Fetal inheritance of chromosomally integrated HHV-6 predisposes to preeclampsia in the mother

Francesca Gaccioli[1,2,12], Ph.D.

Susanne Lager[1,2,3,12], Ph.D.

Marcus C de Goffau[4,12], Ph.D.

Ulla Sovio[1,2], Ph.D.

Justyna Dopierala[1,2,11], Ph.D.

Sungsam Gong[1], Ph.D.

Emma Cook[1], B.Sc.

Andrew Sharkey[5,2], Ph.D.

Ashley Moffett[5,2], M.D.

Wai Kwong Lee[6], Ph.D.

Christian Delles[6], Ph.D.

Cristina Venturini[7], Ph.D.

Judith Breuer[7,8], M.D.

Julian Parkhill[4], Ph.D.

Sharon J Peacock[9,10], Ph.D.

D Stephen Charnock-Jones[1,2,13], Ph.D.

Gordon CS Smith[1,2,13,*], D.Sc.

[1]Department of Obstetrics and Gynaecology

[2]Centre for Trophoblast Research, University of Cambridge, UK

[3]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[4]Department of Veterinary Medicine, University of Cambridge, UK

[5]Department of Pathology, University of Cambridge, UK

[6]Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK

[7]Division of Infection and Immunity, University College London

[8]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

[9]Department of Medicine, University of Cambridge, Cambridge, UK

[10]London School of Hygiene and Tropical Medicine, London, UK

[11]Current address: Functional Genomics, GlaxoSmithKline Limited, Gunnels Wood Road, Stevenage, Hertfordshire, UK

[12]These authors contributed equally to this work

[13]These authors contributed equally to this work

[*]Corresponding author:

Professor Gordon C. S. Smith, Department of Obstetrics and Gynaecology, University of Cambridge, Box 223 The Rosie Hospital, Cambridge, CB2 0SW, UK. Tel: +44 (0)1223 336871, Fax: +44 (0)1223 215327, E-mail: gcss2@cam.ac.uk

Abstract

Preeclampsia (typically characterised by new onset hypertension and proteinuria in the second half of pregnancy) represents a major determinant of the global burden of disease[1,2]. Its pathophysiology involves placental dysfunction, but the mechanism is unclear. Viral infection can cause organ dysfunction but its role in placentally-related disorders of human pregnancy is unknown[3]. We addressed this using RNA-seq metagenomics[4-6] of placental samples from normal and complicated pregnancies. Here we show that human herpes virus 6 (HHV-6, A or B) RNA was detected in 6.1% of cases of preeclampsia and 2.2% of other pregnancies. Fetal genotyping demonstrated that 70% of samples with HHV-6 RNA in the placenta exhibited inherited, chromosomally integrated HHV-6 (iciHHV-6). We genotyped 467 preeclampsia cases and 3,854 controls and found an excess of iciHHV-6 in cases (odds ratio (OR) 2.8, 95% CI: 1.4 to 5.6, P=0.008). We validated this finding, comparing iciHHV-6 in a further 740 cases with controls from large-scale population studies (OR=2.5, 95% CI: 1.4 to 4.4, P=0.0013). We conclude that iciHHV-6 results in transcription of viral RNA in the human placenta and predisposes to preeclampsia.

Main

Preeclampsia and fetal growth restriction (FGR) are major causes of maternal and perinatal morbidity and mortality. We hypothesized that viral infection could be an underlying cause of the placental dysfunction which characterises these pregnancy complications. In order to identify viral sequences from both RNA viruses and actively replicating DNA viruses, we analyzed non-human reads in 279 RNA-seq datasets of placental samples from 99 cases of preeclampsia, 48 cases of FGR and 132 controls (Extended Data Figure 1). Using this discovery-based approach, the only clear viral signal was HHV-6, which was detected in 10 samples (Table 1 and Extended Data Figure 2). Targeted qPCR-based analysis of 12 viruses including HHV-6, confirmed these findings (Supplementary Information). 6.1% (6/99) of the preeclampsia cases and 2.2% (4/180) of the patients without preeclampsia were HHV-6 positive (Table 2).

Table 1. Detection of HHV-6 RNA and/or DNA in infant and parental samples.

Patient	Status	Placental HHV-6 RNA (RNA-seq)		Placental HHV-6 DNA (qPCR)	Parental HHV-6 DNA (qPCR)	
		Virus	*Reads**	*Virus*	*Maternal*	*Paternal*
Infant_1	CON	HHV-6B	331	**iciHHV-6B**	**iciHHV-6B**	negative
Infant_2	CON	HHV-6B	259	**iciHHV-6B**	negative	**iciHHV-6B**
Infant_3	FGR	HHV-6A	77	**iciHHV-6A**	**iciHHV-6A**	negative
Infant_4	CON	HHV-6B	43	**iciHHV-6B**	negative	**iciHHV-6B**
Infant_5	PE	HHV-6A	22	negative	negative	negative
Infant_6	PE	HHV-6B	9	**iciHHV-6B**	negative	**iciHHV-6B**
Infant_7	PE	HHV-6B	7	**iciHHV-6B**	**iciHHV-6B**	negative
Infant_8	PE	HHV-6A	3	**iciHHV-6A**	negative	**iciHHV-6A**
Infant_9	PE	HHV-6A	1	negative	negative	negative
Infant_10	PE	HHV-6B	1	negative	negative	negative

iciHHV-6 corresponds to a high HHV-6 DNA signal in the sample measured by qPCR, i.e. within 4 cycles of the RNase P signal. Negative indicates viral DNA not detected. Placental HHV-6 denotes HHV-6 studied in placental samples; parental HHV-6 denotes HHV-6 studied in parental samples; CON denotes a healthy pregnancy without FGR or preeclampsia (see Methods); FGR denotes fetal growth restriction; PE denotes a patient with preeclampsia; iciHHV-6 denotes inherited chromosomally integrated human herpesvirus 6; HHV-6A denotes human herpesvirus 6, variant A; HHV-6B denotes human herpesvirus 6, variant B. *Reads indicates the number of sequencing reads identified by the Kraken software as aligning to the HHV-6 genome.

Between 0.2% and 1% of humans carry a copy of HHV-6 integrated into the telomeric region of a chromosome in every cell of their body, including the germ cells. Inherited, chromosomally integrated HHV-6A and HHV-6B (iciHHV-6A and iciHHV-6B, respectively) can be transmitted from either the mother or father to the fetus in Mendelian fashion[7,8]. Hence, we investigated whether HHV-6 DNA was detected in the placental and parental DNA samples of the 10 HHV-6 positive patients identified by RNA-seq. Of these 10 placental samples, 7 were strongly positive for HHV-6 DNA, consistent with viral chromosomal integration (Table 1 and Extended Data Figure 3). In all 7

cases with a high placental DNA signal, one of the parental DNA samples was also strongly positive for the virus, consistent with inherited chromosomal integration.

We next performed deep sequencing of the HHV-6B genome in 9 parent and offspring pairs with high levels of HHV-6B DNA (these analyses included some additional HHV-6 positive samples identified by genotyping the entire POP study cohort – see below). We created "barcode" graphics of the viral genome by indicating the presence or absence of 187 informative SNPs as a black or white vertical line, respectively (Figure 1A and Extended Data Figure 4). These analyses demonstrated 100% concordance between the offspring and parent HHV-6B genomes. There were 999 informative SNPs in the HHV-6A genome. These SNPs were 100% concordant between 2 parent and offspring pairs with ciHHV-6A (not shown). As a proof of principle of chromosomal integration, we analyzed one parent and offspring pair in detail and identified the ciHHV-6B integration site at the telomeric side of 4p16.3 in both the parental and offspring genomes (Supplementary Information). This integration site has an identical nucleotide sequence to a previously described insertion site for ciHHV-6A[9]. Relative qPCR quantification and sequencing reads spanning the HHV-6 genome and the flanking human telomere demonstrated viral chromosomal integration in the placental samples analyzed.

We also compared the SNP concordance between the chromosomally integrated DNA in the parent and the RNA detected in the placenta, using a case where there was high RNA-seq coverage of the U100 viral gene (Figure 1B). There was concordance between the SNPs present in the placental RNA and the parent's DNA, demonstrating that the placental viral RNA was encoded by the inherited viral genome. This concordance could not be explained by maternal contamination as the case in question involved fetal inheritance of ciHHV-6B from the father. There were 16 offspring with iciHHV-6 on the basis of analysis of umbilical cord DNA (described below) and an associated RNA sample, and viral RNA was detected in the placenta of 11 (69%) of these (Extended Data Figure 3).

Hence, the RNA-seq and qPCR analyses led us to conclude that HHV-6 was the only clear viral signal observed in a large number of case and control placentas and that there was no evidence suggesting an association between preeclampsia or fetal growth restriction and any virus other than HHV-6. Furthermore, SNP analysis indicated that the viral genomes in the parent and the child were identical and that the viral RNA expressed in the placenta also contained the same SNPs observed in the parent. This suggests that, when present, RNA from the virus was usually, but not invariably, associated with iciHHV-6.

The initial analysis of cases and controls using RNA-seq suggested an association between HHV-6 RNA positivity in the placenta and preeclampsia, but replication was required to strengthen the evidence. As we had offspring DNA samples available from 92% of the POP study cohort and from an external case control study, we then sought to determine the association between fetal inheritance of ciHHV-6 (which also refers to placental inheritance of the virus as the placental genome comes from the zygote) and the risk of preeclampsia in a separate study group, which excluded the patients analyzed by RNA-seq. The two sources included 368 cases of preeclampsia and 3,674 pregnancies without preeclampsia. The proportions where the umbilical cord DNA was positive for iciHHV-6 were 1.9% (7/368) and 0.7% (26/3,674), respectively (P=0.022) (Table 2). When we analyzed the whole of the POP study cohort and the external case control study combined, iciHHV-6 positive samples were 2.1% (10/467) and 0.8% (30/3,854) in cases and non-cases, respectively (P=0.008). Among the women from the POP and the case control studies (n=4,321), we had a 74% power to detect a difference of 1.36% in the proportions of iciHHV-6 positive samples between cases and controls at alpha=0.05. Fetal inheritance of ciHHV-6 was associated with a three-fold increased risk of preeclampsia (odds ratio (OR) 2.8, 95% CI: 1.4 to 5.6, P=0.008, Figure 2A-B).

Table 2. Datasets described in the current work.

Dataset description	Dataset name	Study	Method	Total	PE status	HHV-6 positive (%)

RNA-seq	**RNA-seq**	POP study	Placental RNA-seq	279	**Case**	99	6	(6.1)
					Non-case	180	4	(2.2)
POP study minus RNA-seq		POP study	Cord DNA qPCR	3,568	**Case**	150	2	(1.3)
					Non-case	3,418	25	(0.7)
POP study all		POP study	Cord DNA qPCR	3,847	**Case**	249	5*	(2.0)
					Non-case	3,598	29	(0.8)
Case control		Case control	Cord DNA qPCR	474	**Case**	218	5	(2.3)
					Non-case	256	1	(0.4)
POP study minus RNA-seq + Case control	**Other**	POP study + Case control	Cord DNA qPCR	4,042	**Case**	368	7	(1.9)
					Non-case	3,674	26	(0.7)
POP study + Case control	**All**	POP study + Case control	Cord DNA qPCR	4,321	**Case**	467	10	(2.1)
					Non-case	3,854	30	(0.8)
GOPEC	**GOPEC**	GOPEC	Cord DNA qPCR	740	**Case**	740	12	(1.6)
					Non-case	n/a		
Healthy/control patients from large population studies	**Total**	Various**	Various**		**Case**	n/a		
					Non-case	61,549	403	(0.7)

Datasets “POP study minus RNA-seq”, “POP study all” and “Case control” were not analyzed as separate groups. The column “Dataset name” refers to the datasets described in Figure 2. POP study denotes the Pregnancy Outcome Prediction study; GOPEC denotes the Genetics of Preeclampsia Consortium; PE denotes preeclampsia; Case denotes a pregnancy affected by preeclampsia; Non-case denotes a pregnancy without preeclampsia; n/a denotes not applicable. *The sum of the two rows above for cases is 6+2=8 whereas this cell states 5. The difference between 8 and 5 is explained by the 3 cases of preeclampsia where the placenta was positive for HHV-6 by RNA-seq but there was no ciHHV-6. **The studies are listed in the Methods.

We performed a further replication of the analysis by genotyping cord DNA samples from 740 cases of preeclampsia recruited by the Genetics of Preeclampsia Consortium (GOPEC)[10]. As the GOPEC cohort did not include preeclamspia non-cases, the proportion with iciHHV-6 in this cohort was compared with a meta-analysis of other large-scale, population-based studies of ciHHV-6 in control patients. There was no overlap between the cases or controls used in the prior experiments and the GOPEC cases or the meta-analysis. The proportion of iciHHV-6 positive samples in the GOPEC study was 1.6% (12/740) and the summary proportion of ciHHV-6 derived from the meta-analysis of large-scale population studies was 0.7% (403/61,549) (Table 2 and Figure 2C). This analysis again demonstrated a two to three fold risk of preeclampsia associated with iciHHV-6 (OR=2.5, 95% CI: 1.4 to 4.4, P=0.0013). Although the meta-analysis indicated heterogeneity in the background rate of ciHHV-6 between the different studies, the proportion of iciHHV-6 in the GOPEC cohort fell outside

the 95% CI of all the individual studies employed. Comparable results were obtained after exclusion of the two Japanese populations (Tanaka-Taya 2014 and Miura 2018; see Supplementary Information). Moreover, the iciHHV-6 positive cases within the GOPEC study were geographically dispersed, i.e. from 5 of the 9 recruitment centres. Collectively, the analysis of the GOPEC study and the meta-analysis of population studies provided strong confirmatory evidence. Further replication studies will require large sample sizes. We estimate that a case control study drawn from a single population would require approximately 1,900 cases and 1,900 controls to identify an OR=2.5 (80% power and alpha=0.05).

Given the association between iciHHV-6 and the risk of preeclampsia in three groups, the POP study, the Cambridge case control study and the GOPEC study, we believe that the current analysis provides strong evidence for an association between placental HHV-6 and the risk of preeclampsia. Equally importantly, our study did not identify any other viral associations. Hence we conclude that a small proportion of cases of preeclampsia is likely to be due to inherited or de novo HHV-6, however, viral infection of the placenta is not a major determinant of the pathophysiology of the condition. Interestingly, the proportional increase in the risk of preeclampsia associated with iciHHV-6 was similar to the association with angina pectoris reported by a study of almost 20,000 Canadian adults[11].

We have identified a number of areas for further study. The majority of cases of preeclampsia are associated with delivery at term and only a small proportion of pregnancies with preeclampsia result in preterm delivery[12]. There is some evidence that the pathophysiology of early onset and late onset disease may differ, hence future studies could address whether the association between preterm preeclampsia and placental HHV-6 may be stronger. However, this would require very large sample sizes to ensure adequate statistical power. Identification of placental HHV-6 RNA by RNA-seq relied on some samples expressing the viral RNA at very low levels (e.g. 3 samples had fewer than 5 reads).

Importantly, RNA expression at term might not reflect levels earlier in gestation, which is relevant as a large body of evidence places the onset of the placental dysfunction associated with preeclampsia and FGR early in pregnancy. Moreover, expression levels in a particular area of the placenta, which reflects placental sampling at the time of collection, might not be representative of the whole organ. Therefore, we have viewed the presence of HHV-6 RNA in a binary fashion. The RNA-seq analysis also identified three placentas containing HHV-6 RNA in the absence of chromosomally integrated virus, and this may represent direct de-novo infection of the placenta by the virus, which could be the result of reactivation of HHV-6 in the mother. It was notable that all three were from cases of preeclampsia. However, the numbers were too small to draw reliable conclusions and two out of the three cases had only a single RNA read. Hence further studies are warranted to address the association between preeclampsia and HHV-6 infection of the placenta in the absence of chromosomally inherited virus. Analysis of cord blood IgM against HHV-6 might be a useful way to address this. Further research will also be required to determine the mechanism of association with iciHHV-6. The analysis of angina pectoris patients demonstrated that subjects with ciHHV-6 had shorter telomeres[11]. Although placental senescence might have a role in the pathophysiology of some pregnancy complications[13], there is currently no direct evidence of an association[14]. Viral infection can lead to a number of derangements of cellular function associated with preeclampsia, such as endoplasmic reticulum stress[15,16], and the potential for viral infection to adversely affect placental function has previously been reviewed[17]. Further studies may also address the timing of expression of the virus in the placenta and the risk of disease, for example by studying viral nucleic acids in the maternal circulation in cases where iciHHV-6 was inherited from the father. We only studied placental expression of the virus following birth, but the clinical consequences of iciHHV-6 may depend on viral replication earlier in pregnancy. Chromosomally integrated HHV-6 appears to be capable of viral reactivation[18,19] and we also detected multiple viral transcripts in the placentas with HHV-6 (Extended Data Figure 2). The presence of HHV-6 during extravillous trophoblast invasion of the maternal vessels could derange the cross talk between trophoblast and maternal

immune cells, leading to abnormal placentation. A recent study has demonstrated a humoral response to the proteins encoded by the genes transcribed by iciHHV-6[19].

A direct effect of viral integration on the expression of adjacent genes involved in the pathogenesis of preeclampsia is an unlikely explanation for the association as the virus inserts in the telomeres, which lack genes, and the exact site of integration is variable. It is possible there is some perturbation of telomere function and gene repression by heterochromatin, and again this warrants further study. Previous studies have demonstrated an association between HHV-6A in the endometrium in women with primary infertility[20]. Moreover, infection of endometrial cells by HHV-6A was associated with increased expression of cytokines by uterine natural killer cells[21]. However, analysis of the parental samples from the POP study and GOPEC cohorts demonstrated that the virus was inherited from the mother in 41% of cases with iciHHV-6 and from the father in 47%, hence, there was no relationship between the parent of origin of iciHHV-6 and the risk of preeclampsia. This observation makes it unlikely that the association between fetal inheritance of ciHHV-6 and preeclampsia is mediated by the presence of ciHHV-6 in the mother.

The present study may help explain, at least in part, some epidemiological characteristics of preeclampsia. Previous studies have indicated the possible existence of the "dangerous father"[22], i.e. certain groups of men are more likely to father a pregnancy complicated with preeclampsia. As ciHHV-6 is inherited in a Mendelian pattern, 50% of the offspring of men with ciHHV-6 will inherit ciHHV-6. Given the three-fold risk of preeclampsia with inherited ciHHV-6, it would be predicted that there would be an overall excess of preeclampsia in pregnancies fathered by men who have ciHHV-6. We and others have also previously shown that mothers experiencing preeclampsia have an increased risk of ischemic heart disease in later life[23,24]. Women with ciHHV-6 would be expected to be at increased risk of preeclampsia based on the present analysis and of angina pectoris based on

the Canadian study, hence, common associations with ciHHV-6 could explain some of the increased risk of later heart disease.

In conclusion, applying placental RNA-seq and metagenomics to cases of placentally-related complications of pregnancy, we identified a single clear viral signal, HHV-6 (A or B). Viral RNA was usually related to fetal inheritance of ciHHV-6 and this was associated with a two to three-fold risk of preeclampsia.

Methods

Study design

We employed samples from three studies. The first was the Pregnancy Outcome Prediction (POP) study, a prospective cohort study of unselected nulliparous women with a singleton pregnancy attending the Rosie Hospital (Cambridge, UK) between January 2008 and July 2012, as previously described[25-27]. Briefly, participants had phlebotomy and fetal biometry at 12, 20, 28 and 36 weeks of gestational age (wkGA). At the 20wkGA visit, maternal blood and paternal saliva were obtained for genotyping the parents. At the time of delivery, the placenta was systematically biopsied and a sample of umbilical cord was obtained for genotyping the offspring. Pregnancy and birth outcome data were ascertained by review of each woman's paper case record by research midwives and by record linkage to clinical electronic databases of ultrasonography (Astraia), delivery (Protos), biochemical tests (Meditech) and neonatal intensive care (Badgernet). Preeclampsia was diagnosed and classified based on the objective criteria of the 2013 American College of Obstetricians and Gynecologists guideline, as previously described[12,28].

The second study was a case control study which was also previously described[29]. In brief, women were recruited from three UK maternity hospitals: St. James University Hospital, Leeds; John Radcliffe Hospital, Oxford; and Rosie Hospital, Cambridge. Preeclampsia was defined as hypertension (>140/90mmHg) and proteinuria (>300mg/24h) presenting for the first time in the second half of pregnancy. There was no overlap in the participants recruited to the two studies and both had ethical approval from the Local Research Ethics Committees and written informed consent was obtained from all participants. Genotyping of the infant was performed using analysis of umbilical cord DNA.

The Genetics of Preeclampsia Consortium (GOPEC) recruited women affected by preeclampsia and their families from 10 UK recruitment centers in the UK (Birmingham, Cambridge, Glasgow, Leeds,

Leicester, London, Newcastle, Nottingham, Oxford, Stoke) between 2000 and 2003[10]. Samples from Cambridge (n=98) were excluded from the current analysis to avoid possible overlap with the POP and case control cohorts. Participants gave informed consent for the study, which was approved by the Trent Multicentre Research Ethics Committee. Women were eligible for the study if aged 18 or over, of white Western European ancestry (by grandparental ethnicity), with a singleton pregnancy and new onset hypertension and proteinuria in or after 20th week of pregnancy. Hypertension was defined as systolic blood pressure ≥140mm Hg and diastolic blood pressure ≥90mm Hg measured on two occasions within a 24 hour period. Proteinuria was defined as >500mg/24 hours or 2+ (1 g/liter) on dipstick testing of urine. Women with hypertension or proteinuria before 20wkGA, with essential hypertension, renal or cardiac disease, or diabetes were excluded. Fetal DNA was isolated from umbilical cord tissue and parental DNA from venous blood. As the GOPEC cohort did not include preeclamspia non-cases, large-scale population studies (n>1,000) of healthy/control patients were used as a comparison group for this cohort. The references for the studies analyzed are the following: Tanaka-Taya 2004[30]; Pellet 2012[31]; Gravel 2015[11]; Hill 2017[32]; Moustafa 2017[33]; Miura 2018[34]; Peddu & Mouammine 2019[19,35]; heterogeneity test, I^2=87%, P=5x10^{-08} (2-sided). The samples used in the current work are listed in Table 2.

RNA-sequencing (RNA-seq) analysis: POP study

Of the 4,212 women followed from their dating ultrasound scan through to delivery in the POP study, the placenta was obtained from 3,870 (92%) and 1,480 (35%) of these were collected within 30 minutes of birth and yielded samples suitable for analysis of RNA. We performed RNA-seq on 279 placentas, including 99 preeclampsia cases (8 with a fetal growth restricted [FGR] infant), 48 FGR without preeclampsia, and 132 pregnancies without either condition. FGR was defined as a customized birth weight <5th percentile[36]. Healthy controls were defined as pregnancies resulting in a live-born infant with a birth weight in the normal range (20-80th percentile) with no evidence of slowing in fetal growth on prenatal scans, and no evidence of pre-existing or acquired hypertensive

disorders, gestational diabetes or diabetes mellitus type I or type II or other obstetric complications. Total placental RNA was isolated from approximately 5mg of tissue stored in RNAlater. After lysis of the tissues in Lysing Matrix D tubes, extraction was performed using the mirVana miRNA Isolation Kit (Ambion) followed by DNase treatment (DNA-free DNA Removal Kit, Ambion). Libraries were prepared from 300-500ng of total placental RNA (RIN values ≥ 7.0) with the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat (Illumina), a method that did not employ selection of RNAs with a polyadenylated tail. Briefly, the protocol includes ribosomal RNA (rRNA) depletion, cDNA synthesis, adapters ligation to discriminate samples and libraries amplification by PCR. Libraries were then pooled and sequenced (single-end, 125bp) using a Single End V4 cluster kit and Illumina HiSeq2500 and HiSeq4000 instruments. At the end of the sequencing process reads were obtained, which are inferred nucleotide sequences corresponding to all or part of a single RNA transcript. If an RNA is highly expressed there will be more correspondent reads and vice versa. The obtained reads were trimmed (using cutadapt and Trim Galore!) and mapped to the primary chromosomal assemblies of the GRCh38.p3 version of the human reference genome using TopHat2, a splice-aware mapper built on top of Bowtie2 short-read aligner[37-39]. The initial 'unmapped' reads were filtered out to remove poor quality reads, based on the following conditions: 1) base-quality score (i.e. Phred score) < 30, 2) read-length < 50bp, 3) undetermined base (i.e. Ns) > 5bp, and 4) poly A/T > 5bp, and 5) low-complexity reads defined by the dust score > 7. In order to remove as many reads of human origin as possible, additional human reads were subtracted if they aligned to sequences present in the following databases: 1) GRCh38.p5, 2) human RefSeq, and 3) all human contigs and clone sequences from NCBI NT. The remaining reads of each sample were mapped to a custom Kraken reference database, including the default bacterial and viral genomes and few additional eukaryotic genomes to remove residual unmapped human reads. Kraken (v0.10.6)[40] was run using the metagm_run_kraken option and identified 10 placental RNA samples containing HHV-6A or HHV-6B reads (Table 1). Mapping of the viral reads against the

GCF_000845685.1 (HHV-6A) or the GCF_000846365.1 (HHV-6B) reference genomes was performed using BWA (v0.7.17-r1188)[41] and visualized using Artemis (v.16.0.0)[42].

DNA isolation and qPCR analyses

Of the 4,212 women followed from their dating ultrasound scan through to delivery in the POP study, samples were collected for DNA isolation from 4,060 (96%) mothers, 3,965 (94%) fathers and 3,869 (92%) offspring. Of the offspring samples, 3,847 were available for the current analysis, after exclusion of the following: 4 with unknown preeclampsia status, 7 miscarriages, and 11 terminations of pregnancy. Of the parental samples, 64 were analyzed in the current study. In the case control study we analyzed only umbilical cord DNA samples: 218 cases of preeclampsia and 256 non-cases. The analysis of the samples from the GOPEC cohort included 740 umbilical cord DNA samples after exclusion of 98 samples from Cambridge and 8 samples with no/low DNA. We also analyzed 22 parental DNA samples corresponding to the 12 ici-HHV6 positive cord DNAs (2 paternal samples were not available).

Placental DNA

Placental DNA was isolated with the Fast DNA Spin kit (MP Biomedical) from 25mg of villous tissue. The DNA isolation procedure has previously been described in detail[43]. Briefly, the tissue was homogenized by bead-beating twice for 40 seconds at a speed of 6.5m/sec on a FastPrep-24 (MP Biomedical). DNA concentrations were determined by Nanodrop Lite (Thermo Fisher Scientific). qPCR was performed on a QuantStudio 6 Flex system (ThermoFisher Scientific). The qPCR reactions were prepared using the TaqMan Multiplex Master Mix (ThermoFisher Scientific). The presence of the following viruses was investigated in placental samples using the primers and probes listed in Supplementary Table 1: Adenovirus, Cytomegalovirus (CMV), Epstein-Barr virus (EBV), Human Papillomavirus types 6, 11, 16 and 18 (HPV-6, HPV-11, HPV-16, HPV-18), Herpes Simplex Viruses types 1 and 2 (HSV-1 and HSV-2), Parvovirus, and Varicella Zoster Virus (VZV); detection of HHV-6 was based on a custom-made TaqMan assay targeting the U67/68 gene and designed to discriminate

HHV-6A and HHV-6B (HHV-6 U67/68 described in Supplementary Table 1). The pre-designed TaqMan RNase P Assay, ABY dye/QSY probe (ThermoFisher Scientific) was used to detect the human positive control gene RNase P (RPPH1).

Cord and parental DNA isolation

Umbilical cord tissues (approximately 100mg) were incubated with Cell Lysis Solution (Qiagen) at 65°C for 1 hour, followed by proteinase K digestion at 55°C overnight using Puregene Proteinase K (Qiagen). After incubation with RNase A Solution (Qiagen) at 37°C for 30 minutes, samples were precipitated using the Protein Precipitation Solution (Qiagen). In each sample, genomic DNA present in the supernatant was then precipitated with 100% isopropanol, washed with 70% ethanol and dissolved in IDTE pH 8.0 (10mM Tris, 0.1mM EDTA; Integrated DNA Technologies). Cord DNA samples from the case control and the GOPEC studies were prepared as previously described[10,29]. Maternal DNA was isolated from whole blood samples. Blood cells were pelleted and lysed with SE buffer (75mM NaCl, 25mM EDTA), Pronase (0.2mg/ml), and 0.9% SDS. Paternal DNA was isolated from saliva with the Oragene DNA collection kit (DNA Genotek Inc.). Genomic maternal and paternal DNA samples were purified by ethanol precipitation and re-suspended in TE buffer (10mM Tris, 1mM EDTA, pH 8.0). Parental DNAs from the GOPEC study were prepared as previously described[10]. Cord and parental DNA genotyping was performed using a multiplex qPCR approach including the custom-made TaqMan assay targeting the U67/68 gene described above. Relative quantitation of HHV-6 signals was achieved by comparison with the human RNase P gene (RPPH1), measured with the TaqMan RNase P Assay (ThermoFisher Scientific). The relative abundance of the target is expressed by the Ct (cycle threshold) value, which is inversely associated with the signal. High HHV-6 signals (i.e. low Ct, close to the signal for RNase P) suggested viral chromosomal integration (Extended Figure 1).

Placental RNA isolation and RT-qPCR analysis

Total RNA was isolated from approximately 10mg of placental villous tissues stored in RNAlater (Ambion), including 16 ciHHV-6 positive and 32 ciHHV-6 negative samples on the basis of the cord gDNA analysis. After lysis in Lysing Matrix D tubes (MP Biomedicals), RNA extraction was performed using the Rneasy Plus Mini Kit with genomic DNA removal (Qiagen) according to the manufacturer's instructions. RNA quantity and quality were assessed with the Agilent RNA 6000 Nano Kit (Agilent Technologies) on an Agilent 2100 Bioanalyzer System. Reverse transcriptase qPCR (RT-qPCR) was employed to assess the presence of viral RNA in placental samples. The RT reaction was performed using 500ng of total placental RNA from each sample and the SuperScript IV VILO Master Mix (ThermoFisher Scientific). Six reactions lacking the reverse transcriptase enzyme were included to rule out genomic DNA amplification. The multiplex qPCR reactions included a custom-designed TaqMan assay detecting both HHV-6A and HHV-6B transcripts encoded by the U100 gene[44] (HHV-6 U100, Supplementary Table 1), and the TaqMan RNase P Assay (both by ThermoFisher Scientific).

DNA-sequencing (DNA-seq) analysis

Deep sequencing of the HHV-6 genome was performed using a target enrichment method. Specifically, overlapping 120-mer RNA baits spanning the length of 7 complete HHV-6A and HHV-6B reference genomes (GenBank accession numbers: AB021506, AF157706, KC465951, KJ123690, KP257584, KT355575, NC001664) were prepared as previously described[45]. Libraries (n=36) were prepared according to the SureSelectXT Illumina paired-end sequencing library protocol, pooled and run on an Illumina MiSeq sequencing platform. Paired-end sequencing results were analyzed as follow.

Analysis of viral genome sequencing by SureSelect target enrichment

Forward and reverse reads from each sample were filtered with KneadData (v0.6.1) (http://huttenhower.sph.harvard.edu/kneaddata), a tool which performs quality control on metagenomic sequencing data and removes human and low quality reads. The following trimmomatic options were used: SLIDINGWINDOW:4:20, LEADING:30, TRAILING 30, MINLEN:75.

Filtered Fastq files were analyzed using Kraken to identify HHV-6A and HHV-6B reads, which were then mapped with BWA (v0.7.17-r1188)[41] to their respective reference genomes GCF_000845685.1 (HHV-6A) or GCF_000846365.1 (HHV-6B). These reads were assembled using Spades (v3.11.0)[46] and visualized using Artemis (v.16.0.0)[42] and samples were considered to be successfully sequenced if >99% coverage of either the HHV-6A or HHV-6B genome (~160kb) was achieved.

Metagenomic sequencing by Illumina HiSeq X Ten platform

Whole genome sequencing data obtained with the Illumina HiSeq X Ten platform from 2 HHV-6B positive placental DNA samples allowed for the comparison of the results from the two DNA-seq workflows. Libraries were sequenced (150bp, paired end) on a HiSeq X Ten (Illumina). Sample processing and library preparation for the metagenomics analysis was performed as previously described[43]. The sequencing coverage was designed to generate >30-fold coverage of the human genomic DNA in each sample. Reads generated in this experiment were analyzed with KneadData to remove most of the human reads, and forward and reverse reads from each sample were filtered with the following trimmomatic options: HEADCROP9, SLIDINGWINDOW:4:20, MINLEN: 100. Approximately 10,000 HHV-6 read pairs were identified using Kraken and assembled using Spades. Coverage of the HHV-6B genome was checked using BWA and Artemis which showed almost complete genome coverage (>99%), similar to the samples successfully sequenced by full viral genome sequencing using SureSelect target enrichment. Furthermore, complete agreement was observed between the Illumina HiSeq X Ten data and the SureSelect data obtained from the two samples analyzed with both methods.

Analysis of single nucleotide polymorphisms (SNPs)

Analysis of SNPs was performed using the HHV-6 reads obtained with the DNA-seq and RNA-seq workflows to investigate direct parental transmission. Nine parent and child pairs with HHV-6B in the placenta and in one of the parents were studied. In order to facilitate both processing and mapping visualization, 50,000 forward and reverse paired reads were subsampled from each sample successfully sequenced by SureSelect target enrichment. All reads were used in the case of one

sample with just 37,597 read pairs. All HHV-6B paired reads generated by HiSeq X Ten sequencing or by RNA-seq were used. Nucleotides that differ among the reconstructed HHV-6 genomes and in comparison to the reference genome were identified by comparing pairs of samples to the reference sequence. Many of the SNPs were uninformative, i.e. all 18 samples shared the identical polymorphism when compared to the reference genome. However, 187 SNPs, varying among the 18 samples analyzed, were identified. This iterative process was performed in a blinded fashion in respect to parent and child pair information. Repeat regions R0, R1, R2 and R3 were excluded from this analysis as mapping and/or sequencing errors occur at high frequency at these locations. A similar analysis was performed for the 2 HHV-6A positive parent and child pairs, identifying 999 informative SNPs. Therefore, single nucleotide polymorphism (SNP) concordance between parental and offspring sequences was assessed on the basis of 999 and 187 informative SNPs in the HHV-6A and HHV-6B genomes, respectively.

Analysis of viral integration sites

The HHV-6 genome is bounded by two identical long repeat regions which are in turn flanked by variable length telomere-like repeat regions (T1 and T2). The T2 region, which consists of a perfect array of telomere-like repeats (TTAGGG)n[47], is essential for viral integration[48]. Due to the relatively short length of the reads generated in this study and the presence these repeats in the regions of interest, it is not possible to use an assembly-based approach, i.e. assembling a contig from the viral sequence and extending into the flanking human sequence. Therefore, an approach was developed to target human reads close to the site of viral integration and possibly a small part of the flanking repetitive region of the HHV-6B genome. The raw reads were mapped to the viral reference genome and read pairs were analyzed when only one read mapped to the viral genome. The non-viral reads of these discordant read pairs were subsequently searched using the NCBI online BLAST interface (blastn, database: Others, http://blast.ncbi.nlm.nih.gov)[49]. Non-viral reads that mapped to the same region within the human genome were grouped, together with their corresponding viral-mapped reads and contigs assembled.

Statistical analysis

Associations were quantified using odds ratios and 95% confidence intervals (CI). The latter were calculated using the recommended quasi-exact Baptista-Pike mid-p method, which performs well across a wide range of sample sizes and proportions[50]. P values (2-sided) from 2x2 tables were calculated using the Fisher-Boschloo unconditional exact test[51], although all analyses were repeated using the Fisher's exact test which is more widely recognized. Metaprop command[52] in Stata was used to perform random effects meta-analysis of proportions and to estimate heterogeneity. Analyses were performed using Stata 15.1 and R 3.2.5.

Data availability

The sequencing data have been deposited in the European Genome-phenome Archive (EGA) with the following accession numbers: EGAD00001003457 (RNA-Seq controls), EGAD00001003507 (RNA-Seq FGR cases), EGAD00001003508 (RNA-Seq PE cases), EGAD00001004197 (metagenomics) and EGAD00001004592 (SureSelect). Custom scripts for the bioinformatics analyses are available at https://github.com/sung/HHV6-Nat-Micr-2020.

Contributions

GCSS had the original idea. GCSS, DSC-J, JB, JP and SJP supervised the study. GCSS, DSC-J, JB, JP, SJP, FG, SL and MCdG designed experiments. FG and JD performed RNA-seq experiments. FG and EC performed cord and parental DNA genotyping experiments. FG analyzed placental RNA samples by RT-qPCR. SL performed qPCRs on placental DNA samples and DNA-seq experiments. MCdG, SG and CV analyzed RNA-seq and DNA-seq data. US performed statistical analyses. AM and AS provided samples and clinical information for the case control study. CD and WKL provided samples and clinical information for the GOPEC study. All authors reviewed the paper prior to submission.

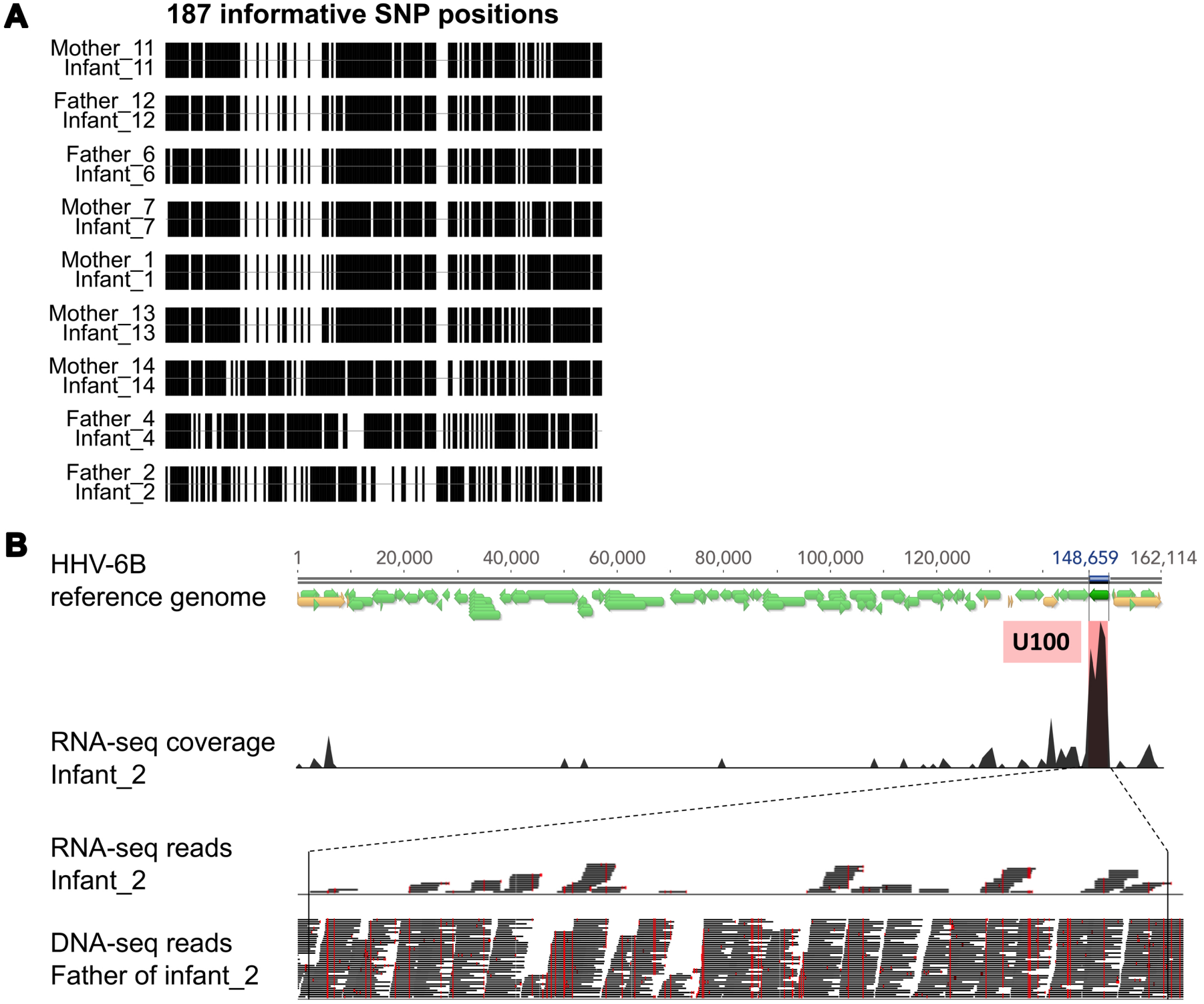
Competing interests

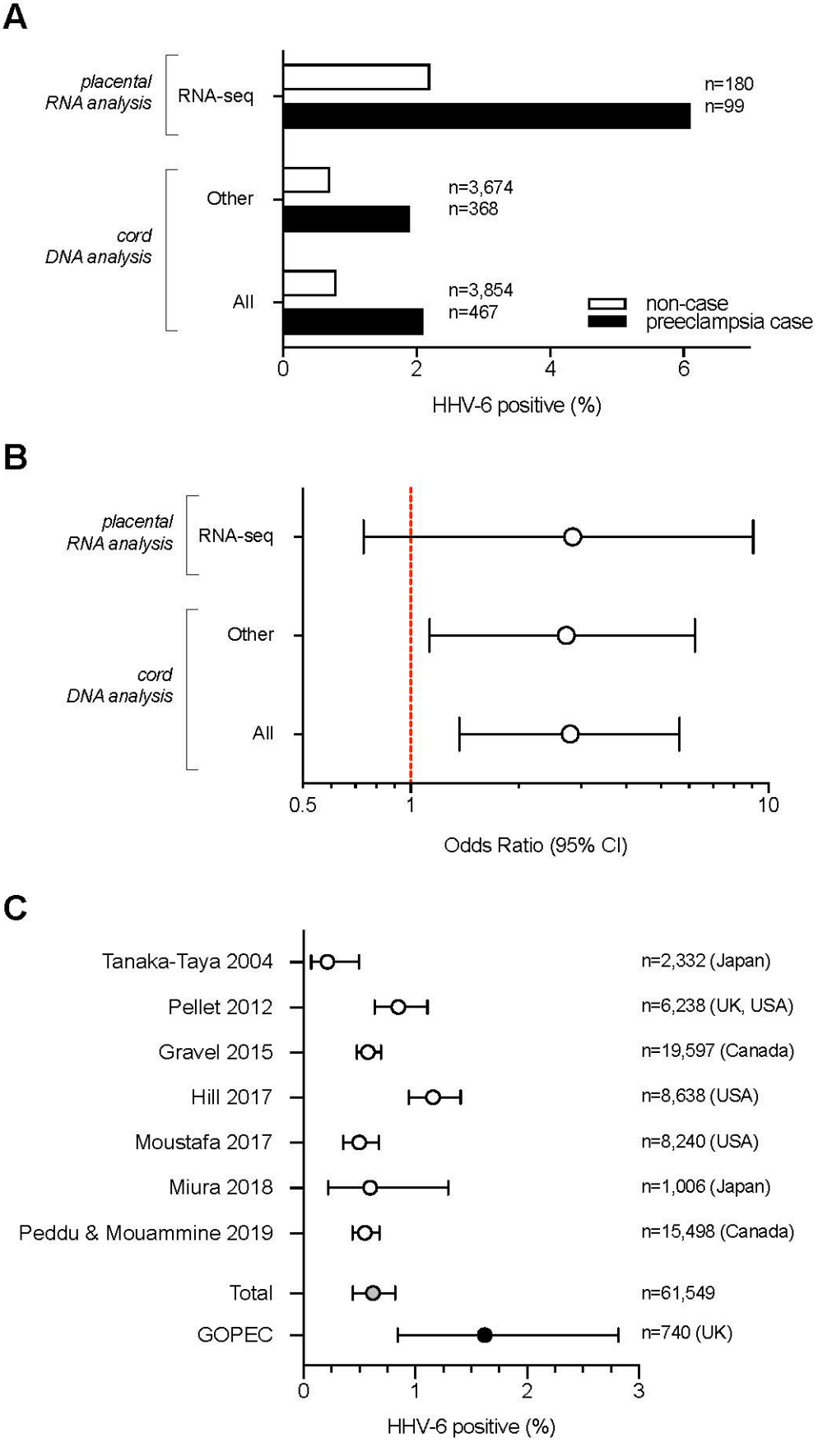
SL, MCdG, JD, SJP, JP, DSC-J, and GCSS report grants from Medical Research Council (UK); FG, US, SG, EC, DSC-J, and GCSS report grants from National Institute for Health Research (UK); US reports grants from Stillbirth & Neonatal Death Society (Sands); JD reports being an employee of GlaxoSmithKline and AS reports being an employee of Robinson College (Cambridge, UK); JP reports grants from Wellcome Trust, grants from Pfizer, personal fees from Next Gen Diagnostics Llc, outside the submitted work; SJP reports personal fees from Specific, personal fees from Next Gen Diagnostics, outside the submitted work; DSC-J reports grants from GlaxoSmithKline Research and Development Limited, outside the submitted work; GCSS reports grants and personal fees from GlaxoSmithKline Research and Development Limited, personal fees and non-financial support from Roche Diagnostics Ltd, outside the submitted work; DSC-J and GCSS report grants from Sera Prognostics Inc, non-financial support from Illumina inc, outside the submitted work. AM, WKL, CD, CV and JB have nothing to disclose.

Correspondence and requests for materials should be addressed to GCSS, AM or CD.

Acknowledgments

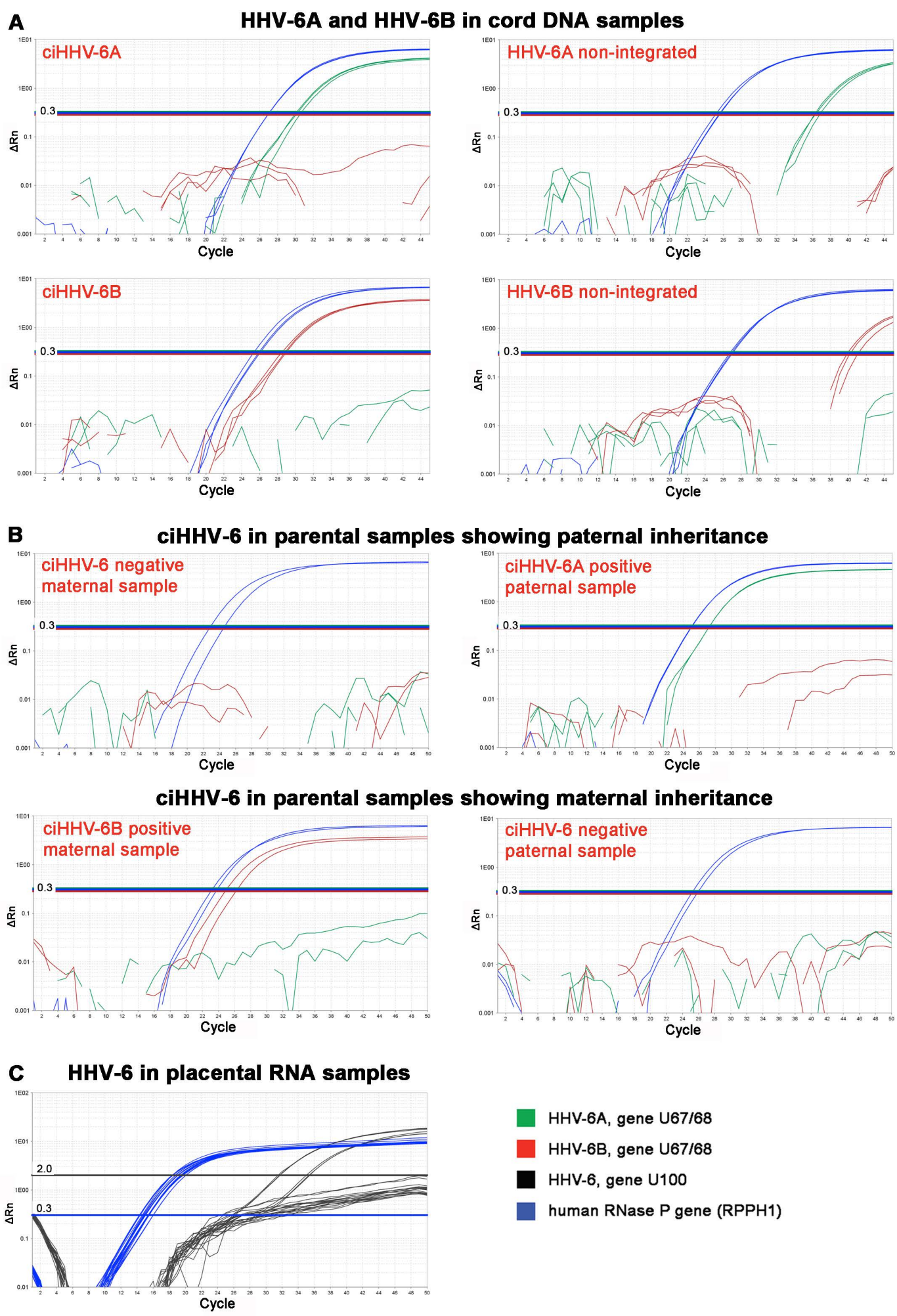
We are grateful to: the participants of the POP, case control and GOPEC studies; Leah Bibby, Samudra Ranawaka, Katrina Holmes, Josephine Gill, Chien Lin Soh, and Ryan Millar for technical assistance; the UCL Pathogen Genomics Unit; the Sanger Institute core sequencing facility and Pathogen Informatics team. This work was supported by the Women's Health theme of the NIHR Cambridge Biomedical Research Centre, the Medical Research Council (MR/K021133/1 to GCSS, DSCJ, JP & SP and G1100221 to GCSS and DSC-J) and the UCL/UCLH NIHR Biomedical Research Centre (to JB and CV).

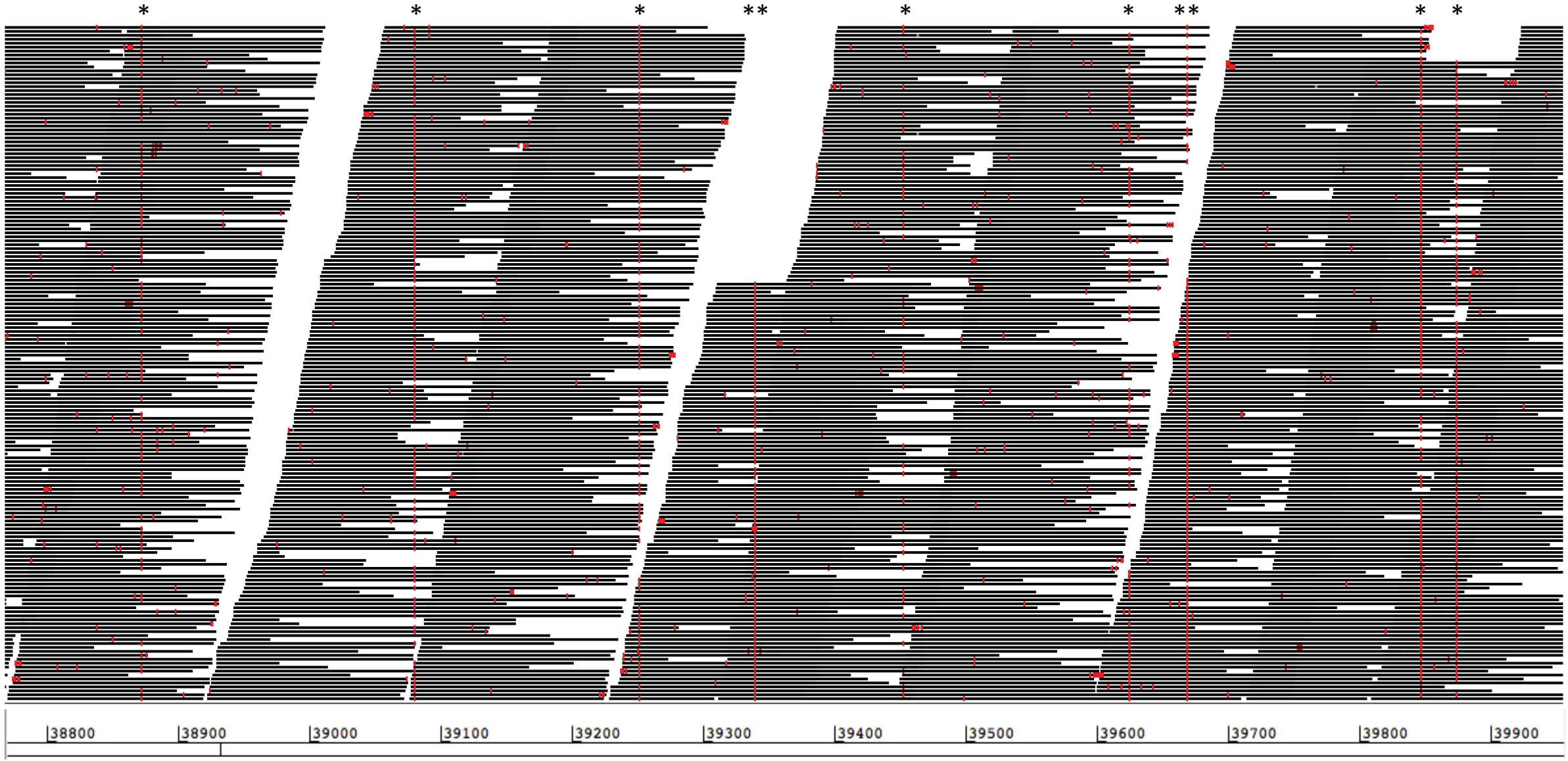

A
187 informative SNP positions
Mother_11
Infant_11
Father_12
Infant_12
Father_6
Infant_6
Mother_7
Infant_7
Mother_1
Infant_1
Mother_13
Infant_13
Mother_14
Infant_14
Father_4
Infant_4
Father_2
Infant_2
B
HHV-6B
reference genome
1
20,000
40,000
60,000
80,000
100,000
120,000
148,659
162,114
U100
RNA-seq coverage
Infant_2
RNA-seq reads
Infant_2
DNA-seq reads
Father of infant_2


A
placental RNA analysis
RNA-seq
n=180
n=99
cord DNA analysis
Other
n=3,674
n=368
All
n=3,854
n=467
non-case
preeclampsia case
0
2
4
6
HHV-6 positive (%)
B
placental RNA analysis
RNA-seq
cord DNA analysis
Other
All
0.5
1
10
Odds Ratio (95% CI)
C
Tanaka-Taya 2004
n=2,332 (Japan)
Pellet 2012
n=6,238 (UK, USA)
Gravel 2015
n=19,597 (Canada)
Hill 2017
n=8,638 (USA)
Moustafa 2017
n=8,240 (USA)
Miura 2018
n=1,006 (Japan)
Peddu & Mouammine 2019
n=15,498 (Canada)
Total
n=61,549
GOPEC
n=740 (UK)
0
1
2
3
HHV-6 positive (%)

Characteristic	RNA-seq		cord DNA, excluding RNA-seq	
	non-case (n=180)	preeclampsia (n=99)	non-case (n=3,418)	preeclampsia (n=150)
Maternal characteristics				
Age, years	31 (28 to 34)	29 (25 to 33)	30 (27 to 33)	29 (26 to 33)
Age stopped FTE (years)	21 (18 to 22)	20 (17 to 21)	21 (18 to 23)	19 (18 to 23)
Missing	4 (2%)	7 (7%)	100 (3%)	4 (3%)
Height, cm	165 (160 to 171)	165 (161 to 168)	165 (161 to 170)	164 (160 to 168)
BMI, kg/m^2	25 (23 to 29)	28 (24 to 34)	24 (22 to 27)	27 (24 to 31)
Missing	0 (0%)	0 (0%)	1 (<1%)	0 (0%)
Smoker at booking	26 (14%)	14 (14%)	430 (13%)	22 (15%)
Missing	0 (0%)	0 (0%)	4 (<1%)	0 (0%)
Any alcohol consumption	9 (5%)	4 (4%)	153 (5%)	3 (2%)
Missing	0 (0%)	0 (0%)	2 (<1%)	0 (0%)
White ethnicity	167 (93%)	94 (95%)	3,163 (93%)	144 (96%)
Missing	3 (2%)	1 (1%)	58 (2%)	1 (1%)
Married	108 (60%)	59 (60%)	2,355 (69%)	102 (68%)
Diabetes				
Type 1 or type 2 DM	0 (0%)	1 (1%)	9 (<1%)	4 (3%)
Gestational DM	1 (1%)	6 (6%)	158 (5%)	7 (5%)
Missing	0 (0%)	0 (0%)	6 (<1%)	0 (0%)
Birth outcomes				
Birth weight, g	3,383 (2,900 to 3,650)	3,480 (3,050 to 3,800)	3,430 (3,120 to 3,745)	3,293 (2,870 to 3,670)
Missing	0 (0%)	0 (0%)	1 (<1%)	0 (0%)
Gestational age, weeks	40.3 (39.3 to 41.1)	39.9 (38.3 to 40.7)	40.3 (39.3 to 41.1)	39.4 (38.0 to 40.7)
Female fetal sex	92 (51%)	39 (39%)	1,704 (50%)	74 (49%)
Missing	0 (0%)	0 (0%)	1 (<1%)	0 (0%)
Induction of labour	41 (23%)	56 (57%)	1,023 (30%)	92 (61%)
Mode of delivery				
Spontaneous vaginal	82 (46%)	22 (22%)	1,716 (50%)	50 (33%)
Assisted vaginal	39 (22%)	32 (32%)	809 (24%)	29 (19%)
Intrapartum caesarean	36 (20%)	31 (31%)	563 (16%)	43 (29%)
Pre-labour caesarean	23 (13%)	14 (14%)	321 (9%)	28 (19%)
Missing	0 (0%)	0 (0%)	9 (<1%)	0 (0%)

Patient	Status	Virus	Genes (number of reads)	Comments
Infant_1	CON	HHV-6B	*DR_L1/DR_L6* (7), U3 (2), U8 (4), U10 (2), U11 (1), U12 (4), U15 (2), U17 (12), U18 (5), U19 (2), U20 (2), U21 (8), U23 (5), U27 (6), U28 (8), U30 (10), U31 (18), U32 (1), U33 (1), U34 (1), U36 (3), U37 (2), U39 (9), U40 (13), U41 (2), U43 (4), U45 (8), U47 (10), U48 (5), U49 (3), U54 (12), U55 (33), U56 (1), U57 (9), U58 (3), U59 (4), U65 (4), U66 (2), U70 (2), U71 (2), U73 (1), U74 (2), U76 (1), U77 (2), Between B6 and B7 (10), U79 (6), U82 (1), U84 (8), U85 (2), U86 (16), *R2A* (10), U90 (1), U91 (3), U94 (5), U95 (22), U100 (3), B9 (1), *DR_R1/DR_R6* (5)	Reads across the whole genome
Infant_2	CON	HHV-6B	*DR_L1/DR_L6* (10), U31 (2), U34 (2), U48 (2), U72 (2), U75 (2), U77 (1), Between B6 and B7 (10), U79 (3), U86 (10), *R2A* (1), U90 (3), U91 (1), *R3* (9), U94 (7), U95 (7), U100 (81), *DR_R1/DR_R6* (12)	Reads across the whole genome, but mainly towards the DR_R termini; large U100 signal
Infant_3	FGR	HHV-6A	between U7 and U11 (2), U12 (1), U19 (1), U21 (3), U22-U23 (3), U28 (1), U30 (4), U31 (2), U36 (4), U38 (2), U39 (1), U40 (3), U42 (1), U44 (1), U48 (1), U49 (3), U54 (2), U57 (1), U58 (2), U70 (1), U75 (2), U77 (2), U78 (3), U82 (3), U84 (1), U86 (1), between U86 and U90 (1), U90 (3), U94 (1), *DR_R1/DR_R6* (3)	Reads across the whole genome
Infant_4	CON	HHV-6B	U58 (1), U73 (3), U76 (1), U79 (8), U86 (4), U90 (3), *R3* (2), U94 (2), U95 (1), U100 (4), *DR_R1/DR_R6* right (4)	Reads towards the DR_R termini
Infant_5	PE	HHV-6A	*DR_L1/DR_L6* (8), *DR_R1/DR_R6* (9)	Reads only from repetitive regions
Infant_6	PE	HHV-6B	U17 (2), U28 (3), U31 (1), U76 (3)	
Infant_7	PE	HHV-6B	U50 (1), U57 (1), U90 (1), *R3* (2), U95 (1)	Few reads towards the DR_R termini
Infant_8	PE	HHV-6A	U60 (1), U94 (2)	Few reads towards the DR_R termini
Infant_9	PE	HHV-6A	U78 (1)	
Infant_10	PE	HHV-6B	U94 (1)	


A
HHV-6A and HHV-6B in cord DNA samples
ciHHV-6A
HHV-6A non-integrated
ciHHV-6B
HHV-6B non-integrated
0.3
ΔRn
Cycle
B
ciHHV-6 in parental samples showing paternal inheritance
ciHHV-6 negative maternal sample
ciHHV-6A positive paternal sample
ciHHV-6 in parental samples showing maternal inheritance
ciHHV-6B positive maternal sample
ciHHV-6 negative paternal sample
C
HHV-6 in placental RNA samples
2.0
HHV-6A, gene U67/68
HHV-6B, gene U67/68
HHV-6, gene U100
human RNase P gene (RPPH1)


38800
38900
39000
39100
39200
39300
39400
39500
39600
39700
39800
39900

Supplementary Information

Supplement to: Gaccioli F, Lager S, de Goffau MC, et al. Fetal inheritance of chromosomally integrated HHV-6 predisposes to preeclampsia in the mother.

Supplementary Results

Detection of viral DNAs in placental samples

We analyzed placentas collected during the POP study from the following pregnancy complications: FGR customized birth weight <5th percentile[1], n=100), preeclampsia (diagnosed according to the 2013 ACOG Guidelines[2], n=100), and pre-term deliveries (<37wkGA, n=100). Pregnancies with FGR and preeclampsia were matched to healthy controls (n=198; two controls were used twice) based on the following criteria: mode of delivery (absolute match), maternal BMI, maternal age, gestational age, sample collection time, maternal smoking, and fetal sex. All matched cases and controls were term deliveries (≥37wkGA). In total, the study cohort included 498 unique pregnancies and the clinical characteristics of the patients are presented in Supplementary Table 2. No placental DNA sample was positive for Adenovirus, CMV, EBV, HPV-6, HPV-11, HPV-16, HPV-18, HSV-1, or HSV-2. Two placentas were positive for Parvovirus (one FGR and one preeclampsia), one placenta was positive for VZV (one pre-term), and HHV-6 was detected in 9 placental samples.

HHV-6 genomic integration: proof of principle

Contigs assembled from discordant read pairs were analyzed using BLAST and they typically resulted in contigs that were completely human. The reason for this is that most discordant read pairs have one read which is mapped erroneously to the HHV-6B genome due to the very high degree of similarity between some of the repetitive regions of the HHV-6B genome and the various repetitive parts of the human genome. However, in one parent and child pair, the ciHHV-6B integration site was identified at the telomeric side of 4p16.3 (https://www.ncbi.nlm.nih.gov/nucleotide/Z95704.1). In both samples, discordant read pairs that mapped to the 4p16.3 region either perfectly mapped to the 4p16.3 region, extending into the telomeric repetitive sequence of chromosome 4p16.3, or there was a gap close to the telomeric repetitive sequence. The sequence on either side of this gap was identical to the reference human genome. This integration site is identical to a previously described insertion site for ciHHV-6A (https://www.ncbi.nlm.nih.gov/nucleotide/KF366418.1): GATCCTTCCTCTTTGCAGCC-GAGAATAATGAGGGTTGGGGTTAGGG[3].

Analysis of the association in women of non-European ethnicity

The POP study cohort was 93% white European (Table S1). Only 9 of the cases of preeclampsia occurred in women of non-European ancestry. Before exclusion, the odds ratio (95% CI) in the POP study (n=3,847) was 2.52 (1.05 to 6.19) whereas after the exclusion (n=3,847-216=3,631) it was 2.56 (1.06 to 6.32), i.e. the results were virtually identical with or without this exclusion. The ethnicity of the participants was not available for the case control study and the women recruited by the GOPEC consortium were all of white Western European ancestry. However, at this stage it is not possible to rule out population stratification due to the low number of women of non-European ethnicity. Further studies are needed to address this issue.

Analysis of the association with stillbirth and spontaneous preterm birth (sPTB)

The analysis of cord DNAs from the POP study cohort included 8 stillbirth cases. This group was not analysed separately as the numbers were too small to demonstrate the presence or absence of an association. Our analysis included 100 pregnancies with sPTB and there were no iciHHV-6 positive samples in this group (the expected number under the null was one), but the number of sPTB cases is too small to draw any meaningful conclusion. Further large studies are needed to evaluate the association between iciHHV-6 and stillbirth or sPTB.

Meta-analysis of large-scale population studies excluding Japanese populations

After exclusions of the two Japanese studies (Tanaka-Taya 2014 and Miura 2018) from the analysis reported in Figure 2C, the summary proportion of ciHHV-6 derived from the meta-analysis of large-scale population studies was 0.7% (392/58,211). Similarly to the main analysis, this sensitivity analysis demonstrated a two to three fold risk of preeclampsia associated with iciHHV-6 (OR=2.4, 95% CI: 1.4 to 4.3, P=0.0019). Heterogeneity between the remaining studies was also similar to what was observed in the main analysis (I^2=89%, 2-sided P=4x10^{-07}).

Supplementary Tables

Supplementary Table 1. Custom-made qPCR primers and probes.

Custom-made Taqman assay	Primer/Probe	Sequence (5' to 3')	Final concentration
Adenovirus	Forward	GCC ACS GTG GGG TTT CTA AAC TT	600 nM
	Probe	[JUN]-TGC ACC AGA CCC GGR CTC AGG TAC TCC GA-[QSY]	400 nM
	Reverse	GCC CCA GTG GKC TTA CAT GCA CAT C	600 nM
CMV	Forward	GCA TGC GCG AGT GTC AAG AC	600 nM
	Probe	[ABY]-TGC GCC GTA TGC TGC TCG ACA-[QSY]	400 nM
	Reverse	GTT ACT TTG AGC GCC ATC TGT TCC T	600 nM
EBV	Forward	CCG GTG TGT TCG TAT ATG GAG	600 nM
	Probe	[JUN]-TGC CCT TGC TAT TCC ACA ATG TCG T-[QSY]	400 nM
	Reverse	GGG AGA CGA CTC AAT GGT GTA	600 nM
HPV-6	Forward	TGG GGT AAT CAA CTG TTT GTT ACT GTG GTA	400 nM
	Probe	[ABY]-GAC ATT ATG TGC ATC CGT AAC TAC-[QSY]	200 nM
	Reverse	GCA TGT ACT CTT TAT AAT CAG AAT TGG TGT ATG TG	400 nM
HPV-11	Forward	CTG GGG AAA CCA CTT GTT TGT TAC TGT G	400 nM
	Probe	[JUN]-GAC ACT ATG TGC ATC TGT GTC TAA-[QSY]	800 nM
	Reverse	CGC ATG TAT TCC TTA TAA TCT GAA TTA GTG TAT GTA	400 nM
HPV-16	Forward	TTG TTG GGG TAA CCA ACT ATT TGT TAC TGT T	400 nM
	Probe	[FAM]-GTC ATT ATG TGC TGC CAT ATC TAC TTC-[QSY]	400 nM
	Reverse	CCT CCC CAT GTC TGA GGT ACT CCT TAA AG	400 nM
HPV-18	Forward	GCA TAA TCA ATT ATT TGT TAC TGT GGT AGA TAC CAC T	400 nM
	Probe	[VIC]-AAC AAT ATG TGC TTC TAC ACA GTC TCC TGT-[QSY]	100 nM
	Reverse	GCT ATA CTG CTT AAA TTT GGT AGC ATC ATA TTG C	400 nM
HSV-1	Forward	TTC TGC AGC TCG CAC CAC	600 nM
	Probe	[FAM]-CGA TGG CAA CGC GGC CCA ACA TAT CGT TGA C-[QSY]	300 nM
	Reverse	GGA GCG CAT CAA GAC CAC C	600 nM
HSV-2	Forward	TGC GTG GCG TTG TAC TT	800 nM
	Probe	[ABY]-CAA ACA TGC GGT CGA TGG CCT C-[QSY]	300 nM
	Reverse	CCA TCT CGA CCA CCT TCA C	800 nM
Parvovirus	Forward	TCC CTG GAA TTA ATG CAG ATG C	1200 nM
	Probe	[FAM]-ACC TCC AAA CCA CCC CAA TTG TCA CA-[QSY]	400 nM
	Reverse	CAC TGC TGC TGA TAC TGG TGT CT	200 nM
VZV	Forward	CAC GTA TTT TCA GTC CTC TTC AAG TG	940 nM
	Probe	[VIC]-TAC CGC CCG TGG AGC GCG-[QSY]	400 nM
	Reverse	TTA GAC GTG GAG TTG ACA TCG TTT	940 nM
HHV-6 U67/68[4]	Forward	TTC CGG TAT ATG ACC TTC GTA AGC	300nM
	Probe HHV-6A	[6-FAM]-ACA TTA TAT GTC GAA CTT GAC ACT ACC TTC CG-[QSY]	250nM
	Probe HHV-6B	[VIC]-CAT TAT ATA TCG AAT CTG ACG CTA CCT TCC G-[QSY]	250nM
	Reverse	GAT GTC TCA CCT CCA AAT CTT TAG AAA T	300nM
HHV-6 U100	Forward	CGA CTT GCC TCA CAA ATA TTG TC	18µM
	Probe	[FAM]-CCA TCG TAA GCG CAT TGT GGC ACT C-[BHQ-1]	5µM
	Reverse	ATG GTG CAT AAT GCG GGA	18µM

The Taqman assay HHV-6 U67/68 was designed to discriminate between the HHV-6A and HHV-6B genes. The Taqman assay HHV-6 U100 was designed to detect transcripts coded by the U100 gene of both HHV-6A and HHV-6B, due to the high homology of the 2 sequences (gene ID: 1487972 for HHV-6A and gene ID: 1497092

for HHV-6B). CMV: Cytomegalovirus; EBV: Epstein-Barr virus; HPV-6, HPV-11, HPV-16, HPV-18: Human Papillomavirus types 6, 11, 16 and 18; HSV-1 and HSV-2: Herpes Simplex Viruses types 1 and 2; VZV: Varicella Zoster Virus; HHV-6A: human herpesvirus 6, variant A; HHV-6B: human herpesvirus 6, variant B; ciHHV-6 U67/68: human herpesvirus 6, U67/68 gene; HHV-6 U100: human herpesvirus 6, U100 gene[5]; 6-FAM: 6-carboxyfluorescein reporter dye; QSY: QSY quencher; VIC: VIC reporter dye; BHQ-1: Black Hole Quencher-1 quencher.

Supplementary Table 2. Clinical characteristics of the patients analyzed by qPCR for detection of placental viral DNAs.

	FGR	Control (FGR)	PE	Control (PE)	Pre-term
N	100	100	100	100	100
Maternal BMI (kg/m^2)	24.9 (22.7 to 27.8)	24.7 (22.6 to 26.6)	25.7 (22.7 to 30.1)	24.1 (22.6 to 27.5)	25.0 (22.8 to 28.5)
Maternal age (years)	30.9 (25.9 to 34.1)	30.6 (27.4 to 32.7)	29.7 (26.5 to 33.6)	29.9 (27.4 to 33.9)	30.8 (27.0 to 34.1)
Gestational age (weeks)	40.3 (39.2 to 41.3)	40.3 (39.4 to 41.0)	39.9 (38.6 to 41.1)	40.0 (39.1 to 40.9)	34.1 (31.9 to 35.6)
Sample collection time (hours)	2.2 (0.3 to 7.7)	2.6 (0.2 to 7.4)	4.1 (0.4 to 9.2)	3.3 (0.3 to 8.1)	0.8 (0.2 to 9.1)
Fetal sex: female	57 (57%)	54 (54%)	44 (44%)	47 (47%)	42 (42%)
Smoking at booking (yes/no)	28 (28%)	18 (18%)	10 (10%)	7 (7%)	15 (15%)
Age stopped FTE (years)	20.0 (18.0 to 23.0)	21.0 (18.0 to 23.0)	21.0 (18.0 to 23.0)	21.0 (18.0 to 23.0)	21.0 (18.0 to 23.0)
Missing	2 (2%)	4 (4%)	4 (4%)	2 (2%)	5 (5%)
Maternal height (cm)	165 (160 to 170)	165 (161 to 169)	165 (161 to 169)	164 (160 to 170)	164 (160 to 167)
Deprivation quartile					
1 (lowest)	28 (28%)	23 (23%)	32 (32%)	26 (26%)	25 (25%)
2	19 (19%)	17 (17%)	21 (21%)	20 (20%)	22 (22%)
3	25 (25%)	32 (32%)	22 (22%)	25 (25%)	31 (31%)
4 (highest)	21 (21%)	23 (23%)	20 (20%)	22 (22%)	19 (19%)
Missing	7 (7%)	5 (5%)	5 (5%)	7 (7%)	3 (3%)
Ethnicity					
Non-white	4 (4%)	9 (9%)	2 (2%)	4 (4%)	7 (7%)
White	94 (94%)	91 (91%)	97 (97%)	92 (92%)	91 (91%)
Missing	2 (2%)	0 (0%)	1 (1%)	4 (4%)	2 (2%)
Married	59 (59%)	73 (73%)	69 (69%)	72 (72%)	66 (66%)
Any alcohol consumption	3 (3%)	6 (6%)	2 (2%)	6 (6%)	4 (4%)
Type I or type II diabetes	1 (1%)	0 (0%)	1 (1%)	0 (0%)	2 (2%)
UtA Doppler mean PI highest decile	25 (25%)	7 (7%)	20 (20%)	7 (7%)	22 (22%)
Missing	5 (5%)	1 (1%)	2 (2%)	3 (3%)	7 (7%)
Birth weight (g)	2660 (2429 to 2810)	3530 (3325 to 3705)	3493 (3182 to 3790)	3495 (3263 to 3680)	2110 (1660 to 2525)
Induction of labor	37 (37%)	20 (20%)	63 (63%)	26 (26%)	5 (5%)
Mode of delivery					
Vaginal	80 (80%)	80 (80%)	56 (56%)	56 (56%)	56 (56%)
Intrapartum Caesarean	13 (13%)	13 (13%)	29 (29%)	29 (29%)	5 (5%)
Prelabor Caesarean	7 (7%)	7 (7%)	15 (15%)	15 (15%)	39 (39%)

Median (interquartile range) or number (%) are given as appropriate. Smoking, maternal age and BMI were recorded at the booking appointment (~12 weeks of gestation) and other maternal characteristics were obtained from the 20 weeks questionnaire. Two women are controls for both an FGR case and a PE case. FGR: fetal

growth restriction, PE: preeclampsia, BMI: body mass index, FTE: full-time education, UtA: uterine artery, PI: pulsatility index.

Supplementary References

1 Gardosi, J., Mongelli, M., Wilcox, M. & Chang, A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* **6**, 168-174, doi:10.1046/j.1469-0705.1995.06030168.x (1995).

2 ACOG. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol* **122**, 1122-1131, doi:10.1097/01.AOG.0000437382.03963.88 (2013).

3 Huang, Y. *et al.* Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic acids research* **42**, 315-327, doi:10.1093/nar/gkt840 (2014).

4 Sedlak, R. H. *et al.* Detection of Human Herpesvirus 6B (HHV-6B) Reactivation in Hematopoietic Cell Transplant Recipients with Inherited Chromosomally Integrated HHV-6A by Droplet Digital PCR. *J Clin Microbiol* **54**, 1223-1227, doi:10.1128/JCM.03275-15 (2016).

5 Mori, Y., Akkapaiboon, P., Yang, X. & Yamanishi, K. The human herpesvirus 6 U100 gene product is the third component of the gH-gL glycoprotein complex on the viral envelope. *Journal of virology* **77**, 2452-2458 (2003).

References

1 Global, regional, and national levels of maternal mortality, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. GBD 2015 Maternal Mortality Collaborators. *Lancet* **388**, 1775-1812, doi:10.1016/S0140-6736(16)31470-2 (2016).

2 McClure, E. M. *et al.* Global Network for Women's and Children's Health Research: probable causes of stillbirth in low- and middle-income countries using a prospectively defined classification system. *BJOG* **125**, 131-138, doi:10.1111/1471-0528.14493 (2018).

3 Brosens, I., Pijnenborg, R., Vercruysse, L. & Romero, R. The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. *Am J Obstet Gynecol* **204**, 193-201, doi:10.1016/j.ajog.2010.08.009 (2011).

4 Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* **19**, 9-20, doi:10.1038/nrg.2017.88 (2018).

5 Houldcroft, C. J., Beale, M. A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol* **15**, 183-192, doi:10.1038/nrmicro.2016.182 (2017).

6 Depledge, D. P. *et al.* Specific capture and whole-genome sequencing of viruses from clinical samples. *PloS one* **6**, e27805, doi:10.1371/journal.pone.0027805 (2011).

7 Gravel, A., Hall, C. B. & Flamand, L. Sequence analysis of transplacentally acquired human herpesvirus 6 DNA is consistent with transmission of a chromosomally integrated reactivated virus. *J Infect Dis* **207**, 1585-1589, doi:10.1093/infdis/jit060 (2013).

8 Pantry, S. N. & Medveczky, P. G. Latency, Integration, and Reactivation of Human Herpesvirus-6. *Viruses* **9**, 194, doi:10.3390/v9070194 (2017).

9 Huang, Y. *et al.* Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic acids research* **42**, 315-327, doi:10.1093/nar/gkt840 (2014).

10 Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. Gopec Consortium. *Am J Hum Genet* **77**, 127-131, doi:10.1086/431245 (2005).

11 Gravel, A. *et al.* Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. *Proc Natl Acad Sci U S A* **112**, 8058-8063, doi:10.1073/pnas.1502741112 (2015).

12 Sovio, U. *et al.* Prediction of Preeclampsia Using the Soluble fms-Like Tyrosine Kinase 1 to Placental Growth Factor Ratio: A Prospective Cohort Study of Unselected Nulliparous Women. *Hypertension* **69**, 731-738, doi:10.1161/HYPERTENSIONAHA.116.08620 (2017).

13 Sultana, Z., Maiti, K., Dedman, L. & Smith, R. Is there a role for placental senescence in the genesis of obstetric complications and fetal growth restriction? *Am J Obstet Gynecol* **218**, S762-S773, doi:10.1016/j.ajog.2017.11.567 (2018).

14 Broady, A. J. *et al.* Protective proteins and telomere length in placentas from patients with pre-eclampsia in the last trimester of gestation. *Placenta* **50**, 44-52, doi:10.1016/j.placenta.2016.12.018 (2017).

15 Sano, R. & Reed, J. C. ER stress-induced cell death mechanisms. *Biochim Biophys Acta* **1833**, 3460-3470, doi:10.1016/j.bbamcr.2013.06.028 (2013).

16 Burton, G. J., Yung, H. W., Cindrova-Davies, T. & Charnock-Jones, D. S. Placental endoplasmic reticulum stress and oxidative stress in the pathophysiology of unexplained intrauterine growth restriction and early onset preeclampsia. *Placenta* **30(Suppl A)**, S43-48, doi:10.1016/j.placenta.2008.11.003 (2009).

17 Racicot, K. & Mor, G. Risks associated with viral infections during pregnancy. *J Clin Invest* **127**, 1591-1599, doi:10.1172/JCI87490 (2017).

18 Zhang, E. *et al.* Inherited Chromosomally Integrated Human Herpesvirus 6 Genomes Are Ancient, Intact, and Potentially Able To Reactivate from Telomeres. *Journal of virology* **91**, doi:10.1128/JVI.01137-17 (2017).

19 Peddu, V. *et al.* Inherited chromosomally integrated HHV-6 demonstrates tissue-specific RNA expression in vivo that correlates with increased antibody immune response. *Journal of virology*, doi:10.1128/JVI.01418-19 (2019).

20 Marci, R. *et al.* Presence of HHV-6A in Endometrial Epithelial Cells from Women with Primary Unexplained Infertility. *PloS one* **11**, e0158304, doi:10.1371/journal.pone.0158304 (2016).

21 Caselli, E. *et al.* HHV-6A Infection of Endometrial Epithelial Cells Induces Increased Endometrial NK Cell-Mediated Cytotoxicity. *Front Microbiol* **8**, 2525, doi:10.3389/fmicb.2017.02525 (2017).

22 Dekker, G., Robillard, P. Y. & Roberts, C. The etiology of preeclampsia: the role of the father. *J Reprod Immunol* **89**, 126-132, doi:10.1016/j.jri.2010.12.010 (2011).

23 Smith, G. C., Pell, J. P. & Walsh, D. Pregnancy complications and maternal risk of ischaemic heart disease: a retrospective cohort study of 129,290 births. *Lancet* **357**, 2002-2006, doi:10.1016/S0140-6736(00)05112-6 (2001).

24 Bellamy, L., Casas, J. P., Hingorani, A. D. & Williams, D. J. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *Bmj* **335**, 974, doi:10.1136/bmj.39335.385301.BE (2007).

25 Pasupathy, D. *et al.* Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* **8**, 51, doi:10.1186/1471-2393-8-51 (2008).

26 Sovio, U., White, I. R., Dacey, A., Pasupathy, D. & Smith, G. C. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *Lancet* **386**, 2089-2097, doi:10.1016/S0140-6736(15)00131-2 (2015).

27 Gaccioli, F., Lager, S., Sovio, U., Charnock-Jones, D. S. & Smith, G. C. S. The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal

ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. *Placenta* **59**, S17-S25, doi:10.1016/j.placenta.2016.10.011 (2017).

28 ACOG. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol* **122**, 1122-1131, doi:10.1097/01.AOG.0000437382.03963.88 (2013).

29 Hiby, S. E. *et al.* Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med* **200**, 957-965, doi:10.1084/jem.20041214 (2004).

30 Tanaka-Taya, K. *et al.* Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. *J Med Virol* **73**, 465-473, doi:10.1002/jmv.20113 (2004).

31 Pellett, P. E. *et al.* Chromosomally integrated human herpesvirus 6: questions and answers. *Rev Med Virol* **22**, 144-155, doi:10.1002/rmv.715 (2012).

32 Hill, J. A. *et al.* Outcomes of hematopoietic cell transplantation using donors or recipients with inherited chromosomally integrated HHV-6. *Blood* **130**, 1062-1069, doi:10.1182/blood-2017-03-775759 (2017).

33 Moustafa, A. *et al.* The blood DNA virome in 8,000 humans. *PLoS Pathog* **13**, e1006292, doi:10.1371/journal.ppat.1006292 (2017).

34 Miura, H. *et al.* Chromosomally integrated human herpesvirus 6 in the Japanese population. *J Med Virol* **90**, 1636-1642, doi:10.1002/jmv.25244 (2018).

35 Mouammine, A. *et al.* rs73185306 C/T is not a predisposing risk factor for inherited chromosomally-integrated human herpesvirus 6A/B. *J Infect Dis*, doi:10.1093/infdis/jiz527 (2019).

36 Gardosi, J., Mongelli, M., Wilcox, M. & Chang, A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* **6**, 168-174, doi:10.1046/j.1469-0705.1995.06030168.x (1995).

37 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

38 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

39 Krueger, F. *Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available online at:* http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, 2012).

40 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).

41 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

42 Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464-469, doi:10.1093/bioinformatics/btr703 (2012).

43 Lager, S. *et al.* Detecting eukaryotic microbiota with single-cell sensitivity in human tissue. *Microbiome* **6**, 151, doi:10.1186/s40168-018-0529-x (2018).

44 Mori, Y., Akkapaiboon, P., Yang, X. & Yamanishi, K. The human herpesvirus 6 U100 gene product is the third component of the gH-gL glycoprotein complex on the viral envelope. *Journal of virology* **77**, 2452-2458 (2003).

45 Brown, J. R. *et al.* Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. *J Clin Microbiol* **54**, 2530-2537, doi:10.1128/JCM.01052-16 (2016).

46 Nurk, S. *et al.* Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* **20**, 714-737, doi:10.1089/cmb.2013.0084 (2013).

47 Telford, M., Navarro, A. & Santpere, G. Whole genome diversity of inherited chromosomally integrated HHV-6 derived from healthy individuals of diverse geographic origin. *Sci Rep* **8**, 3472, doi:10.1038/s41598-018-21645-x (2018).

48 Wood, M. L. & Royle, N. J. Chromosomally Integrated Human Herpesvirus 6: Models of Viral Genome Release from the Telomere and Impacts on Human Health. *Viruses* **9**, doi:10.3390/v9070184 (2017).

49 Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic acids research* **41**, W29-33, doi:10.1093/nar/gkt282 (2013).

50 Fagerland, M. W., Lydersen, S. & Laake, P. Recommended confidence intervals for two independent binomial proportions. *Stat Methods Med Res* **24**, 224-254, doi:10.1177/0962280211415469 (2015).

51 Lydersen, S., Fagerland, M. W. & Laake, P. Recommended tests for association in 2 x 2 tables. *Stat Med* **28**, 1159-1175, doi:10.1002/sim.3531 (2009).

52 Nyaga, V. N., Arbyn, M. & Aerts, M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Arch Public Health* **72**, 39, doi:10.1186/2049-3258-72-39 (2014).

53 Dominguez, G. *et al.* Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. *Journal of virology* **73**, 8040-8052 (1999).

Figure Legends

Figure 1. SNP analysis of HHV-6B genome sequenced in fetal and parental samples. A) Comparison of HHV-6B genomes sequenced in different samples revealed 187 informative SNPs, represented as black or white lines in the "barcode" graph if concordant or discordant to the HHV-6B reference genome (GCF_000846365.1), respectively. B) Comparison of placental and parental SNPs based on RNA-seq and DNA-seq reads aligning to the HHV-6B gene U100, which codes for glycoprotein Q (gQ).[44] The genes of the HHV-6B reference genome are in light green in the upper part of the panel, and the U100 gene is indicated in dark green; repeat regions, including the two large direct repeat regions on both termini, are indicated in amber. The RNA-seq coverage of the U100 gene (highlighted in light-red and enlarged) is shown for one placental sample (Infant 2 in Table 1 and Extended Data Figure 2) and represented by the surface area of the black peaks. The DNA-seq reads obtained from the corresponding paternal sample are shown in the bottom part of the panel. Red vertical lines indicate positions where both the infant RNA and the paternal DNA concordantly differ from the HHV-6B reference genome.

Figure 2. Fetal inheritance of ciHHV-6 and the risk of preeclampsia. A) Proportion of iciHHV-6 positive samples in preeclampsia cases (black bars) and non-cases (white bars) in the following datasets: placental RNA samples within the POP study analyzed by RNA-seq ("RNA-seq", n=279); all samples analyzed by cord DNA genotyping, excluding those in the first group, i.e. with RNA-seq data available ("Other", n=4,042); the combined study population, i.e. all the genotyped cord DNA samples from both studies ("All", n=4,321). n represents the number of patients analyzed in each group. For the 3 datasets "RNA-seq", "Other" and "All" the P values for the association between fetal iciHHV-6 and the risk of preeclampsia were 0.14, 0.022 and 0.008, respectively, using the recommended Fisher-Boschloo unconditional exact test[51] (2-sided); and 0.17, 0.026 and 0.008, respectively, using the more widely used Fisher's exact test (2-sided). **B)** Odds ratios and 95% confidence intervals (CI) for preeclampsia by the presence of iciHHV-6 calculated in each dataset and

in the pooled study population. **C)** Proportion of iciHHV-6 positive samples in the following datasets: large-scale population studies (n>1,000) of healthy/control patients currently available in the literature (white symbols; see Methods); pooled values of the healthy/control populations (dataset "Total", grey symbol); GOPEC samples (dataset "GOPEC", black symbol). Symbols represent the proportion of iciHHV-6 positive samples with 95% CI; n represents the number of patients involved in each study; the location where the studies were conducted is in parenthesis.