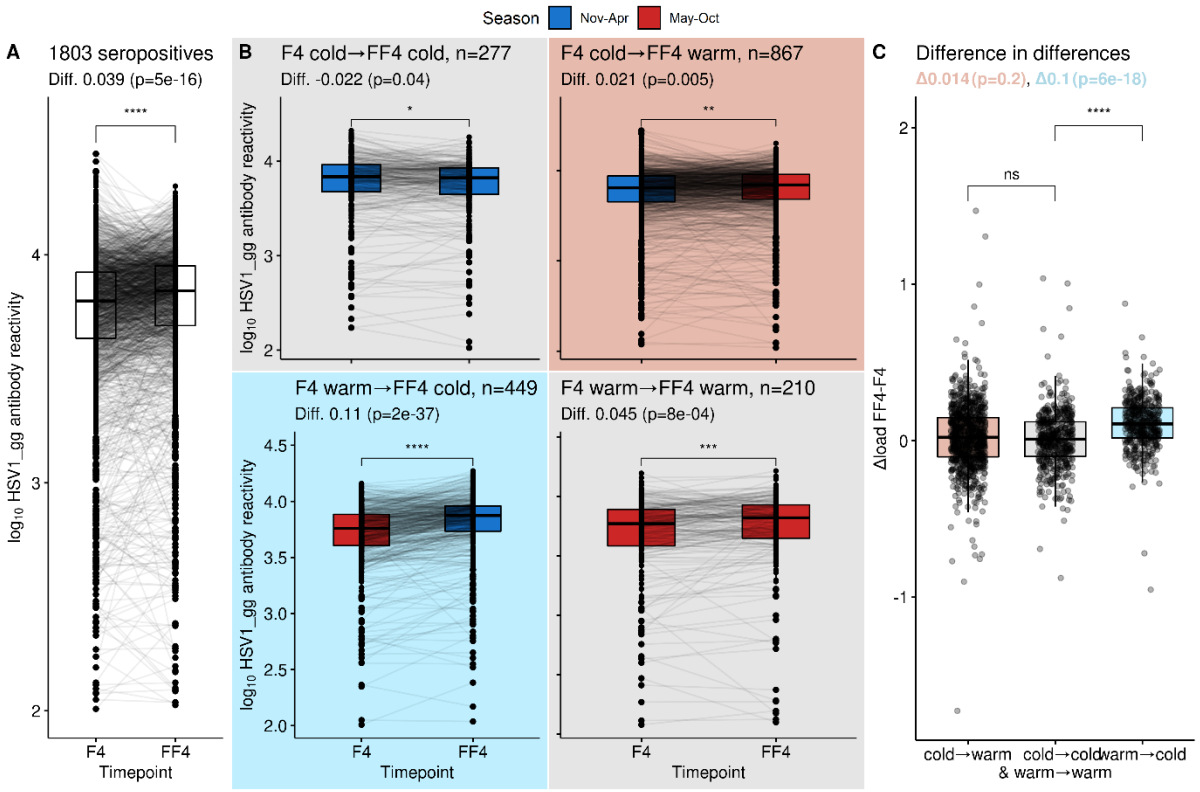


Figure 13: Pairwise analysis of HSV–1 seropositive participants (A), stratified by "seasonal switch" from F4 to FF4 (B) and difference in differences analysis (C)

In Figure 13A, one can see that the overall average antibody reactivity is significantly higher for HSV–1 seropositives at FF4 compared to F4. This is not too surprising, as in chapter 3.1.4 it has been concluded that antibody reactivity is positively correlated with age for HSV–1 (and many of the other viruses) and there are on average 6.5 years between F4 and FF4. Now if the hypothesis that viral antibody reactivity is increased in the cold season were true, one would expect the following two observations in this difference in differences analysis: (1) Participants that "switch" from the warm to the cold season would be expected to have a higher increase in viral antibody reactivity ($\Delta_{F4warm \rightarrow FF4cold}$, light blue box) than participants not switching season ($\Delta_{no-seasonal-switch}$, grey boxes combined). (2) On the other hand, participants that "switch" from the cold to the warm season would be expected to have a lower increase in viral antibody reactivity ($\Delta_{F4cold \rightarrow FF4warm}$, light red box) than participants not switching season ($\Delta_{no-seasonal-switch}$, grey boxes combined).

Interestingly, for many viral antigens the first expectation is reflected in the data while the second one is not: Participants who have been examined between May–October at F4 (warm) and between November–April at FF4 (cold) have indeed a significantly higher increase in viral antibody reactivity of HSV–1 at FF4 ($\Delta_{F4warm \rightarrow FF4cold} = 0.11$, light blue boxplot) than participants who have not switched season ($\Delta_{no-seasonal-switch} = 0.007$, grey boxes combined). The difference of these differences is highly significant: $\Delta_{F4warm \rightarrow FF4cold} - \Delta_{no-seasonal-switch} = 0.10$, p=6e-18.

Table 8 shows this warm–to–cold difference in differences analysis for all 17 viral antigens. It can be seen that the seropositive participants who switch from warm to cold have significantly higher viral antibody reactivities at FF4 than the other seropositives for most viral antigens, suggesting indeed an average increase of viral antibody reactivities during the cold season (except for HSV–2, possibly due to low number of seropositives, and HHV–6). Detailed figures for each viral antigen can be found in Supplementary Figures 35–50.

Table 8: Difference in differences analysis of viral antigens for paired comparison of seropositives with warm to cold "seasonal switch" against no–seasonal–switch (corresponding to light blue box in figure above)

	mean $\Delta_{F4warm \rightarrow FF4cold}$ (paired t–test p–value)	**mean $\Delta_{no\ switch}$ (paired t–test p–value)**	**mean $\Delta_{F4warm \rightarrow FF4cold}$ - mean $\Delta_{no–seasonal–switch}$ (two sample t–test p)**
HSV–1 gg	**0.11 (p=1.7e-37)**	0.0067 (p=0.4)	**0.10 (p=6e-18)**
HSV–2 mgg	**0.14 (p=0.00098)**	0.074 (p=0.1)	0.069 (p=0.3)
VZV giorf67	-0.035 (p=0.25)	-0.021 (p=0.6)	-0.014 (p=0.8)
VZV georf68	**0.14 (p=2.5e-12)**	-0.024 (p=0.3)	**0.16 (p=2e-08)**
EBV ead	**0.19 (p=1.5e-57)**	**0.094 (p=4e-16)**	**0.099 (p=8e-11)**
EBV ebna1	**0.044 (p=5.1e-07)**	**-0.024 (p=0.004)**	**0.068 (p=2e-08)**
EBV vcap18	**0.12 (p=1.9e-53)**	**0.027 (p=4e-04)**	**0.095 (p=1e-19)**
EBV zebra	**0.19 (p=5.1e-63)**	**0.080 (p=3e-14)**	**0.11 (p=3e-14)**
CMV pp28	**0.14 (p=2.9e-27)**	**0.028 (p=0.04)**	**0.11 (p=5e-10)**
CMV pp52	**0.099 (p=2.6e-21)**	**0.025 (p=0.05)**	**0.074 (p=4e-06)**
CMV pp65	**0.17 (p=2.6e-21)**	**0.068 (p=7e-05)**	**0.099 (p=2e-05)**
CMV pp150	**0.11 (p=6.2e-18)**	0.016 (p=0.2)	**0.089 (p=8e-08)**
HHV–6 ie1atr	**0.18 (p=3.5e-29)**	**0.16 (p=5e-22)**	0.022 (p=0.3)
HHV–6 p100tr	0.053 (p=0.47)	-0.031 (p=0.6)	0.084 (p=0.4)
HHV–6 ie1btr	**0.055 (p=0.034)**	0.045 (p=0.08)	0.0094 (p=0.8)
HHV–6 p101ktr	0.043 (p=0.084)	-0.0071 (p=0.8)	0.050 (p=0.2)
HHV–7 u14	**0.17 (p=1.7e-48)**	**0.059 (p=8e-10)**	**0.11 (p=2e-15)**

However, the expected opposite cold–to–warm effect is not observed: participants who have been examined in the cold season at F4 and in the warm season at FF4 do not have the expected lower than no–seasonal–switch increase in viral antibody reactivity, compare the light red box in Figure 13. Indeed, for most viral antigens, this difference in differences is not significant as can be seen in rightmost column of Table 9 below. The only two significant crude p–values are around 0.02 and would easily be lost after any kind of multiple–testing–correction.

Table 9: Difference in differences analysis of viral antigens for paired comparison of seropositives with cold to warm "seasonal switch" against no–seasonal–switch (corresponding to light red box in figure above)

	mean $\Delta_{F4cold \rightarrow FF4warm}$ (paired t–test p–value)	**mean $\Delta_{no\ switch}$ (paired t–test p–value)**	**mean $\Delta_{F4cold \rightarrow FF4warm}$ - mean $\Delta_{no–seasonal–switch}$ (two sample t–test p)**
HSV–1 gg	**0.021 (p=0.005)**	0.0067 (p=0.4)	0.014 (p=0.2)
HSV–2 mgg	**0.1 (p=0.006)**	0.074 (p=0.1)	0.03 (p=0.6)

VZV giorf67	-0.039 (p=0.3)	-0.021 (p=0.6)	-0.018 (p=0.7)
VZV georf68	-0.0026 (p=0.9)	-0.024 (p=0.3)	0.021 (p=0.4)
EBV ead	**0.11 (p=2e-32)**	**0.094 (p=4e-16)**	0.013 (p=0.4)
EBV ebna1	0.0032 (p=0.7)	**-0.024 (p=0.004)**	**0.028 (p=0.02)**
EBV vcap18	**0.032 (p=5e-07)**	**0.027 (p=4e-04)**	0.0055 (p=0.6)
EBV zebra	**0.1 (p=6e-32)**	**0.08 (p=3e-14)**	0.022 (p=0.1)
CMV pp28	**0.046 (p=6e-05)**	**0.028 (p=0.04)**	0.018 (p=0.3)
CMV pp52	**0.027 (p=0.004)**	**0.025 (p=0.05)**	0.0017 (p=0.9)
CMV pp65	**0.076 (p=9e-08)**	**0.068 (p=7e-05)**	0.0081 (p=0.7)
CMV pp150	**0.041 (p=8e-05)**	0.016 (p=0.2)	0.025 (p=0.1)
HHV–6 ie1atr	**0.11 (p=1e-12)**	**0.16 (p=5e-22)**	**-0.048 (p=0.02)**
HHV–6 p100tr	-0.066 (p=0.1)	-0.031 (p=0.6)	-0.035 (p=0.6)
HHV–6 ie1btr	-0.0062 (p=0.8)	0.045 (p=0.08)	-0.052 (p=0.1)
HHV–6 p101ktr	0.036 (p=0.1)	-0.0071 (p=0.8)	0.044 (p=0.2)
HHV–7 u14	**0.064 (p=4e-18)**	**0.059 (p=8e-10)**	0.0047 (p=0.7)

3.1.7 Combining Confounders

It has been concluded that sex, age and potentially season are influencing viral status and viral antibody reactivity for at least some of the viruses and their antigens. Now, their combined effects in multivariate logistic regressions for binary viral status (seronegative / seropositive) are examined. Table 10 shows the ORs of the three covariates for both F4 and FF4 with 95% CIs and Bonferroni significant cells in bold (corrected for 7 viruses × 2 timepoints = 14 tests).

Table 10: Combined effects of sex, age and season on viral status (95% CI in brackets, Bonferroni significant cells bold)

	F4			**FF4**		
	$OR_{10\ years}$	**OR_{female}**	**$OR_{cold\ season}$**	**$OR_{10\ years}$**	**OR_{female}**	**$OR_{cold\ season}$**
HSV–1	**1.46 (1.3–1.6)**	**1.36 (1.1–1.7)**	1.06 (0.8–1.3)	**1.52 (1.4–1.7)**	1.28 (1.0–1.7)	0.86 (0.7–1.1)
HSV–2	**1.19 (1.1–1.3)**	1.20 (1.0–1.5)	0.94 (0.7–1.2)	1.12 (1.0–1.2)	1.26 (1.0–1.6)	0.98 (0.8–1.3)
VZV	0.96 (0.9–1.0)	**0.64 (0.5–0.8)**	1.33 (1.1–1.6)	1.05 (1.0–1.2)	0.72 (0.6–0.9)	1.04 (0.8–1.3)
EBV	1.40 (1.1–1.7)	2.42 (1.4–4.3)	1.17 (0.7–2.0)	1.44 (1.1–1.9)	2.50 (1.3–5.1)	1.08 (0.6–2.1)
CMV	**1.34 (1.3–1.4)**	**1.64 (1.4–1.9)**	1.01 (0.9–1.2)	**1.29 (1.2–1.4)**	**1.44 (1.2–1.7)**	1.07 (0.9–1.3)
HHV–6	1.01 (1.0–1.1)	1.22 (1.0–1.4)	**0.76 (0.6–0.9)**	0.98 (0.9–1.1)	1.05 (0.9–1.2)	**1.50 (1.3–1.8)**
HHV–7	0.91 (0.8–1.0)	**1.54 (1.3–1.9)**	1.39 (1.1–1.7)	1.06 (0.9–1.2)	**1.76 (1.3–2.5)**	1.26 (0.9–1.8)

It can be seen that the effects are overall very similar to those described earlier in Table 5, Table 6 and Table 7, suggesting there is very little correlation among the three confounders themselves. F4 is overall more powerful as it has roughly 800 more samples than FF4, which is reflected by a higher number of Bonferroni significant findings (bold cells). The overall correlation of effect sizes between F4 and FF4 is R=0.92 for age, R=0.95 for sex and R=-0.29 for season, reflecting that the results for season have to be viewed with caution, as already

established in 3.1.6. However, there is evidence for robust effects of age on viral status for HSV–1, HSV–2, EBV and CMV (positive association) and robust effects of sex on viral status for all viruses but HHV–6 with generally higher odds for females except for VZV, which has higher odds for males.

Table 11 shows the combined effects of sex, age and season on the log–transformed viral antibody reactivities of seropositives of all 17 antigens as determined by multivariate linear regression models on the respective MFIs. Bold cells are significant after Bonferroni correction for 34 tests (17 antigens × 2 timepoints). (Crude significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001)

Table 11: Combined effects of sex, age and season on antibody reactivity of seropositives (Bonferroni significant cells bold)

	F4			FF4		
	$beta_{10\ years}$	**$beta_{female}$**	**$beta_{cold}$**	**$beta_{10\ years}$**	**$beta_{female}$**	**$beta_{cold}$**
HSV–1 gg	**0.03 ******	**-0.05 *** **	**0.05 *** **	**0.03 ******	-0.04 *	0.01 ns
HSV–2 mgg	**0.10 ******	**0.25 ******	0.01 ns	**0.12 ******	0.11 ns	-0.02 ns
VZV giorf67	-0.01 ns	0.03 ns	-0.08 *	-0.01 ns	0.04 ns	0.04 ns
VZV georf68	0.02 *	**-0.08 *** **	0.07 **	0.02 *	**-0.14 ******	0.00 ns
EBV ead	0.01 ns	0.06 **	0.02 ns	0.02 **	0.05 *	0.07 **
EBV ebna1	-0.01 *	-0.05 **	0.04 *	-0.01 ns	-0.04 *	-0.04 *
EBV vcap18	**0.01 *** **	-0.00 ns	**0.06 ******	**0.01 *** **	-0.01 ns	0.00 ns
EBV zebra	0.02 **	**0.11 ******	0.03 ns	**0.03 ** **	**0.12 ******	0.05 *
CMV pp28	-0.00 ns	**0.09 ******	**0.09 ******	-0.00 ns	0.06 *	0.02 ns
CMV pp52	**0.04 ******	**0.14 ******	**0.09 *** **	**0.04 ** **	**0.16 ******	-0.01 ns
CMV pp65	0.02 ns	0.06 *	-0.00 ns	0.04 **	0.01 ns	0.06 ns
CMV pp150	0.01 ns	**0.13 ******	0.07 **	0.03 **	**0.11 ******	-0.01 ns
HHV–6 ie1atr	-0.01 ns	-0.03 ns	-0.04 ns	0.01 ns	0.02 ns	0.03 ns
HHV–6 p100tr	0.01 ns	-0.01 ns	-0.03 ns	0.01 ns	-0.05 ns	0.01 ns
HHV–6 ie1btr	-0.02 *	0.06 *	0.06 *	-0.02 ns	**0.11 *** **	-0.03 ns
HHV–6 p101ktr	0.00 ns	0.06 ns	-0.06 ns	0.02 ns	0.04 ns	0.07 ns
HHV–7 u14	**-0.03 ******	**0.12 ******	0.04 *	**-0.03 ******	**0.11 ******	-0.00 ns

Again one can see that the effect directions and significances are overall very similar to those described earlier in Table 5, Table 6 and Table 7, suggesting again that there is very little correlation among the three confounders themselves. The correlation of effect sizes is highest for age with R=0.97, followed by sex with R=0.86 and very low for season with R=-0.53. This fairly strong negative correlation maybe does indeed support the idea stated in 3.1.6 that many people with intrinsically high antibody reactivities have switched from having been examined in the cold season at F4 to the warm season at FF4.

3.2 Characterisation of Type 2 Diabetes in KORA

3.2.1 Prevalence and Incidence of T2D

At F4, 845 out of 2,950 participants have IFG/IGT (prediabetes) and 337 have T2D as verified by OGTT, which corresponds to prevalences of 29% and 11% respectively. At FF4, 754 out of 2,129 participants have prediabetes and 310 have T2D, which increases prevalences to 34% and 15% respectively. In this thesis, the prevalences and incidences of the two subgroups pre-diabetes and T2D are often examined as a combined "(pre)diabetes" outcome in order to obtain a binary outcome variable.

Figure 14 concentrates on the 1,967 overlapping samples between the two timepoints and demonstrates incidence, which can be partitioned in several ways: Out of 1,257 people with normal OGTTs at F4, there are 364 incident cases of IFG/IGT (incidence rate (IR) = 28.9% per 6.5 years or 4.5% per year) and 17 incident cases of T2D (combined incidence rate = 30.3% per 6.5 years or 4.7% per year). Among 542 people with IFG/IGT at F4, there are 113 incident cases of T2D (incidence rate = 20.8% per 6.5 years or 3.2% per year). Some patients are also improving: 10 of originally 168 T2D participants recover to IFG/IGT and 1 of them even re-covers to normal OGTT (faint downward–facing yellow and blue lines in Figure 14). Out of originally 542 prediabetics, 90 recover to normal OGTT (thick downward–facing blue line).

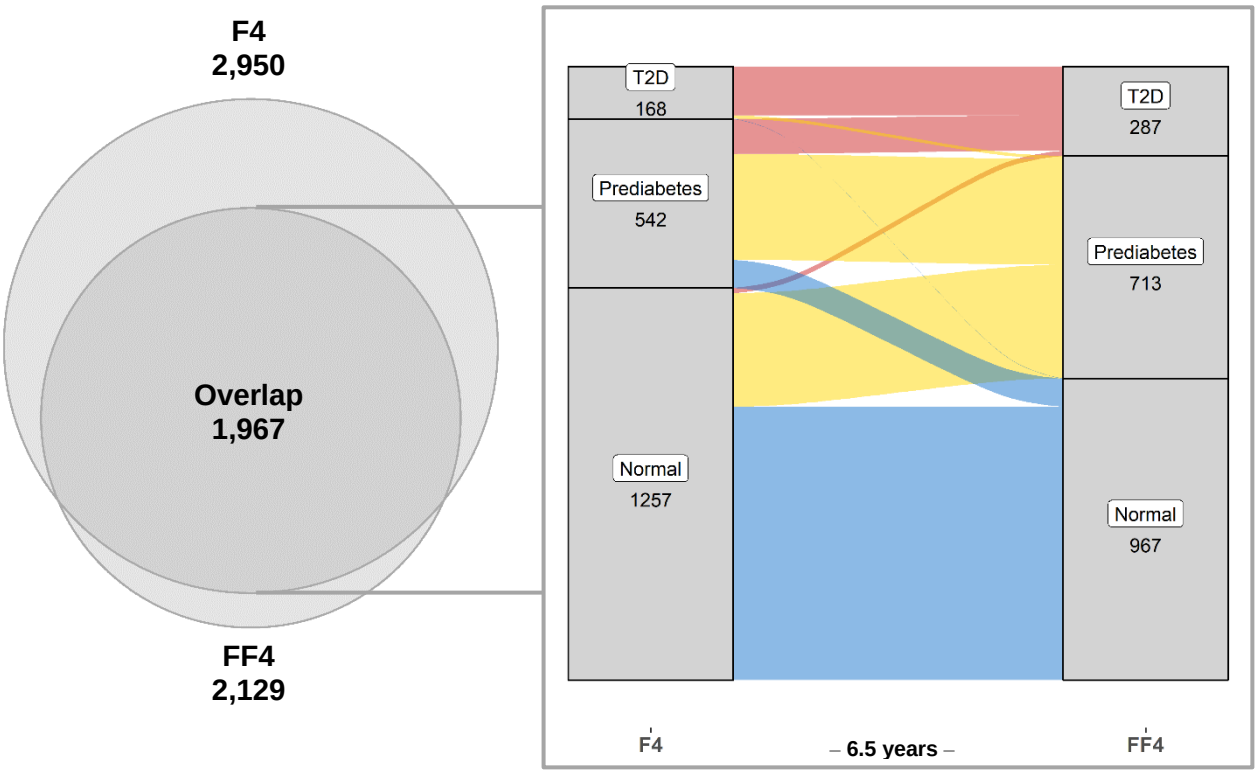


Figure 14: Venn diagram showing overlap of F4 / FF4 participants and alluvial plot showing incidence of pre-diabetes / T2D

3.2.2 Known Risk Factors

The conventional environmental risk factors discussed in the introduction were all highly significant in both the F4 group and the FF4 group of KORA, which can be seen for F4 in Figure 15 (significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001). Old age, overweight, smoking and low education–levels are known to increase risk of type 2 diabetes.

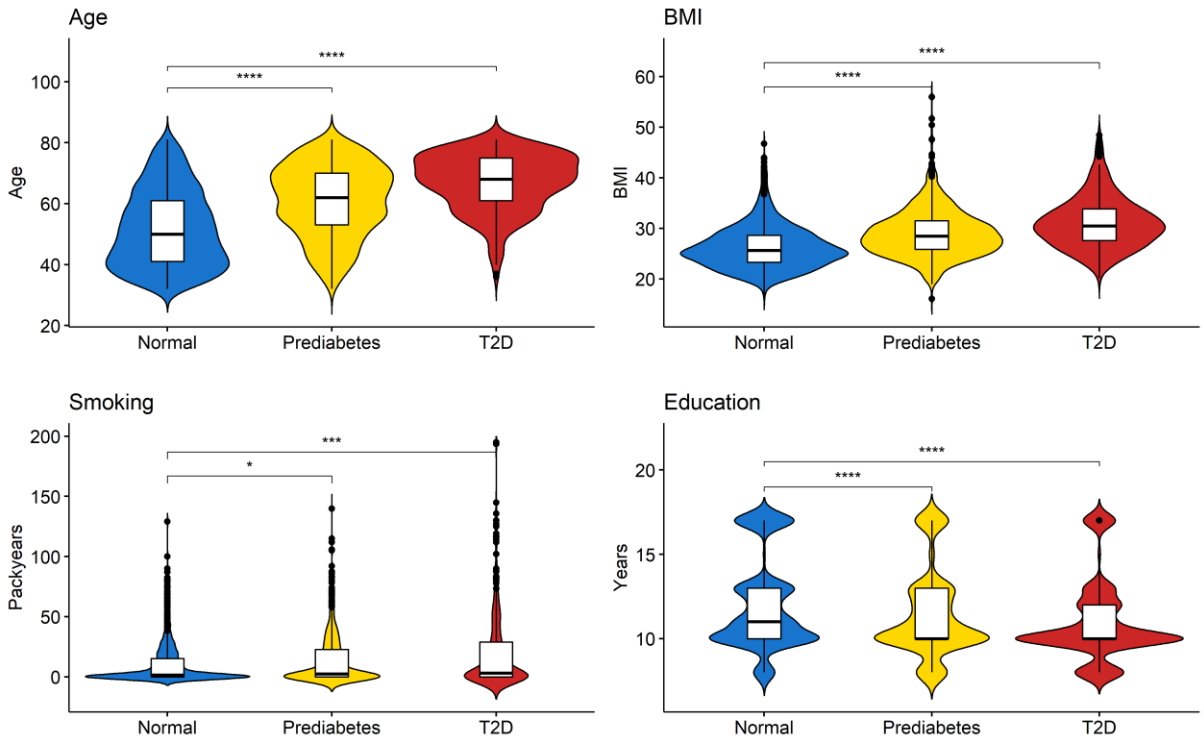


Figure 15: T2D risk factors age, overweight, smoking, poor–education at F4 timepoint

Interestingly, there was also a highly significant difference in the proportion of males to females, with more males suffering from prediabetes and T2D in this cohort. This as well as other demographic findings are discussed in detail by Meisinger et al. [65].

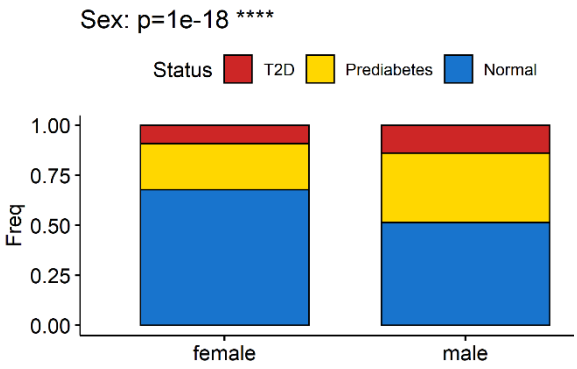


Figure 16: sex vs. diabetes at F4 timepoint

3.3 Association of Herpesviruses with T2D in KORA

3.3.1 Association of Herpesviruses with Diabetes Prevalence

At first, the associations of herpesvirus serostatus with diabetes prevalence at the two timepoints F4 & FF4 will be examined separately, and then combined with a GEE model for maximum statistical power (see 2.3.4 in the Methods section). In order to get a binary outcome variable,

prediabetes (IFG / IGT) and T2D have been grouped together as "(pre)diabetes". Consider for example HSV–1 in Figure 17: The boxes show contingency tables, below them are summary statistics for the confounders sex and age. On the right–hand side is a forest plot, demonstrating how the OR changes with adjustments for said confounders.

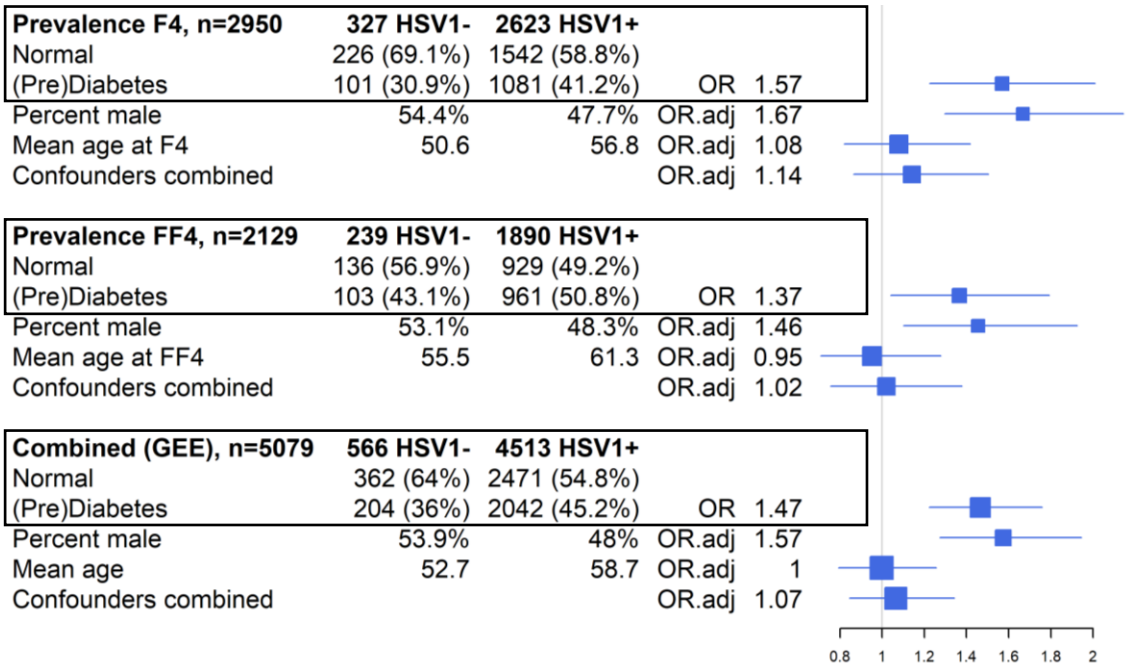


Figure 17: Associations between HSV–1 seropositivity and (pre)diabetes prevalence, separated by F4 and FF4 as well as jointly analysed.

The crude odds of having prediabetes or T2D are 1.57 (95% CI 1.23–2.01) times higher for carriers of HSV–1 versus noncarriers in F4 and 1.37 (95% CI 1.04–1.79) times higher in FF4. Combining data from both timepoints in a GEE model, the overall crude odds are increased by 47% in carriers (OR = 1.47, 95% CI 1.22–1.76).

There is considerable confounding by sex and age. Sex is a negative confounder (the effect increases after adjustment): Women are more likely to be viral carriers (see 3.1.5) but less likely to be (pre)diabetic (see 3.2.2). Age on the other hand is a positive confounder (the effect decreases after adjustment): As carriers of HSV–1 tend to be older (see 3.1.4) and older people are more likely to be (pre)diabetic (see 3.2.2), the effect decreases and the significance is lost when calculated in a logistic regression model with adjustment for age.

Table 12 shows ORs for all 7 viruses with significant results in bold. As can be seen above, crude ORs for HSV–1 and also for CMV are significantly positively associated with diabetic prevalence at both F4 and FF4 whereas HSV–2 and VZV are only significantly positively associated at FF4. The crude OR of HHV–7 is significantly negatively associated at F4. However, only two of these significant associations of viral status versus diabetes prevalence remain significant after adjustment for confounders.

Table 12: Crude and adjusted ORs for associations of viral status vs. (pre)diabetes prevalence (bold cells significant, 95% CI in brackets)

	(Pre)Diabetes Prevalence F4		(Pre)Diabetes Prevalence FF4		(Pre)Diabetes Prevalence F4 & FF4 Combined (GEE)	
	Crude OR	**Adj. OR**	**Crude OR**	**Adj. OR**	**Crude OR**	**Adj. OR**
HSV–1	**1.57 (1.23–2.01)**	1.14 (0.87–1.5)	**1.37 (1.04–1.79)**	1.02 (0.76–1.38)	**1.47 (1.22–1.76)**	1.07 (0.85–1.34)
HSV–2	1.03 (0.82–1.3)	0.87 (0.67–1.12)	**1.36 (1.05–1.75)**	**1.37 (1.04–1.82)**	1.18 (0.99–1.4)	1.01 (0.82–1.23)
VZV	1.1 (0.91–1.34)	1.08 (0.87–1.33)	**1.33 (1.05–1.69)**	1.21 (0.93–1.57)	**1.22 (1.05–1.42)**	1.15 (0.97–1.37)
EBV	1.25 (0.73–2.13)	1.01 (0.56–1.81)	0.55 (0.29–1.07)	**0.44 (0.22–0.9)**	0.91 (0.61–1.36)	0.69 (0.43–1.11)
CMV	**1.35 (1.16–1.56)**	1.11 (0.94–1.31)	**1.28 (1.08–1.52)**	1.14 (0.94–1.38)	**1.3 (1.17–1.46)**	1.13 (0.98–1.3)
HHV–6	0.93 (0.8–1.08)	0.93 (0.79–1.1)	1.03 (0.87–1.22)	1.06 (0.88–1.28)	1.01 (0.91–1.13)	0.98 (0.87–1.11)
HHV–7	**0.8 (0.66–0.98)**	0.96 (0.76–1.2)	1.03 (0.74–1.42)	1.13 (0.79–1.62)	0.93 (0.78–1.1)	1.07 (0.89–1.3)

HSV–2 remains significant after adjusting for sex and age with OR.adj = 1.37 (95% CI 1.04–1.82). However, while not significant, the HSV–2 result from F4 is pointing in the opposition direction and thus the combined GEE model is fairly balanced with OR.adj = 1.01 (95% CI 0.82–1.23), see Figure 18.

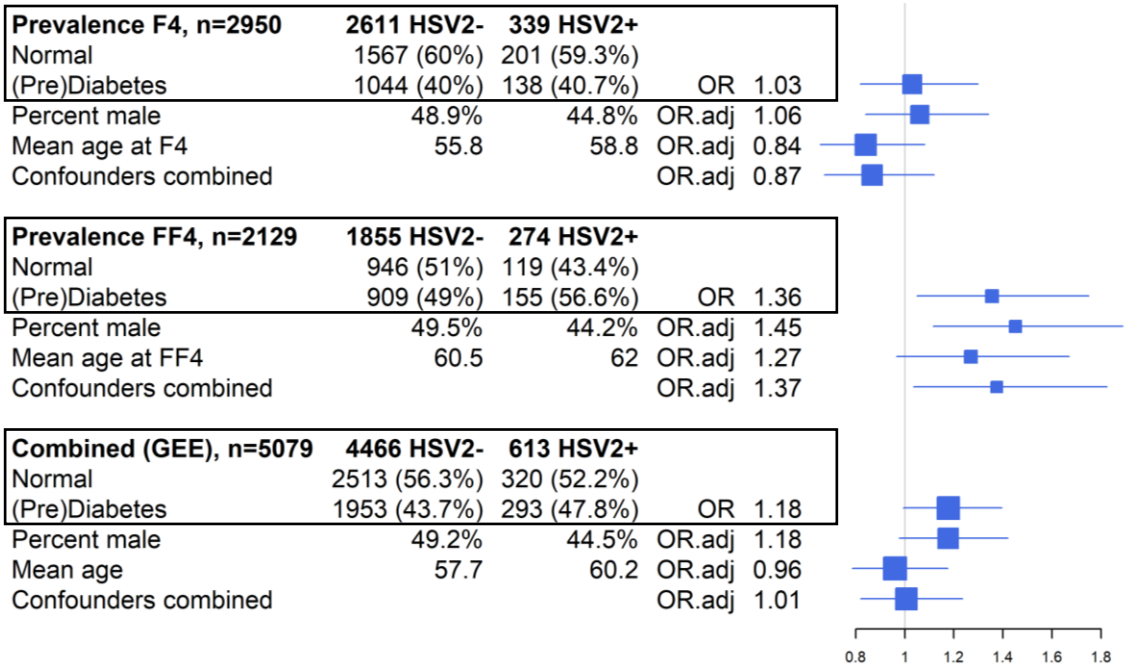


Figure 18: Associations between HSV–2 seropositivity and (pre)diabetes prevalence, separated by F4 and FF4 as well as jointly analysed.

The second significant adjusted result is that for EBV in FF4, which can be seen in Figure 19. Because the virus is highly prevalent (98%), only few participants are seronegative (60 at F4 & 39 at FF4). As EBV is seemingly protective for (pre)diabetic prevalence at FF4 with an adjusted OR of 0.44 (95% CI 0.22–0.9), sex is now the positive confounder and age the negative confounder. However, the result remains ambiguous as the effect direction points slightly in the opposite direction at F4.

Prevalence F4, n=2950	60 EBV-	2890 EBV+		
Normal	39 (65%)	1729 (59.8%)		
(Pre)Diabetes	21 (35%)	1161 (40.2%)	OR	1.25
Percent male	68.3%	48.1%	OR.adj	1.44
Mean age at F4	50.9	56.3	OR.adj	0.86
Confounders combined			OR.adj	1.01
Prevalence FF4, n=2129	**39 EBV-**	**2090 EBV+**		
Normal	14 (35.9%)	1051 (50.3%)		
(Pre)Diabetes	25 (64.1%)	1039 (49.7%)	OR	0.55
Percent male	69.2%	48.5%	OR.adj	0.66
Mean age at FF4	55.8	60.8	OR.adj	0.37
Confounders combined			OR.adj	0.44
Combined (GEE), n=5079	**99 EBV-**	**4980 EBV+**		
Normal	53 (53.5%)	2780 (55.8%)		
(Pre)Diabetes	46 (46.5%)	2200 (44.2%)	OR	0.91
Percent male	68.7%	48.2%	OR.adj	1.06
Mean age	52.8	58.2	OR.adj	0.6
Confounders combined			OR.adj	0.69

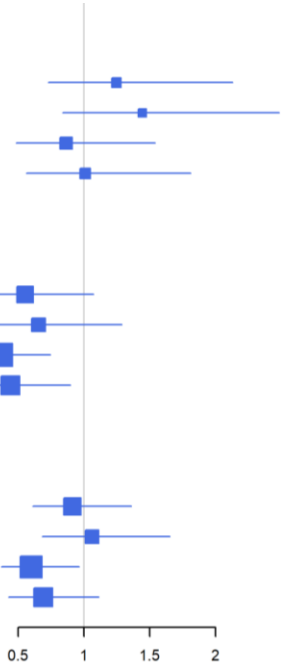


Figure 19: Associations between EBV seropositivity and (pre)diabetes prevalence, separated by F4 and FF4 as well as jointly analysed.

For CMV the overall picture is similar to HSV–1, see Figure 20: The positive association of viral serostatus with (pre)diabetes prevalence is negatively confounded by sex and positively by age. However, adjusting for both confounders and combining the data from both timepoints in the GEE model in order to maximise power, the adjusted effect size only slightly misses significance with an OR.adj = 1.13 (95% CI 0.98–1.3).

Prevalence F4, n=2950	1511 CMV-	1439 CMV+		
Normal	958 (63.4%)	810 (56.3%)		
(Pre)Diabetes	553 (36.6%)	629 (43.7%)	OR	1.35
Percent male	53.9%	42.8%	OR.adj	1.48
Mean age at F4	53.8	58.6	OR.adj	1.01
Confounders combined			OR.adj	1.11
Prevalence FF4, n=2129	**1126 CMV-**	**1003 CMV+**		
Normal	596 (52.9%)	469 (46.8%)		
(Pre)Diabetes	530 (47.1%)	534 (53.2%)	OR	1.28
Percent male	52.6%	44.7%	OR.adj	1.4
Mean age at FF4	58.9	62.6	OR.adj	1.03
Confounders combined			OR.adj	1.14
Combined (GEE), n=5079	**2637 CMV-**	**2442 CMV+**		
Normal	1554 (58.9%)	1279 (52.4%)		
(Pre)Diabetes	1083 (41.1%)	1163 (47.6%)	OR	1.3
Percent male	53.3%	43.6%	OR.adj	1.46
Mean age	56	60.3	OR.adj	1.03
Confounders combined			OR.adj	1.13

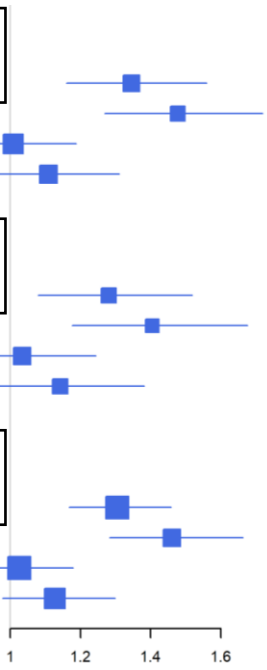


Figure 20: Associations between CMV seropositivity and (pre)diabetes prevalence, separated by F4 and FF4 as well as jointly analysed.

Detailed contingency tables and forest plots for VZV, HHV–6 and HHV–7 can be found in Supplementary Figures 51–53.

3.3.2 Association of Herpesviruses with Diabetes Incidence

The following section focuses on incidence instead of prevalence, after all the longitudinal nature of this study is one of its strengths. 1,257 of the total 1,967 overlapping participants have a normal fasting glucose and glucose tolerance at F4 and are thus still at risk for developing (pre)diabetes (Figure 14). A graphical overview exemplarily for HSV–2 including contingency table, forest plot and adjustments for the confounders sex and age similarly to the prevalence–figures in the chapter above can be seen in Figure 21.

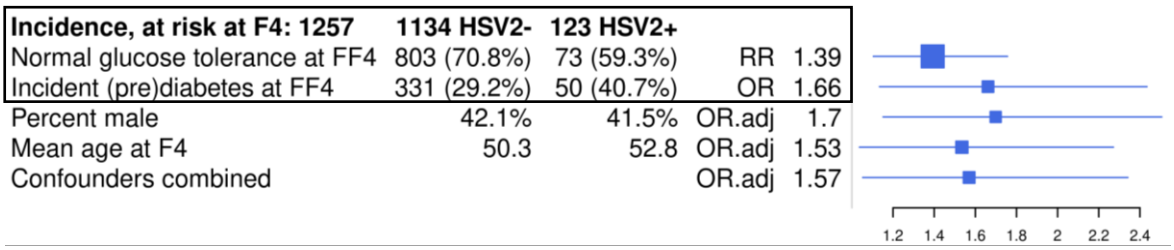


Figure 21: Associations between HSV–2 seropositivity and (pre)diabetes incidence.

Participants seropositive for HSV–2 have 39% higher risk (95% CI 10%–76%) and 66% higher crude odds (OR = 1.66, 95% CI 1.14–2.44) to develop (pre)diabetes in the 6.5 years between F4 and FF4. There is slight confounding by sex and age but adjusting for these only changes the result marginally: OR.adj = 1.57 (95% CI 1.05–2.35), potentially hinting at a contribution of HSV–2 to (pre)diabetes development.

Table 13 shows the crude and adjusted ORs for the association of all 7 viruses with (pre)diabetes incidence. Only HSV–2 and CMV reach significance, both for the crude as well as for the adjusted logistic model, potentially being risk factors for (pre)diabetes incidence. Carriers of HSV–1 and HHV–6 also have slightly higher adjusted odds of developing (pre)diabetes than non–carriers, albeit not reaching significance. Seropositives for EBV have lower odds than the 27 seronegatives among the at risk population. The adjusted ORs for VZV and HHV–7 are nearly 1, showing little to no association.

Table 13: Crude and adjusted ORs for associations of viral status vs. (pre)diabetes incidence (bold cells significant, 95% CI in brackets)

	(Pre)Diabetes Incidence	
	Crude OR	**Adjusted OR**
HSV–1	1.31 (0.90–1.89)	1.14 (0.77–1.68)
HSV–2	**1.66 (1.14–2.44)**	**1.57 (1.05–2.35)**
VZV	1.01 (0.73–1.39)	0.94 (0.67–1.32)
EBV	0.54 (0.25–1.15)	0.55 (0.25–1.22)
CMV	**1.46 (1.15–1.86)**	**1.36 (1.05–1.76)**
HHV–6	1.13 (0.89–1.45)	1.21 (0.93–1.56)
HHV–7	0.82 (0.58–1.14)	0.93 (0.66–1.32)

When examining the picture for CMV, a by this time well–known pattern emerges: The effect is negatively confounded by sex and positively by age, just as was observed for the prevalence models (compare Figure 20). However, while the combined GEE prevalence model for CMV slightly missed significance after adjustments with an OR.adj of 1.13 (95% CI 0.98–1.3), this is not the case in the incidence model with an OR.adj of 1.37 (95% CI 1.05–1.76). Details for the other 5 viruses can be found in Supplementary Figures 54–58.

Incidence, at risk at F4: 1257	**697 CMV-**	**560 CMV+**		
Normal glucose tolerance at FF4	511 (73.3%)	365 (65.2%)	RR	1.3
Incident (pre)diabetes at FF4	186 (26.7%)	195 (34.8%)	OR	1.47
Percent male	46.3%	36.6%	OR.adj	1.62
Mean age at F4	48.5	53	OR.adj	1.25
Confounders combined			OR.adj	1.37

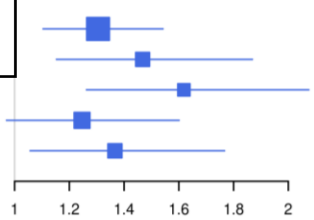


Figure 22: Associations between CMV seropositivity and (pre)diabetes incidence.

3.3.3 Association of Number of Viruses with Diabetes Prevalence & Incidence

Next, the questions is investigated, whether certain viral co–occurrence patterns might influence (pre)diabetes prevalence or incidence. Figure 23 shows the 10 most co–occurring viral patterns at FF4, similar to Figure 7 in 3.1.3 (F4) (see Supplementary Figure 59 for F4). There is no obvious effect of the different groups on diabetic prevalence and the chi–square test for the underlying 3×10 table is not significant with p=0.1. If anything, there seems to be a slight trend that more viruses tend to go along with higher prevalence of (pre)diabetes, as exemplified by the two red ellipses. This trend will be explored more formally now.

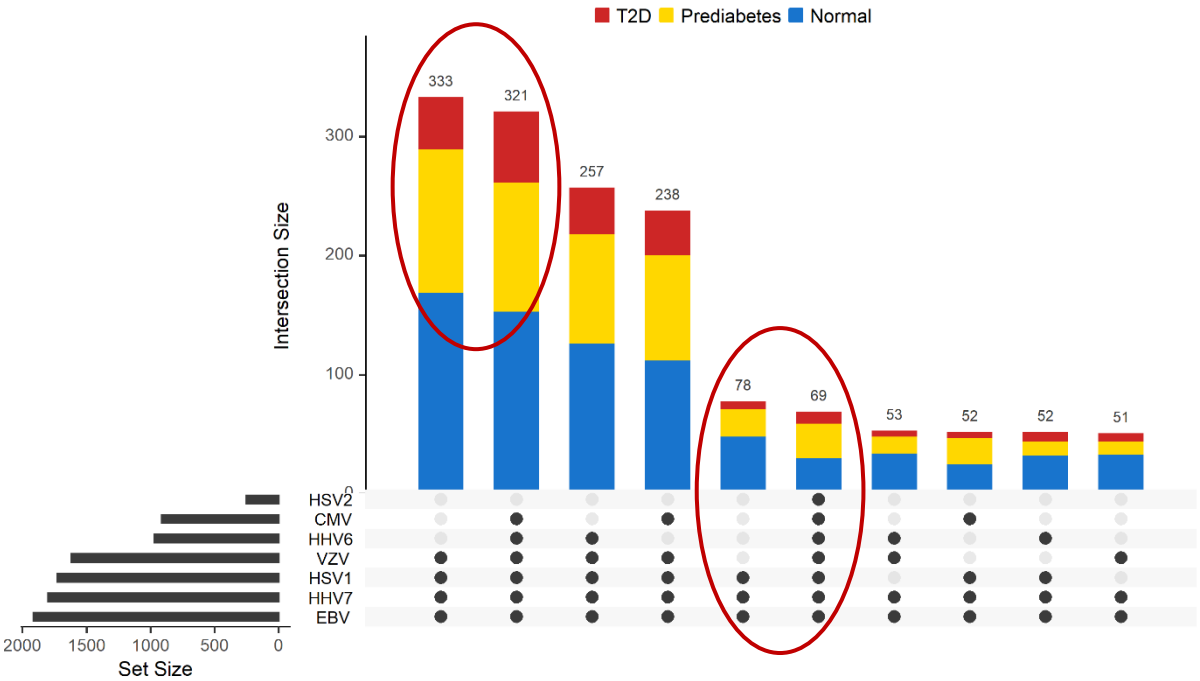


Figure 23: Co–occurrence of viruses at FF4 with groups coloured by (pre)diabetic prevalence (n=1,947, right tail cut off)

In section 3.1.3 it got apparent that the number of viruses per person is on average 4 or larger at both F4 and FF4, increasing significantly between the two timepoints. Indeed, there seems to be an association of the number of viruses per person and their diabetic prevalence as demonstrated by Figure 24: While only 42% of participants positive for 3 viruses at FF4 are (pre)diabetic, this number increases to 46% for 4 viruses, 52% for 5 viruses, 54% for 6 viruses and finally 57% for 7 viruses (the groups with only and one and two viruses contain very few participants).

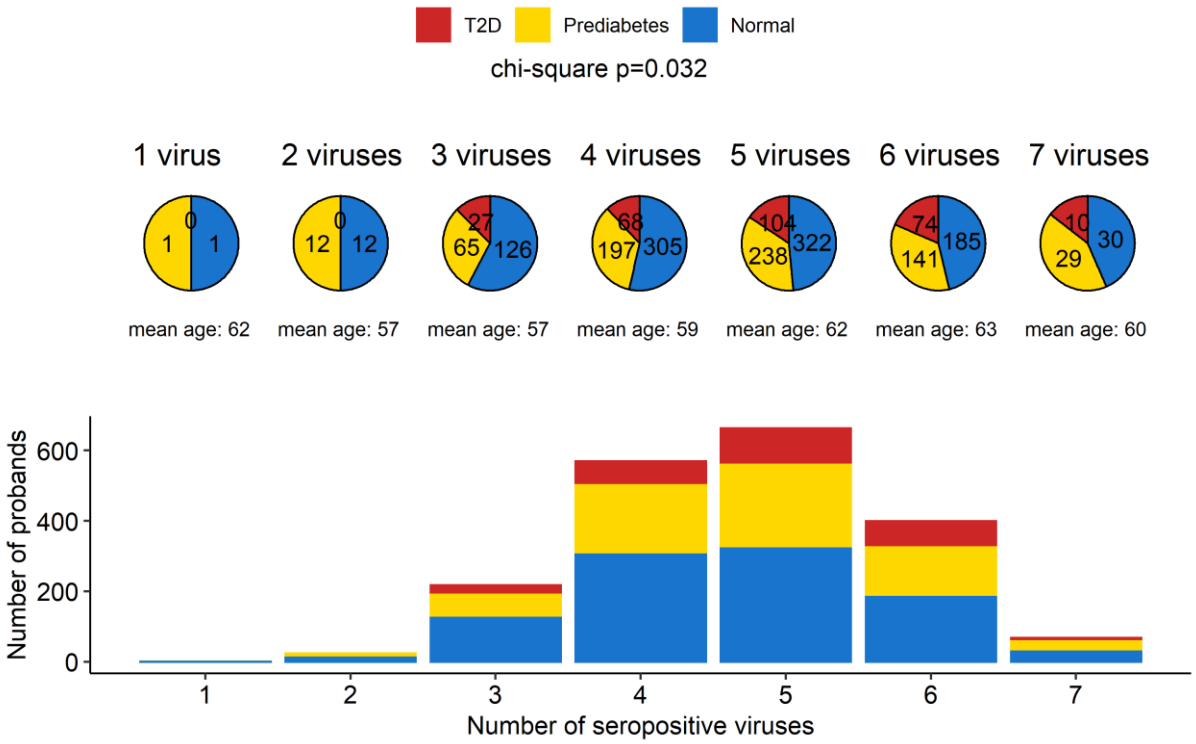


Figure 24: Number of seropositive viruses vs. diabetic prevalence in FF4 (n=1,947)

The chi–square test for the underlying 3×7 table is significant with p=0.032. However, while the trend is similar at the F4 timepoint, it misses significance with chi–square p=0.17 (see Supplementary Figure 60). Additionally, there seems to be some confounding by age, which is illustrated in Figure 24: Carriers of only few viruses tend to be younger than carriers of more viruses. In order to examine these relationships more thoroughly, logistic regression models for both F4 and FF4 with a binary dependent variable indicating normal vs. (pre)diabetic prevalence have been fitted. As independent variable I have used the number of viruses per person and as covariates I have included age and sex.

Table 14 shows the results of these logistic regression models: The unadjusted coefficients for number of viruses are significant at both F4 and FF4 with an OR of 1.15 at FF4, indicating that the odds of being (pre)diabetic increase by 15% per seropositive virus. The ORs and their significance increase when adjusting for sex only, suggesting slight negative confounding by sex.

However, the coefficients lose their significance when adjusted for age. It remains unclear whether the whole effect is explained by age or whether the sample sizes are merely too small to detect the smaller effect remaining after adjusting for age.

Table 14: ORs of (pre)diabetics over normal per additional number of viruses (significant cells bold)

	Prevalence F4 (n=2,520)	**Prevalence FF4 (n=1,947)**	**Incidence (n=986 at risk)**
Number of viruses only	**OR 1.08 (p = 0.04)**	**OR 1.15 (p = 0.001)**	OR 1.12 (p = 0.07)
Num. vir. + sex	**OR 1.11 (p = 0.005)**	**OR 1.19 (p = 7e-05)**	**OR 1.15 (p = 0.03)**
Num. vir. + age	OR 0.99 (p = 0.8)	OR 1.04 (p = 0.4)	OR 1.06 (p = 0.3)
Num. vir. + sex + age	OR 1.02 (p = 0.7)	OR 1.08 (p = 0.1)	OR 1.09 (p = 0.2)

3.3.4 Association of Antibody Reactivities with (Pre)Diabetes Prevalence & Incidence

So far, the focus has been on binary viral serostatus. However, there might be dose effects of the continuous viral antibody reactivities among seropositives as measured by MFI. This is particularly interesting for viruses with very high prevalence, namely EBV, HSV–1, HHV–7 and VZV (see 3.1.2), as they have very few seronegatives and thus their binary status has little power. Figure 25 shows this exemplarily for HHV–7 *u14*, Supplementary Figures 61–76 for the other herpesvirus antigens.

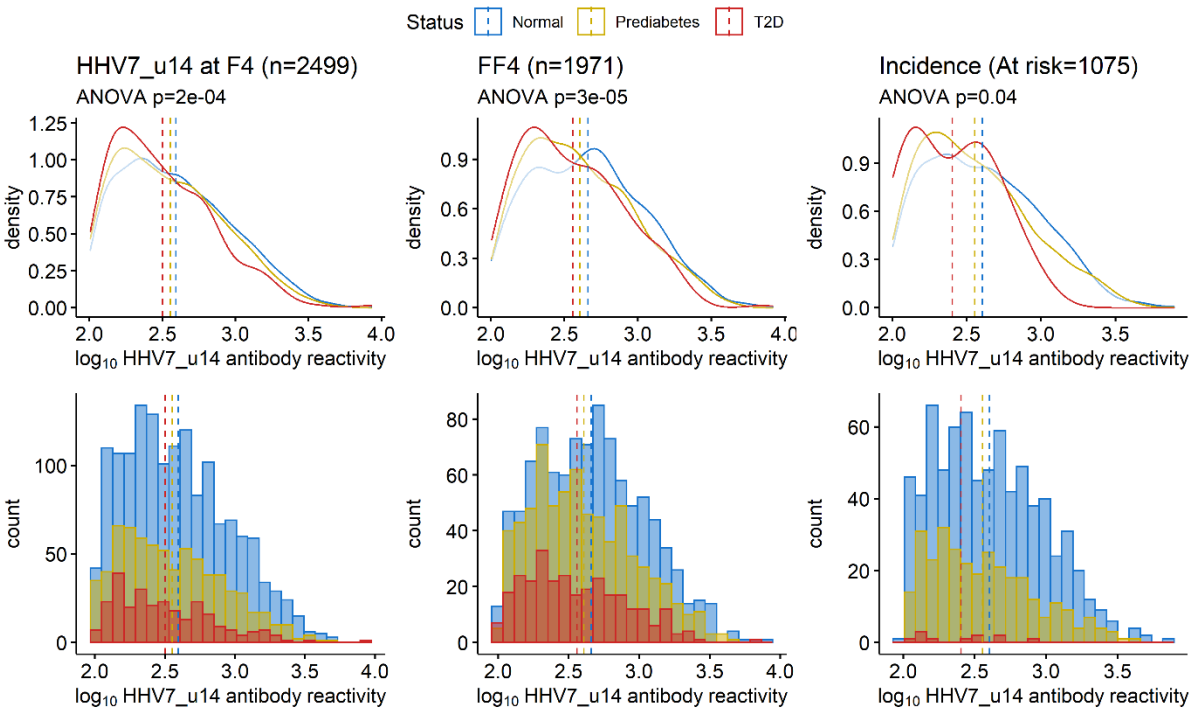


Figure 25: Association of HHV–7 viral antibody reactivity among seropositives with diabetic prevalence / incidence

One can see that the viral antibody reactivity is on average highest among people with normal OGTT, followed by prediabetics and lowest among people with type 2 diabetes. This interesting

association is highly significant for both prevalence at F4 and FF4 as well as (pre)diabetes incidence from normal OGTT at F4 and has the same effect direction as the association of binary HHV–7 status with diabetic prevalence (OR 0.8, see Table 12 in 3.3.1).

However, while it is tempting to interpret this as some biological effect, it is again more likely due to confounding by age: As indicated in Figure 9, both HHV–7 status as well as antibody reactivity among seropositives are negatively associated with age. In fact, in a linear regression model with log–transformed viral antibody reactivity of HHV–7 as dependent variable and diabetic status (combining prediabetes and diabetes) and age and sex as independent variables, age was highly significant, but diabetes status was not anymore. These p–values can be found in Table 15 for all 17 viral antigens. The crude p–values are slightly lower as the ANOVA p–values from the viral antibody reactivity figures (e.g. Figure 25) because they are based on coefficients combining prediabetes and T2D in linear regression models but they remain in the same magnitude.

Table 15: P–values of (pre)diabetes–coefficient in linear regressions modelling log–transformed viral antibody reactivity among seropositives (Bonferroni significant cells bold)

	(Pre)Diabetes Prevalence F4		**(Pre)Diabetes Prevalence FF4**		**(Pre)Diabetes Incidence**	
	Crude p	**Adjusted p**	**Crude p**	**Adjusted p**	**Crude p**	**Adjusted p**
HSV–1 gg	******0.0001**	0.2928	** 0.0044	0.6295	* 0.0128	0.0751
HSV–2 mgg	0.0525	0.1291	0.0714	0.3990	0.6460	0.3242
VZV giorf67	0.2142	0.2906	0.1324	0.2883	0.1295	0.1785
VZV georf68	** 0.0038	0.0993	* 0.0425	0.9812	0.5639	0.5228
EBV ead	0.5358	0.3833	0.8210	0.7212	0.8549	0.7329
EBV ebna1	0.0960	0.1580	0.9854	0.9714	0.4044	0.3791
EBV vcap18	* 0.0356	0.3863	0.2714	0.6791	0.9652	0.6224
EBV zebra	0.3893	0.4384	0.4743	0.5520	0.5812	0.4709
CMV pp28	0.2069	0.6417	0.7500	0.8903	0.5876	0.6509
CMV pp52	0.0604	0.1170	0.5462	0.5834	0.6579	0.9214
CMV pp65	0.6711	0.9155	0.1546	0.5581	0.8147	0.9999
CMV pp150	0.3923	0.0989	0.3054	0.2786	0.5722	0.6393
HHV–6 ie1atr	0.5790	0.4788	0.5885	0.3235	* 0.0462	* 0.0406
HHV–6 p100tr	0.9446	0.6880	0.4169	0.7231	0.5084	0.5424
HHV–6 ie1btr	** 0.0013	* 0.0347	** 0.0072	0.1113	0.0944	0.1596
HHV–6 p101ktr	0.7698	0.5796	0.3781	0.4714	0.4624	0.3589
HHV–7 u14	***** 0.0003**	0.7678	******0.0000**	0.4557	* 0.0284	0.1262

Only few antigens have significant crude associations of antibody reactivity with (pre)diabetes and among those, only 3 remain significant after Bonferroni correction: HSV–1 at F4 and HHV–7 at both F4 and FF4. However, none of them remain significant after correcting for the confounders age and sex. To summarise, it seems like all associations of herpesvirus antibody

reactivity among seropositives with (pre)diabetes are confounded and mostly explained by age. No convincing dose–effects seem to exist.

3.3.5 Association of Herpesviruses with HbA1c

Finally, the association of herpesviruses with a laboratory measure highly related to (pre)diabetes will be examined: HbA1c, the proportion of glycated haemoglobin in the blood, indicating long–term hyperglycaemia. Figure 26 shows the association of HSV–2 status and viral antibody reactivity among seropositives with HbA1c at F4 and FF4. The HbA1c is on average 0.13 percentage points higher for seropositives than for seronegatives (p = 2e-04) at F4 and 0.11 percentage points at FF4 (p = 0.01). While the viral status is significantly associated, the viral antibody reactivity among seropositives does not reach significance, even though there is a slight positive trend with R = 0.094 at F4 and R = 0.066 at FF4. Plots for the other viruses can be found in Supplementary Figures 77–85.

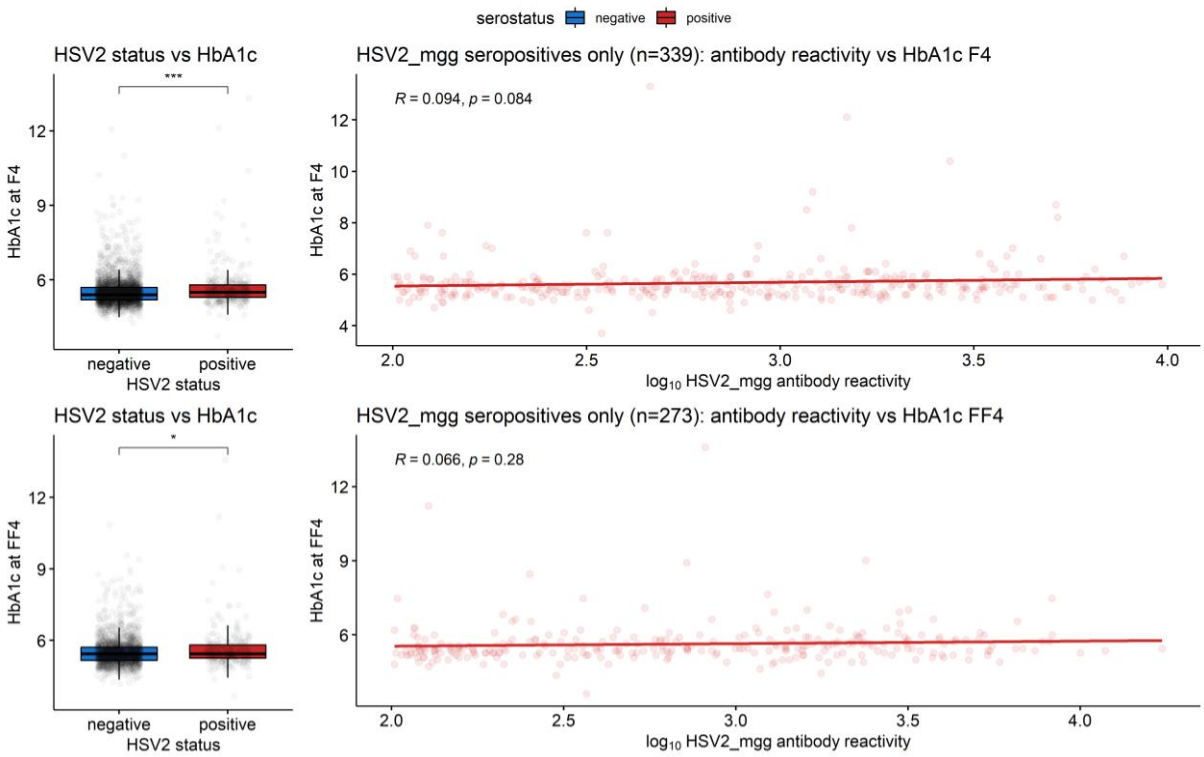


Figure 26: Association of HSV–2 status and antibody reactivity among seropositives with HbA1c at F4 and FF4

The question arises, how strongly is this association explained by the main confounders sex and age? To find an answer, three linear regression models with HbA1c as dependent variable and 1st) viral status only, 2nd) viral status + sex + age and 3rd) viral status + sex + age + (pre)diabetes status as independent variables have been fitted. This progression of model complexity with the first reflecting the same crude association as the figure above helps to find effects independent of sex, age and potentially even diabetes status itself. The beta coefficients and their respective p–values are depicted in Table 16.

Table 16: Modelling HbA1c without and with confounders in linear regression (beta–coefficients are for viral status, significant results in bold)

	F4			FF4		
	Viral status only	**+ Sex + Age**	**+ Sex + Age + (Pre)Diab.**	**Viral status only**	**+ Sex + Age**	**+ Sex + Age + (Pre)Diab.**
HSV–1	**ß = 0.08 (p=0.03)**	ß = -0.02 (p=0.5)	ß = -0.03 (p=0.2)	**ß = 0.1 (p=0.02)**	ß = 0.00 (p=1)	ß = 0.01 (p=0.8)
HSV–2	**ß = 0.13 (p=2e-04)**	**ß = 0.08 (p=0.01)**	**ß = 0.11 (p=5e-05)**	**ß = 0.11 (p=0.01)**	**ß = 0.08 (p=0.04)**	ß = 0.04 (p=0.2)
VZV	ß = 0.01 (p=0.6)	ß = 0.02 (p=0.5)	ß = 0.01 (p=0.8)	**ß = 0.08 (p=0.04)**	ß = 0.06 (p=0.09)	ß = 0.01 (p=0.7)
EBV	ß = 0.09 (p=0.2)	ß = 0.01 (p=0.9)	ß = 0.06 (p=0.3)	ß = 0.12 (p=0.2)	ß = 0.05 (p=0.6)	ß = 0.12 (p=0.1)
CMV	**ß = 0.09 (p=5e-05)**	ß = 0.02 (p=0.5)	ß = 0.00 (p=0.8)	**ß = 0.06 (p=0.02)**	ß = 0.01 (p=0.8)	ß = 0.01 (p=0.8)
HHV–6	ß = 0.00 (p=0.9)	ß = 0.00 (p=0.9)	ß = 0.00 (p=0.8)	ß = 0.03 (p=0.3)	ß = 0.04 (p=0.2)	ß = 0.02 (p=0.4)
HHV–7	ß =-0.05 (p=0.09)	ß = -0.01 (p=0.6)	ß = 0.00 (p=0.9)	ß = 0.07 (p=0.2)	ß = 0.06 (p=0.2)	ß = 0.00 (p=1)

As can be seen, HSV–1, HSV–2 and CMV have significant crude associations at both timepoints but only HSV–2 remains significant after correcting for age and sex. Interestingly, at the F4 timepoint, the serostatus of HSV–2 remains significant (and so with an even lower p–value) when further adjusting for diabetes status itself, potentially suggesting a role of viral influence on the long term blood sugar independent of diabetes (or vice–versa, as cross–sectional modelling cannot distinguish causal effect–directions). Figure 27 demonstrates the beta coefficients of the three models for HSV–2 (see Supplementary Figure 86 for CMV).

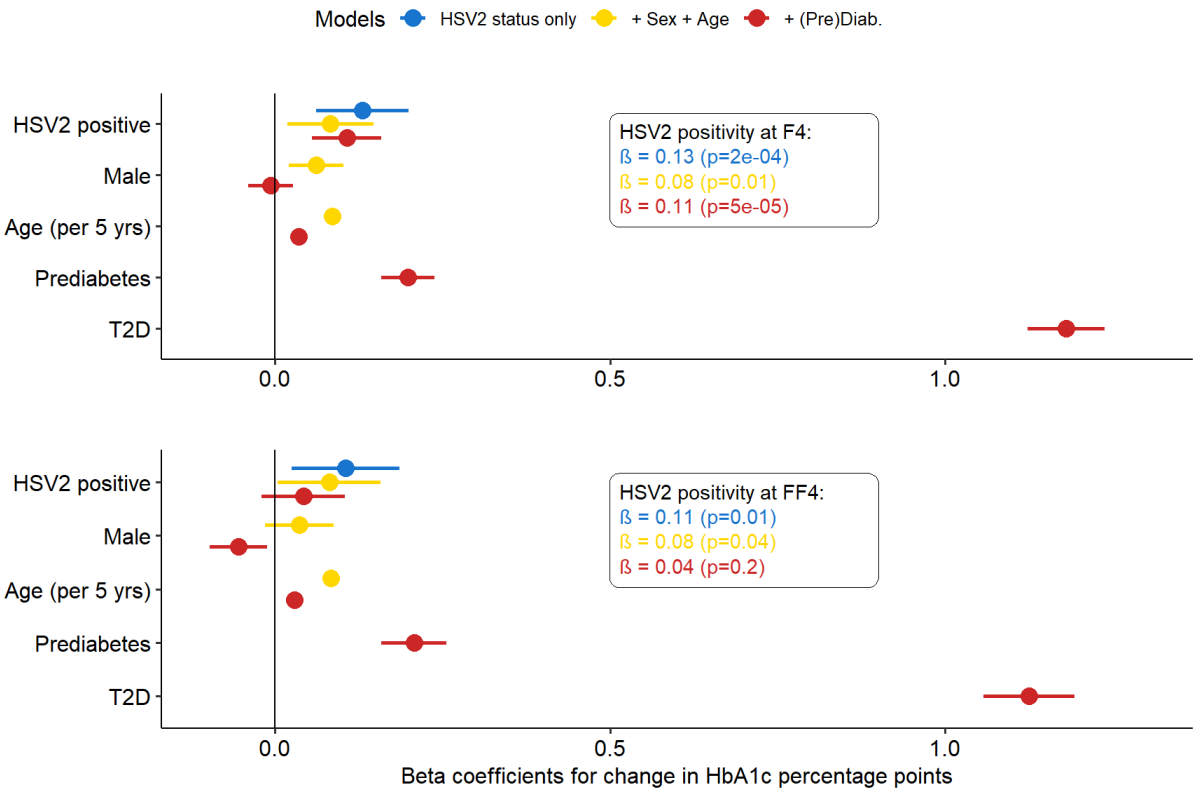


Figure 27: Beta coefficients in 3 increasingly complex linear regression models for HSV–2 and HbA1c at F4 and FF4

4 Discussion

4.1 Age and Sex Stratified Prevalence of Viruses

The epidemiological findings concerning HSV–1 & HSV–2 are in line with the literature: The seroprevalence of HSV–1 in Germany was reported to be 91% in women and 86% in men (>30 years old) and that of HSV–2 13.9% (>18 years old) by Pebody et al. [66], based on data from the "Bundes–Gesundheitssurvey" of 1998. This is similar to this study's 88% overall prevalence for HSV–1 and 11% for HSV–2 at F4 (all KORA F4 & FF4 participants were over 30 years old). The OR for women compared to men adjusted for age was reported to be 1.39 (95% CI 1.16–1.65) for HSV–1 and 1.64 (95% CI 1.35–1.98) for HSV–2 [66], similar to this study's OR for HSV–1 of 1.36 (95% CI 1.1–1.7) and slightly higher than this study's OR for HSV–2 of 1.20 (95% CI 1.0–1.5), see Table 10. In addition, there was a clear and significant increase of seroprevalence for both HSV–1 & HSV–2 with age [66], as was the case in this study. Also in other populations besides Germany, HSV–1 is much more common than HSV–2 and both viruses are more common in the elderly and in women (especially HSV–2) [7,67].

For VZV, the prevalences found in the literature are generally higher than the 78% at F4 found in this study: Wutzler et al. found prevalences of >99% in Germany in the age group >40 years [68], which is the adequate comparison group given that the median age at F4 is already 56 years. Similarly, Nardone et al. found a prevalence of 97.7% in the age group 20–29 years in Germany [69], quite a bit younger than the KORA sample population. It is generally established that most infections occur in early childhood [68–70] and interestingly, median antibody levels seem to stay more or less stable with increasing age [68]. It should be noted that the routine VZV vaccination program, which has been recommended in Germany since 2004 [71], did not affect the much older KORA study population. Unfortunately, vaccination status was not part of the questionnaire data but it can be assumed that only a small fraction would have undergone adult VZV vaccination. Interestingly, while in this study women had significantly lower odds of being VZV seropositive (OR_{female} = 0.7, 95% CI 0.5–0.8, see Table 6), this is not a general trend, and on the population level sex does not seem to be a risk factor [70].

EBV typically has a bimodal incidence–by–age distribution in developed countries, with most infections occurring in early childhood and adolescence due to exchange of saliva with parents and intimate partners [13]. While no specific seroprevalence estimates for Germany were found, studies from the US and UK suggest prevalences >80% for adults >18 years old [72,73]. This study's prevalence of 98% is close to the recent UK Biobank estimate of 95% using a similar multiplex serology [15]. Just like in this study, females generally tend to have higher prevalences [15,72] and antibody titres [13] than males (OR_{female} = 2.3, 95% CI 1.2–4.4, see Table

6). Hjalgrim and colleagues remark that this sex–difference has also been found with other viruses and that it is likely at least partly due to an overall stronger immune response in women than in men [13,74].

For CMV, this study's prevalence estimate of 46% at F4 is a bit smaller than in a German 1998 representative adult sample, which lead to an overall seroprevalence of 57% composed of 62% in women and 51% in men and with a clear age–dependence [75]. The sex difference is also present in this study (OR_{female} = 1.6, 95% CI 1.4–1.8, see Table 6), as is the age–dependence ($OR_{10\ years}$ = 1.34, 95% CI 1.3–1.4, see Table 10). The comparably high proportion of seronegative women in reproductive age is particularly of importance as congenital CMV infection represents a high burden of disability [75]. Interestingly, while the age–dependency seems to be universal internationally, the overall prevalence is much higher in developing than in Western countries [76].

As already mentioned in 2.2, the results for HHV–6 and –7 have to be regarded with caution, as validation of their detection in the multiplex assay used in KORA against reference assays is still ongoing [15]. In this light, the prevalence of HHV–6 (including both A and B subtypes) of only 37% at F4 is surprisingly and suspiciously low compared to a reported near–universal prevalence >80% in the general population [22,77]. This is further complicated by the observed correlation of the HHV–6 *p100tr* and CMV *pp28* antigens described in chapter 3.1.1.

The 85% prevalence of HHV–7 on the other hand is much more on par with a published prevalence of >85% in the US population [78], but the observed negative association with older age is unprecedented. Whether this has to do with immune senescence, viral clearance or is merely an artifact of an unvalidated assay remains to be determined by further research.

4.2 Seasonal Influences on Herpesviruses

Many viruses exhibit seasonal patterns because of optimal temperature and humidity levels. Influenza for example is more common in the cold months and polio used to be more common in the warm months before mass vaccination [79]. VZV is the only herpesvirus for which a marked seasonality has been described: Internationally, the peak of incidence is usually in the cold and dry season [80–82] and in Scotland for example, the difference in the number of incident cases between the warm and the cold season was roughly 4 fold each year from 1990–1998 [70]. In Germany, this tendency was observed as well: More than 4 times as many cases were reported in March 2006 (~4300) compared to September 2005 (~1000). The seasonal influence is very clear in this German dataset from 2005–2009, even with the overall number of cases dropping because of the general vaccination program starting 2004 [71]. The main driver

for this latitude–dependent seasonality seems to be temperature [70,82,83], but other factors like the timing of school–holidays have been hypothesized as well [84].

Of course, this study cannot answer questions about seasonality of incidence, the KORA population is much older than early childhood where herpesvirus incidence usually occurs. However, this study can answer questions about differences in viral antibody reactivity depending on season, as the month of the study examination at which the blood of the participant was drawn is known. Paired analysis of participants measured at both timepoints allows a kind of natural experiment because the month of examination was not in any way guided by viral status of the patients and can thus be considered quasi–random. Difference in differences analyses revealed that for nearly all viruses (except HSV–2 and HHV–6), participants that switched from being examined in the warm season at F4 to the cold season at FF4 had a significantly higher than average increase in antibody reactivity, suggesting indeed a slightly higher viral antibody reactivity and / or immune system activity in the cold months.

However, the opposite effect was not observed: Participants that switched from being examined in the cold season at F4 to the warm season at FF4 did not have a lower than average increase in antibody reactivity, which would have been expected under the hypothesis of a higher viral antibody reactivity and / or immune system activity in the cold season. Overall, this ambiguity renders the results highly exploratory and hard to interpret.

Additionally, some significant ORs of binary viral status by season were found, albeit not consistently between F4 and FF4 at all (see Table 7). Most of them can be considered to be false–positive findings and most of them do not remain significant after multiple testing correction with Bonferroni (see Table 10), except for HHV–6 and HHV–7. The former has completely opposite results between the two timepoints and is thus not very convincing. HHV–7 status on the other hand is significantly associated with being tested in the cold season at F4, surviving Bonferroni correction, and shows a clear trend in the same direction at FF4. This could be a direct effect of the higher viral antibody reactivities in the cold season hypothesised above, potentially meaning that participants that have below–threshold “baseline” viral antibody reactivities have higher chances of being tested positive in the cold season (compare Figure 6 in 3.1.2).

4.3 Association of Herpesviruses with T2D

One major strength of this study is its longitudinal character. As already pointed out in 1.3, cross–sectional associations of viral and diabetes prevalence suffer from a reverse causality

issue. This study found significant associations of HSV–2 and CMV seropositivity and incidence of (pre)diabetes, leaving no doubt about the chronology of events.

Some studies have demonstrated an association of CMV with T2D prevalence in general populations [85–87], but the results have been partly confounded by age and other demographic factors [88]. There have also been some niche studies on CMV and the risk of developing T2D in post–transplantation patients [89,90], who are particularly susceptible to severe CMV infection, see 1.1.4.

Concerning incidence, at the beginning of this study in 2019 I was only aware of one study linking CMV to increased incidence of diabetic atherosclerosis in T2D patients but not with T2D itself [91]. In December 2019, a Korean study was published by Yoo et al. linking CMV to T2D incidence rather than prevalence [92], to my knowledge the first of its kind. It reported an OR adjusted for many demographic confounders of 2.1 (95% CI 1.3–3.2), quite a bit larger than this study's adjusted OR of 1.36 (95% CI 1.05–1.76). This difference might be explained by the fact that Yoo et al. were looking at history of manifest CMV disease as evidenced by insurance claims, rather than CMV serostatus, leading to only 576 adult cases in a database encompassing the entire South Korean population of 50 million. They indicate that manifest CMV disease has a higher impact on the overall immune system and inflammatory state than mere subclinical CMV infection [92], with serostatus capturing both CMV disease and subclinical infection.

In addition to these epidemiological studies, CMV has been found histopathologically in the islets of Langerhans in the pancreas, the location of insulin production, in T2D patients but not in controls [93]. While this further increases the plausibility of a causal contribution to T2D, the exact pathomechanism remains unclear. Both direct pancreatic damage through CMV and indirect perturbances on the glucose metabolism pathways through inflammation have been proposed [92].

Concerning the other herpesviruses examined in this study (for HHV–8 see 1.3), no association as clear as with CMV has been described. Of note, the incidence of herpes zoster disease seems to be increased in T2D patients [94–96]. One recent Polish study by Dworzański et al. found significantly increased prevalences of EBV in T2D patients but not of CMV and HSV–1 [97]. Haeseker et al. found an association of T2D with high IgG titres of HHV–6 and EBV but not of CMV [98]. Piras et al. have examined differences in the viral DNA counts as well as antibody titres for EBV, CMV, HHV–6 and HHV–7 and have found no differences between diabetics and controls [99]. These diverse and sometimes contradicting results show once again that

cross–sectional designs are not optimal for the robust linkage of viruses and T2D and that many demographic confounders as well as nuanced differences in populations can affect the results.

Interestingly, I could not find any studies even examining the relation of HSV–2 and T2D in a general population, let alone show a significant increase in T2D incidence among seropositives or an association with HbA1c independent of many demographic factors, like this study did. My research has led me only to one paper finding no association of HSV–2 with IFG / IGT in a group of HIV patients with antiretroviral therapy, exhibiting limited generalisability [100].

4.4 Limitations

Some important limitations have to be taken into account when interpreting the results. First, serology does not always fully capture past infections. Viral antibody concentrations are influenced by severity and strength of the immune reaction upon primary infection, state of the immune system, virus–host interaction and potential recurrent infections, among others. A certain instability of serostatus between the two timepoints F4 and FF4 has been found in chapter 3.1.2, which cannot only be explained by incident cases. Unfortunately, we do not have any information on acute herpesvirus manifestations in the KORA study (e.g. prevalence and frequency of orolabial and genital herpes or zoster disease, history of varicella or infectious mononucleosis, etc.), making it hard to untangle the potential reasons for the observed seroconversions.

A limiting factor of serology in general but multiplex serology in particular is the need to use somewhat arbitrary thresholds for detection. Brenner et al. state the following [53]:

> *"Depending on the reference panel and corresponding reference assay, different cut–offs were found to optimize statistical characteristics per antigen. Thus, we conclude that cut–offs might not be directly transferable between studies. This might have multiple potential reasons such as differences in the underlying study population, differential blood collection conditions and storage of serum specimens before testing, as well as potential assay drift and reagent performance over time."*

They go on to recommend standard quality control and normalisation procedures between studies in order to overcome these problems [53] and as Dr. Tim Waterboer from the same group has created and implemented the multiplex strategy for the KORA studies, the cut–offs can be considered sound. Nonetheless, population–specific determination of cut–offs with gold–standard monoplex serology in a subset of samples might improve the results especially for borderline participants, who are discussed in 3.1.2.

Another limitation of this study is the lack of validation of the viral multiplex assay for HHV–6 and HHV–7 with classic monoplex serology, partly because no universally agreed upon gold standards exist as discussed by Brenner et al. [53]. The results for HHV–6 and 7 have thus to be regarded with caution, as discussed in 4.1.

For diagnosing diabetes, the KORA study has already used the gold standard OGTT, so there is no additional uncertainty with diabetes prevalence and incidence. However, the mean 6.5 years between F4 and FF4 are quite short and lead to moderate event numbers in terms of (pre)diabetes incidence. I have used the diagnostic criteria defined by the American Diabetes Association for (pre)diabetes, which are slightly stricter than those of the World Health Organisation for IFG [31], thus leading to a higher event rate. A sensitivity analysis could shed light on the generalisability of the results with respect to different diagnostic criteria and thresholds.

Lastly, the medium sample size and the imperfect overlap between F4 and FF4 are limiting the statistical power of this study. About ⅔ of the 2,950 participants with both OGTT and viral multiplex serology at F4 have participated at FF4 as well.

4.5 Outlook

The work presented here calls for additional analyses within the KORA cohort. For example, it could be investigated whether further known risk factors for diabetes like BMI, smoking and socioeconomic status are also associated with viral status and thus might confound their association beyond age and sex. Moreover, the impact of herpesviruses on different anti–diabetic medications like metformin or insulin and their therapeutic effectiveness could be examined.

If more resources were available, some of the above–mentioned limitations could also be mitigated by performing new analyses on stored frozen blood specimens of study participants, in particular monoplex serology. Additionally, more viruses for which associations with diabetes have been described could be examined by means of monoplex serology, e.g. HHV–8 (see 1.3). Now that SARS–CoV–2 has infected parts of the population – albeit with locally widely ranging prevalences – it might be worthwhile to also include it in the viral panel at some point, especially as a bidirectional relationship with T2D has already been described [1].

Polymerase chain reaction (PCR) could have been used in addition to serology to determine the true viral load by number of copies of viral genomes. Herpesvirus PCR is routinely used in clinical contexts (e.g. cerebrospinal fluid) and has been performed in many different tissues in experimental setups (e.g. nerve biopsies, plasma, urine, tears, etc.) [5,22,101,102]. Saliva and whole blood seem to be the two most promising and easily accessible tissues which could be used for PCR in a context like KORA. However, while beta- and gammaherpesviruses are latent

in leukocytes, alphaherpesviruses are only detectable in blood during viremia, which mainly occurs in primary infection. Thus, an advantage of serology is that insights are gained independently of the primary site of infection.

Moving beyond KORA, other population cohorts can be valuable resources to confirm or reject the longitudinal findings from this thesis in other populations. The UK Biobank for example is planning viral multiplex serology for all of its ~500,000 participants; the data for a pilot of 9,695 participants is already available [15]. Additionally, the German National Cohort (NAKO) will eventually also have viral serology and OGTT, at least in subgroups of the ~200,000 participants [103].

Finally, a systematic literature review seems warranted for this topic. While I have tried to cover as much of the existing literature as possible in the discussion, I have realised there is already more than expected and it would be better to summarise it in a more formal way. Also, at least for CMV and HHV–8 it might be already worth doing a meta–analysis of the existing evidence on their associations with T2D (see chapter 4.3 and 1.3 respectively).

5 Conclusions

This study has demonstrated a significant association of HSV–2 and CMV seropositivity with incidence of (pre)diabetes after adjustment for age and sex. Similar results have recently been reported for CMV from a Korean group, albeit for history of clinical CMV disease rather than serostatus. For HSV–2, this is a novel finding, further strengthened by the significant association of HSV–2 serostatus with HbA1c, independent of sex, age and diabetes status. More research is needed, but once the evidence is clear it should be presented to policymakers to increase efforts in prevention strategies and possibly vaccine–development.

In line with the literature, this study has conclusively shown that all herpesviruses but VZV are more prevalent in women than in men and at least HSV–1, HSV–2, EBV and CMV are more prevalent in the older than in the younger. Consequently, the number of co–occurring herpesviruses is significantly larger with older age.

References

1. Rubino F, Amiel SA, Zimmet P, Alberti G, Bornstein S, Eckel RH, et al. New-Onset Diabetes in Covid-19. N Engl J Med. 2020 Jun 12.
2. Epstein MA, Achong BG, Barr YM. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. The Lancet. 1964 Mar 28;283(7335):702–3.
3. King AMQ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Gorbalenya AE, et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). Arch Virol. 2018 Sep 1;163(9):2601–31.
4. Davison AJ. Overview of classification. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
5. Whitley R, Kimberlin DW, Prober CG. Pathogenesis and disease. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
6. Vyse AJ, Gay NJ, Slomka MJ, Gopal R, Gibbs T, Morgan-Capner P, et al. The burden of infection with HSV-1 and HSV-2 in England and Wales: implications for the changing epidemiology of genital herpes. Sex Transm Infect. 2000 Jun 1;76(3):183–7.
7. Xu F, Sternberg MR, Kottiri BJ, McQuillan GM, Lee FK, Nahmias AJ, et al. Trends in Herpes Simplex Virus Type 1 and Type 2 Seroprevalence in the United States. JAMA. 2006 Aug 23;296(8):964–73.
8. Khadr L, Harfouche M, Omori R, Schwarzer G, Chemaitelly H, Abu-Raddad LJ. The Epidemiology of Herpes Simplex Virus Type 1 in Asia: Systematic Review, Meta-analyses, and Meta-regressions. Clin Infect Dis. 2019 Mar 1;68:757–72.
9. Moffat J, Ku C-C, Zerboni L, Sommer M, Arvin A. VZV: pathogenesis and the disease consequences of primary infection. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
10. Cohen JI. VZV: molecular basis of persistence (latency and reactivation). In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
11. Ambinder RF, Cesarman E. Clinical and pathological aspects of EBV and KSHV infection. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
12. Moss DJ, Burrows SR, Khanna R. EBV: immunobiology and host response. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
13. Hjalgrim H, Friborg J, Melbye M. The epidemiology of EBV and its association with malignant disease. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
14. Gilden DH. Infectious causes of multiple sclerosis. Lancet Neurol. 2005 Mar 1;4(3):195–202.
15. Mentzer AJ, Brenner N, Allen N, Littlejohns TJ, Chong AY, Cortes A, et al. Identification of host-pathogen-disease relationships using a scalable Multiplex Serology platform in UK Biobank. medRxiv. 2019 Aug 29;19004960.

16. Maple PAC. Cytomegalovirus and Epstein–Barr Virus Associations with Neurological Diseases and the Need for Vaccine Development. Vaccines. 2020 Mar;8(1):35.
17. Britt W. Virus entry into host, establishment of infection, spread in host, mechanisms of tissue damage. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
18. Ishii T, Sasaki Y, Maeda T, Komatsu F, Suzuki T, Urita Y. Clinical differentiation of infectious mononucleosis that is caused by Epstein-Barr virus or cytomegalovirus: A single-center case-control study in Japan. J Infect Chemother. 2019 Jun 1;25(6):431–6.
19. Salahuddin SZ, Ablashi DV, Markham PD, Josephs SF, Sturzenegger S, Kaplan M, et al. Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. Science. 1986 Oct 31;234(4776):596–601.
20. Frenkel N, Schirmer EC, Wyatt LS, Katsafanas G, Roffman E, Danovich RM, et al. Isolation of a new herpesvirus from human CD4+ T cells. Proc Natl Acad Sci U S A. 1990 Jan;87(2):748–52.
21. Wang F-Z, Pellett PE. HHV-6A, 6B, and 7: immunobiology and host response. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
22. Emery VC, Clark DA. HHV-6A, 6B, and 7: persistence in the population, epidemiology and transmission. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
23. Mori Y, Yamanishi K. HHV-6A, 6B, and 7: pathogenesis, host response, and clinical disease. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
24. Aimola G, Beythien G, Aswad A, Kaufer BB. Current understanding of human herpesvirus 6 (HHV-6) chromosomal integration. Antiviral Res. 2020 Apr;176:104720.
25. Bartolini L, Theodore WH, Jacobson S, Gaillard WD. Infection with HHV-6 and its role in epilepsy. Epilepsy Res. 2019 Jul 1;153:34–9.
26. Leibovitch EC, Jacobson S. Viruses in chronic progressive neurologic disease. Mult Scler Houndmills Basingstoke Engl. 2018;24(1):48–52.
27. Tamayo T, Brinks R, Hoyer A, Kuß O, Rathmann W. The Prevalence and Incidence of Diabetes in Germany. Dtsch Ärztebl Int. 2016 Mar;113(11):177–82.
28. Jacobs E, Hoyer A, Brinks R, Kuss O, Rathmann W. Burden of Mortality Attributable to Diagnosed Diabetes: A Nationwide Analysis Based on Claims Data From 65 Million People in Germany. Diabetes Care. 2017 Dec 1;40(12):1703–9.
29. American Diabetes Association. 2. Classification and Diagnosis of Diabetes. Diabetes Care. 2015 Jan 1;38(Supplement 1):S8–16.
30. Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, Zhao Y. From Pre-Diabetes to Diabetes: Diagnosis, Treatments and Translational Research. Medicina (Mex). 2019 Sep;55(9):546.
31. Kowall B, Rathmann W, Bongaerts B, Kuss O, Stang A, Roden M, et al. Incidence Rates of Type 2 Diabetes in People With Impaired Fasting Glucose (ADA vs. WHO Criteria) and Impaired Glucose Tolerance: Results From an Older Population (KORA S4/F4/FF4 Study). Diabetes Care. 2019 Feb 1;42(2):e18–20.
32. Ramachandran A, Snehalatha C, Shetty AS, Nanditha A. Trends in prevalence of diabetes in Asian countries. World J Diabetes. 2012 Jun 15;3(6):110–7.
33. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016 04;536(7614):41–7.

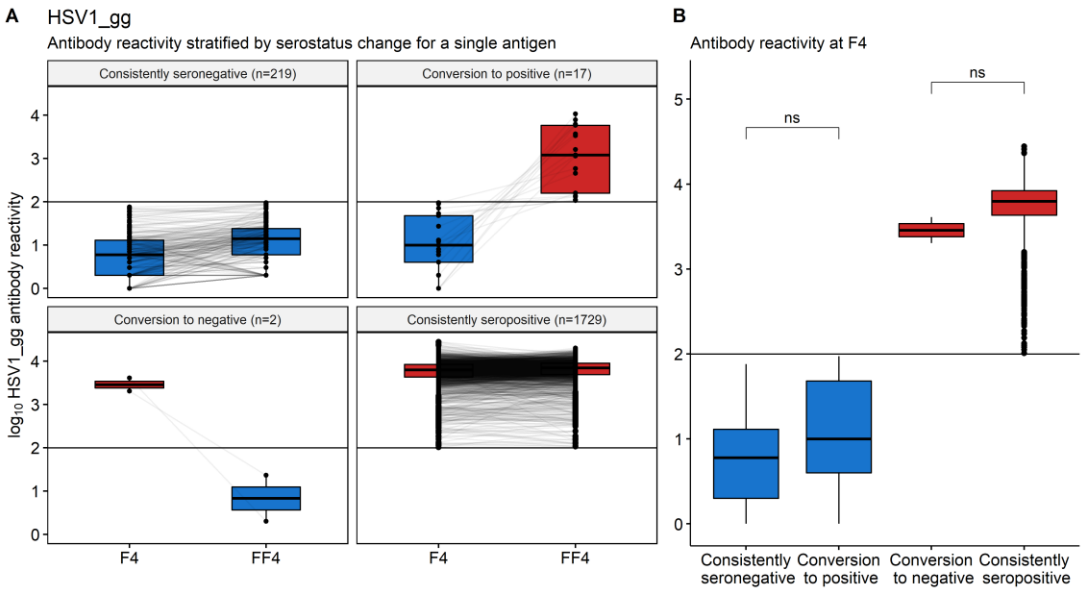
34. Chester B, Babu JR, Greene MW, Geetha T. The effects of popular diets on type 2 diabetes management. Diabetes Metab Res Rev. 2019 May 23;0(0):e3188.
35. Glechner A, Keuchel L, Affengruber L, Titscher V, Sommer I, Matyas N, et al. Effects of lifestyle changes on adults with prediabetes: A systematic review and meta-analysis. Prim Care Diabetes. 2018 Oct 1;12(5):393–408.
36. Garber AJ. Obesity and type 2 diabetes: which patients are at risk? Diabetes Obes Metab. 2012;14(5):399–408.
37. Konduracka E. A link between environmental pollution and civilization disorders: a mini review. Rev Environ Health. 2019;34(3):227–33.
38. Tsalamandris S, Antonopoulos AS, Oikonomou E, Papamikroulis G-A, Vogiatzi G, Papaioannou S, et al. The Role of Inflammation in Diabetes: Current Concepts and Future Perspectives. Eur Cardiol. 2018 Dec 21.
39. Barbarroja N, López-Pedrera R, Mayas MD, García-Fuentes E, Garrido-Sánchez L, Macías-González M, et al. The obese healthy paradox: is inflammation the answer? Biochem J. 2010 Aug 15;430(1):141–9.
40. Korbel L, Easterling RS, Punja N, Spencer JD. The burden of common infections in children and adolescents with diabetes mellitus: A Pediatric Health Information System study. Pediatr Diabetes. 2018;19(3):512–9.
41. Zoppini G, Fedeli U, Schievano E, Dauriz M, Targher G, Bonora E, et al. Mortality from infectious diseases in diabetes. Nutr Metab Cardiovasc Dis. 2018 May 1;28(5):444–50.
42. Zhu H, Wang Y, Yu L, Xu Y, Zhou H, Ding Y, et al. Serological and molecular analysis on the relationships between type 2 diabetes mellitus and hepatitis B virus infection. J Infect Dev Ctries. 2016 Aug 31;10(08):837–44.
43. Fallahi P, Ferrari SM, Colaci M, Ruffilli I, Vita R, Azzi A, et al. Hepatitis C virus infection and type 2 diabetes. Clin Ter. 2013;164(5):e393-404.
44. Ba-Essa EM, Mobarak EI, Al-Daghri NM. Hepatitis C virus infection among patients with diabetes mellitus in Dammam, Saudi Arabia. BMC Health Serv Res. 2016 Jul 27;16(1):313.
45. Hober D, Alidjinou EK. Enteroviral pathogenesis of type 1 diabetes: queries and answers. Curr Opin Infect Dis. 2013 Jun;26(3):263.
46. Fallahi P, Di AD, Mazzi V, Santini F, Fabiani S, Sebastiani M, et al. Hepatitis C virus and type 1 diabetes. Clin Ter. 2013;164(5):e437-44.
47. Yoo SG, Han KD, Lee KH, La Y, Kwon DE, Han SH. Impact of Cytomegalovirus Disease on New-Onset Type 2 Diabetes Mellitus: Population-Based Matched Case-Control Cohort Study. Diabetes Metab J. 2019 Dec;43(6):815–29.
48. Pompei R. The Role of Human Herpesvirus 8 in Diabetes Mellitus Type 2: State of the Art and a Medical Hypothesis. Adv Exp Med Biol. 2016;901:37–45.
49. Ingianni A, Carta F, Reina A, Manai M, Desogus A, Pompei R. Prevalence of Herpesvirus 8 Infection in Type 2 Diabetes Mellitus Patients. Am J Infect Dis. 2007 Sep 30;3(3):123–7.
50. Sobngwi E, Choukem SP, Agbalika F, Blondeau B, Fetita L-S, Lebbe C, et al. Ketosis-Prone Type 2 Diabetes Mellitus and Human Herpesvirus 8 Infection in Sub-Saharan Africans. JAMA. 2008 Jun 18;299(23):2770–6.
51. Cui M, Fang Q, Zheng J, Shu Z, Chen Y, Fan Y, et al. Kaposi's sarcoma-associated herpesvirus seropositivity is associated with type 2 diabetes mellitus: A case–control study in Xinjiang, China. Int J Infect Dis. 2019 Mar 1;80:73–9.
52. Holle R, Happich M, Löwel H, Wichmann HE, MONICA/KORA Study Group. KORA - A Research Platform for Population Based Health Research. Gesundheitswesen. 2005 Aug;67(S 1):19–25.

53. Brenner N, Mentzer AJ, Butt J, Michel A, Prager K, Brozy J, et al. Validation of Multiplex Serology detecting human herpesviruses 1-5. PLoS ONE. 2018 Dec 27;13(12):e0209379.
54. Bland M. An Introduction to Medical Statistics. 4. Auflage. Oxford: Oup Oxford; 2015. 448 p.
55. Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). Am J Epidemiol. 1990 Feb;131(2):373–5.
56. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. 1st ed. 2013, Corr. 7th printing 2017. New York: Springer; 2013. 440 p.
57. Hanley JA, Negassa A, Edwardes MD deB, Forrester JE. Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. Am J Epidemiol. 2003 Feb 15;157(4):364–75.
58. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2019. Available from: https://www.R-project.org
59. Wei T, Simko V. R package 'corrplot': Visualization of a Correlation Matrix. 2017. Available from: https://github.com/taiyun/corrplot
60. Brunson JC. ggalluvial: Alluvial Plots in 'ggplot2'. 2019. Available from: https://CRAN.R-project.org/package=ggalluvial
61. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. 2019. Available from: https://CRAN.R-project.org/package=ggpubr
62. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017 Sep 15;33(18):2938–40.
63. Vincent J Carey. Ported to R by Thomas Lumley and Brian Ripley. Files src/dgedi.f and src/dgefa.f are for LINPACK authored by Cleve Moler. gee: Generalized Estimation Equation Solver. 2019. Available from: https://CRAN.R-project.org/package=gee
64. Lumley T. rmeta: Meta-Analysis. 2018. Available from: https://CRAN.R-project.org/package=rmeta
65. Meisinger C, Strassburger K, Heier M, Thorand B, Baumeister SE, Giani G, et al. Prevalence of undiagnosed diabetes and impaired glucose regulation in 35–59-year-old individuals in Southern Germany: the KORA F4 Study. Diabet Med. 2010;27(3):360–2.
66. Pebody RG, Andrews N, Brown D, Gopal R, De HM, François G, et al. The seroepidemiology of herpes simplex virus type 1 and 2 in Europe. Sex Transm Infect. 2004 Jun;80(3):185–91.
67. Smith JS, Robinson NJ. Age-Specific Prevalence of Infection with Herpes Simplex Virus Types 2 and 1: A Global Review. J Infect Dis. 2002 Oct 15;186(Supplement_1):S3–28.
68. Wutzler P, Färber I, Wagenpfeil S, Bisanz H, Tischer A. Seroprevalence of varicella-zoster virus in the German population. Vaccine. 2001 Oct 12;20(1):121–4.
69. Nardone A, de Ory F, Carton M, Cohen D, van Damme P, Davidkin I, et al. The comparative sero-epidemiology of varicella zoster virus in 11 countries in the European region. Vaccine. 2007 Nov 7;25(45):7866–72.
70. Seward J, Jumaan A. VZV: persistence in the population. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
71. Siedler A, Arndt U. Impact of the routine varicella vaccination programme on varicella epidemiology in Germany. Eurosurveillance. 2010 Apr 1.
72. Dowd JB, Palermo T, Brite J, McDade TW, Aiello A. Seroprevalence of Epstein-Barr virus infection in U.S. children ages 6-19, 2003-2010. PLoS ONE. 2013.

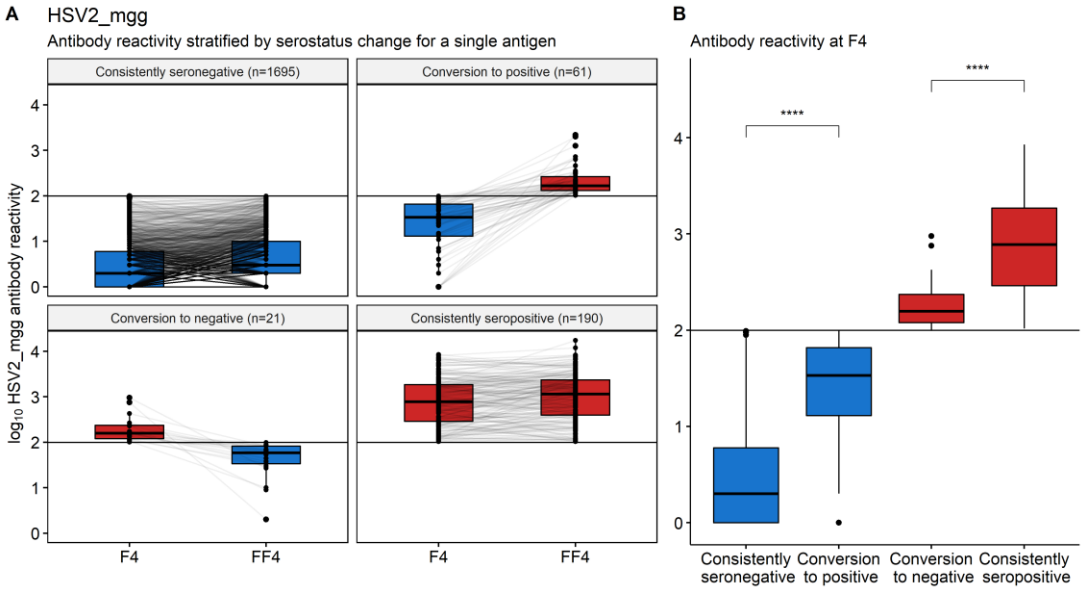
73. Morris MC, Edmunds WJ, Hesketh LM, Vyse AJ, Miller E, Morgan-Capner P, et al. Sero-epidemiological patterns of epstein-barr and herpes simplex (HSV-1 and HSV-2) viruses in England and Wales. J Med Virol. 2002;67(4):522–7.
74. Beagley KW, Gockel CM. Regulation of innate and adaptive immunity by the female sex hormones oestradiol and progesterone. FEMS Immunol Med Microbiol. 2003;38(1):13–22.
75. Lachmann R, Loenenbach A, Waterboer T, Brenner N, Pawlita M, Michel A, et al. Cytomegalovirus (CMV) seroprevalence in the adult population of Germany. PLoS ONE. 2018 Jul 25.
76. Boppana SB, Fowler KB. Persistence in the population: epidemiology and transmisson. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
77. Ward KN. The natural history and laboratory diagnosis of human herpesviruses-6 and -7 infections in the immunocompetent. J Clin Virol. 2005 Mar 1;32(3):183–93.
78. Ablashi DV, Berneman ZN, Kramarsky B, Whitman J, Asano Y, Pearson GR. Human herpesvirus-7 (HHV-7): current status. Clin Diagn Virol. 1995 Jul 1;4(1):1–13.
79. Hemmes JH, Winkler KC, Kool SM. Virus survival as a seasonal factor in influenza and poliomyelitis. Antonie Van Leeuwenhoek. 1962 Dec 1;28(1):221–33.
80. Lee. Review of varicella zoster seroepidemiology in India and South-east Asia. Trop Med Int Health. 1998;3(11):886–90.
81. Seward JF, Watson BM, Peterson CL, Mascola L, Pelosi JW, Zhang JX, et al. Varicella disease after introduction of varicella vaccine in the United States, 1995-2000. JAMA. 2002 Feb 6;287(5):606–11.
82. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. Proc Natl Acad Sci. 2016 Jun 14;113(24):6689–94.
83. Kokaze A, Yoshida M, Sekine Y, Ishikawa M, Kurokochi T, Uchida Y, et al. The magnitude of variation in temperature within a year has an effect on the seasonal variations of chickenpox incidence in Japan. Epidemiol Infect. 2001 Apr;126(2):269–77.
84. Deguen S, Chau NP, Flahault A. Epidemiology of Chickenpox in France (1991-1995). J Epidemiol Community Health 1979-. 1998;52:46S-49S.
85. Blum A, Peleg A, Weinberg M. Anti-cytomegalovirus (CMV) IgG antibody titer in patients with risk factors to atherosclerosis. Clin Exp Med. 2003 Nov;3(3):157–60.
86. Bertram W. Roberts, Irina Cech. Association of type 2 diabetes mellitus and seroprevalence for cytomegalovirus. South Med J. 2005 Jul;98(7):686–92.
87. Chen S, JM de Craen A, Raz Y, Derhovanessian E, Vossen CTM A, Westendorp GJ R, et al. Cytomegalovirus seropositivity is associated with glucose regulation in the oldest old. Results from the Leiden 85-plus Study. Immun Ageing. 2012 Aug 28;9(1):18.
88. Schmidt L, Nelson HH, Thyagarajan B, Hunter-Schlichting D, Pankow JS, Capistrant B, et al. Association between cytomegalovirus seropositivity and Type 2 diabetes is explained by age and other demographic characteristics: the National Health and Nutrition Examination Survey. Diabet Med J Br Diabet Assoc. 2018;35(12):1722–6.
89. Hjelmesæth J, Sagedal S, Hartmann A, Rollag H, Egeland T, Hagen M, et al. Asymptomatic cytomegalovirus infection is associated with increased risk of new-onset diabetes mellitus and impaired insulin release after renal transplantation. Diabetologia. 2004 Sep 1;47(9):1550–6.
90. Hjelmesæth J, Müller F, Jenssen T, Rollag H, Sagedal S, Hartmann A. Is there a link between cytomegalovirus infection and new-onset posttransplantation diabetes mellitus? Potential mechanisms of virus induced β-cell damage. Nephrol Dial Transplant. 2005 Nov 1;20(11):2311–5.

91. Zhang J, Liu Y, Sun H, Li S, Xiong H, Yang Z, et al. High Human Cytomegalovirus IgG Level is Associated with Increased Incidence of Diabetic Atherosclerosis in Type 2 Diabetes Mellitus Patients. Med Sci Monit Int Med J Exp Clin Res. 2015 Dec 30;21:4102–10.
92. Yoo SG, Han KD, Lee KH, La Y, Kwon DE, Han SH. Impact of Cytomegalovirus Disease on New-Onset Type 2 Diabetes Mellitus: Population-Based Matched Case-Control Cohort Study. Diabetes Metab J. 2019;43(6):815–29.
93. Lohr JM, Oldstone MBA. Detection of cytomegalovirus nucleic acid sequences in pancreas in type 2 diabetes. The Lancet. 1990 Sep 15;336(8716):644–8.
94. Muñoz-Quiles C, López-Lacort M, Ampudia-Blasco FJ, Díez-Domingo J. Risk and impact of herpes zoster on patients with diabetes: A population-based study, 2009-2014. Hum Vaccines Immunother. 2017 02;13(11):2606–11.
95. Lai S-W, Lin C-L, Liao K-F. Real-world database investigating the association between diabetes mellitus and herpes zoster in Taiwan. Medicine (Baltimore). 2019 May;98(18):e15463.
96. Saadatian-Elahi M, Bauduceau B, Del-Signore C, Vanhems P. Diabetes as a risk factor for herpes zoster in adults: A synthetic literature review. Diabetes Res Clin Pract. 2020 Jan;159:107983.
97. Dworzański J, Drop B, Kliszczewska E, Strycharz-Dudziak M, Polz-Dacewicz M. Prevalence of Epstein-Barr virus, human papillomavirus, cytomegalovirus and herpes simplex virus type 1 in patients with diabetes mellitus type 2 in south-eastern Poland. PLoS One. 2019;14(9):e0222607.
98. Haeseker MB, Pijpers E, Dukers-Muijrers NH, Nelemans P, Hoebe CJ, Bruggeman CA, et al. Association of cytomegalovirus and other pathogens with frailty and diabetes mellitus, but not with cardiovascular disease and mortality in psycho-geriatric patients; a prospective cohort study. Immun Ageing A. 2013 Jul 23;10:30.
99. Piras E, Madeddu MA, Palmieri G, Angius F, Contini P, Pompei R, et al. High Prevalence of Human Herpesvirus 8 Infection in Diabetes Type 2 Patients and Detection of a New Virus Subtype. In: Donelli G, editor. Advances in Microbiology, Infectious Diseases and Public Health: Volume 7. Cham: Springer International Publishing; 2017.
100. Tan DHS, Raboud JM, Szadkowski L, Yi TJ, Shannon B, Kaul R, et al. Herpes Simplex Virus Type 2 Serostatus Is Not Associated with Inflammatory or Metabolic Markers in Antiretroviral Therapy-Treated HIV. AIDS Res Hum Retroviruses. 2015 Mar 1;31(3):276–81.
101. Griffiths PD, Boeckh M. Antiviral therapy for human cytomegalovirus. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.
102. Kaufman HE, Azcuy AM, Varnell ED, Sloop GD, Thompson HW, Hill JM. HSV-1 DNA in Tears and Saliva of Normal Adults. Invest Ophthalmol Vis Sci. 2005 Jan;46(1):241–7.
103. Ahrens W, Greiser KH, Linseisen J, Pischon T, Pigeot I. Erforschung von Erkrankungen in der NAKO Gesundheitsstudie. Die wichtigsten gesundheitlichen Endpunkte und ihre Erfassung. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2020 Apr 1;63(4):376–84.

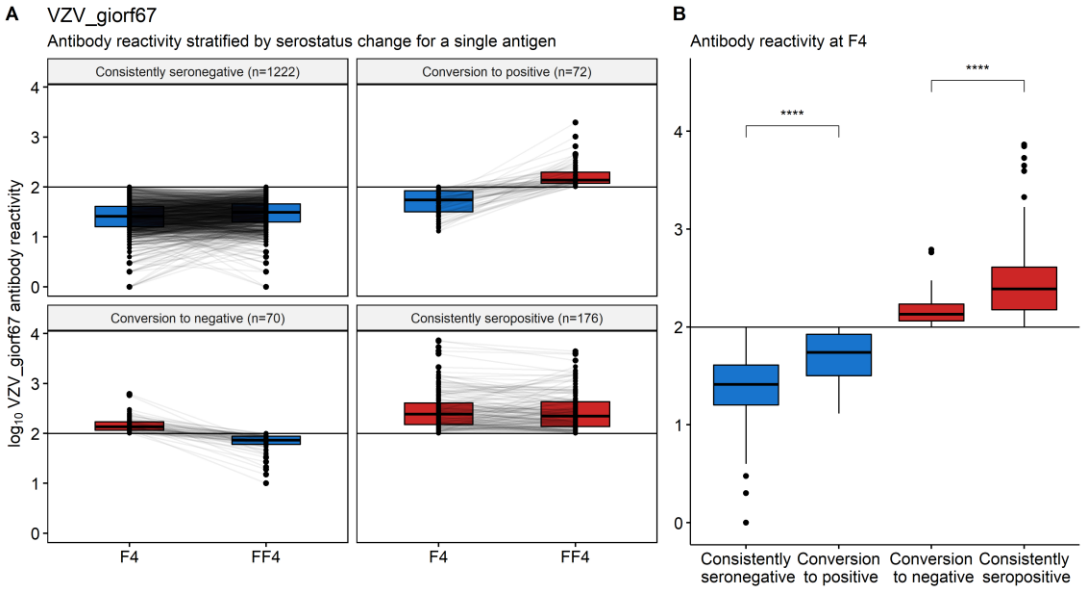
Supplementary Figures



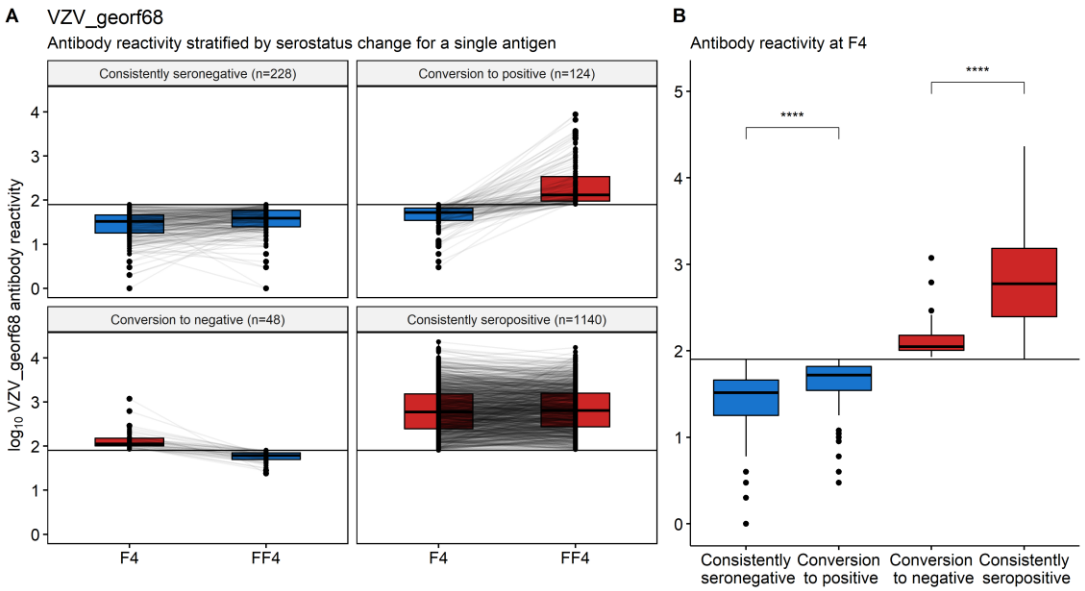
Supplementary Figure 1: HSV–1 gg reactivity by serostatus change



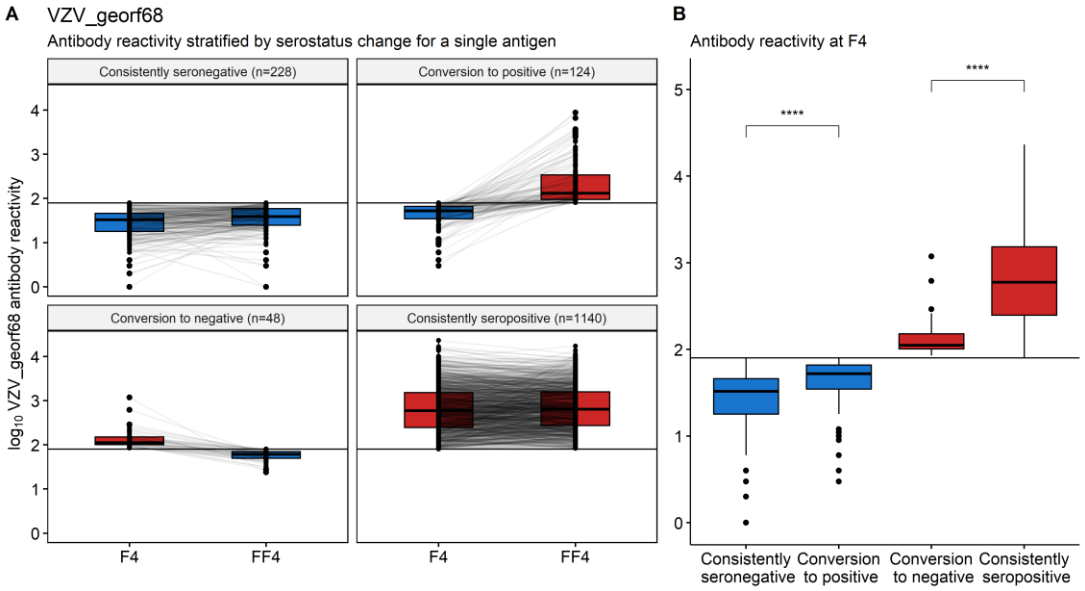
Supplementary Figure 2: HSV–2 mgg reactivity by serostatus change



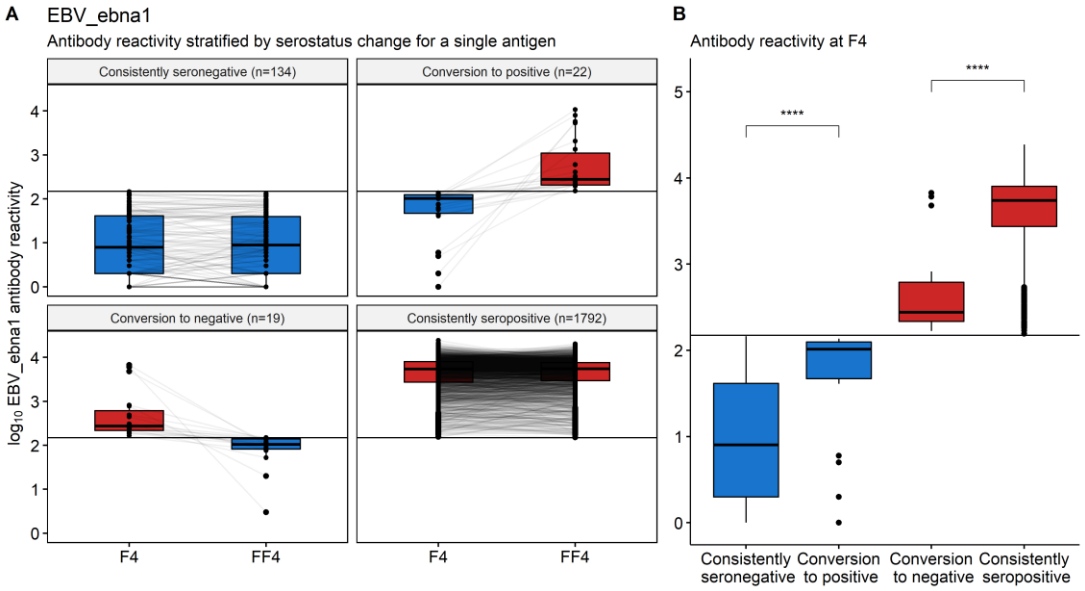
Supplementary Figure 3: VZV giorf67 reactivity by serostatus change



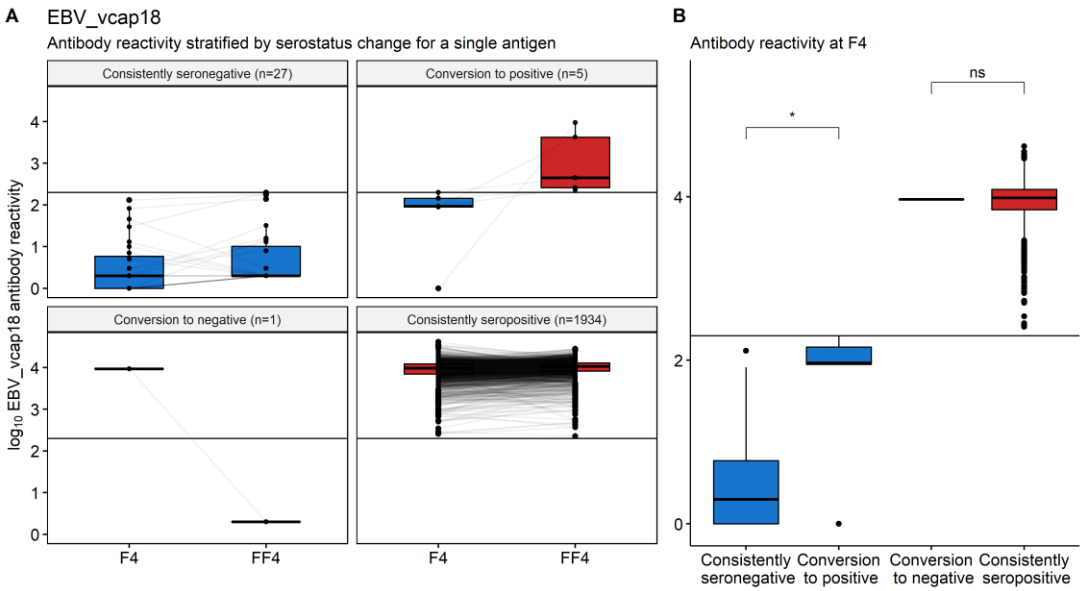
Supplementary Figure 4: VZV georf68 reactivity by serostatus change



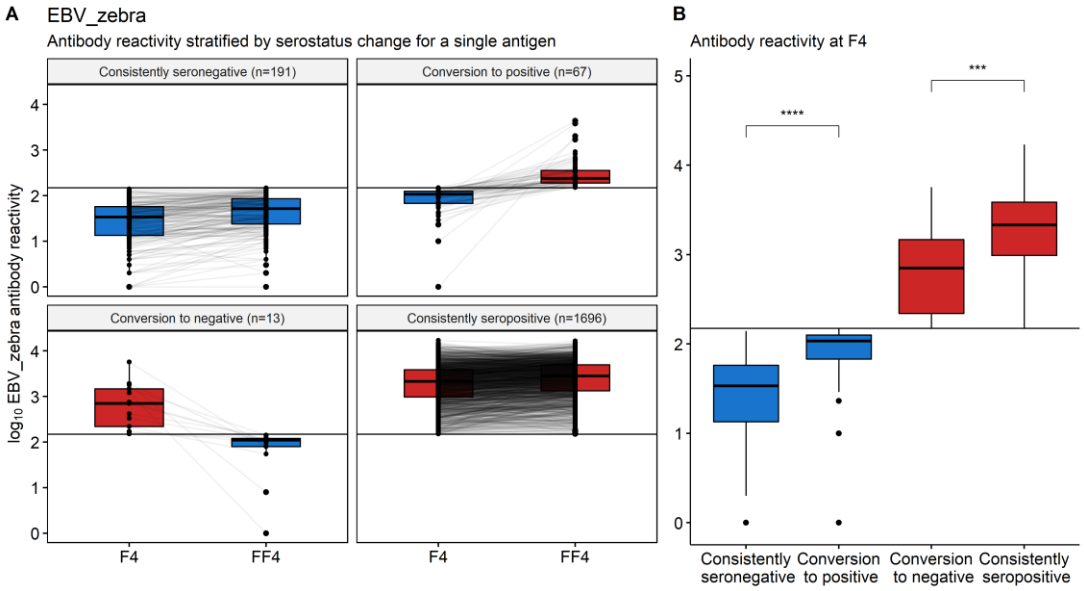
Supplementary Figure 5: EBV ead reactivity by serostatus change



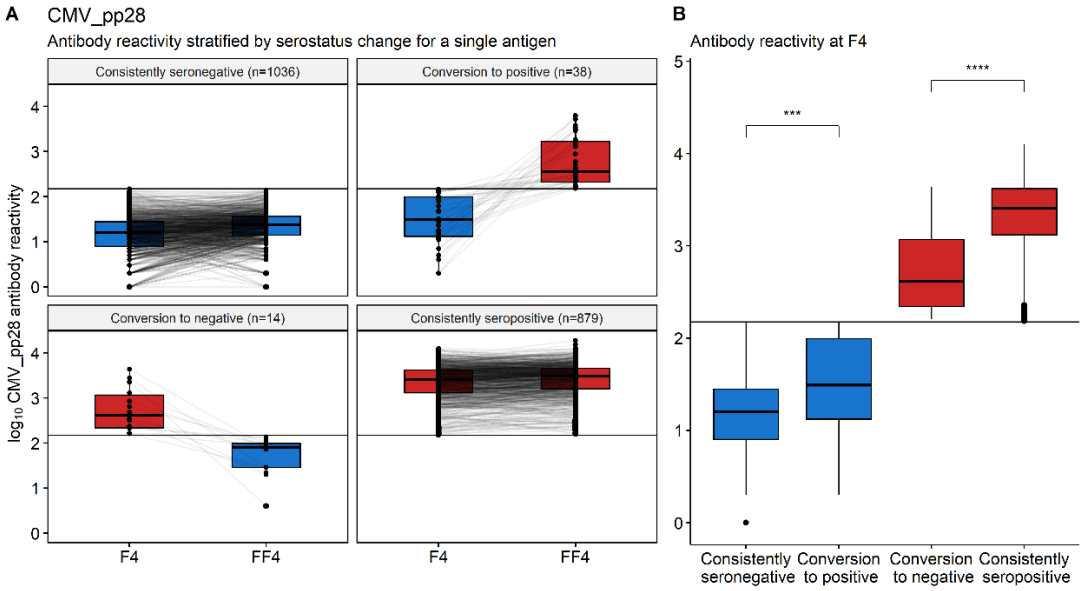
Supplementary Figure 6: EBV ebna1 reactivity by serostatus change



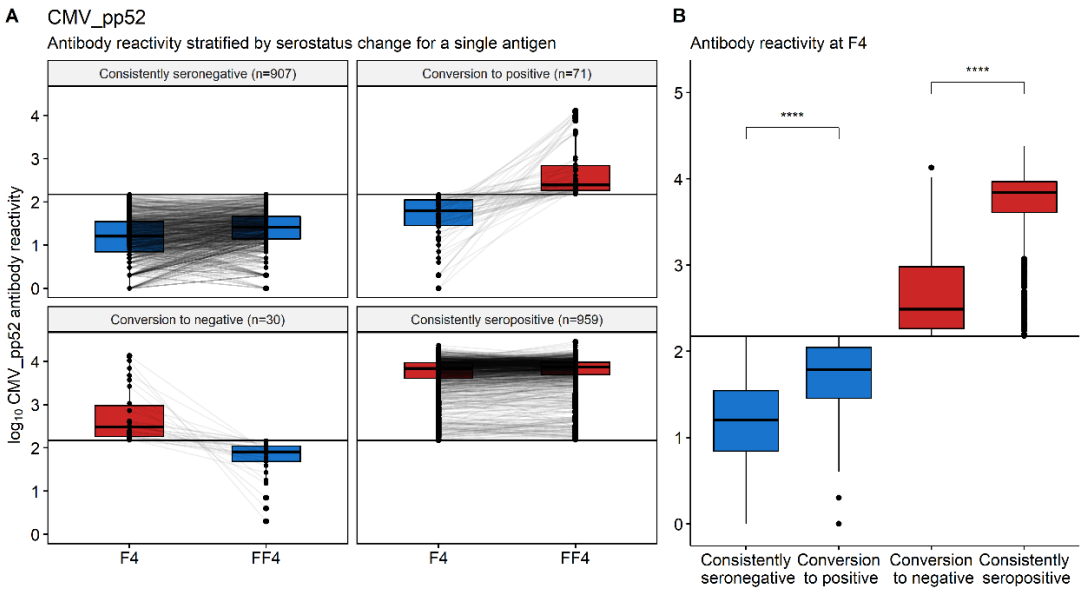
Supplementary Figure 7: EBV vcap18 reactivity by serostatus change



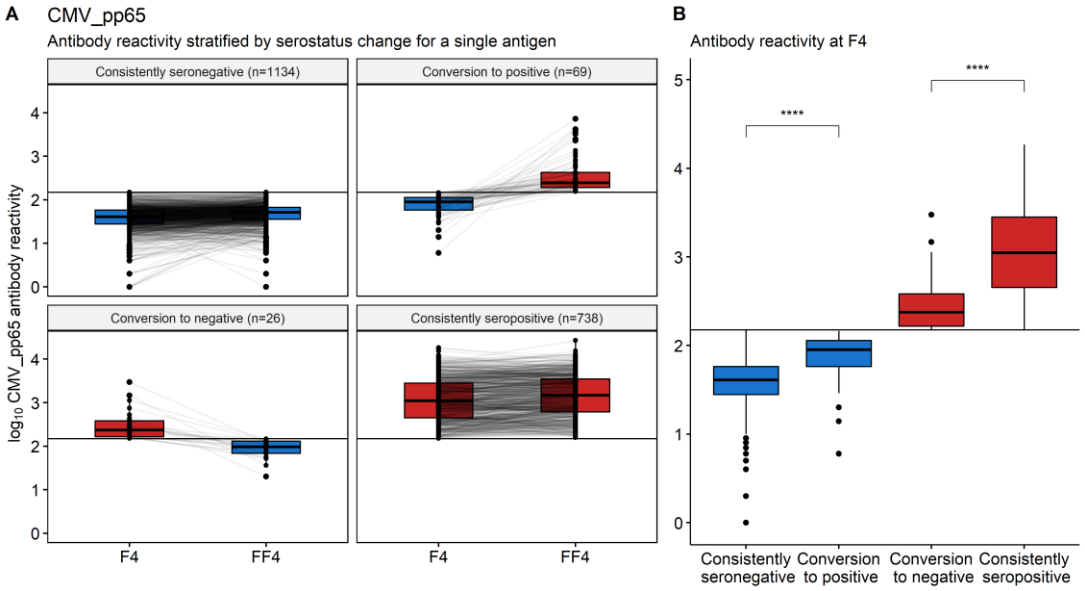
Supplementary Figure 8: EBV zebra reactivity by serostatus change



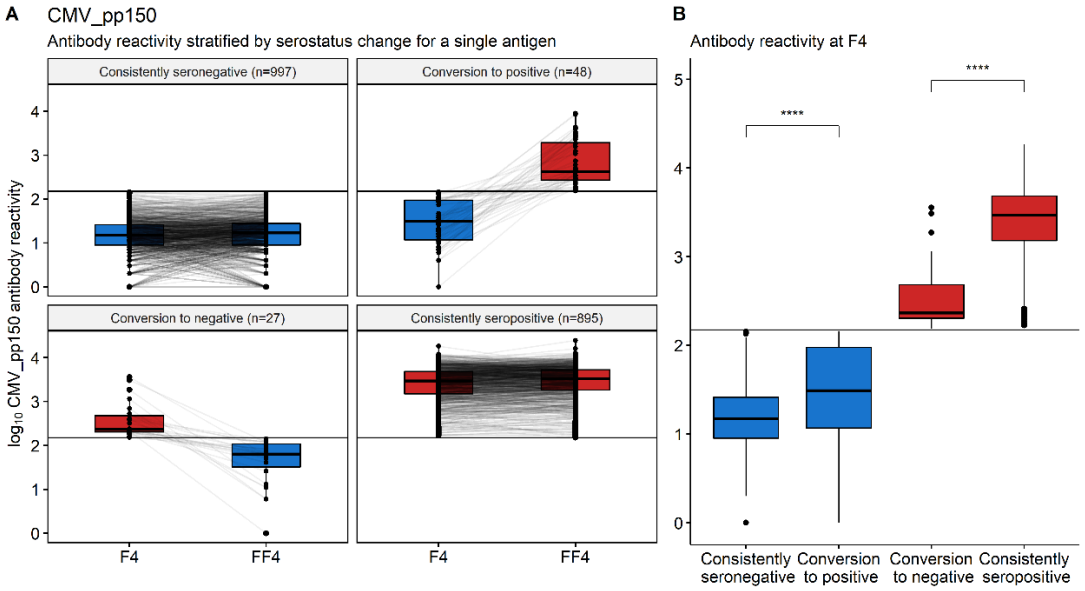
Supplementary Figure 9: CMV pp28 reactivity by serostatus change



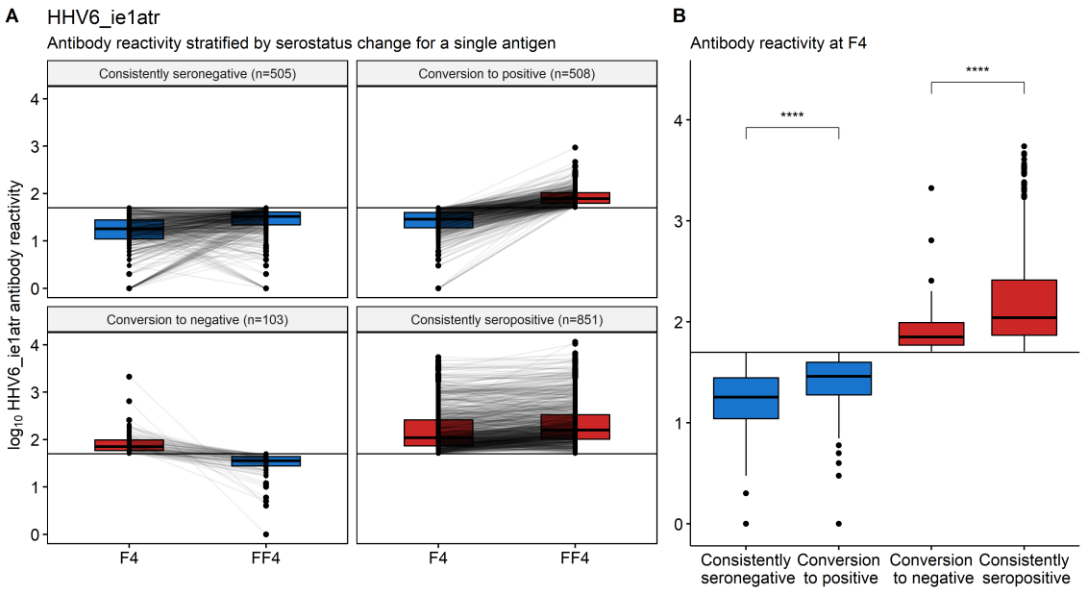
Supplementary Figure 10: CMV pp52 reactivity by serostatus change



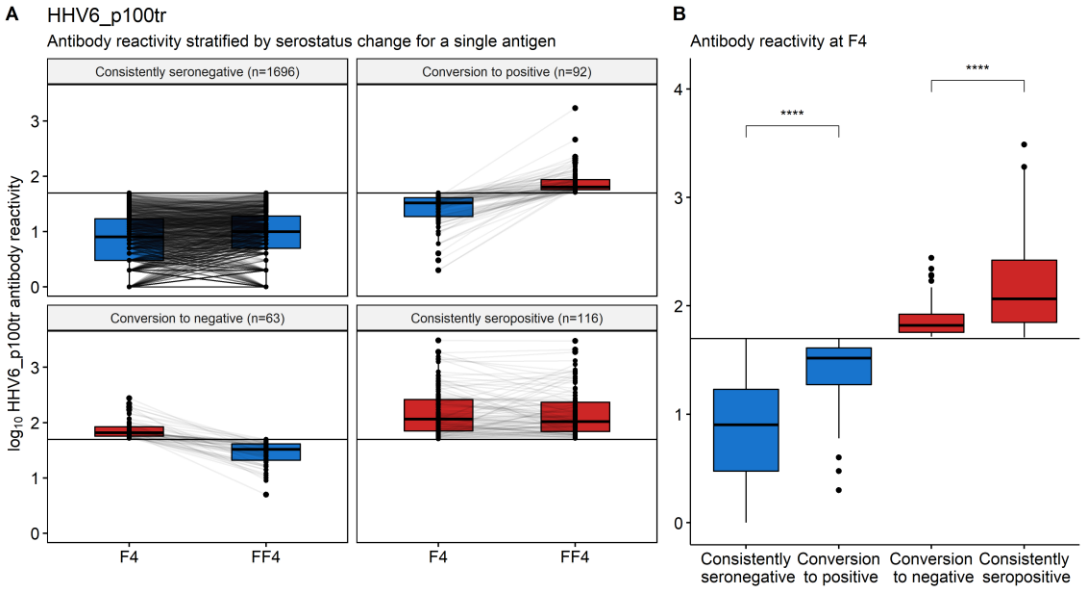
Supplementary Figure 11: CMV pp65 reactivity by serostatus change



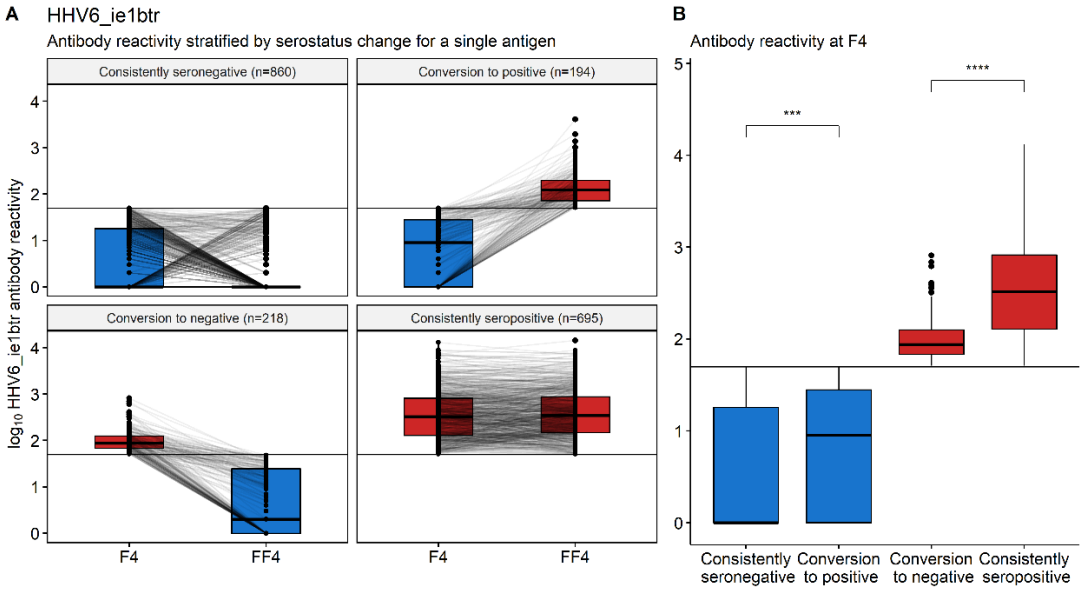
Supplementary Figure 12: CMV pp150 reactivity by serostatus change



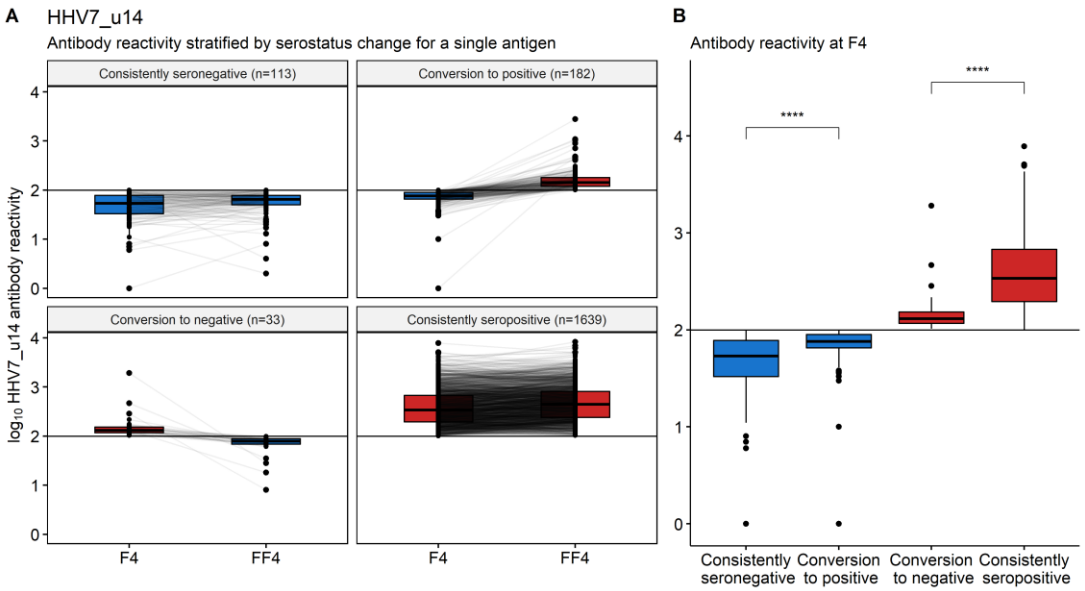
Supplementary Figure 13: HHV–6 ie1atr reactivity by serostatus change



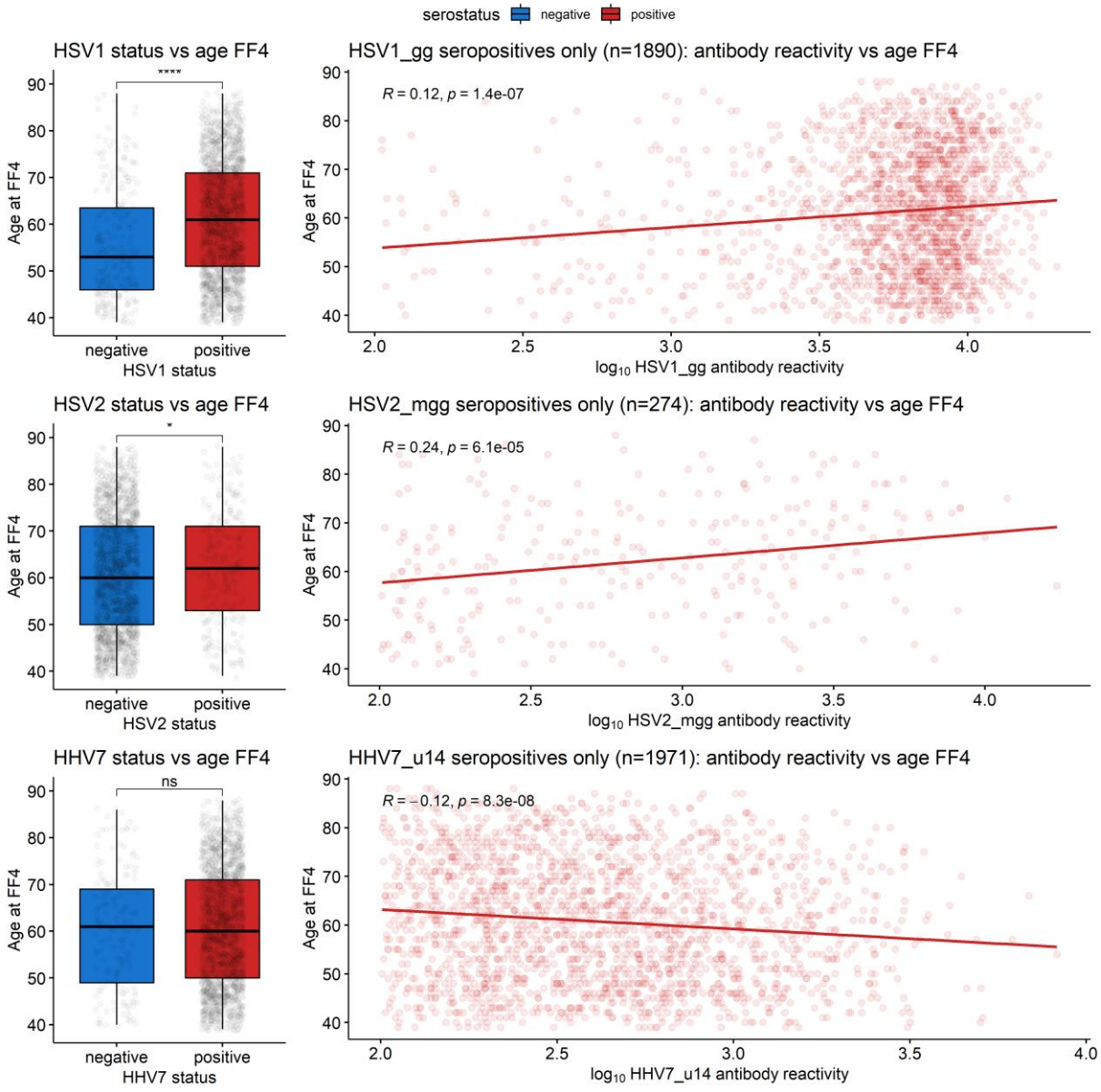
Supplementary Figure 14: HHV–6 p100tr reactivity by serostatus change



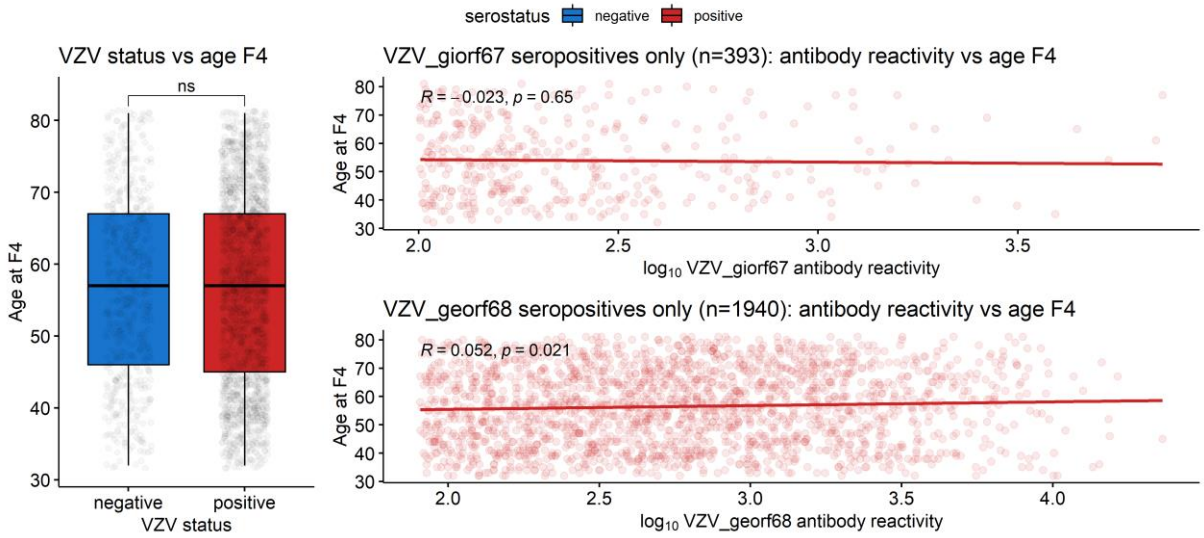
Supplementary Figure 15: HHV–6 ie1btr reactivity by serostatus change



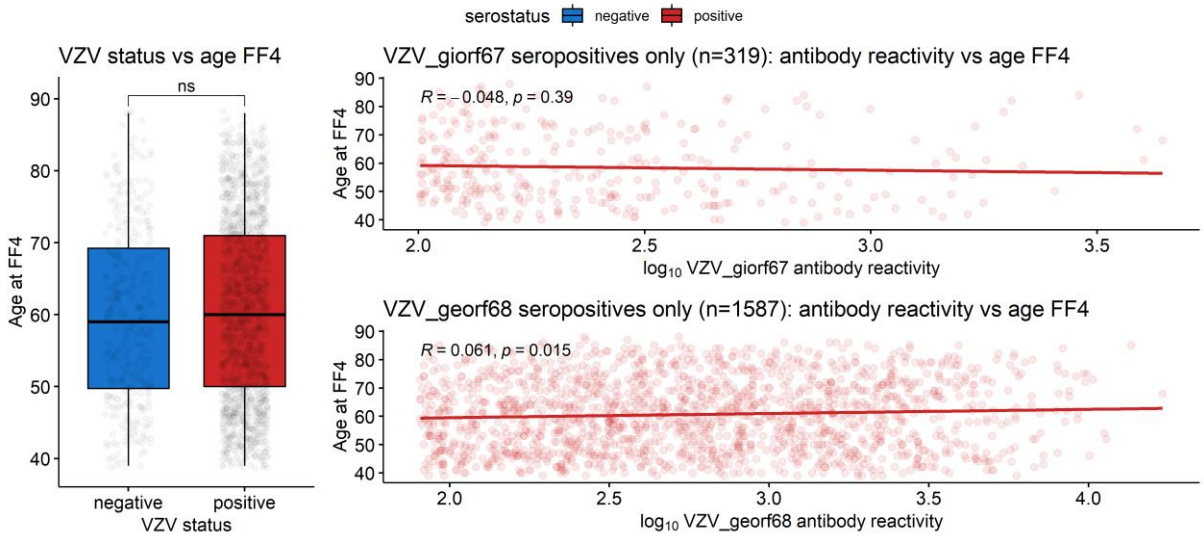
Supplementary Figure 16: HHV–7 u14 reactivity by serostatus change



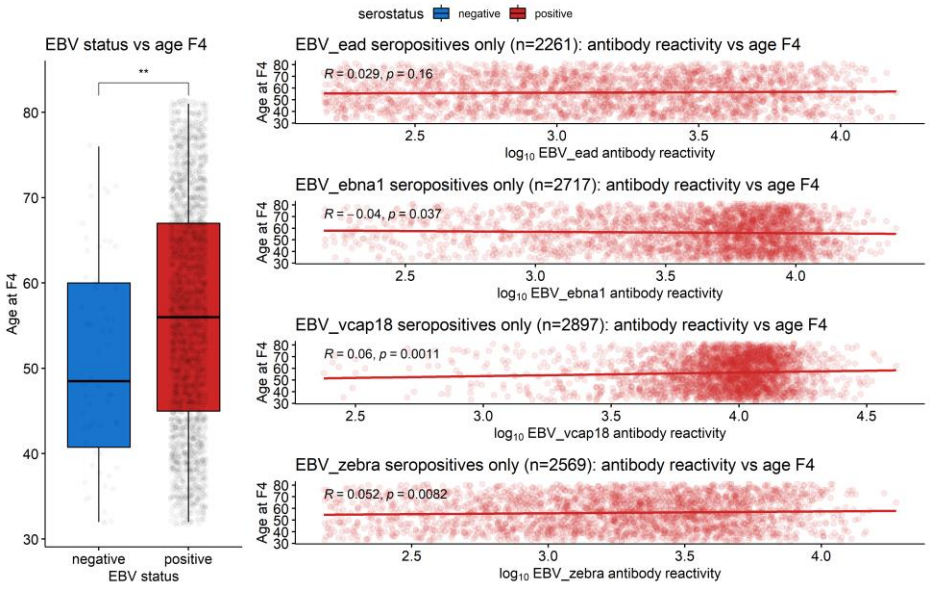
Supplementary Figure 17: HSV–1, HSV–2, HHV–7 viral status / reactivity vs age at FF4



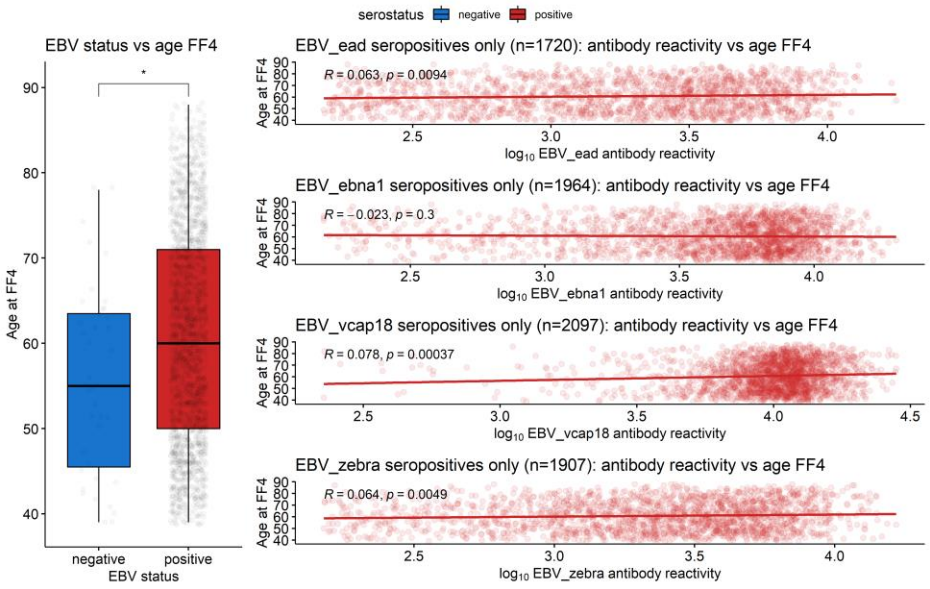
Supplementary Figure 18: VZV viral status / reactivity vs age at F4



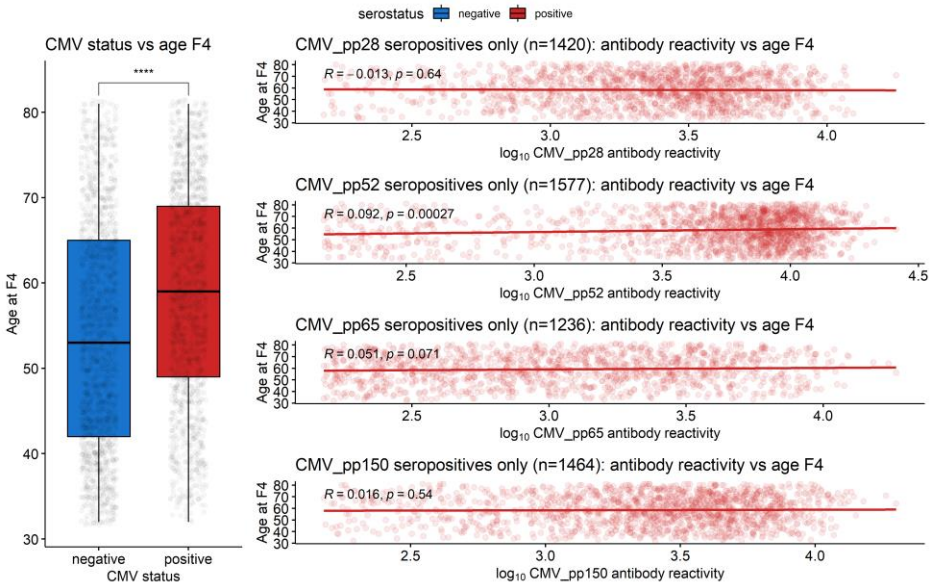
Supplementary Figure 19: VZV viral status / reactivity vs age at FF4



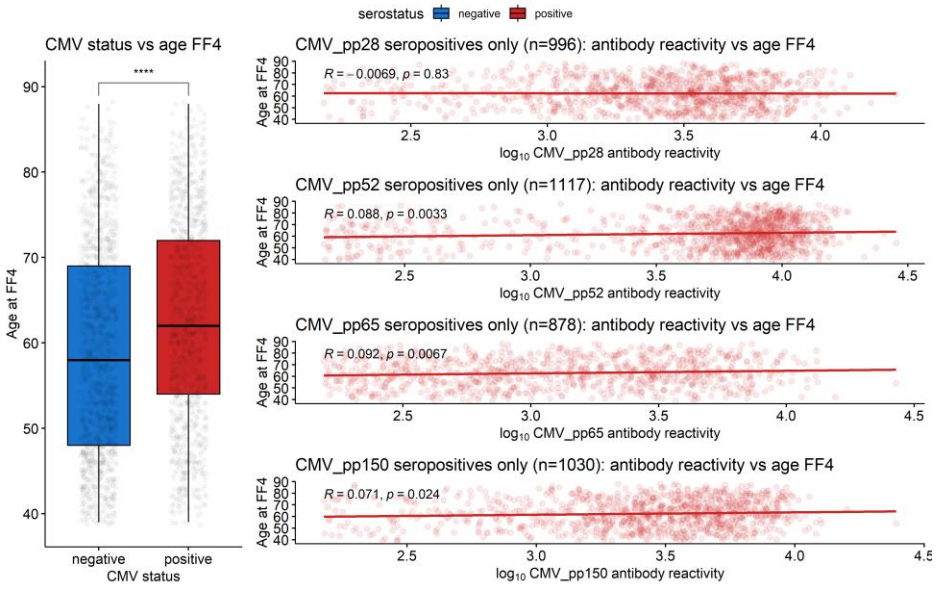
Supplementary Figure 20: EBV viral status / reactivity vs age at F4



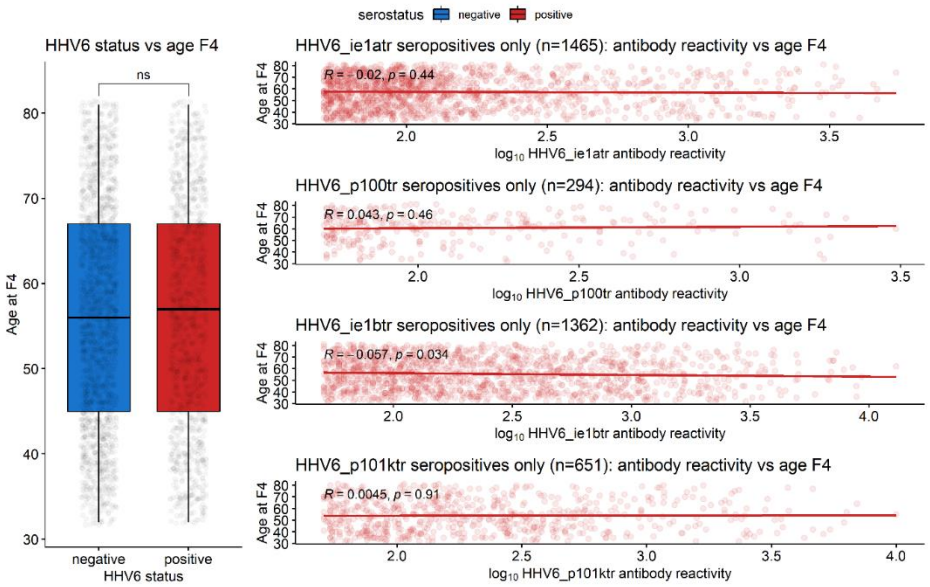
Supplementary Figure 21: EBV viral status / reactivity vs age at FF4



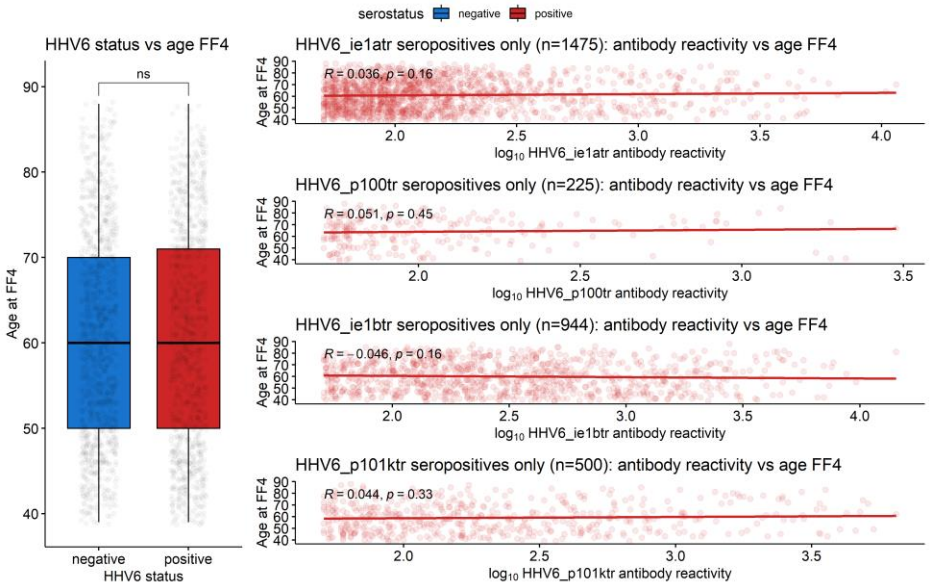
Supplementary Figure 22: CMV viral status / reactivity vs age at F4



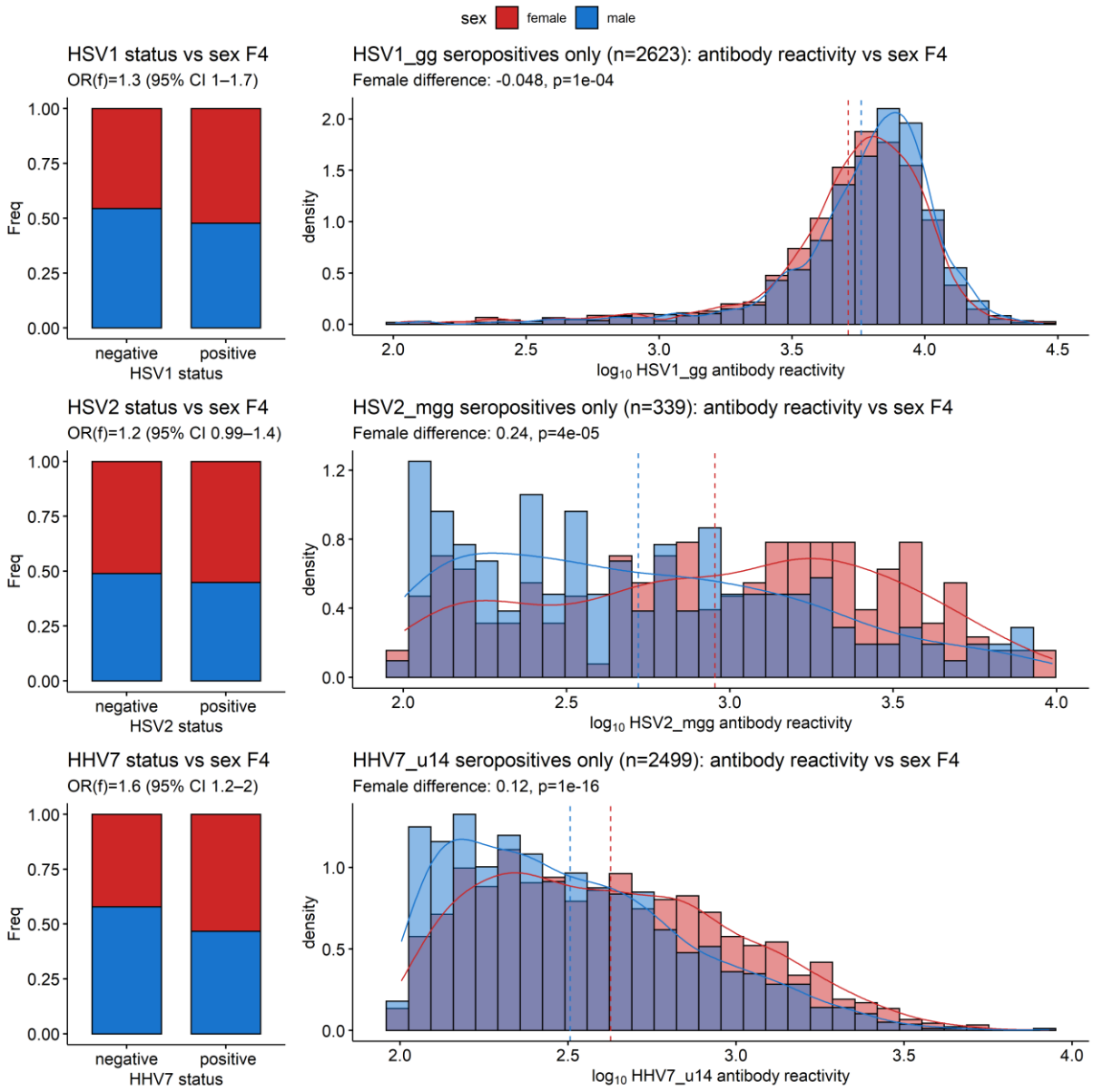
Supplementary Figure 23: CMV viral status / reactivity vs age at FF4



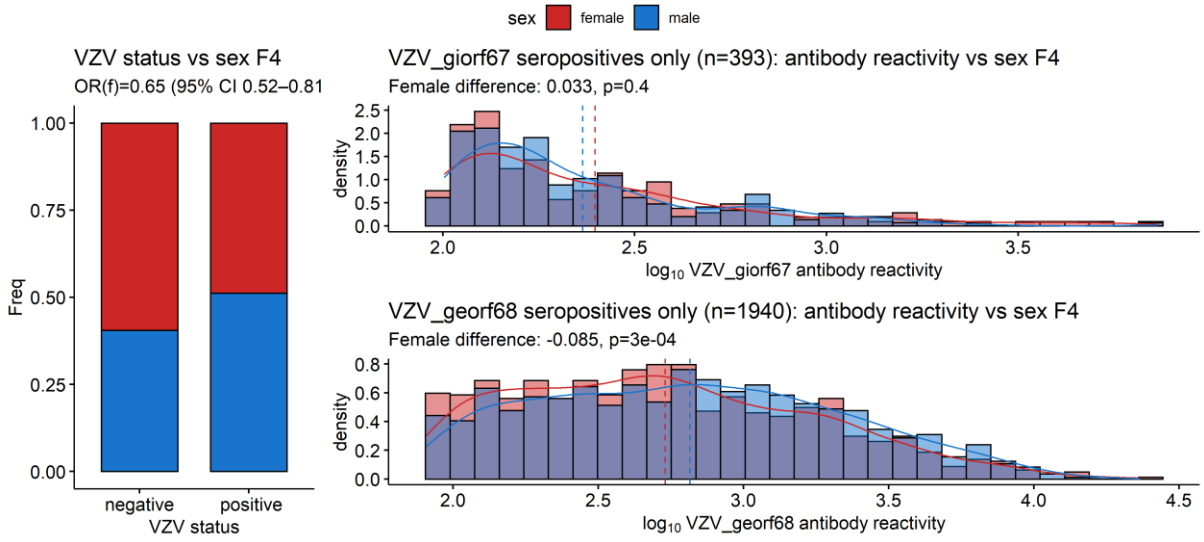
Supplementary Figure 24: HHV–6 viral status / reactivity vs age at F4



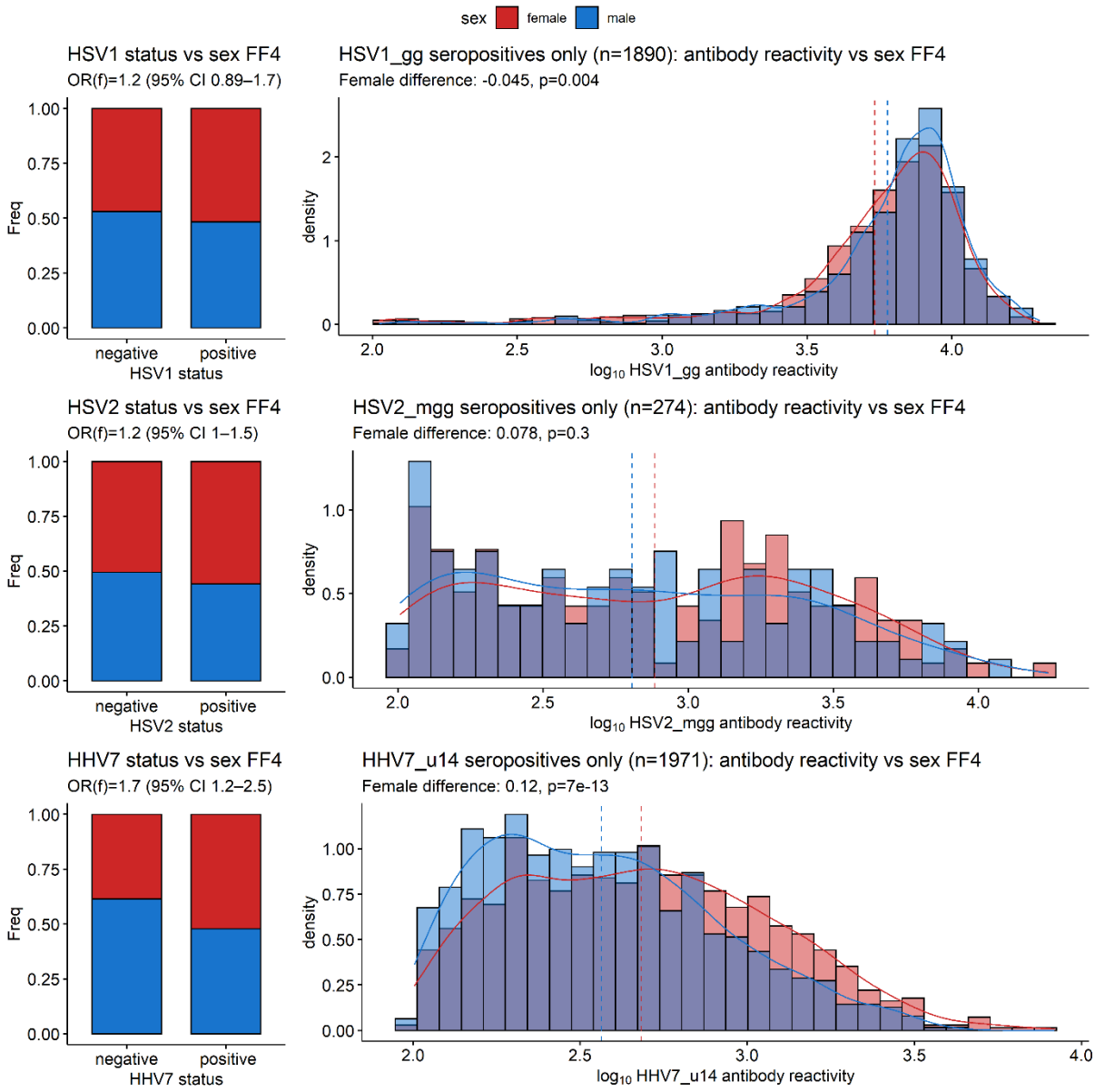
Supplementary Figure 25: HHV–6 viral status / reactivity vs age at FF4



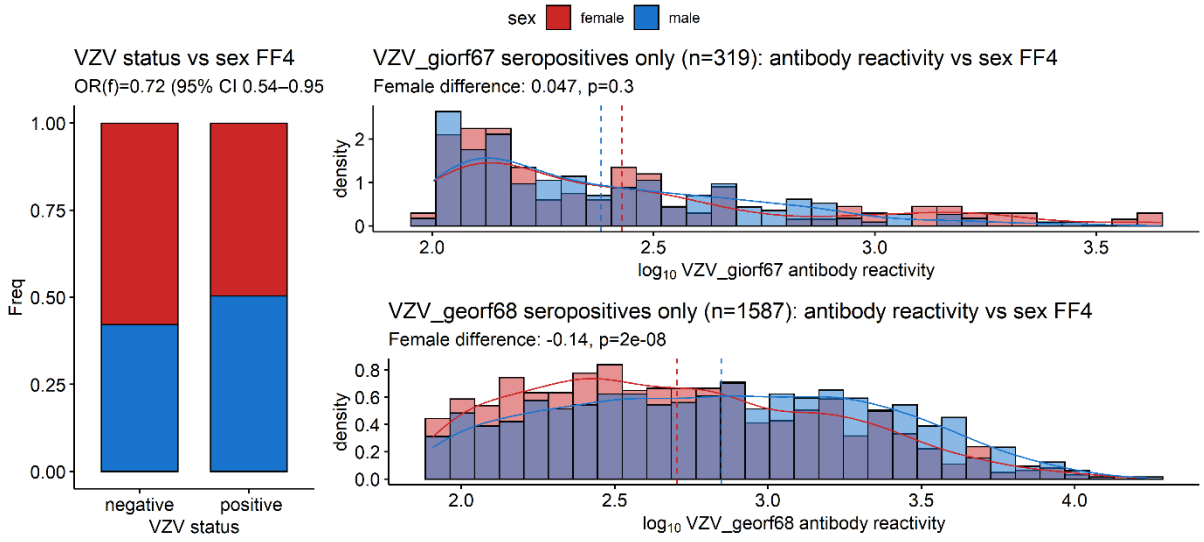
Supplementary Figure 26: Distribution of HSV–1, HSV–2 & HHV–7 status / reactivity by sex at F4



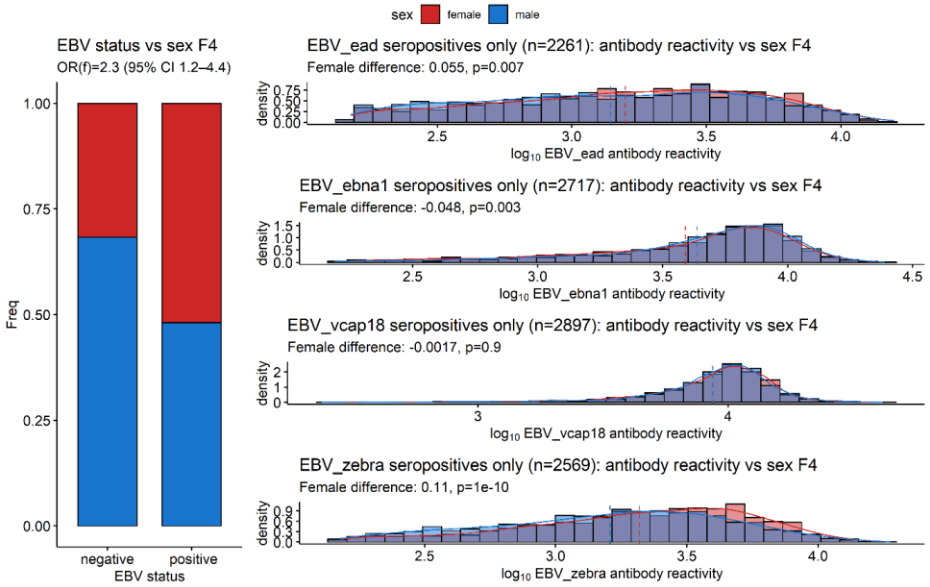
Supplementary Figure 27: Distribution of VZV status / reactivity by sex at F4



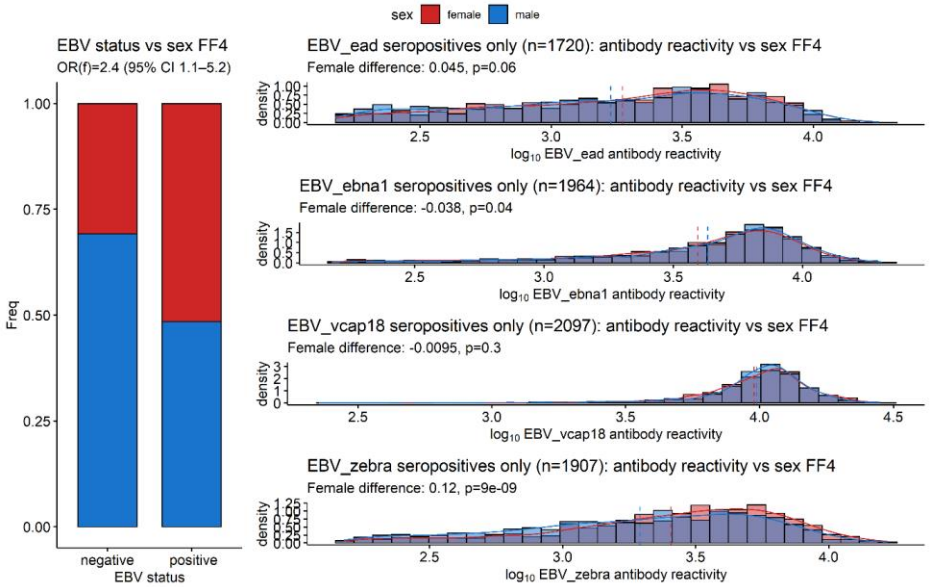
Supplementary Figure 28: Distribution of HSV–1, HSV–2 & HHV–7 status / reactivity by sex at FF4



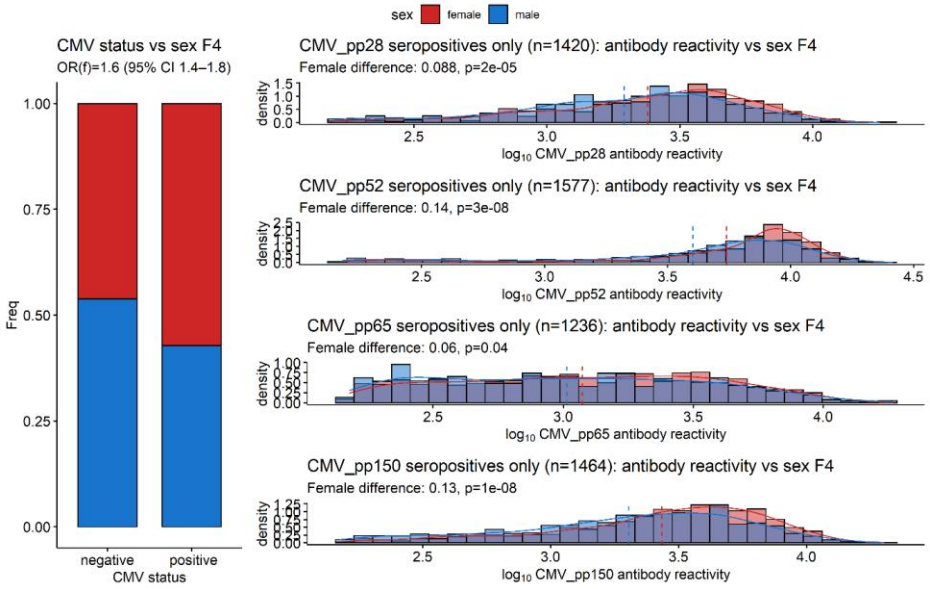
Supplementary Figure 29: Distribution of VZV status / reactivity by sex at FF4



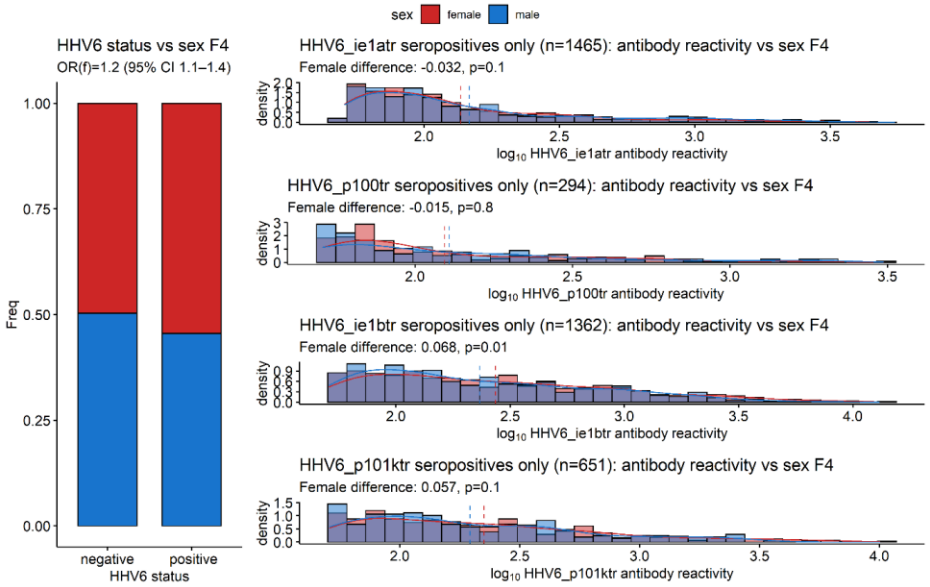
Supplementary Figure 30: Distribution of EBV status / reactivity by sex at F4



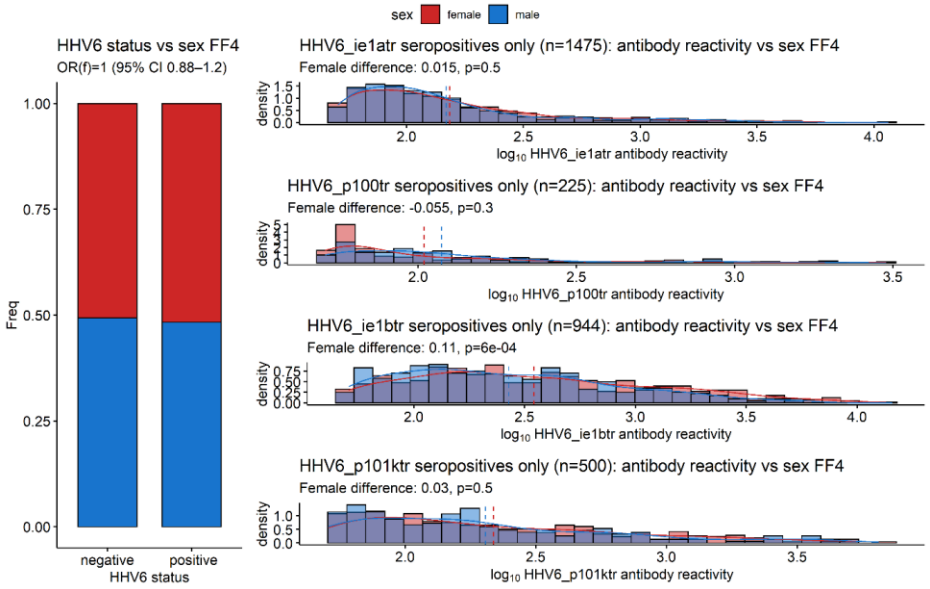
Supplementary Figure 31: Distribution of EBV status / reactivity by sex at FF4



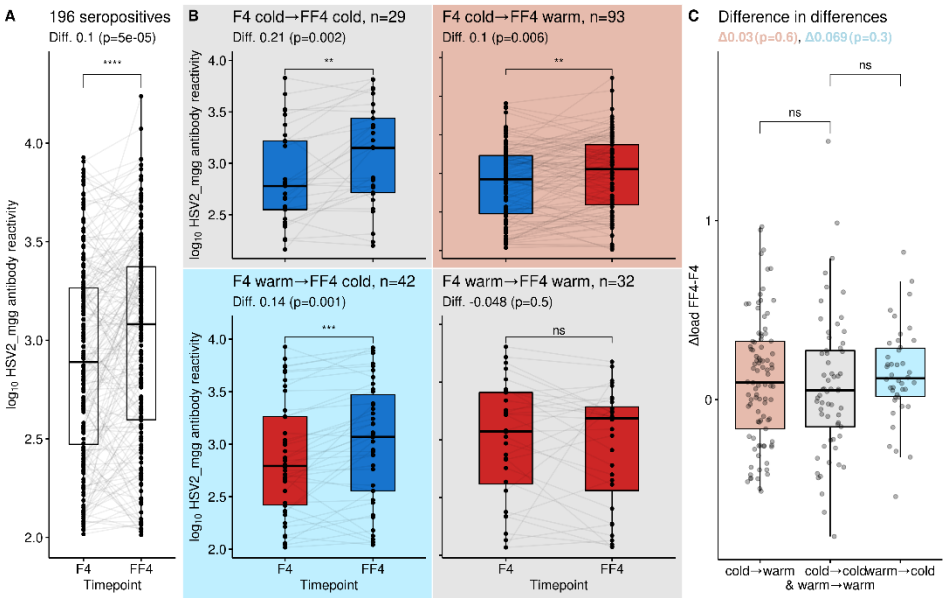
Supplementary Figure 32: Distribution of CMV status / reactivity by sex at FF4



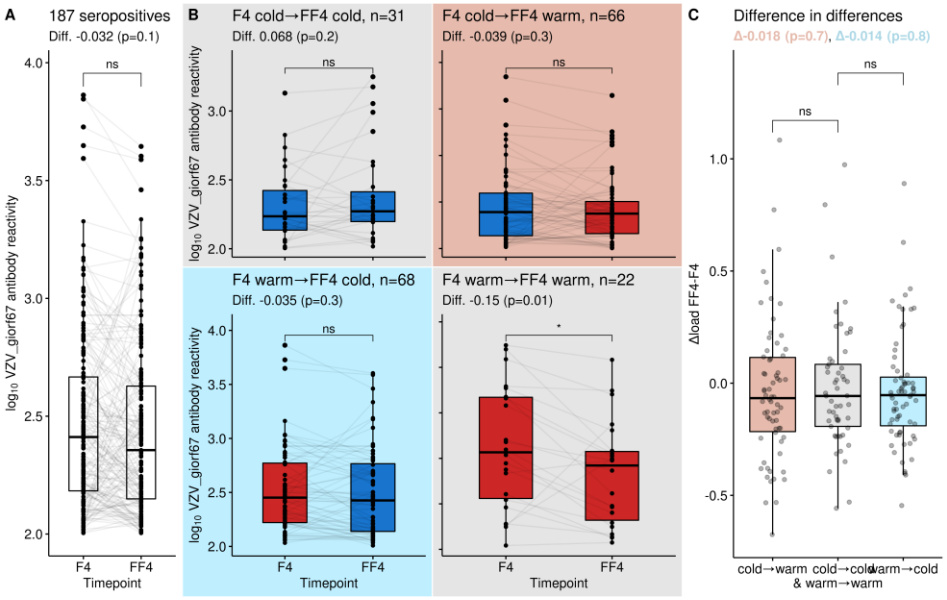
Supplementary Figure 33: Distribution of HHV–6 status / reactivity by sex at F4



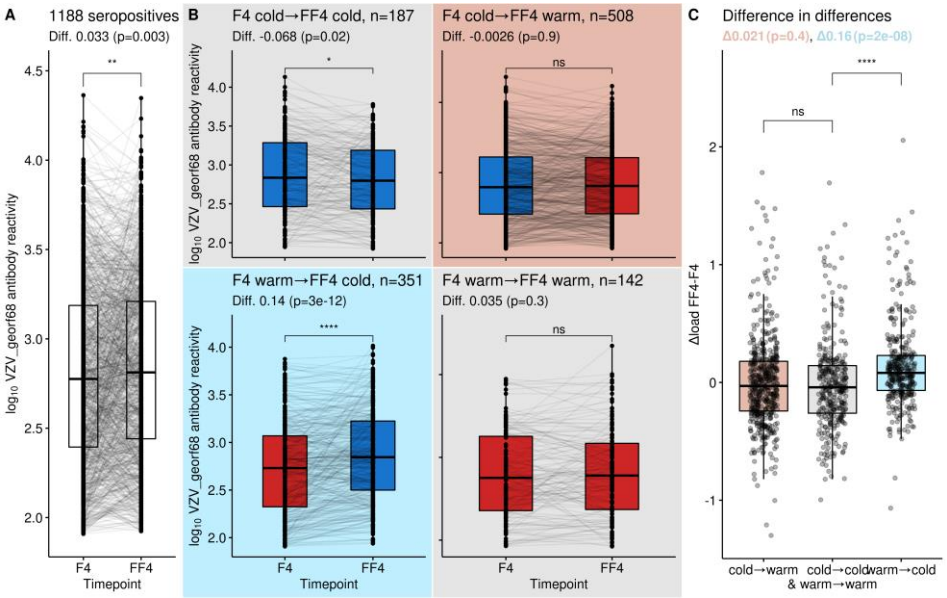
Supplementary Figure 34: Distribution of HHV–6 status / reactivity by sex at FF4



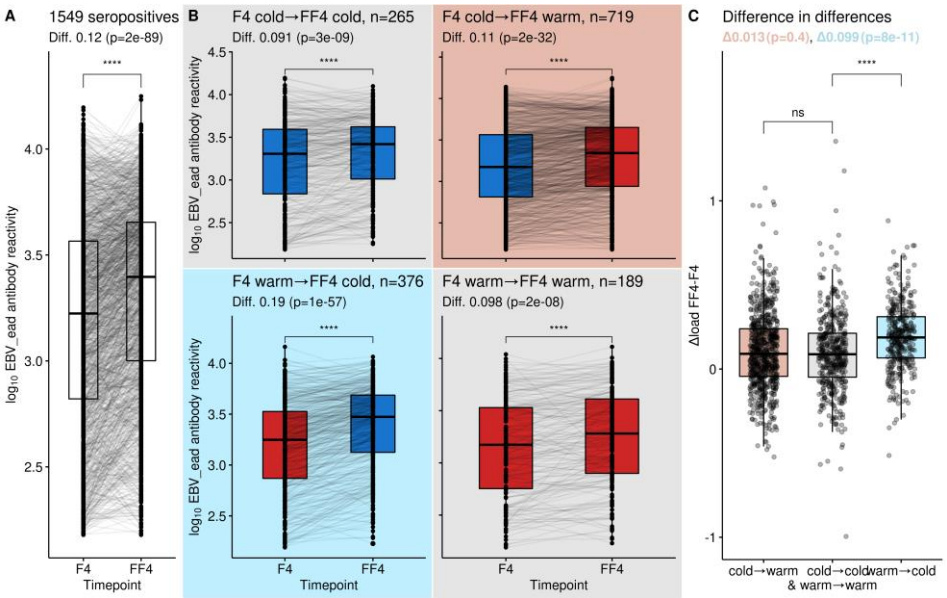
Supplementary Figure 35: Difference in differences analysis for season for HSV–2 mgg



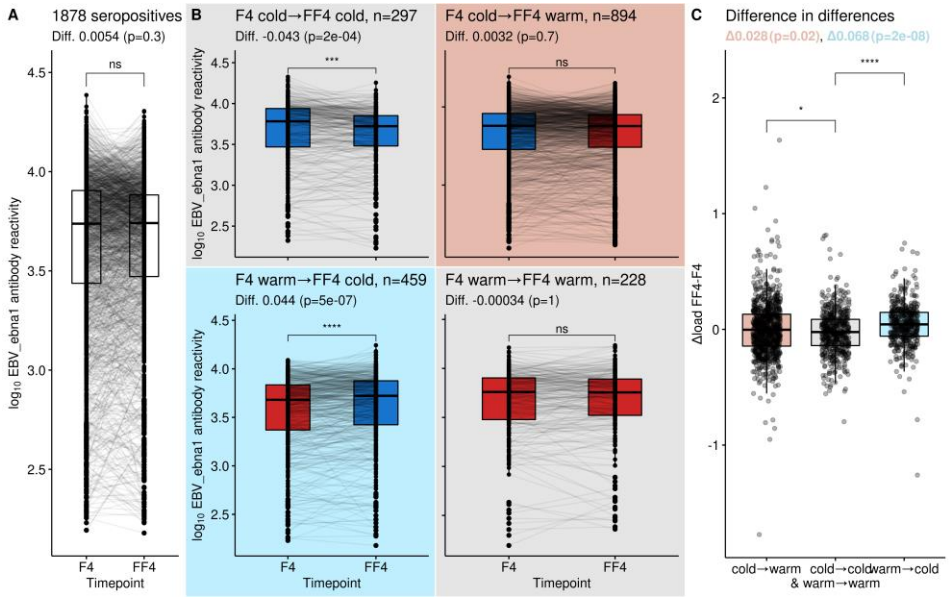
Supplementary Figure 36: Difference in differences analysis for season for VZV giorf67



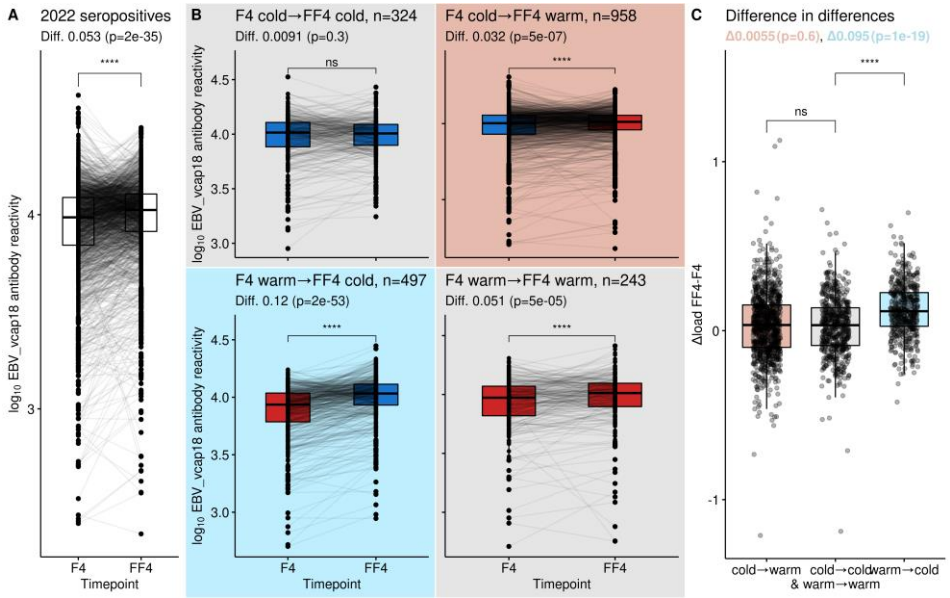
Supplementary Figure 37: Difference in differences analysis for season for VZV georf68



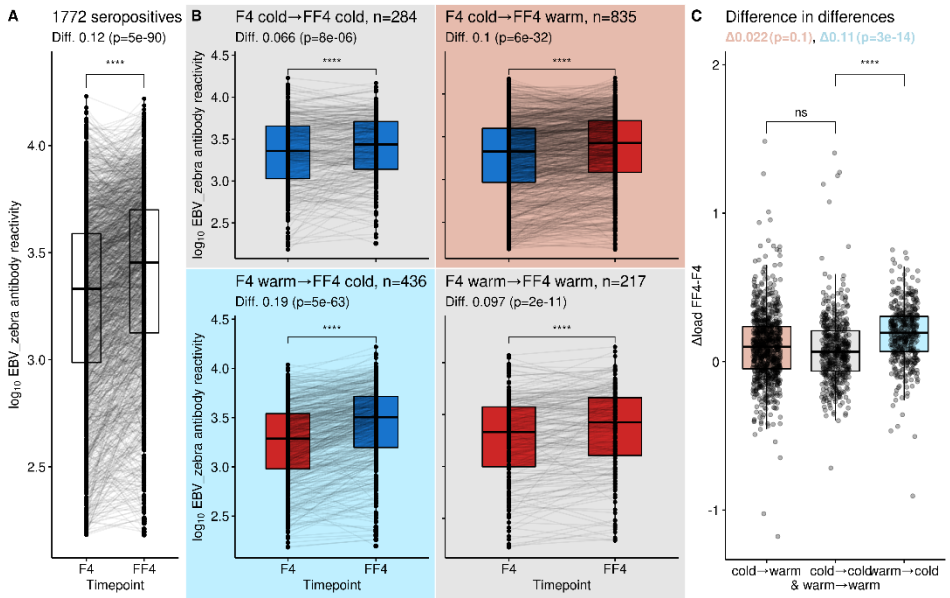
Supplementary Figure 38: Difference in differences analysis for season for EBV ead



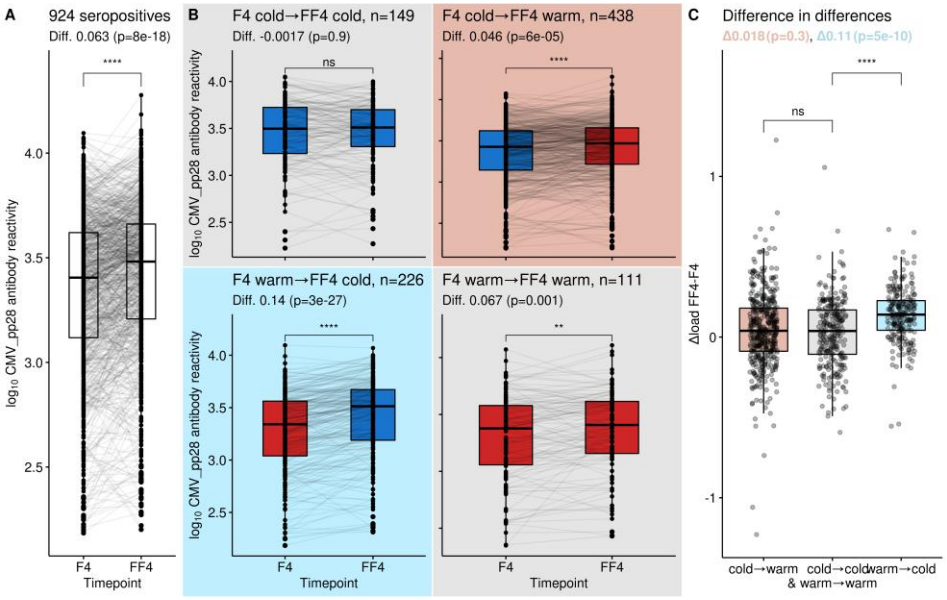
Supplementary Figure 39: Difference in differences analysis for season for EBV ebna1



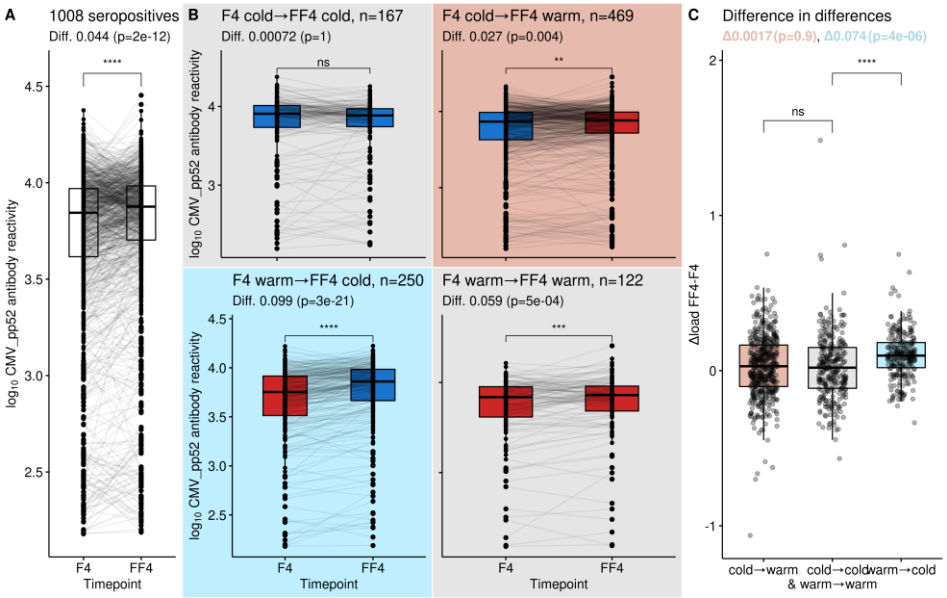
Supplementary Figure 40: Difference in differences analysis for season for EBV vcap18



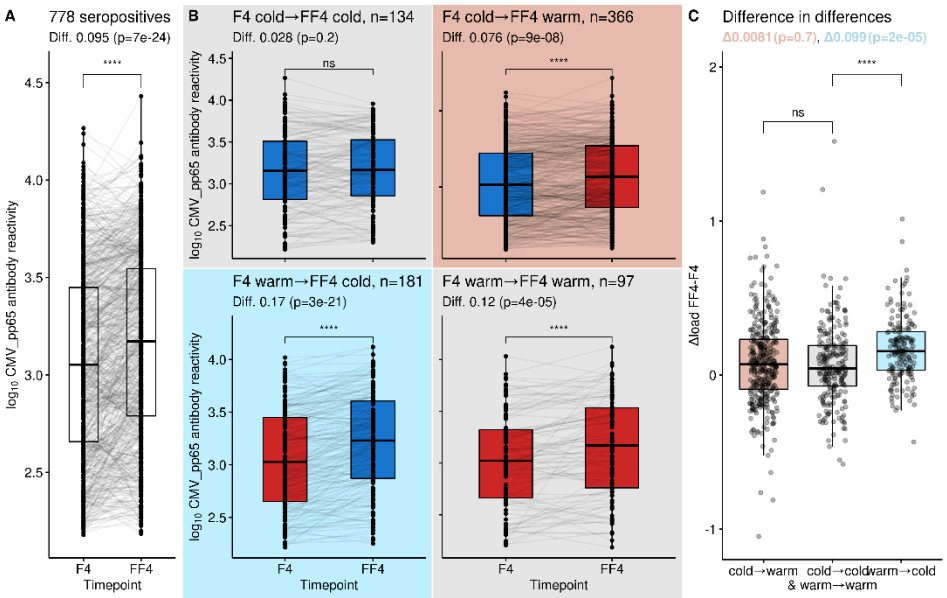
Supplementary Figure 41: Difference in differences analysis for season for EBV zebra



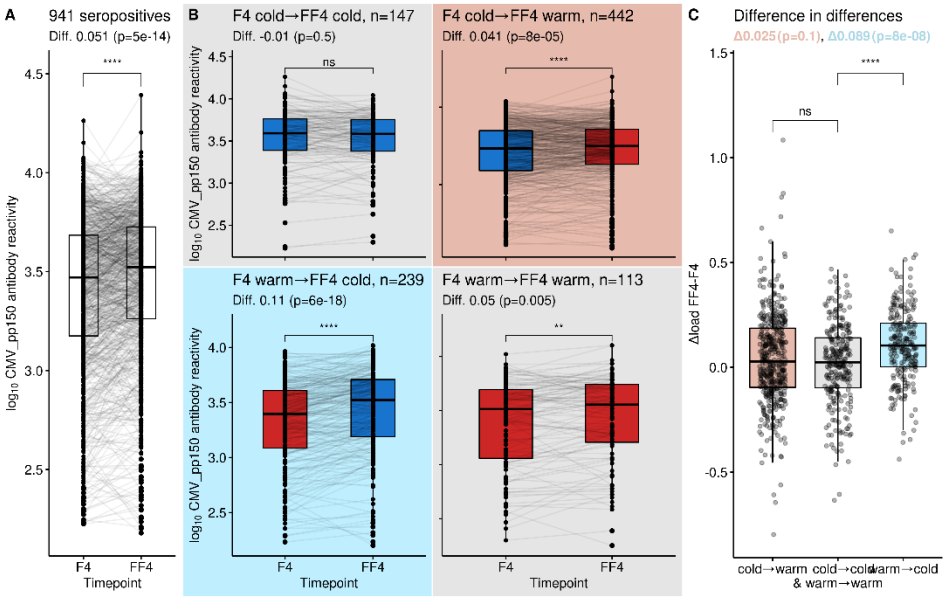
Supplementary Figure 42: Difference in differences analysis for season for CMV pp28



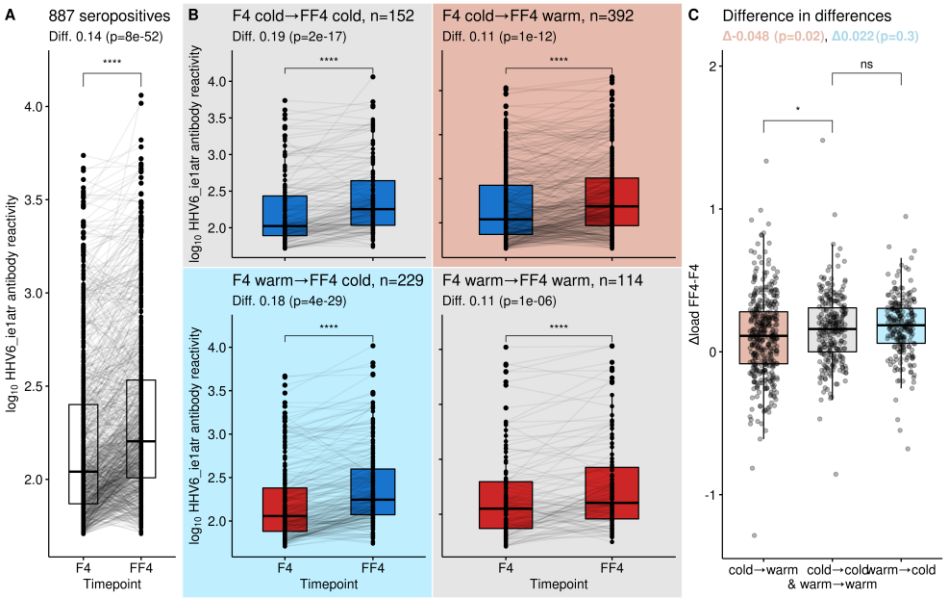
Supplementary Figure 43: Difference in differences analysis for season for CMV pp52



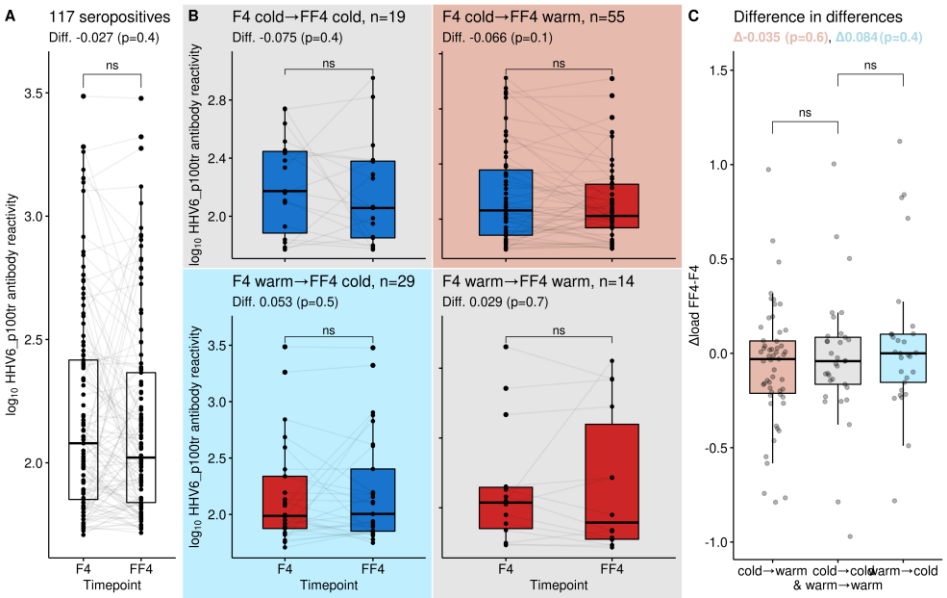
Supplementary Figure 44: Difference in differences analysis for season for CMV pp65



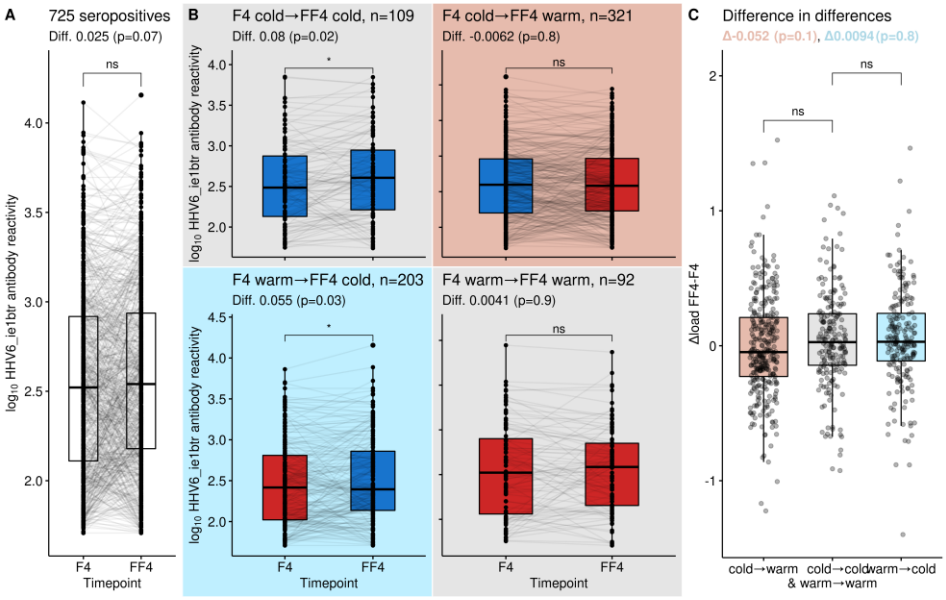
Supplementary Figure 45: Difference in differences analysis for season for CMV pp150



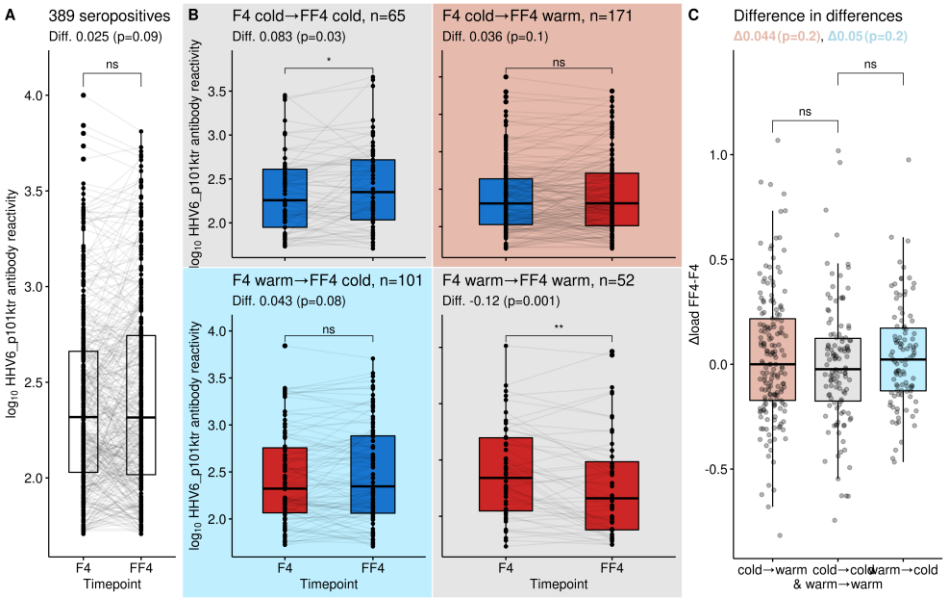
Supplementary Figure 46: Difference in differences analysis for season for HHV–6 ie1atr



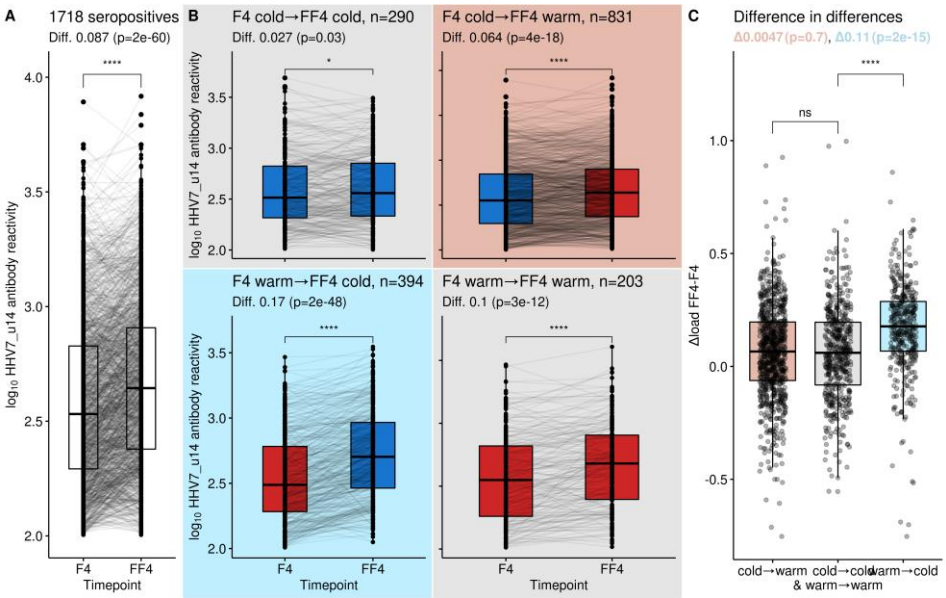
Supplementary Figure 47: Difference in differences analysis for season for HHV–6 p100tr



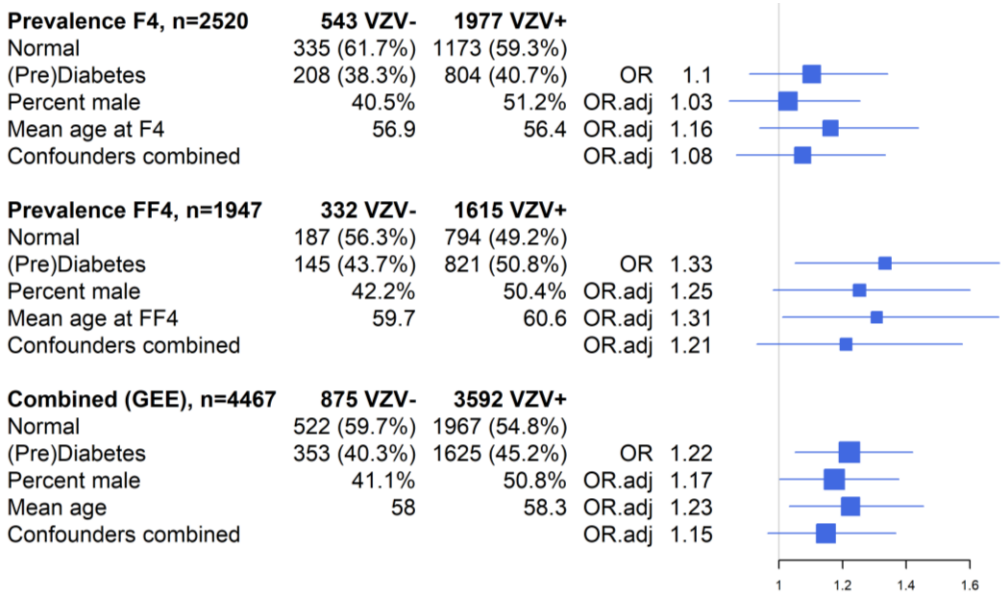
Supplementary Figure 48: Difference in differences analysis for season for HHV–6 ie1btr



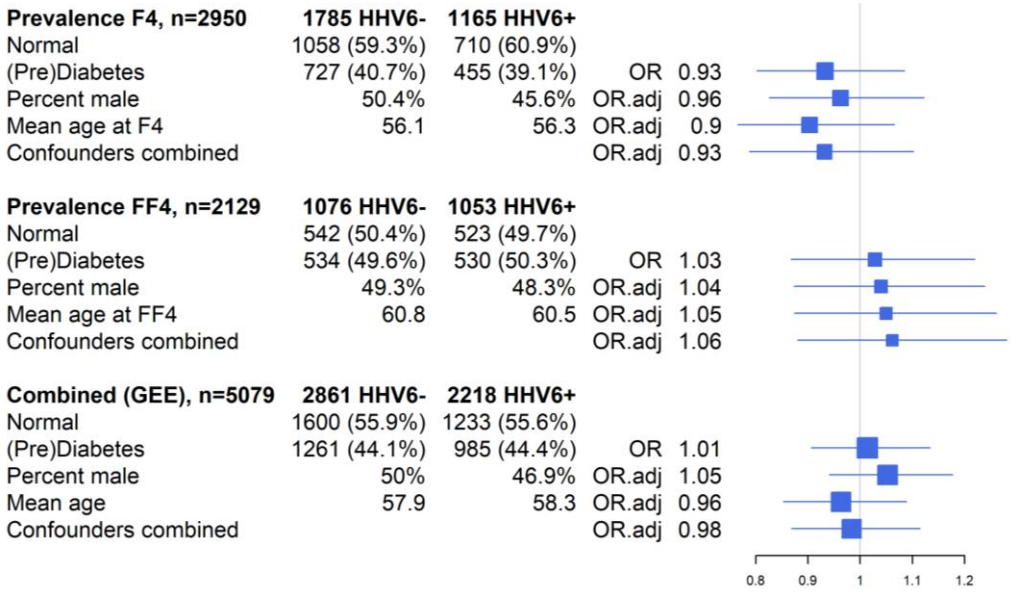
Supplementary Figure 49: Difference in differences analysis for season for HHV–6 p101ktr



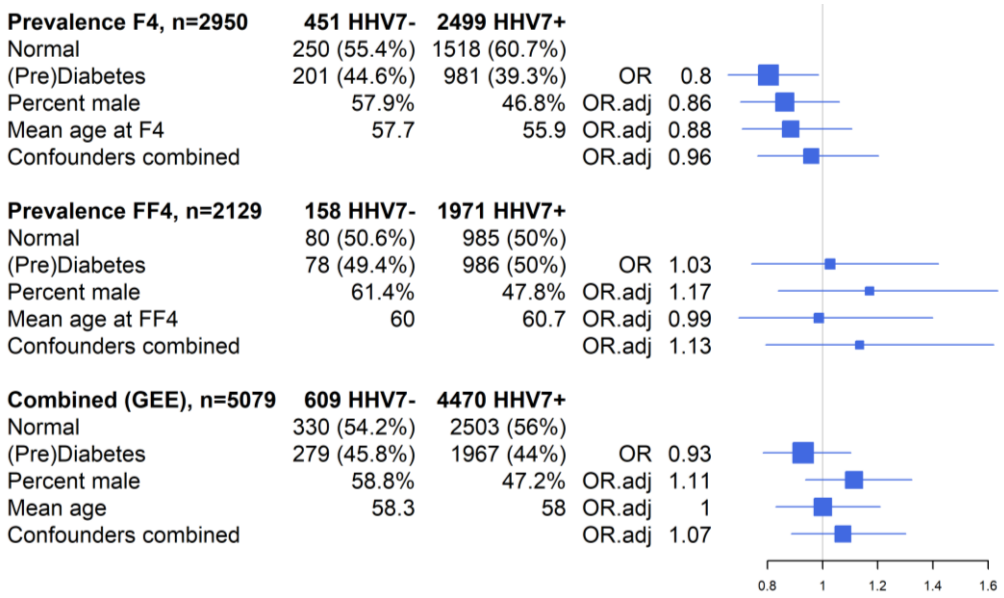
Supplementary Figure 50: Difference in differences analysis for season for HHV–7 u14



Supplementary Figure 51: VZV seropositivity and (pre)diabetes prevalence

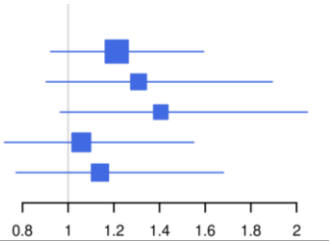


Supplementary Figure 52: HHV–6 seropositivity and (pre)diabetes prevalence



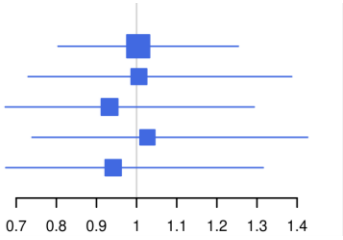
Supplementary Figure 53: HHV–7 seropositivity and (pre)diabetes prevalence

Incidence, at risk at F4: 1257	168 HSV1-	1089 HSV1+		
Normal glucose tolerance at FF4	125 (74.4%)	751 (69%)	RR	1.21
Incident (pre)diabetes at FF4	43 (25.6%)	338 (31%)	OR	1.31
Percent male	48.8%	41%	OR.adj	1.41
Mean age at F4	46	51.2	OR.adj	1.06
Confounders combined			OR.adj	1.14



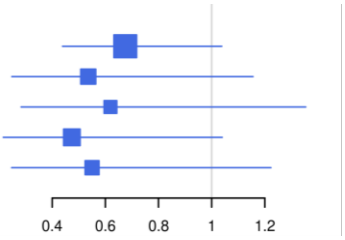
Supplementary Figure 54: HSV–1 seropositivity and (pre)diabetes incidence

Incidence, at risk at F4: 986	227 VZV-	759 VZV+		
Normal glucose tolerance at FF4	157 (69.2%)	524 (69%)	RR	1
Incident (pre)diabetes at FF4	70 (30.8%)	235 (31%)	OR	1.01
Percent male	34.8%	44.8%	OR.adj	0.93
Mean age at F4	51.1	50.5	OR.adj	1.03
Confounders combined			OR.adj	0.94



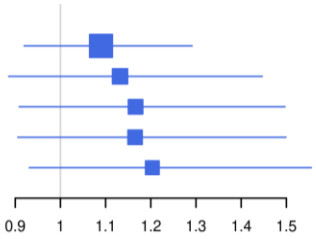
Supplementary Figure 55: VZV seropositivity and (pre)diabetes incidence

Incidence, at risk at F4: 1257	27 EBV-	1230 EBV+		
Normal glucose tolerance at FF4	15 (55.6%)	861 (70%)	RR	0.68
Incident (pre)diabetes at FF4	12 (44.4%)	369 (30%)	OR	0.54
Percent male	63%	41.5%	OR.adj	0.62
Mean age at F4	48.5	50.6	OR.adj	0.47
Confounders combined			OR.adj	0.55



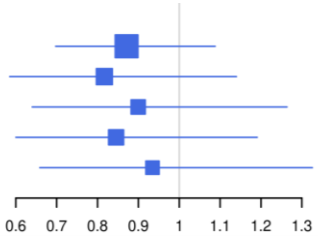
Supplementary Figure 56: EBV seropositivity and (pre)diabetes incidence

Incidence, at risk at F4: 1257	765 HHV6-	492 HHV6+		
Normal glucose tolerance at FF4	541 (70.7%)	335 (68.1%)	RR	1.09
Incident (pre)diabetes at FF4	224 (29.3%)	157 (31.9%)	OR	1.13
Percent male	43.3%	40%	OR.adj	1.17
Mean age at F4	50.7	50.2	OR.adj	1.17
Confounders combined			OR.adj	1.2

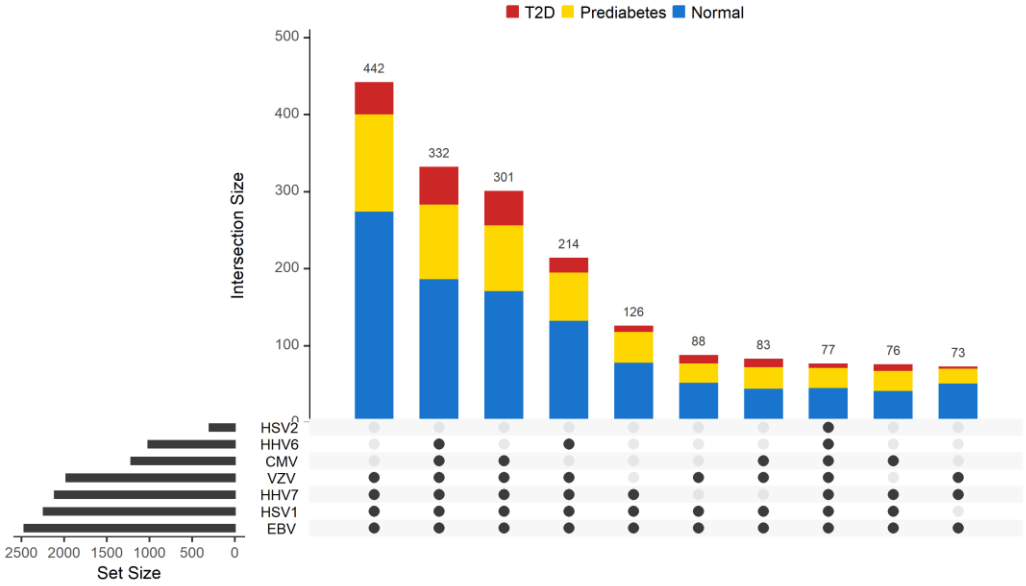


Supplementary Figure 57: HHV–6 seropositivity and (pre)diabetes incidence

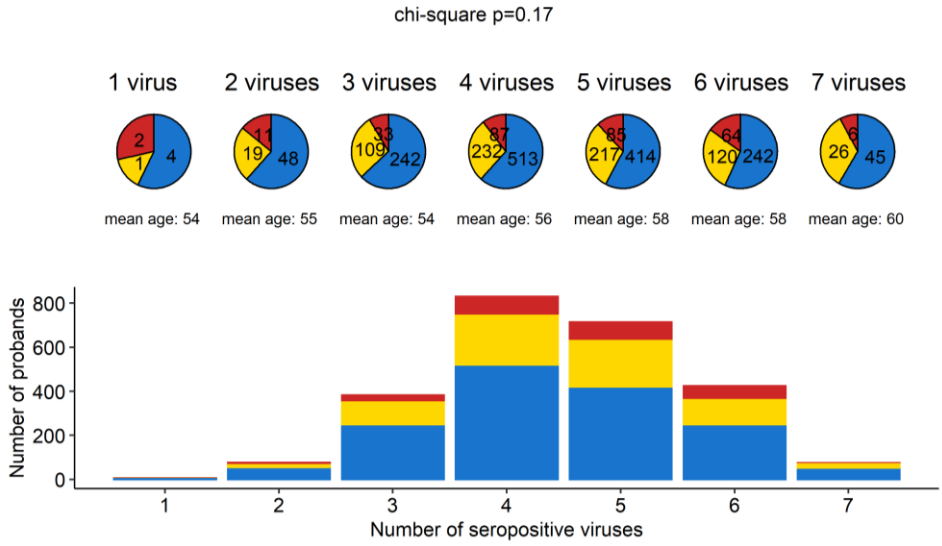
Incidence, at risk at F4: 1257	182 HHV7-	1075 HHV7+		
Normal glucose tolerance at FF4	120 (65.9%)	756 (70.3%)	RR	0.87
Incident (pre)diabetes at FF4	62 (34.1%)	319 (29.7%)	OR	0.82
Percent male	53.3%	40.1%	OR.adj	0.9
Mean age at F4	51.4	50.4	OR.adj	0.85
Confounders combined			OR.adj	0.93



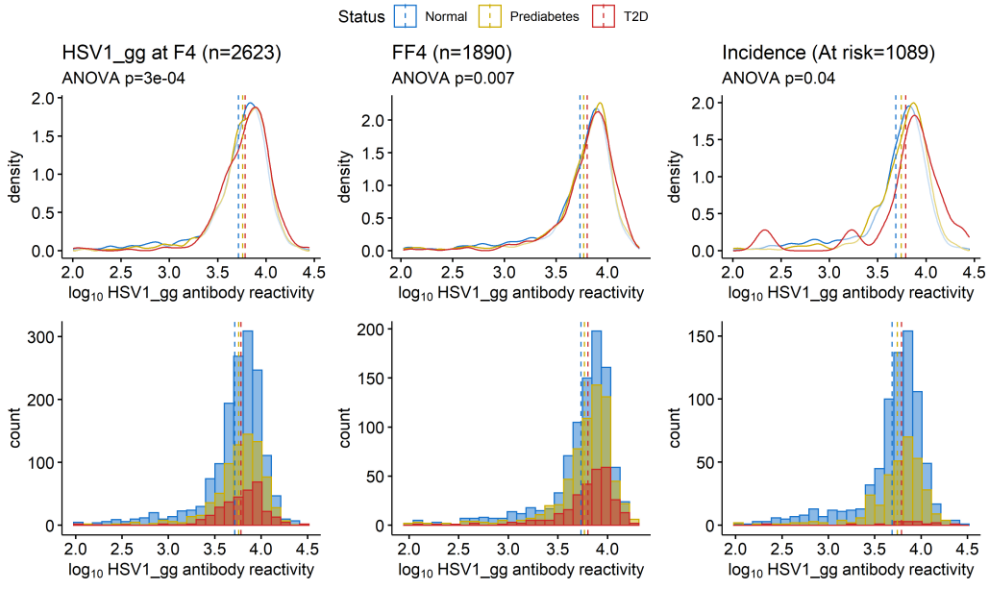
Supplementary Figure 58: HHV–7 seropositivity and (pre)diabetes incidence



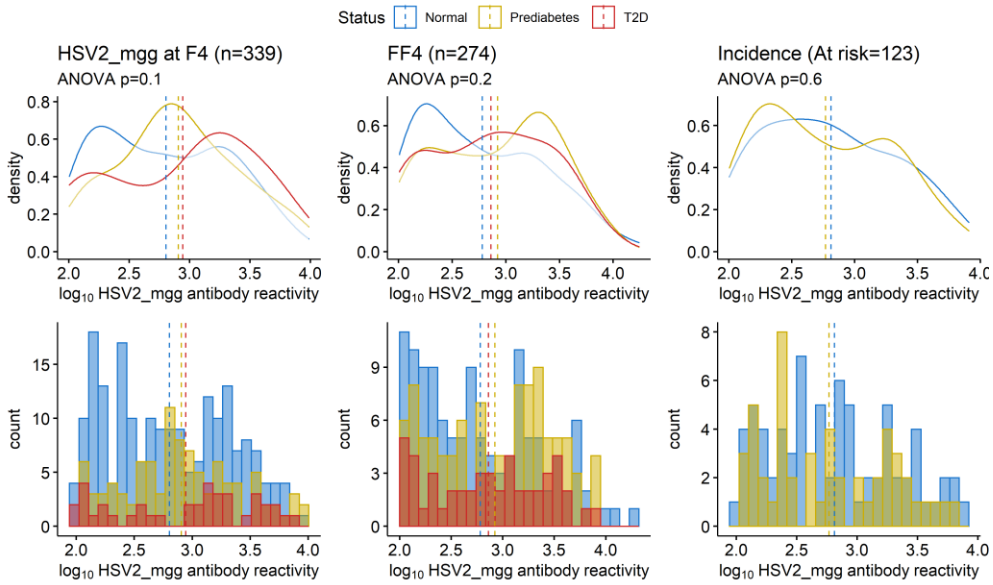
Supplementary Figure 59: Co–occurrence of viruses at F4 with groups coloured by (pre)diabetic prevalence (n=2,520, right tail cut off)



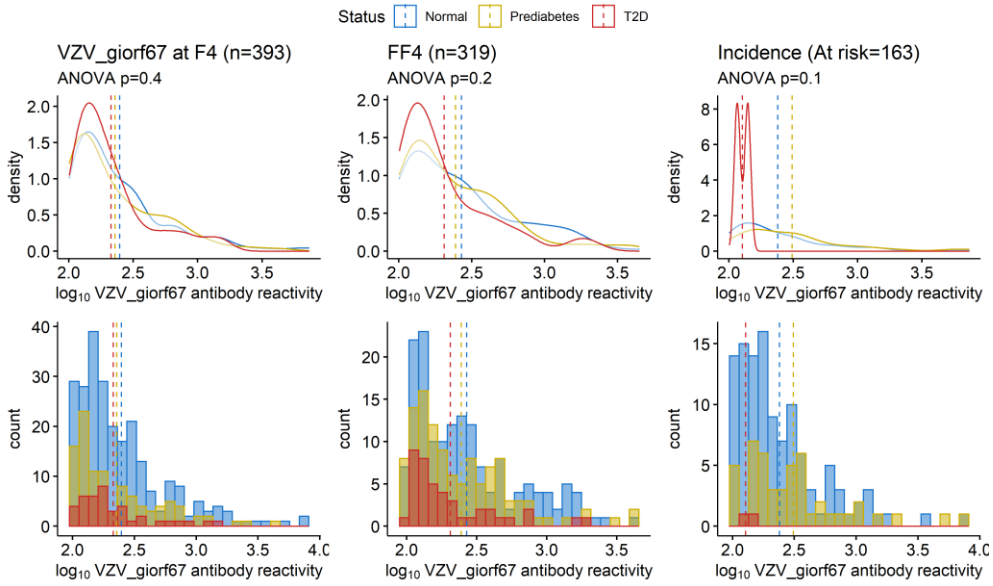
Supplementary Figure 60: Number of seropositive viruses vs. diabetic prevalence in F4 (n=2,520)



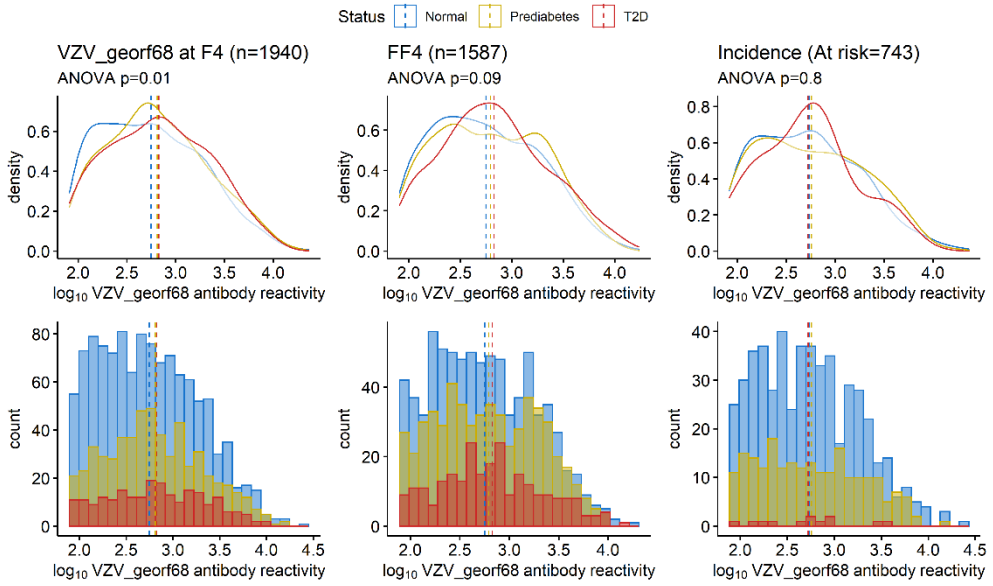
Supplementary Figure 61: HSV–1 gg reactivity in seropositives and (pre)diabetes



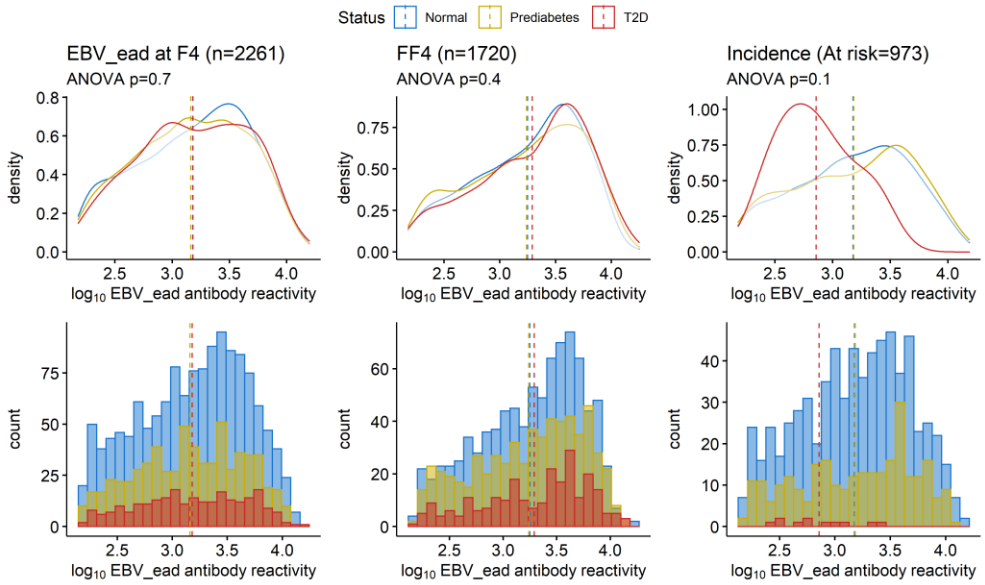
Supplementary Figure 62: HSV–2 mgg reactivity in seropositives and (pre)diabetes



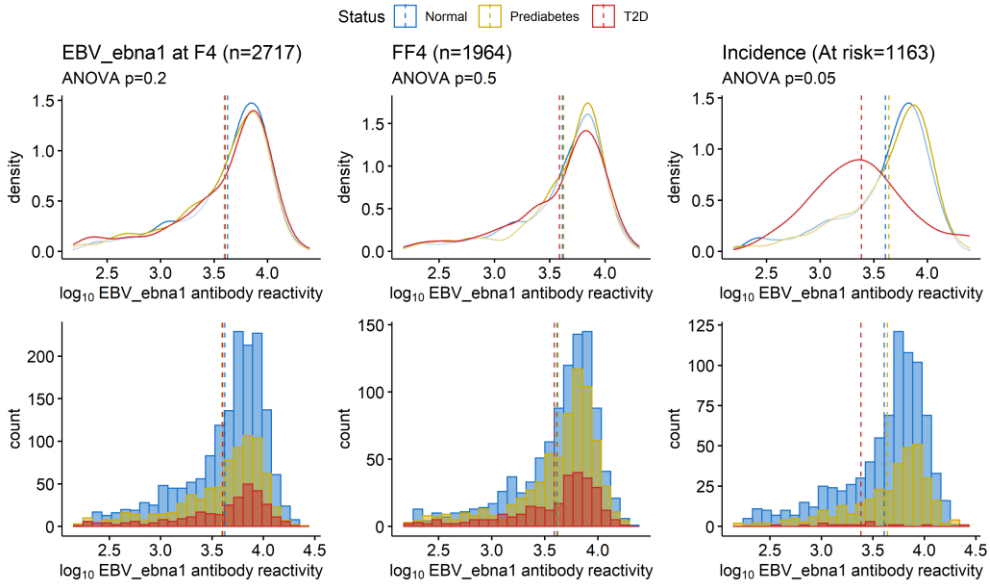
Supplementary Figure 63: VZV giorf67 reactivity in seropositives and (pre)diabetes



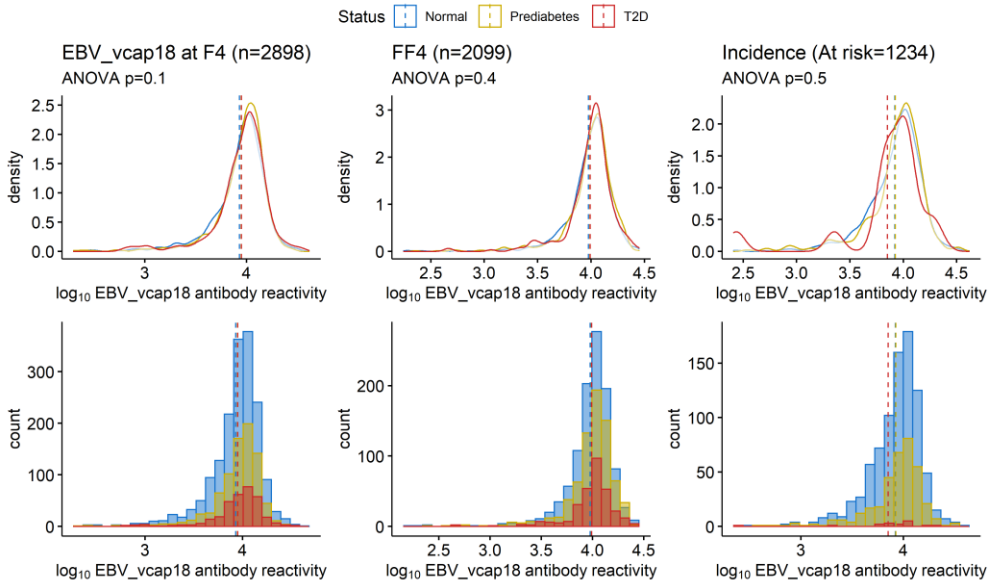
Supplementary Figure 64: VZV georf68 reactivity in seropositives and (pre)diabetes



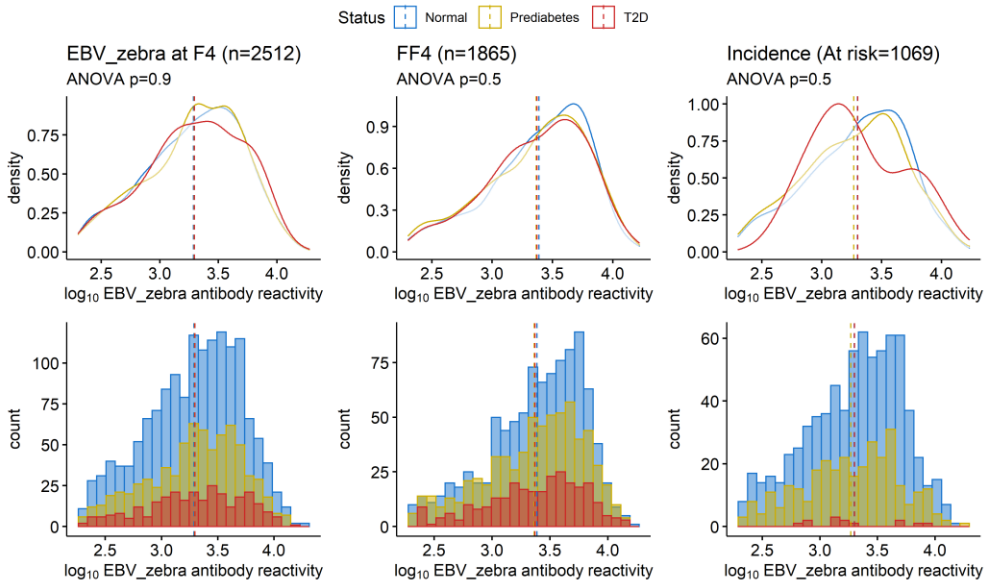
Supplementary Figure 65: EBV ead reactivity in seropositives and (pre)diabetes



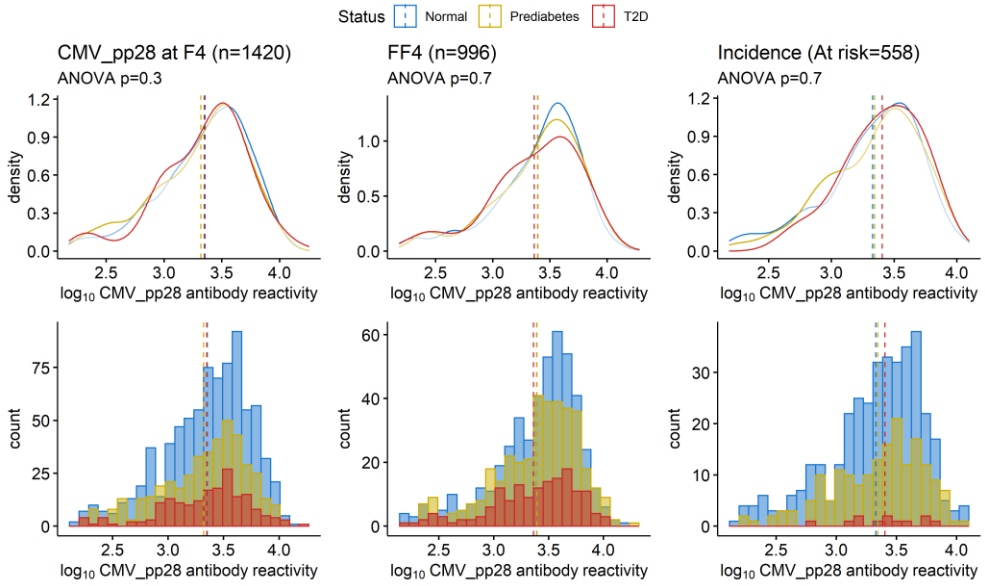
Supplementary Figure 66: EBV ebna1 reactivity in seropositives and (pre)diabetes



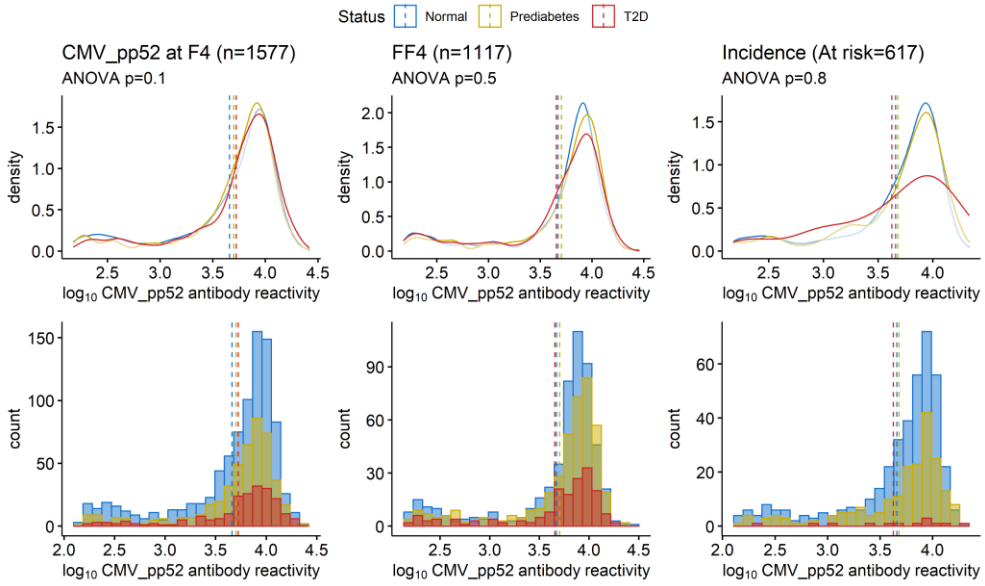
Supplementary Figure 67: EBV vcap18 reactivity in seropositives and (pre)diabetes



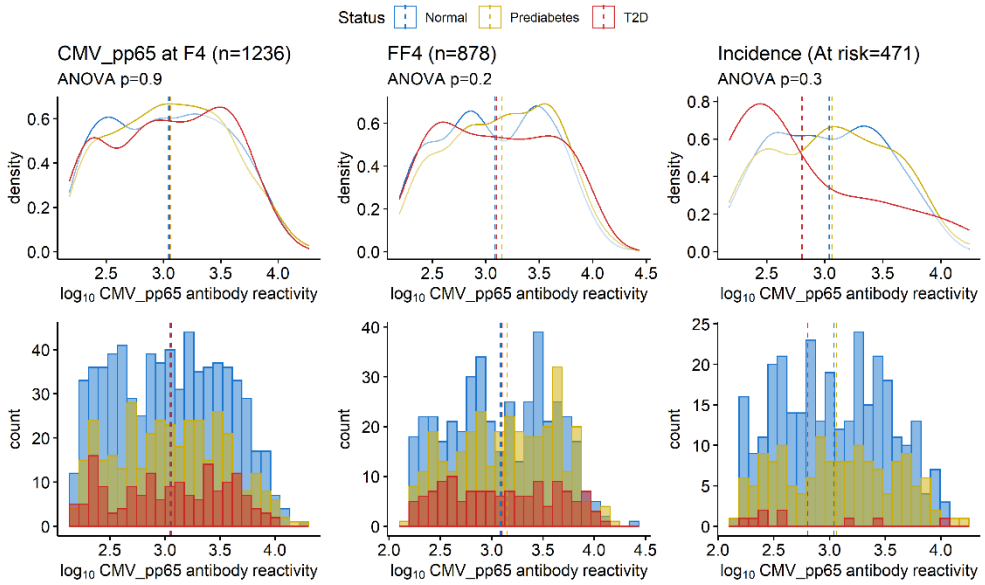
Supplementary Figure 68: EBV zebra reactivity in seropositives and (pre)diabetes



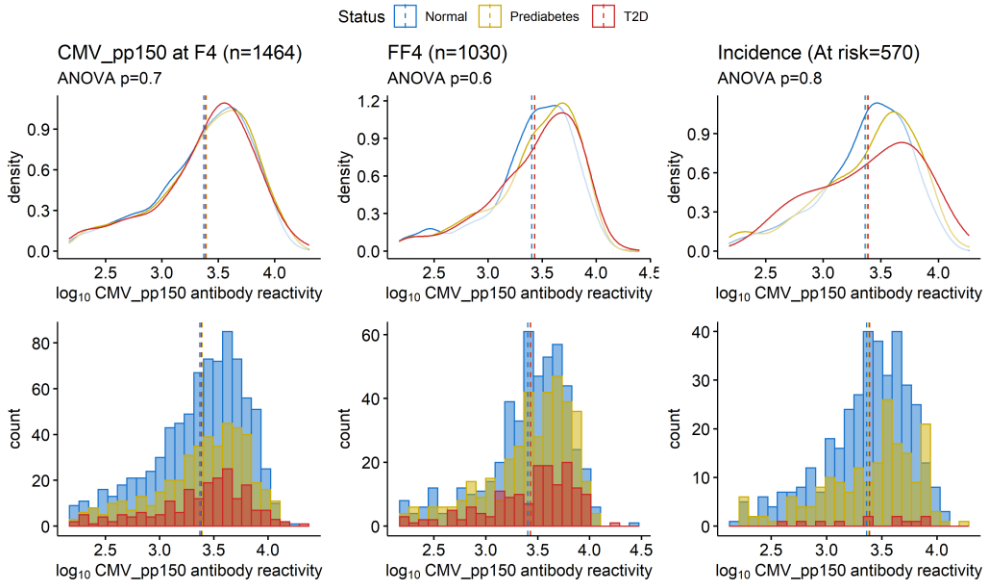
Supplementary Figure 69: CMV pp28 reactivity in seropositives and (pre)diabetes



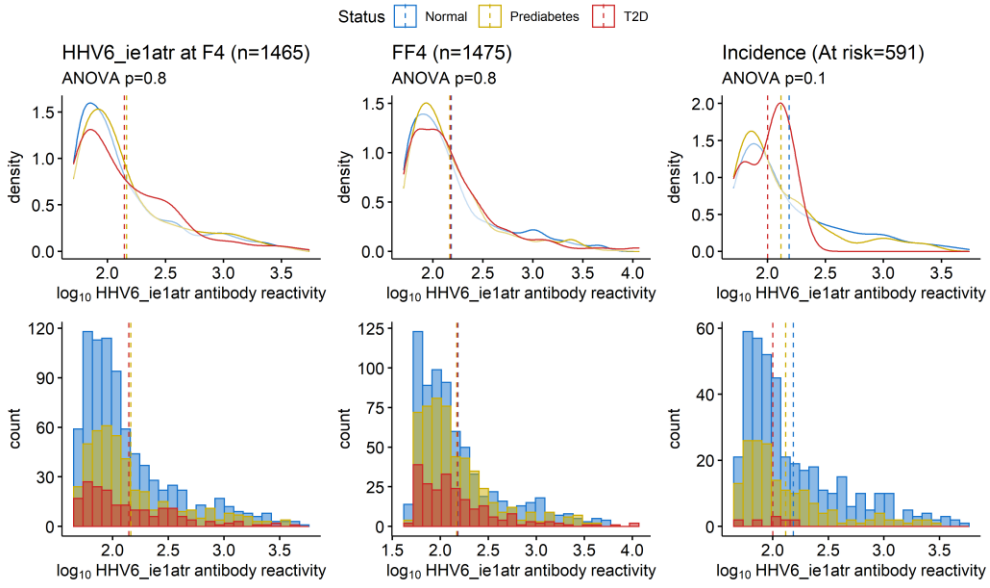
Supplementary Figure 70: CMV pp52 reactivity in seropositives and (pre)diabetes



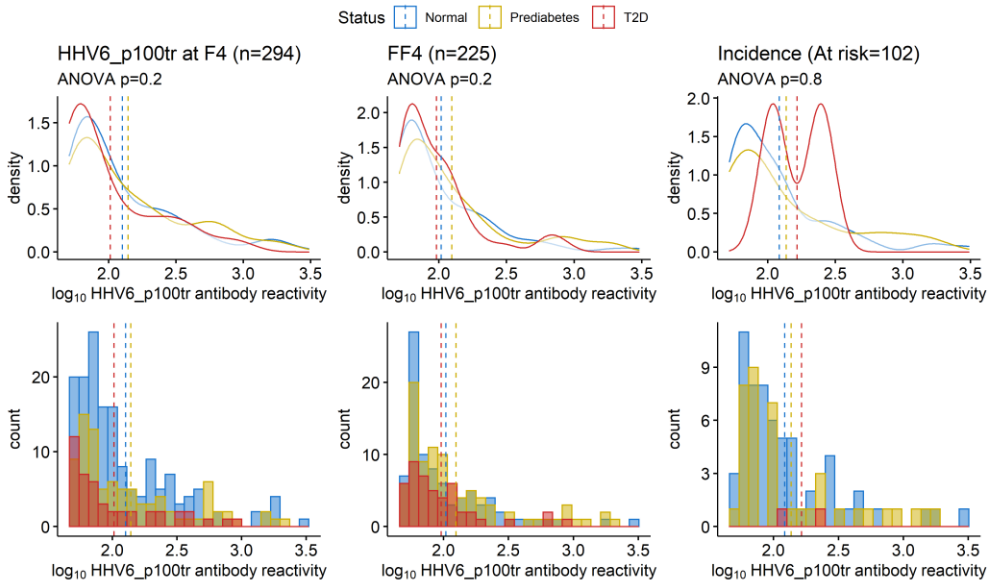
Supplementary Figure 71: CMV pp65 reactivity in seropositives and (pre)diabetes



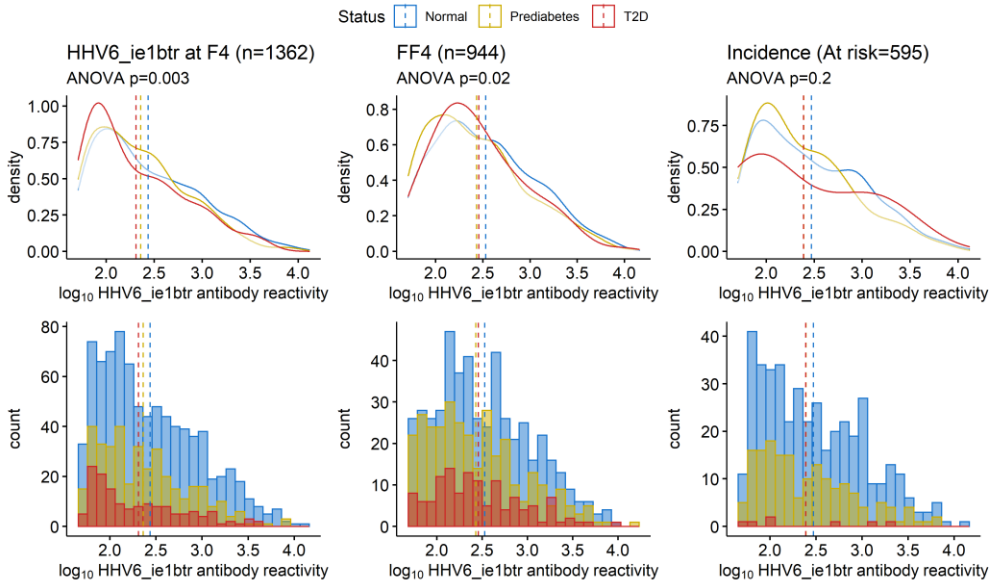
Supplementary Figure 72: CMV pp150 reactivity in seropositives and (pre)diabetes



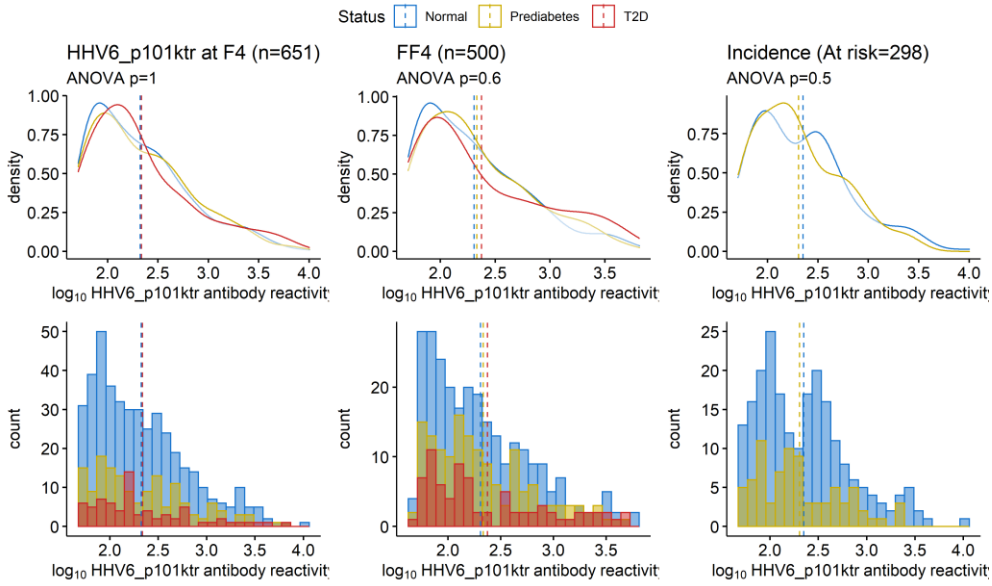
Supplementary Figure 73: HHV–6 ie1atr reactivity in seropositives and (pre)diabetes



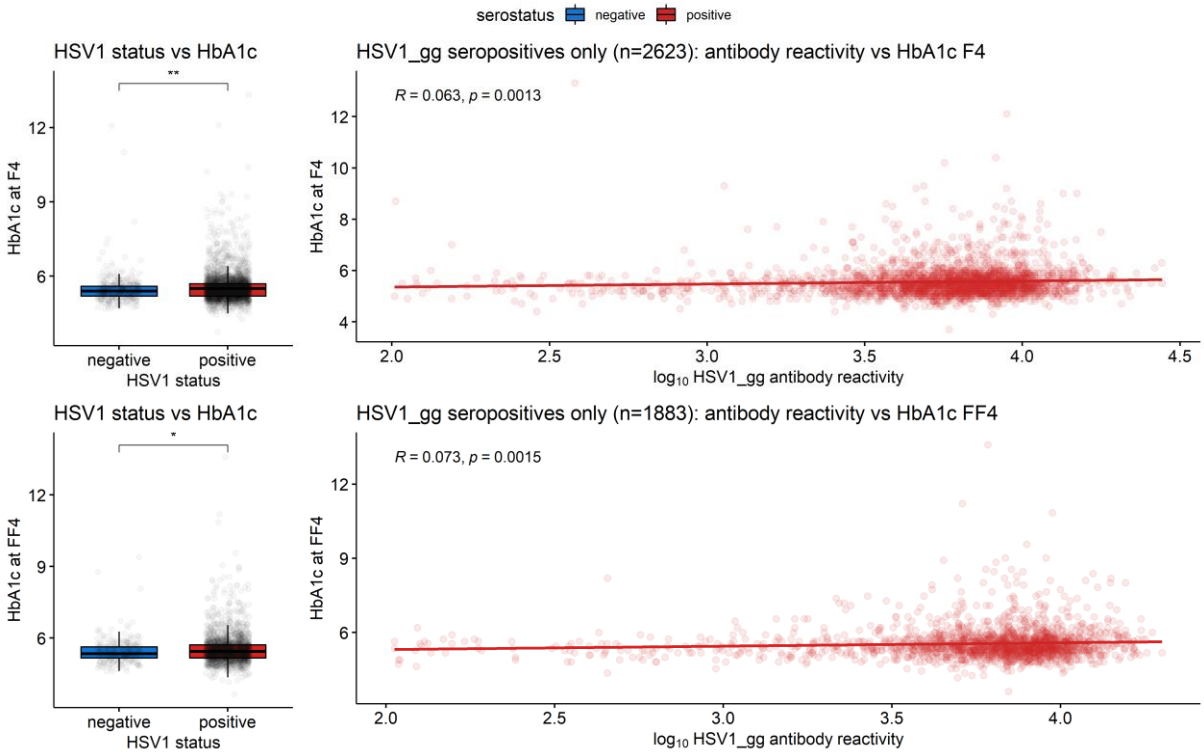
Supplementary Figure 74: HHV–6 p100tr reactivity in seropositives and (pre)diabetes



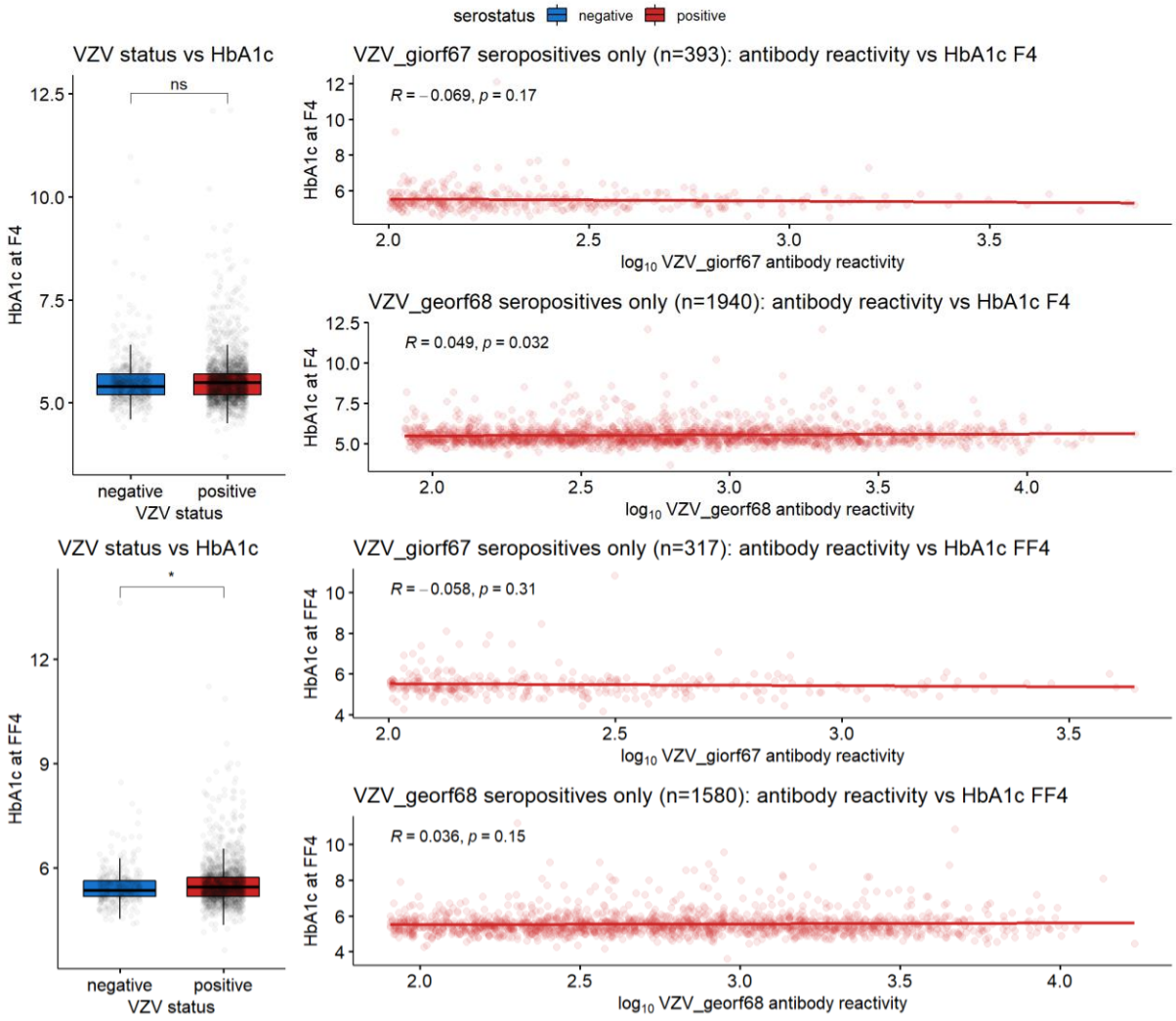
Supplementary Figure 75: HHV–6 ie1btr reactivity in seropositives and (pre)diabetes



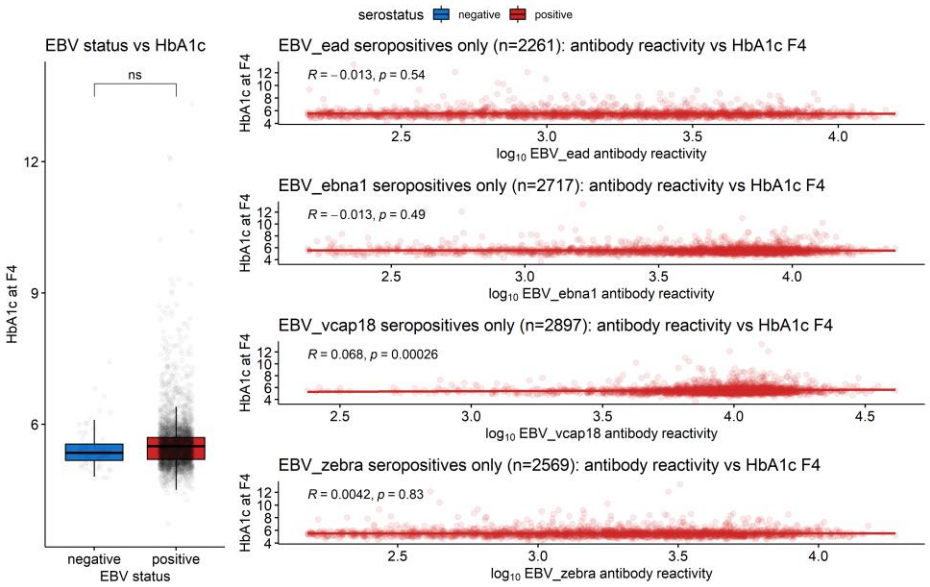
Supplementary Figure 76: HHV–6 p101ktr reactivity in seropositives and (pre)diabetes



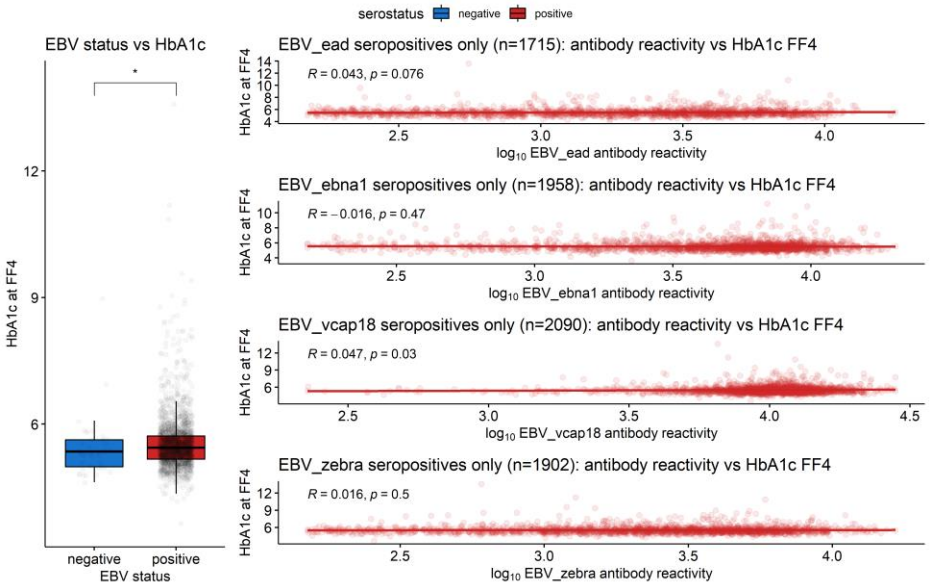
Supplementary Figure 77: HSV–1 and HbA1c at F4 and FF4



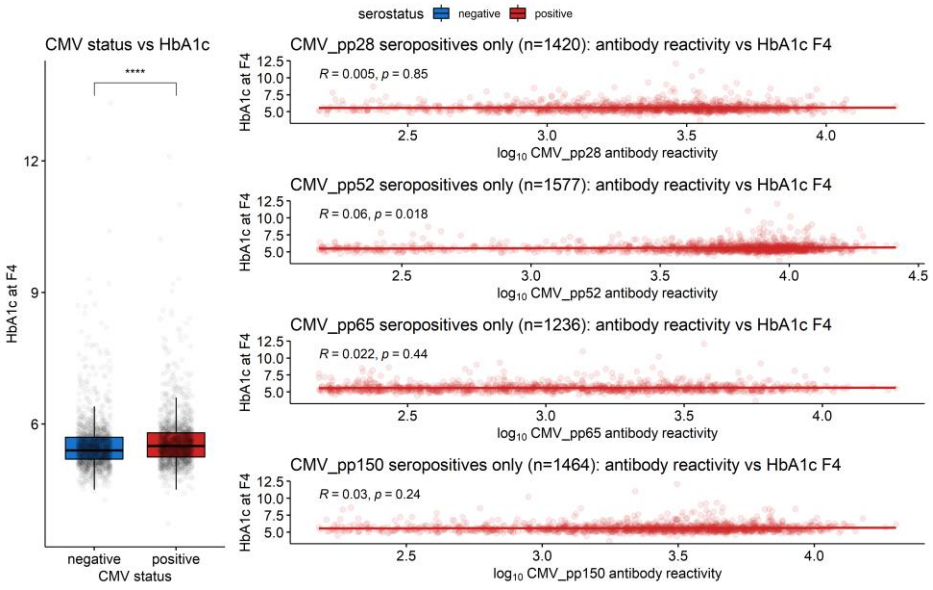
Supplementary Figure 78: VZV and HbA1c at F4 and FF4



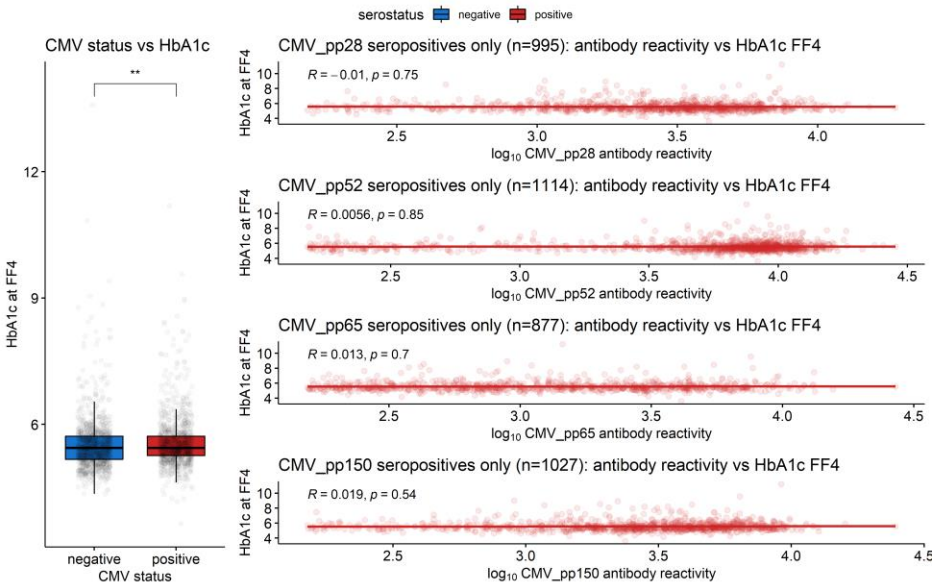
Supplementary Figure 79: EBV and HbA1c at F4



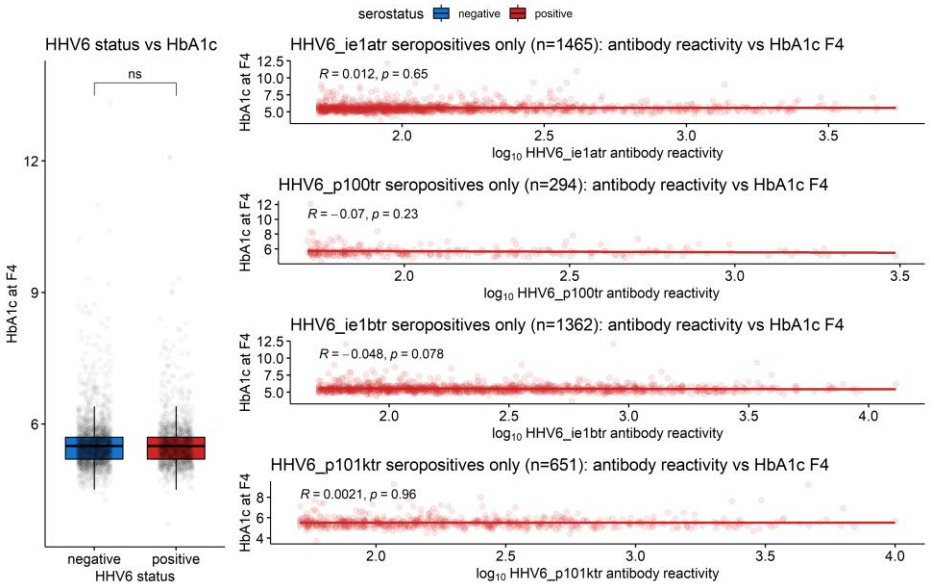
Supplementary Figure 80: EBV and HbA1c at FF4



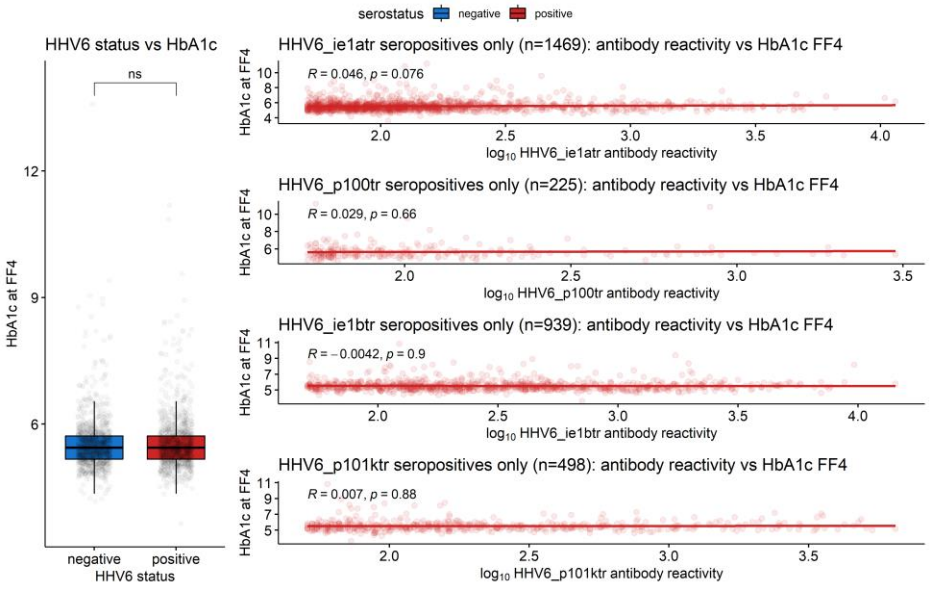
Supplementary Figure 81: CMV and HbA1c at F4



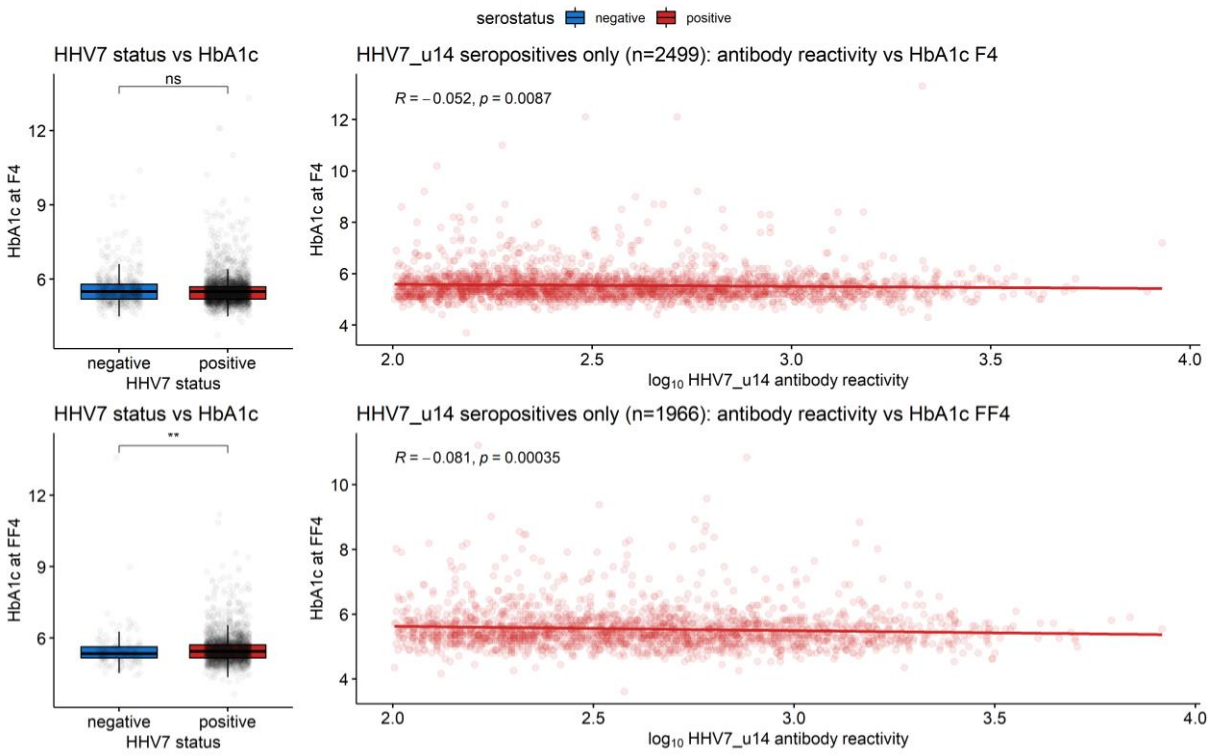
Supplementary Figure 82: CMV and HbA1c at FF4



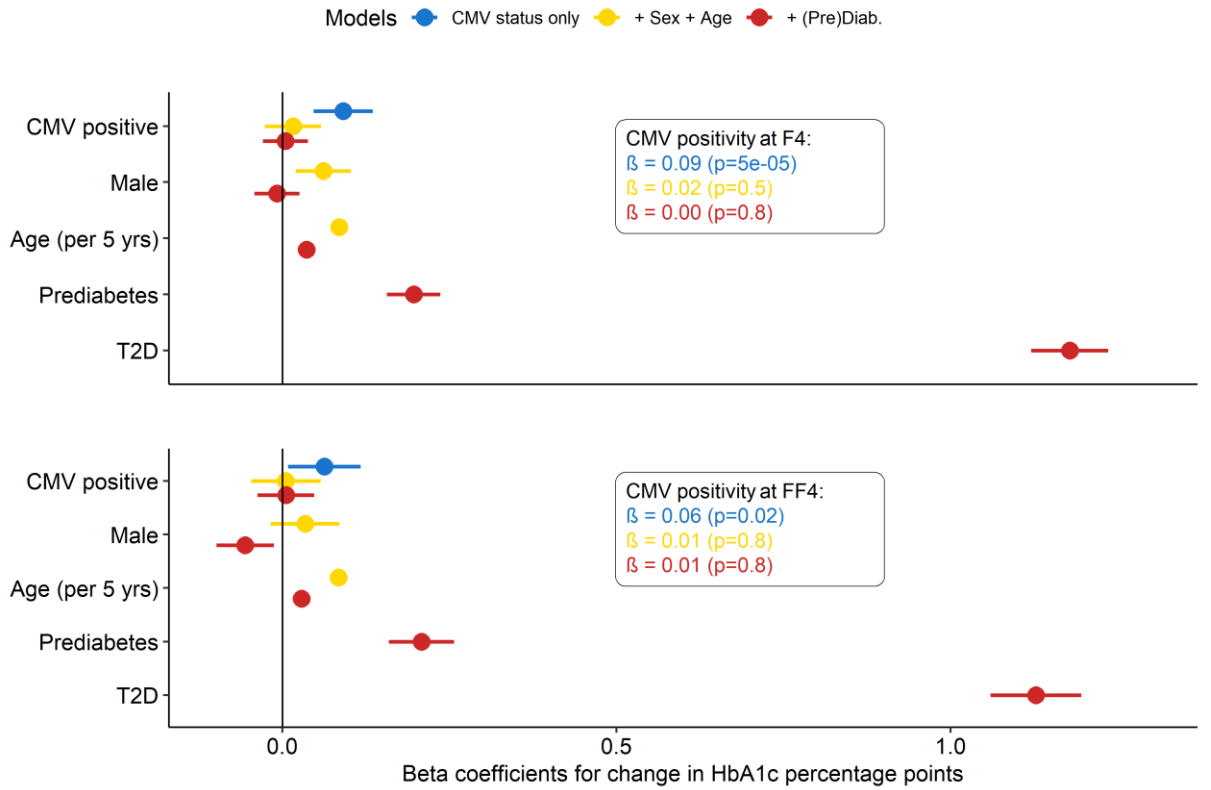
Supplementary Figure 83: HHV–6 and HbA1c at F4



Supplementary Figure 84: HHV–6 and HbA1c at FF4



Supplementary Figure 85: HHV–7 and HbA1c at F4 and FF4



Supplementary Figure 86: Beta coefficients in 3 increasingly complex linear regression models for CMV and HbA1c at F4 and FF4

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Eidesstattliche Versicherung

Wölfle, Tim

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

Characterisation of Herpesviruses in KORA Cohort and Association with Type 2 Diabetes

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Lörrach, 26.10.2021 Tim Wölfle

Ort, Datum Unterschrift Doktorandin bzw. Doktorand