Genome-wide association study (GWAS) of human host factors influencing viral severity of herpes simplex virus type 2 (HSV-2)

Sarah E. Kleinstein1,2 · Patrick R. Shea1 · Andrew S. Allen3 · David M. Koelle4,5,6,7,8 · Anna Wald4,5,9 · David B. Goldstein1

Received: 6 July 2017 / Revised: 24 November 2017 / Accepted: 1 December 2017 / Published online: 28 February 2018
© Macmillan Publishers Limited, part of Springer Nature 2018

Abstract
Herpes simplex virus type 2 (HSV-2) is an incurable viral infection with severity ranging from asymptomatic to frequent recurrences. The viral shedding rate has been shown as a reproducible HSV-2 severity endpoint that correlates with lesion rates. We used a genome-wide association study (GWAS) to investigate the role of common human genetic variation in HSV-2 severity. We performed a GWAS on 223 HSV-2-positive participants of European ancestry. Severity was measured by viral shedding rate, as defined by the percent of days PCR+ for HSV-2 DNA over at least 30 days. Analyses were performed under linear regression models, adjusted for age, sex, and ancestry. There were no genome-wide significant (p < 5E-08) associations with HSV-2 viral shedding rate. The top nonsignificant SNP (rs75932292, p = 6.77E-08) associated with HSV-2 viral shedding was intergenic, with the nearest known biologically interesting gene (ABCA1) ~130 kbp downstream. Several other SNPs approaching significance were in or near genes with viral or neurological associations, including four SNPs in KIF1B. The current study is the first comprehensive genome-wide investigation of human genetic variation in virologic severity of established HSV-2 infection. However, no significant associations were observed with HSV-2 virologic severity, leaving the exact role of human variation in HSV-2 severity unclear.

Introduction
Herpes simplex virus type 2 (HSV-2) is one of the most prevalent sexually transmitted infections worldwide, with a global prevalence estimated at 417 million and ~19.2 million new infections acquired per year [1]. In the United States, the number of HSV-2-infected individuals has stabilized at ~16% of the population, indicating that transmission is a continuing public health problem [2]. HSV-2 establishes lifelong latency upon infection and, to date, there are no vaccines, cures, or even fully efficacious suppressive treatments [3].

Although 75–90% of HSV-2 infections are subclinical and asymptomatic [4], the severity varies widely. Symptomatic, clinical disease presents as painful, recurrent, and often frequent outbreaks of genital lesions [5–7]. Daily treatment with antiviral medications can reduce outbreak frequency, though it does not completely eliminate them [8]. Due to its incurable status and adverse effects on quality of life, HSV-2 diagnosis may be associated with psychological distress [9, 10]. Further, HSV-2 infection during pregnancy, particularly primary and asymptomatic infections, can result in perinatal transmission to the infant, a rare
Genome-wide association study (GWAS) of human host factors influencing viral severity...

but severe outcome [11]. Children who acquire HSV-2 at birth experience significant morbidity and mortality, with survivors risking encephalitis or multi-organ disseminated disease, and over 50% developing central nervous system disease [11]. Of additional public health concern, HSV-2 infection is associated with at least a threefold increased risk of both acquiring and transmitting human immunodeficiency virus type 1 (HIV-1) [12, 13].

There are currently no completely effective methods for interrupting HSV-2 transmission. While standard safe sex practices are recommended [5], condom usage risk reduction estimates have varied by gender and measurement, ranging from 30 to 96% [14, 15], as active lesions and viral replication may occur on unprotected skin. Similarly, while antiviral treatment reduces HSV-2 transmission by ~50%, it does not completely ablate active viral replication as measured by viral shedding [8, 11]. Therefore, it is important to understand the host factors influencing HSV-2 severity in order to elucidate mechanisms to limit its impact on human health.

HSV-2 severity is a complex phenotype that can be measured by symptomatic lesion recurrences or viral shedding. Viral shedding from genital mucosa has previously been identified as a more objective representation of infection severity than the number of symptomatic recurrences [16, 17]. Days with active lesions show higher risk of viral shedding; thus, those with the most severe disease show both increased viral shedding and increased outbreaks [16]. Further, though some asymptomatic individuals eventually recognize lesions, particularly following education on lesion identification, those who remain asymptomatic have the lowest viral shedding [16]. Thus, having quantitative information on viral shedding is valuable in determining virologic HSV-2 severity and limiting confounding from perception biases in lesion detection.

As a complex trait, infection severity is potentially influenced by a combination of viral [18], host [19–21], and environmental factors. The importance of host genetic variation in herpes pathogenesis has previously been demonstrated for certain rare phenotypes, such as herpes simplex virus type 1 encephalitis (HSE) [22, 23]. It is well established that neuron-intrinsic deficiencies in toll-like receptor 3 (TLR3) pathway genes result in severe childhood HSE after primary HSV-1 infection [22, 24]. For HSV-2, candidate gene studies have implicated viral control and immune genes, including TLRs, in both susceptibility [25–28] and severity [19–21] during the chronic phase. Although these studies have detected variants potentially associated with herpes pathogenesis, they were limited to a small number of common allelic variants in candidate genes and none of these candidate gene associations have been replicated by other, independent studies. To date, only two genome-wide studies have investigated the role of human genetic factors in alphaherpes-related diseases: a genome-wide association study (GWAS) investigating herpes zoster susceptibility, which identified an association with HCP5 in the human leukocyte antigen (HLA) region and age of shingles onset [29], and a family-based linkage analysis that identified a 2.5 Mb region on chromosome 21 associated with HSV-1 susceptibility [30]. Subsequent targeted sequencing pinpointed specific variants in the C21orf91 (CSSG1) gene that were responsible for the chromosome 21 association with HSV-1 susceptibility [31]. Thus, while human genetics have been implicated in herpetic diseases, there remains a dearth of validated causal variants for HSV-2 susceptibility and severity.

It is important to utilize unbiased, genome-wide studies to definitively investigate the role of human genetics in disease etiology. The use of GWAS to investigate complex traits is a well-validated standard in human genetics. Though most complex disease traits previously studied relate to inherited diseases, host genetic factors influencing infectious diseases have been detected through genome-wide studies [32–35], including for other alphaherpesviruses [29, 30]. To date, no GWAS has been reported for HSV-2. In this study, we report the first genome-wide investigation of common human genetic variation influencing HSV-2 severity, as measured by the quantitative viral shedding rate.

**Results**

Demographic and clinical characteristics of participants included in the final analyses are described in Table 1. Briefly, most participants were symptomatic (83%), with more female participants than male (61% vs. 39%). Forty-four percent of participants were HSV-1 seropositive. The

| Table 1 | Participant demographics for the subset of the cohort included in the final analyses (N = 223 with genetically confirmed European ancestry) |
|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sex     | N = 223                                                                                                                                                                                            |
| Male, N (%) | 86 (38.57%)                                                                                                                             |
| Female, N (%) | 137 (61.43%)                                                                                                                          |
| Age (years), median (range) | 39.5 (22–76)                                                                                                                             |
| Diagnosis, N (%) |                                                                                                                                     |
| Symptomatic | 186 (83.41%)                                                                                                                             |
| Asymptomatic | 37 (16.59%)                                                                                                                             |
| Days since diagnosisa, median (range) | 3407 (14–12 649)                                                                                                                          |
| Median viral shedding rate, % (range) | 15.3 (0–100%)                                                                                                                             |
| Median lesion rateb, % (range) | 6.3 (0–85.4%)                                                                                                                             |
| HSV-1 positive, N (%) | 97 (43.5%)                                                                                                                             |

aData not available for all samples
viral shedding rate ranged from 0 to 100%, with a median of 15% (Fig. 1). The full cohort had similar demographics to the genetically confirmed European subset used in this study (data not shown).

After multiple testing correction, there were no genome-wide significant associations ($p < 5\times10^{-8}$) with HSV-2 severity, as measured by the quantitative viral shedding rate and adjusted for age, gender, and ancestry (Figs. 2 and 3). The 10 single-nucleotide polymorphisms (SNPs) with the lowest $p$-values are listed in Table 2. The SNP that achieved the lowest $p$-value in our analysis was rs75932292, which was just below statistical significance ($p = 6.77\times10^{-8}$). rs75932292 is intergenic, with the nearest biologically relevant coding gene, ATP-binding cassette subfamily A member 1 (ABCA1), located ~130 kbp downstream. To examine whether rs75932292 might tag functional variation in the ABCA1 gene, we examined linkage disequilibrium (LD) within ±1 Mb of this SNP among 640 whole-genome sequenced (WGS) population controls of European ancestry. We identified several rare, functional variants in ABCA1 that were in high LD with rs75932292 and could potentially account for any association with HSV-2 viral shedding, though their causality cannot be definitively determined (see Supplementary Table 1). As rs75932292 itself is relatively rare, with a minor allele frequency (MAF) of 0.07 in our cohort (MAF = 0.05 for Europeans in the 1000 Genomes Project [36]), there were no individuals homozygous for the variant allele in our data set. However, there was a slight trend toward increased viral shedding among carriers heterozygous for the variant genotype (GA) compared to homozygous wild-type (GG) individuals (data not shown).

Despite the lack of genome-wide significant associations, several potentially biologically interesting SNPs approached statistical significance. Four intronic SNPs in the kinesin family member 1B (KIF1B) gene were observed among the top results. These four SNPs (rs17034615, rs17034775, rs72865926, and rs72867415; all $p = 6.57\times10^{-6}$) were in perfect LD ($r^2 = 1$) among individuals of European ancestry in our cohort, with a MAF of 0.05 (identical to that expected for Europeans in the 1000 Genomes Project [36]), suggesting that they might all be tagging the same underlying variant. When LD was explored among the 640 WGS population controls of European ancestry, several rare, functional KIF1B variants were in high LD with the intronic variants (see Supplementary Table 2). There was also a slight trend toward increased viral shedding with the presence of any intronic SNP minor allele(s); however, this was primarily driven by heterozygous individuals, as there was only a single homozygous variant individual for three of the SNPs and none for rs72867415 (data not shown).

Three other SNPs also approached genome-wide significance and were in or near genes of plausible biological relevance to neurological and/or viral phenotypes, though none had previously been linked to HSV-2. These included two intergenic SNPs, rs56122323 ($p = 3.76\times10^{-7}$), located ~25 kbp downstream of mannosyl (alpha-1,6-)glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase, isozyme B.

**Fig. 1** Shedding rate (percent of days PCR+ for HSV-2 DNA) distribution among individuals of European ancestry ($N = 223$)

**Fig. 2** Manhattan plot of the GWAS for HSV-2 viral severity among individuals of European ancestry ($N = 223$). Linear regression model. The red line indicates the genome-wide statistical significance threshold ($p < 5\times10^{-8}$)
Due to prior associations of HSV pathogenesis with immune-related genes, including a recent association with HLA-A*01 in a portion of this cohort [21], a targeted analysis limited to just the major histocompatibility complex (MHC) region of the genome was conducted. However, there were no suggestive associations with HSV-2 viral shedding when just the MHC region was considered (see Supplementary Figures 1-2). Further, while we did not directly test HLA haplotypes in this analysis, we genotyped five SNPs in nearly perfect LD ($r^2 \sim 0.95$) with HLA-A*0101. While these SNPs failed to reach genome-wide significance ($p = 0.001$) when tested using linear or Poisson regression models, we did observe ~10% higher frequency of these SNPs among the highest viral shedders ($\geq 25\%$ of days with viral shedding) compared to low/no shedders ($<25\%$ of days with viral shedding; Supplementary Table 3), suggesting that the HLA-A*01 haplotype may have a moderate influence on HSV-2 shedding levels, but with an effect that requires a larger sample size to detect.

Additionally, a candidate gene analysis of 10 non-MHC genes previously implicated in HSV-2 pathogenesis was also conducted. Though not statistically significant after Bonferroni correction, $p$-values were slightly lower than expected on the quantile-quantile plot (no genomic inflation (lambda = 1); see Supplementary Figure 3). The vast majority of the SNPs with the lowest $p$-values were on chromosome 10 in the mannose-binding lectin 2 (MBL2) gene region. MBL2 is part of the innate immune system, where it activates the classical complement pathway and can detect viruses, including binding HSV-2 surface glycoproteins [27, 37]. The three SNPs with the lowest $p$-values (rs201381710, $p = 0.01$; rs10824793, $p = 0.02$; and rs4935047, $p = 0.02$) shared nearly perfect LD ($r^2 > 0.99$) and were all intronic in MBL2. The SNP with the next lowest $p$-value in the non-MHC-region candidate gene analysis was rs4696483 ($p = 0.02$), which is intronic in TLR2. This SNP is not in LD with the previous TLR2 candidate SNP rs1898830 in our cohort ($r^2 = 0.03$). The only other SNP with a $p < 0.05$ was rs2147419 ($p = 0.04$), which is intronic in FAS.

**Discussion**

This study represents the first GWAS of any measure of HSV-2 severity. Overall, the results failed to achieve statistical significance and there was no evidence for associations of common host genetic variation and HSV-2 viral shedding rate (as quantitatively measured by percent of days PCR+ for HSV-2 DNA at self-swabbed sites over a period of at least 30 days). Additionally, we were not able to replicate previously observed candidate gene associations with HSV-2 pathogenesis disclosed in targeted investigations, though we did see a slight, nonsignificant increase in frequency of HLA-A*0101 tagSNPs among high viral shedders relative to low viral shedders.

The top SNP (rs75932292, $p = 6.77E-08$) was intergenic and just below the $p < 5E-08$ threshold for genome-wide significance; it showed some evidence of linkage with

---

**Fig. 3** QQ plot of the GWAS for HSV-2 viral severity among individuals of European ancestry ($N = 223$). Linear regression model. No genomic inflation was observed (lambda = 1.02)

| Rank | SNP            | SNP type       | Nearest gene | $p$-value |
|------|----------------|----------------|--------------|-----------|
| 1    | rs75932292     | Intergenic     | —            | 6.77E-08  |
| 2    | rs73664402     | Intergenic     | —            | 1.21E-07  |
| 3    | rs56122323     | Intergenic     | LOC105371899 and MGAT5B | 3.76E-07  |
| 4    | rs62377770     | Intergenic     | CEP120 and CSNK1G3 | 4.29E-07  |
| 5    | rs55963884     | Intron/intergenic | ARHGAP25   | 1.06E-06  |
| 6    | rs4912855      | Intergenic     | LOC101926941 | 1.84E-06  |
| 7    | rs11204209     | Intronic       | ZNF488      | 2.51E-06  |
| 8    | rs4910264      | Intronic       | LOC105376548 and LOC105376550 | 3.28E-06  |
| 9    | rs59849217     | Intergenic     | —            | 3.45E-06  |
| 10   | rs75644638     | Intergenic     | —            | 3.75E-06  |

Linear regression analysis, adjusted for age, sex, and significant PC axes

(MGAT5B), and rs62377770 ($p = 4.29E-07$), located ~58.5 kbp upstream of casein kinase 1 gamma 3 (CSNK1G3); as well as rs117944720 ($p = 5.43E-06$), which is intronic in parkin RBR E3 ubiquitin protein ligase (PARK2).
potentially causal rare, functional variants in the downstream protein-coding gene **ABCA1**. **ABCA1** is a cholesterol/lipid regulator that is associated with several lipoprotein disorders [38–41]. It has also been shown to interact with HIV-1 viral proteins [42–45], Newcastle disease virus [46], and hepatitis C [47]. Rare variants in **ABCA1** might conceivably affect HSV-2 viral reactivation by altering lipids involved in membrane fusion or viral egress, hypotheses that are amenable to testing in vitro, as has been done with the viruses discussed above. There were several additional nonsignificant SNPs in the top results that are potentially of note, as they were present in genes previously associated with viral infections, including four intronic SNPs in **KIF1B**. **KIF1B** is a kinesin motor protein involved in anterograde transport of mitochondria and synaptic vesicle precursors. While other **KIF1B** mutations have been linked to Charcot-Marie-Tooth disease, neuroblastoma, and pheochromocytoma, **KIF1B** has also been identified in several studies related to hepatitis B virus-related hepatocellular carcinoma [48] and may act in early HIV-1 viral trafficking [49]. The role of **KIF1B** in anterograde synaptic transport could conceivably be related to HSV-2 reactivation because during viral reactivation from neurons HSV virion components are actively transported by this mechanism [50]. While several rare, functional protein-coding variants in **KIF1B** were in high LD with the identified intronic SNPs, it is unclear if these variants could affect viral trafficking without resulting in the severe phenotypes mentioned above and linked to known pathogenic **KIF1B** mutations.

In addition, several other SNPs approaching, but not reaching, genome-wide significance were noted for their location in or near genes with plausible biological linkage to viral replication or immunity, including: **MGAT5B**, an acetyl-glucosaminyltransferase isozyme that may be involved in processing HIV-1 protein glycosylation [51]; **CSNK1G3**, a serine/threonine kinase that has been shown to bind the HIV-1 vpu protein in vitro [52]; and finally **PARK2**, which has primarily been associated with Parkinson’s disease [53] but may also act in hepatitis C replication [54].

This study did not replicate any previous HSV-2 associations either in the full analysis or targeted candidate gene and MHC-region analyses. For primary HSV-1, it is well established that deficiencies in the TLR3 pathway lead to the severe phenotype of HSE in children [22, 23]. It is not surprising that there were no associations with **TLR3** or genes, such as **UNC93B** or **TRIF**, that are upstream or downstream of **TLR3** and initiate type I interferon signaling. Mechanistic studies of these genes associated with pediatric primary HSV-1 have shown that they act intrinsically in neurons to reduce HSV-1 replication during the innate phase of the initial response [24], while recurrent shedding, the phenotype examined in the present study, relates to epithelial cell replication and immune cell function. Both HSV-2 viral shedding and lesion rates have previously been linked to two SNPs (rs4696480 and rs1898830) in another TLR, **TLR2** [20]. **APOE** has also been linked to HSV severity, but was only associated with HSV-1 oral lesions, not with HSV-2 genital viral shedding, the severity phenotype under consideration here [19]. None of the top SNPs in the present study were in or near **APOE** or **TLR2**. Further, the **TLR2** SNP rs1898830 identified previously for HSV-2 was directly genotyped in our study and was not significant ($p = 0.57$). Thus, associations with **APOE** or **TLR2**, the two genes previously implicated in candidate gene studies utilizing a portion of this cohort, were not able to be replicated at the genome-wide level in the current cohort.

In the non-MHC candidate gene analysis, though no SNPs reached significance after Bonferroni multiple testing correction, the top SNPs were primarily located in **MBL2**, a component of the innate immune system. A **MBL2** structural variant was previously identified as more common among participants with recurrent (symptomatic) HSV-2 than asymptomatic individuals or healthy controls in a small candidate gene study [27]. The only additional SNPs with a $p < 0.05$ were rs4696483 ($p = 0.02$), which is intronic in **TLR2** and not in LD with the previously identified **TLR2** candidate SNP rs1898830, and rs2147419 ($p = 0.04$), which is located in intron 2 of the **FAS** gene. **FAS** regulates activation-induced cell death and two polymorphisms, 1377G>A (rs2234767) and 670A>G (rs1800682), were previously implicated in HSV-2 susceptibility in a small candidate gene study of South African women that focused on three **FAS** and **FASLG** SNPs [28]. Amongst these, neither the **FAS** SNPs (rs2234767, $p = 0.22$; rs1800682, $p = 0.89$) nor the candidate **FASLG** SNP (rs763110, $p = 0.21$) were significant in our analysis. The lack of replication of previously implicated HSV-2 SNPs underscores the need for rigorous replication of disease-associated genetic variants and the importance of determining biological mechanisms, if at all possible, particularly for variants identified through candidate gene studies.

While some nonsignificant SNPs were identified in potentially biologically plausible genes, we have been unable to demonstrate the presence of any common genetic variants robustly associated with HSV-2 viral shedding rate at the genome-wide level. Though the available sample size was modest for a GWAS, thus limiting our study power, viral shedding is a unique, robust, and quantitative measure of HSV-2 virologic severity, making it important to investigate. Post hoc power calculations indicated that we had >99% power to detect a common SNP at MAF = 5% that accounted for at least 25% of the variance of HSV-2 viral shedding, while we had 71.53% power to detect SNPs that accounted for 15% of the shedding variance, and only...
27.29% power to detect SNPs that accounted for 10% of the shedding variance. Thus, while we had reasonable power to detect SNPs with large effect sizes, our ability to detect weaker effects was limited. However, the lack of statistical significance in this study suggests that there is no single common (MAF > 5%) SNP that explains a large portion of HSV-2 viral shedding, adding valuable information about the genetic architecture of HSV-2 severity. It is possible that multiple common variants of smaller effect sizes act in HSV-2 severity or, as suggested by the post hoc linkage analyses, it may be that rare, rather than common, human genetic variation has a role in HSV-2 severity, as has been the case with many complex diseases [55–57] and which was not within the scope of our study design.

There remains active debate on the best surrogate measure for HSV-2 severity. Some previous studies have focused on active viral lesions or the dichotomy of asymptomatic or symptomatic diagnosis as a measure of HSV-2 severity, and measures of lesion recurrences are commonly used as an end point for drug or vaccine trials. While viral shedding rates correlate with lesion rates, the viral shedding rate has previously been shown to be a more accurate and consistent measure of HSV-2 viral reactivation than lesion rates, as viral shedding can occur at times lacking lesions, and some individuals who are initially asymptomatic may later recognize lesions, implying that the lesion rate is influenced by subject perception [16, 58]. Further, previously conducted candidate gene studies using a portion of this cohort implicated a role for host genetics in HSV-2 severity, including shedding rates [19–21]. Though these associations were not replicated in the full GWAS of the current cohort, this is not uncommon, as most candidate gene associations are not replicated [59]; indeed, we were unable to replicate any previous HSV-2 candidate gene associations in our GWAS analysis.

Despite the lack of evidence for a role of common host genetic variation of larger effect sizes in HSV-2 severity, as measured by the viral shedding rate in this study, it remains likely that host and viral factors interact with the environment to control HSV-2 reactivation, as with other herpesviruses [22, 23, 29, 30]. While these analyses were adjusted for age, gender, and ancestry, it is possible that additional factors might confound genetic associations with HSV-2 shedding severity, such as time since HSV-2 acquisition, HSV-2 inoculum size, or viral strain. HSV-2 severity as measured by both genital lesions and shedding rate can decrease with time since acquisition [60]. In order to interrogate the full range of HSV-2 virologic severity, including asymptomatic individuals, for whom information on time since acquisition is not available, we did not adjust for time since HSV-2 acquisition. However, inclusion of time since HSV-2 acquisition in the main analysis did not dramatically change the results (see Supplementary Table 4). While HSV-1 can cause genital herps, oral HSV-1 seropositivity does not affect genital HSV-2 viral shedding [16]. Thus, we included individuals co-infected with HSV-1, as all individuals in this cohort had western blot-confirmed genital HSV-2 and a majority of the global population has oral herps.

Although we focused our study on a reasonably sized and well-characterized cohort of HSV-2-positive individuals of European ancestry, with unique quantification of HSV-2 viral shedding rate over at least a month, this represents a convenience cohort and may not be representative of the general population [16]. Of note, approximately 80% of HSV-2 seropositive individuals are asymptomatic [4], while our cohort was 83% symptomatic, such that we may have under-represented persons with milder phenotypes. Larger studies of individuals across ethnicities and the HSV-2 severity spectrum will be needed to determine the role of human genetics, both for common and rare variation. In particular, we currently lack large cohorts with robust HSV-2 phenotyping, including information on the viral shedding rate, which will be important to gather for future studies in order to increase study power to detect genetic associations. HSV-2 reactivation remains a complicated and poorly understood process involving both host and viral factors, without a cure in sight. Though the role of human genetics in the rare and extremely severe HSV-1 caused HSE is undisputed, it remains unclear how strongly human genetic variation affects genital HSV-2 severity.

Materials and methods

Participants

Western blot-confirmed HSV-2 seropositive North American participants followed at the University of Washington were included in this study. All participants signed informed consent for genetics studies and institutional review boards (IRBs) approved this study. This cohort has detailed phenotypic and quantitative information available, as has been described previously [16]. Briefly, participants in this cohort were at least age 18, HIV-1-negative, not on antiviral treatments during the study period, and exhibited a wide range of disease severity by viral shedding and lesion rates. As part of the study protocol, quantitative data on daily viral shedding over a period of at least 30 days were collected prospectively, a window within which more than 77% of both asymptomatic and symptomatic individuals show viral shedding [17, 58].

Viral shedding was determined using real-time PCR of self-sampled anogenital swabs with a cutoff of >150 copies of HSV-2 DNA per specimen, which has been validated to
be an accurate measurement of HSV shedding [17]. The viral shedding rate was calculated as the number of days PCR+ for HSV-2 DNA over the period of sampled days and has previously been shown to accurately represent viral behavior [16]. Any symptomatic HSV-2 episodes of lesions were recorded by self-reported diary entries and requested confirmatory clinical visits during the ≥30 day sampling period.

Genome-wide association study

Genotyping

A total of ~4.3 million SNPs in 307 participants were genotyped using the Illumina HumanOmni5Exome array platform (San Diego, CA). Of these, 191 participants were genotyped on the Omni5Exome4v1-1 array and 116 on the Omni5Exome4v1.0 array. The full cohort was combined for data analysis, including only non-monomorphic SNPs that were shared between the two arrays and where the DNA strand could be definitively determined (all symmetric (A/T or G/C) SNPs were excluded). The combined data set included ~4 million SNPs.

Quality control

A series of quality control (QC) checks and all subsequent analyses were carried out to ensure sample integrity using PLINK [61]. Gender was assessed for concordance between genetically inferred and self-reported gender, with two discordant participants removed. Duplicate samples and cryptic relatedness (identity by descent > 0.125) were identified based on genetic data; two pairs of duplicate samples were removed. Data quality was also evaluated at the marker level to remove low-quality genotypes. Four participants with anomalously high or low heterozygosities (−0.07 < F > 0.07) were excluded and markers missing >1% of genotype calls were removed. In addition, rare variants with a MAF < 5% were excluded. Following initial QC, there were 297 individuals of all ethnicities, 245 of whom self-reported as Caucasian. Quantitative estimates of genomic ancestry (principal components analysis (PCA) implemented with EIGENSTRAT software [62]) were then performed and compared with self-reported ethnicity, with outliers removed. Following QC and EIGENSTRAT, full phenotypic information and genotype data were available for 1 539 908 SNPs for 223 PCA-confirmed individuals of European ancestry.

Statistical analysis

As 83% of the genotyped cohort self-reported as Caucasian, analyses were restricted to only the subset of participants PCA-confirmed as having European ancestry (N = 223) in order to reduce population heterogeneity. We used the rate of HSV-2 viral shedding (quantified by the percent of days PCR+ for HSV-2 DNA over the sampling period) to measure HSV-2 virologic severity. The association of each SNP with the quantitative viral shedding was tested by linear regression under an additive model, including all 223 individuals of European ancestry, and adjusted for age, sex, and principal component (PC) axes that significantly contributed to the variance (PC1–3). Statistical models were examined to ensure they were robust and performed well under the assumptions of the tests. We used the standardized value of \( p < 5 \times 10^{-8} \) [63–65] as the threshold for genome-wide statistical significance for our main regression analysis.

To investigate whether the top SNP and biological candidate KIF1B SNPs were tagging functional variation in nearby coding genes, LD within ±1 Mb was tested using D' among 640 previously WGS population controls of European ancestry with IRB permission for use. Functional coding variation included stop gain/loss, frame-shift, start gain/loss, non-synonymous, splice site acceptor/donor, and indel variants.

Post hoc power analysis calculations of the quantitative viral shedding trait were conducted using the Genetic Power Calculator [66] for quantitative traits, under the following assumptions: sample size was N = 223, marker allele frequency = 5%, there was no dominance (additive model), individuals were unrelated, there was perfect LD, the genome-wide statistical significance threshold was 5E-08, and either the power or detectable variance explained was varied.

Targeted analyses of candidate SNPs

Two targeted approaches were used to investigate SNPs with a higher prior probability of association with HSV-2. Given the central role of T cells in several types of chronic viral infection and prior associations with other alphaherpesviridae, a targeted analysis of the MHC gene region (hg19/Ch37 chromosome 6: 29 570 005–33 377 699 [67]) was conducted. Other non-MHC genes previously implicated in HSV pathogenesis from candidate gene studies (FASLG, TLR3, TLR2, MBL2, FAS, UNC93B1, TRAF3, TBX21, APOE, and C21orf91) were tested for enrichment of association beyond that expected under the null hypothesis. In our cohort, there were 8791 SNPs in the MHC gene region and 131 SNPs among the non-MHC candidate genes genotyped with a MAF > 5%. For targeted analyses of candidate SNPs, the threshold for statistical significance was determined using Bonferroni correction for the number of markers tested in each analysis.
Acknowledgements The authors would like to thank Dr. Amalia Magaret for her statistical input. This work was supported by NIH grants P01 AI030731 (AW and DMK) and R01 AI094019 (DMK).

Compliance with ethical standards

Conflict of interest SEK, PRS, and ASA declare that they have no conflict of interest. DMK is a consultant to GlaxoSmithKline and has research grants or contracts from Immune Design Corporation, Admedus Vaccines, Merck, and Sanofi Pasteur. DMK and AW are co-inventors on patents owned by the University of Washington. AW receives research funding from Genocea and Vical and is a consultant for Aicuris. DBG receives research funding from Janssen, AstraZeneca, Gilead, Biogen, and UCB, is a consultant for AstraZeneca, and holds a patent for IL28B findings.

References

1. Looker KJ, Magaret AS, Turner KME, Vickerman P, Gottlieb SL, Newman LM. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. PLoS ONE. 2015;10: e114989.
2. Bradley H, Markowitz LE, Gibson T, McQuillan GM. Seroprevalence of herpes simplex virus types 1 and 2-United States, 1999–2010. J Infect Dis. 2013;209:325–33.
3. Johnston C, Gottlieb SL, Wald A. Status of vaccine research and development of vaccines for herpes simplex virus. Vaccine. 2016;34:2948–52.
4. Wald A, Zeh J, Selke S, Warren T, Ryncarz AJ, Ashley R, et al. Reactivation of genital herpes simplex virus type 2 infection in asymptomatic seropositive persons. N Engl J Med. 2000;342:844–50.
5. Kimberlin DW, Rousse DJ. Clinical practice. Genit herpes. N Engl J Med. 2004;350:1970–7.
6. Centers for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines 2002. MMWR Morb Mortal Wky Rep. 2002;51:1–78.
7. STD facts - genital herpes [Internet] [cited 31 Oct 2013]. http://www.cdc.gov/std/herpes/STDFact-herpes.html. Accessed 13 Jul 2016.
8. Corey L, Wald A, Patel R, Sacks SL, Tyring SK, Warren T, et al. Once-daily valacyclovir to reduce the risk of transmission of genital herpes. N Engl J Med. 2004;350:11–20.
9. Ross K, Johnston C, Wald A. Herpes simplex virus type 2 serological testing and psychosocial harm: a systematic review. Sex Transm Infect. 2011;87:594–600.
10. Carney O, Ross E, Bunker C, Ikkos G, Mindel A. A prospective study of the psychological impact on patients with a first episode of genital herpes. Genitourin Med. 1994;70:40–5.
11. Sacks SL, Griffiths PD, Corey L, Cohen C, Cunningham A, Dusheiko GM, et al. HSV-2 transmission. Antivir Res. 2004;63 Suppl 1:827–35.
12. Freeman EE, Weiss HA, Glynn JR, Cross PL, Whitworth JA, Hayes RJ. Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies. AIDS Lond Engl. 2006;20:73–83.
13. Gray RH, Li X, Wawer MJ, Serwadda D, Sewankambo NK, Wahwire-Mangen F, et al. Determinants of HIV-1 load in subjects with early and later HIV infections, in a general-population cohort of Rakai, Uganda. J Infect Dis. 2004;189:1209–15.
14. Martin ET, Krantz E, Gottlieb SL, Magaret AS, Langenberg A, Stanberry L, et al. A pooled analysis of the effect of condoms in preventing HSV-2 acquisition. Arch Intern Med. 2009;169:1233–40.
15. Magaret AS, Mujugira A, Hughes JP, Lingappa J, Bukusi EA, DeBryun G, et al. Effect of condom use on per-act HSV-2 transmission risk in HIV-1, HSV-2-discordant couples. Clin Infect Dis Publ Infect Dis Soc Am. 2016;62:456–61.
16. Tronstein E, Johnston C, Huang M-L, Selke S, Magaret A, Warren T, et al. Genital shedding of herpes simplex virus among symptomatic and asymptomatic persons with HSV-2 infection. JAMA. 2011;305:1441–9.
17. Magaret AS, Johnston C, Wald A. Use of the designation “shedder” in mucosal detection of herpes simplex virus DNA involving repeated sampling. Sex Transm Infect. 2009;85:270–5.
18. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, et al. Evolution and diversity in human herpes simplex virus genomes. J Virol. 2013;88:1209–27.
19. Koelle DM, Magaret A, Warren T, Schellenberg GD, Wald A. APOE genotype is associated with oral herpetic lesions but not genital or oral herpes simplex virus shedding. Sex Transm Infect. 2010;86:202–6.
20. Bochud P-Y, Magaret AS, Koelle DM, Aderem A, Wald A. Polymorphisms in TLR2 are associated with increased viral shedding and lesional rate in patients with genital herpes simplex virus Type 2 infection. J Infect Dis. 2007;196:505–9.
21. Magaret A, Dong L, John M, Mallal SA, James I, Warren T, et al. HLA class I and II alleles, heterozygosity and HLA-KIR interactions are associated with rates of genital HSV shedding and lesions. Genes Immun. 2016;17:412–8.
22. Zhang S-Y, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, et al. TLR3 deficiency in patients with herpes simplex encephalitis. Science. 2007;317:1522–7.
23. Zhang S-Y, Casanova J-L. Inborn errors underlying herpes simplex encephalitis: from TLR3 to IRF3. J Exp Med. 2015;212:1342–3.
24. Lafaille FG, Pessach IM, Zhang S-Y, Ciancanelli MJ, Herman M, Abhyankar A, et al. Impaired intrinsic immunity to HSV-1 in human iPSC-derived TLR3-deficient CNS cells. Nature. 2012;491:769–73.
25. Svensson A, Tunback P, Nordstrom I, Padyukov L, Eriksson K. Polymorphisms in Toll-like receptor 3 confer natural resistance to human herpes simplex virus type 2 infection. J Gen Virol. 2012;93 (Pt 8):1717–24.
26. Svensson A, Bergin A-MH, Löwöhn G-B, Tunbäck P, Bellner L, Padyukov L, et al. A 3′-untranslated region polymorphism in the TBL2 gene encoding T-bet is a risk factor for genital herpes simplex virus type 2 infection in humans. J Gen Virol. 2008;89:2262–8.
27. Seppäläinen M, Lokki M-L, Lappalainen M, Hiltunen-Back E, Rovio AT, Kares S, et al. Mannose-binding lectin 2 gene polymorphism in recurrent herpes simplex virus 2 infection. Hum Immunol. 2009;70:218–21.
28. Chatterjee K, Dandara C, Gyllensten U, van der Merwe L, Galal U, Hoffman M, et al. A fas gene polymorphism influences herpes simplex virus type 2 infection in South African women. J Med Virol. 2010;82:2082–6.
29. Crosslin DR, Carrell DS, Burt A, Kim DS, Underwood JG, Hanna DS, et al. Genetic variation in the HLA region is associated with susceptibility to herpes zoster. Genes Immun. 2015;16:1–7.
30. Hobbs MR, Jones BB, Otterud BE, Leppert M, Kriesel JD. Identification of a herpes simplex labialis susceptibility region on 15q15.1, in a genome-wide association study of human herpes simplex labialis (cold sore) frequency: description of a cold sore susceptibility gene. J Infect Dis. 2011;204:1654–62.
31. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461:399–401.
