Published in final edited form as:

Int J Cancer. 2014 January 15; 134(2): 397–404. doi:10.1002/ijc.28351.

A candidate gene approach for virally-induced cancer with application to HIV-related Kaposi's sarcoma

Brahim Aissani¹, Howard W. Wiener¹, Kui Zhang², Richard A. Kaslow^{1,3}, Kisani M. Ogwaro⁴, Sadeep Shrestha¹, and Lisa P. Jacobson⁵

¹Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

²Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

³Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

⁴Department of Psychiatry, University of Arizona School of Medicine, Tucson, Arizona 85724, USA

⁵Department of Epidemiology, Johns Hopkins University, Baltimore, MD 21205, USA

Abstract

Like other members of the γ -herpesvirus family, human herpes virus 8 (HHV-8), the etiologic agent of classic and HIV-related Kaposi's sarcoma (HIV-KS) acquired and evolved several human genes with key immune modulatory and cellular growth control functions. The encoded viral homologs substitute for their human counterparts but escape cellular regulation, leading to uncontrolled cell proliferation. We postulated that DNA variants in the human homologs of viral genes that potentially alter the expression or the binding of the encoded factors controlling the antiviral response may facilitate viral interference. To test whether cellular homologs are candidate susceptibility genes, we evaluated the association of DNA variants in 92 immune-related genes including 7 cellular homologs with the risk for HIV-KS in a matched case and control study nested in the Multicenter AIDS Cohort Study. Low- and high-risk gene-by-gene interactions were estimated by multifactor dimensionality reduction and used as predictors in conditional logistic models. Among the most significant gene interactions at risk (OR=2.84–3.92; Bonferroni-adjusted $p=9.9\times 10^{-3}$ – 2.6×10^{-4}), three comprised human homologs of two latently expressed viral genes, cyclin D1 (*CCND1*) and interleukin-6 (*IL-6*), in conjunction with angiogenic genes (*VEGF*, *EDN-I* and *EDNRB*). At lower significance thresholds (adjusted $p < 0.05$), human homologs related to apoptosis (*CFLAR*) and chemotaxis (*CCL2*) emerged as candidates. This “proof of concept” study identified human homologs involved in the regulation of type I interferon-induced signaling, cell cycle and apoptosis potentially as important determinants of HIV-KS

Keywords

Kaposi's sarcoma; Immunodeficiency; Herpes Virus 8; Multifactor Dimensionality Reduction; Polymorphism; Genetic association

INTRODUCTION

Kaposi's sarcoma (KS) is a multifocal angioproliferative inflammatory condition well described in its classical, post-transplantation and African endemic forms before the advent of the epidemic of HIV-1/AIDS.¹ Histologically, KS evolves from a small patch of new blood vessels to a vascular tumor, with a sparse inflammatory infiltrate to nodular lesions with proliferating spindle cells.² KS cell lines have been shown to express markers of endothelial, lymphatic and smooth muscle lineages.³ Seroepidemiological studies have clearly established HHV-8 (also called Kaposi's sarcoma herpes virus, KSHV) as the etiologic agent of all clinical forms of KS. These studies have found increased KSHV seroprevalence rates in individuals at high risk for sexually transmitted diseases⁴ and in the general populations of several central and east African and Mediterranean countries.^{5, 6} Thus, KSHV infection occurs at higher rates among men who have sex with men (MSM) (~40%) and in Mediterranean populations (5%–20%), while it is less common in other unselected populations (1%–5% in North America and Northern Europe). In the context of HIV/AIDS, KS is the most common malignancy⁷, although its incidence decreased precipitously with the advent of highly active anti-retroviral therapy (HAART). In HIV-1-infected individuals other than MSM (i.e. heterosexually infected women, injecting drug users, patients with hemophilia), the incidence of HIV-KS is as low in KSHV endemic countries of Southern Europe as in countries where KSHV infection is uncommon in the general population.⁸ In other words, the high rate of KS in MSM appears to be due to both immunosuppression and shared host or behavioral risk factors.

HIV-1-infected MSM have a 50% risk of developing KS within 10 years when KSHV infection occurs prior to HIV-1 infection and within 5 years when KSHV is acquired following HIV-1 infection.⁹ Replication of these data in the MACS cohort showed that those patients who seroconverted for KSHV after HIV-1 had more than twice the risk (relative hazard 2.19; $p=0.057$) for developing KS. The risk associated with immunosuppression is apparently independent of HIV-1 infection since KSHV-infected transplant recipients have an increased cumulative risk (8–30%) for KS.^{10, 11} Collectively, these data clearly establish the degree of immunosuppression as one of the most important risk factors for KS in KSHV-infected individuals.

Few studies have explored the host genetic determinants other than those of the suspected human leukocyte antigen (HLA) system. Positive associations of HIV-KS with promoter polymorphism G-174C in the *IL-6* gene¹² and with DNA variants in the gene encoding FcγRIIIa¹³ have been reported but not replicated. We postulated that in susceptible hosts (HIV-1 and KSHV seropositive individuals), DNA polymorphisms in immune or cell cycle/apoptosis genes that alter the expression or the binding of the encoded factors involved in the antiviral response may provide favorable conditions for viral homologs to interfere with

that response. To test this hypothesis, we selected 247 SNPs in a set of 92 human homologue and non-homologue of KSHV genes involved in cellular pathways targeted by KSHV and evaluated their effects on the risk of HIV-KS. We present supportive data for the effectiveness of the proposed approach to identify candidate susceptibility genes for HIV-KS.

MATERIALS and METHODS

Study participants

Eligible participants to the Multicenter AIDS Cohort Study (MACS), a prospective longitudinal study of the natural history of HIV-1 infection among 5,622 homosexual men recruited in 1984–1985 and 1987–1990 in major metropolitan US cities.¹⁴ The cut-off date for the follow-up is January 1, 1996, when HAART became more widely available.

Ascertainment

An immunoblot-confirmed positive ELISA defined HIV-1 seropositivity. Standardized T-cell phenotyping was performed at each follow-up visit. KSHV antibodies against KSHV lytic antigens were determined by use of an indirect immunofluorescence assay using 10-Q-tetradecanoyl phorbol 13-acetate induced body cavity B cell lymphoma-1 cells containing the KSHV genome. For each batch of serum samples tested, known KSHV-positive and -negative sera were assayed. All serum samples were tested twice in a blinded fashion and were assessed microscopically for the presence of whole cell immunofluorescence by the same reader. Sera from the enrollment or the immediately subsequent visit and from the most recent visit were tested. Positivity at either sample defined a KSHV-infected individual and KSHV negativity at both visits defined an uninfected individual.

Study design

The study sample consisted of 360 matched pairs of cases and controls predominantly (88%) of European American descent. Cases were defined as dually (HIV-1 and KSHV) infected individuals who later developed KS and controls as dually infected individuals who were free of KS. Controls were matched to cases by HIV/KSHV serostatus, race, KS-free time and CD4+ T lymphocyte cell counts (henceforth CD4+ counts). To account for the reported influence of the temporal order of HIV-1 and KSHV infections on progression to KS, cases were matched by controls within each of the four different serostatus groups [HIV-1-seroprevalent (SP)/KSHV-SP; HIV-1-SP/KSHV-seroconverter (SC); HIV-1-SC/KSHV-SP and HIV-1-SC/KSHV-SC]. Time to KS was estimated as the date of KS diagnosis in index minus the baseline date or minus the date of the seroconversion date, defined as the mid-point between the last HIV-1- or KSHV-negative and the first positive date. Case and control pairs were matched for CD4+ counts at the visit within 1 year prior to index diagnosis.

SNP selection and genotyping

KSHV encodes at least 15 known human homologs implicated in immunoevasive pathways; however, at the time this study was designed, validated SNP assays were not available for all known human counterpart of KSHV genes. We therefore expanded the selection to include

genes with documented relevance to KS pathogenesis (e.g. genes with upregulated or downregulated expression in KS lesions or KS cell lines and/or genes regulating angiogenesis). Overall, screening of Phase I HapMap resource and other resources (NCI SNP500Cancer, Seattle SNPs, Perlegen and NIEHS SNPs) identified an initial set of 284 SNPs from a selected set of 96 genes.

High-throughput genotyping was performed on the Illumina BeadArray® platform (Illumina Inc., San Diego, CA) for most SNPs. For a subset of the SNP selection, typing was carried out using custom TaqMan® assays from ABI (Applied Biosystems, Inc.) or bi-directional Sanger sequencing (Polymorphic DNA Inc., Alameda, CA) when several SNPs lie within short DNA segments such as in gene promoters.

Quality Control (QC)

Reliability in our typing data was assessed by a small set of intra- and inter-plates blind duplicates. SNP calls were checked for adherence to Hardy-Weinberg equilibrium (HWE) in each of the KS outcome categories and only SNPs showing no significant deviation ($p > 0.01$) from HWE in KS-free controls were included for analyses.

Statistical Analysis

Population composition—To minimize type-I error associated with population structure and to allow for a maximum number of cases and controls to be incorporated in our paired analyses, principal component analysis (PCA) and discriminant analysis were used to identify misclassified samples in the self-report of ancestry. A combined dataset based on 743 SNPs (247 SNPs from the present study and 496 SNPs from another investigation of the same study sample) was used to calculate the principal components.

Covariates—We controlled for the confounding effects of age (age at the time of seroconversion to HIV-1 and KSHV infections), of the degree of immunosuppression and of the rate of immunity decline over the exposure time (time from dual infection to index diagnosis). The degree of immunosuppression was measured by the area under the curve (AUC), which was estimated by the trapezoidal method¹⁵ as the time-weighted sum of the CD4+ counts below the cut-off level of 200/mm³ integrated over the exposure time. The slope of the CD4+ count changes over the exposure time measured the rate of immunity decline. Data points with more than 3 standard deviations from the mean were considered outliers and were not included in the calculation of the slope.

Association testing—To detect gene by gene interactions, which we have hypothesized to be the most likely determinants of risk because of the multiple cellular pathways targeted by KSHV, we adopted the reduction strategy used in GMDR (Generalized Multifactor Dimensionality Reduction)¹⁶ and in the original MDR approach.¹⁷ We first imputed the missing genotypes separately for the case and control samples on the basis of pair of haplotypes estimated via the EM-based haplotype inference program HAPLORE¹⁸ (<http://www.soph.uab.edu/Statgenetics/People/KZhang/HAPLORE/>). We then evaluated the effects of all possible 2- and 3-way gene interaction models instead of limiting the analysis to the best gene interaction models as in the MDR and GMDR approaches. An in-house algorithm

in R was developed to classify SNP genotypes in the interaction terms as high or low risk variables and modeled them as dichotomous predictors in conditional logistic regression analysis. We performed the analysis under an additive genetic model and treated the low risk variable as reference in the estimation of the odds ratios (OR). Baseline age, AUC and the slope of CD4+ counts were used as covariates and the affection status as the outcome variable.

For the purpose of comparisons, we also evaluated the marginal effects of the studied SNPs using conditional logistic regression analysis with adjustments for covariates as indicated above. For each of the above tests, we derived the odds ratios and 95% confidence limits, and the p-values for the test of significance after adjustment for multiple testing using Bonferroni correction.

System biology

To validate our hypothesis, we tested whether the most significant genes identified in this study act in immune regulatory pathways relevant to HIV-KS pathogenesis. To this end, we used g:Profiler (<http://biit.cs.ut.ee/gprofiler/index.cgi>), a web based server for functional profiling¹⁹, to query several ontology databases including but not limited to Gene Ontology (GO)²⁰, Human Phenotype Ontology (HPO)²¹ and the biological pathway databases KEGG²² and Reactome.²³ We queried ontology databases with two different sets of candidate genes; those from the most significant (Bonferroni-adjusted $p=10^{-3}$) gene interaction models or those occurring in at least 2 significant (Bonferroni-adjusted $p < 0.05$) interaction models (top 28 genes). A stringent threshold ($p=10^{-4}$) adjusted for multiple testing using the Benjamini-Hochberg FDR approach was set to declare statistical significance. The null hypothesis H_0 was that given a set of input genes Q as query, and genes associated to GO (KEGG, HPO...etc) term T, the number of common elements to Q and T, i.e. the intersection of Q and T, has appeared by random chance.

RESULTS

Quality control of the data indicated that 11 samples (0.9%) failed typing and among the remaining 709 typed samples, the call rate was greater than 98% in more than 92% of the sampled individuals. The overall concordance rate between the duplicates was 99.4%, implying that false finding due to typing errors or mishandling of the samples is unlikely. Of the 284 SNPs assayed, 12 (~ 4%) did not meet the HWE criterion in controls and were excluded.

Concordant with the self-report of race, PCA yielded 2 distinct clusters representing populations of European and African American descents (supplementary Fig. S1; panel A). Discriminant analysis identified 22 misclassified individuals who were then reclassified on the basis of the first two principal components (supplementary Fig. S1; panel B). The majority of these individuals self-reported as Hispanic Europeans indicating that population admixture may explain the misclassification. Re-matching of the cases and controls within this subset of 22 individuals resulted in a loss of 3 additional individuals because no matched cases or controls could be identified.

With our study sample composed of men only, a subset of 23 SNPs from chromosome X and 2 SNPs from chromosome Y were not eligible for MDR analysis because this approach uses genotype data. Based on a final set of 247 eligible SNPs (Supplementary Table S1) in 92 genes, we show that no SNP main effect or 2-way gene interaction remained significant at $p=0.05$ after Bonferroni correction (not shown). However, 60 gene interaction models reached significance in the 3-way analysis, of which the most significant 20 models (Bonferroni-adjusted $p < 0.01$) are depicted in Supplementary Table S2. None of these 20 interaction models involved any of the 8 negative control SNPs composed of null Ancestry Informative Markers (AIM) or SNPs located in large gene deserts. Examination of gene identity in the 11 top gene interaction models (Bonferroni-adjusted $p < 0.001$) revealed *CCND1* (cyclin D1 gene) and *IL-6* (interleukin-6, alias interferon, beta 2, gene), two homologue genes expressed by KSHV, as important determinants of risk for HIV-KS (OR=2.84–3.92; $p=9.9\times 10^{-3}$ – 4×10^{-3}) (Table 1). In several instances the associated gene interactions implicate non-synonymous and functional non-coding SNPs (highlighted in bold, by parentheses or underlined in Table 1) with known variants at risk; for example the IL-6 high-producing gene promoter variant G-174C (rs1800795) and the G870A splice variant (rs603965, merged to rs9344) of *CCND1* influencing the relative production of Cyclin D1b. Consistent with the above findings is the observation that IL-6R, the cognate receptor of IL-6, also emerged in the interaction model 11 implicating the same *CCND1* variant. As can be reasonably predicted for this tumor of endothelial origin, among the most significant interaction models identified in this study are those (models 3, 8, 9 and 11) implicating genes encoding angiogenic factors (VEGF, EDN1, EDNRB) central to HIV-KS.

Because our study is not statistically powered to detect interactions beyond the 3-way interaction order, we used system biology to search for cellular pathways or functions shared by the interacting genes in models 1 to 11. Queries of gene ontology databases with the 20 genes defining models 1 to 11 (Table 1) showed that the most likely cellular pathways are in the order of decreased importance those involved in i) regulation of smooth muscle cell proliferation (Q&T/T=0.057; FDR-adjusted $p=9.0\times 10^{-6}$), ii) leukocyte chemotaxis (Q&T/T=0.04; FDR-adjusted $p=5.6\times 10^{-5}$) and iii) negative regulation of endopeptidase activity (Q&T/T=0.038; FDR-adjusted $p=5.3\times 10^{-5}$) (Table 2). For increased specificity, we then restricted the search to cellular pathways defined by T (the number of genes associated to GO) of less than 100 and found that only cellular pathway # 3 (regulation of smooth muscle cell proliferation) remained significant at FDR-adjusted $p<10^{-4}$.

To ensure that no candidate genes were missed in our stringent sampling of only the most significant (FDR-adjusted $p < 10^{-2}$) gene interaction models, we extended the input query genes (Q) to include all genes that occurred in at least two interaction models with FDR-adjusted $p < 0.05$. This led to the identification of 8 additional genes including 2 human genes acquired by KSHV, *CCL2* (C-C chemokine ligand 2) and *CFLAR* (CASP8 and FADD-like apoptosis regulator) (Supplementary Table S3). Querying GO databases with the entire set of 28 genes ordered in decreased importance (frequency of occurrence in significant interaction models) or unordered identified a third pathway (positive regulation of JAK-STAT cascade, pathway #3, Table 3) in addition to pathways #3 and #4 of Table 2. It is worth emphasizing that at a stringent cut-off ($T = 100$), genes like *CCND1* that function

far downstream in the signal transduction cascade are not among the listed genes in the candidate pathways; this is consistent with the JAK-STAT cascade being an early event in IFN signaling.

DISCUSSION

In this first attempt to test the hypothesis that human homologue of KSHV genes are potential susceptibility genes for HIV-KS, four (*CCND1*, *IL-6*, *CCL2* and *CFLAR*) out of seven assayed cellular homologs, showed specific gene interactions conferring up to 3-fold increase of risk for KS. No simple model can predict how these gene variants favor or abrogate viral interference. Plausible models include different types of competition for common cognate receptors or binding partners, with the control of the immune response by the virus or the host being dictated by the polymorphisms affecting the expression or binding affinity of target host homologs or their binding partners (receptors or co-factors). Furthermore, study of system biology suggested that regulation of smooth muscle cell proliferation and leukocyte chemotaxis are the most likely target cellular processes in HIV-KS, consistent with the current belief that the proliferating spindle-shaped cells in KS lesions are of endothelial, lymphatic and smooth muscle lineages.³

The lack of significant main and two-way gene interaction effects in the presence of three-way effects deserves further comments. A straightforward explanation would be that the joint action, additive or combinatorial, of undetectable and numerous weak main effects predispose to higher risks for HIV-KS. While this is currently considered to be more the rule than the exception in the etiology of cancer and complex diseases in general, understanding how genetic interactions lead to the expression of disease phenotypes is elusive. Nonetheless, since gene effects are context dependent²⁴ and molecular mimicry, as a pathogenic mechanism, interferes and abrogates the regulatory activity of numerous cellular genes, an alternative explanation for the absence of detectable main effects is locus heterogeneity (non-allelic heterogeneity) that diminishes the power to detect them. Under the proposed model, the affected cases carry risk variants in different combinations of homologue and non-homologue genes but acting in the same cellular pathways antagonized by KSHV or relevant to KS such as in angiogenesis. This model is not inconsistent with our detection of several 3-way gene interactions in which the interacting gene(s) function in specific cellular pathways targeted by KSHV. For example, *CCND1* was observed in 4 distinct gene interactions, some of which implicating genes involved in the regulation of type I IFN-induced signaling, an immune response heavily antagonized by KSHV through expression of pleiotropic viral IRF homologs (vIRF-1-3).

Members of the IRF family play central roles in one of the early immune responses to viral infection through transcription of type I IFN genes, IFN-stimulated genes and other cytokines and chemokines. Although none of the four cellular *IRF* genes tested in this study (*IRF-1*, *IRF-4*, *IRF-5* and *IRF-8*) was found in the top 11 gene interaction models, two of them (*IRF-5* and *IRF-8*) occurred in more than two significantly associated interaction models. Human IRF-8 (hIRF-8) is the closest homolog (22% amino-acid identity) of vIRF-1 encoded by *ORFK9*; it is considered as a tumor suppressor and mutations in hIRF8 cause immunodeficiency and chronic myelogenous leukemia-like syndrome in mice.

The classification in latent and lytic replication patterns is relevant to our study because latency is generally assumed to be the state leading to cell proliferation, whereas lytic replication results in cell death and is therefore by definition antitumorigenic. Interestingly, viral homologs of *CCND1*, *vCYC* (ORF72), of *IL-6*, *vIL-6* (*K2*), and of *CFLAR*, *vFLIP* (*K13*) encode latently expressed molecules. If replicated; this would imply that cellular homologs of latently expressed viral genes are potentially important determinants of HIV-KS.

We have hypothesized that the gene encoding the pleiotropic cytokine IL-6 is a strong candidate susceptibility gene because *vIL-6* is present only in the *Rhadinoviruses* and KSHV is the sole human representative of this γ -2 herpesvirus subfamily. In-vitro studies have shown that, despite their similar structure (26% amino-acid identity) and downstream signaling activity, hIL-6 and vIL-6 have different receptor requirements and utilization. The subversive activity of vIL-6 resides in its independence of the IL-6 α receptor (IL-6R; gp80 or CD126), which is down-regulated by IFNs during KSHV infection.²⁵ In classical signaling, hIL-6 first forms a complex with the membrane-bound gp80 subunit which in turns associates with and activates the signal-transducing β -receptor chain gp130 (encoded by *IL-6ST*). In contrast, vIL-6 bypasses IL-6R and directly binds gp130. vIL-6 inhibits Tyk2 and STAT2 phosphorylation by IFN- α thereby blocking downstream signaling and IFN induction of the tumor suppressor p21. The presence of IFN-stimulated response elements (ISREs) in the promoter of vIL-6 further enhances the expression of vIL-6 by IFN- α and provides a novel negative feedback mechanism. hIL-6 cannot reproduce this effect because IFN- α down-regulates surface expression of IL-6R.

Many of the biological activities of hIL-6 are mediated by trans-signaling through a naturally occurring soluble IL-6R (sIL-6R), which is either a splice variant of IL-6R or a product of a proteolytic cleavage. In the trans-signaling mode, signal transduction by hIL-6 relies solely on the extracellular concentration of sIL-6R whereas that of vIL-6 is independent of it. Because trans-signaling is counteracted by a soluble form of gp130 (sgp130, sIL-6ST), which inhibits the activity of the [sIL-6R/hIL-6] complex, competition between vIL-6 and hIL-6 for the membrane-bound IL-6ST is a possible pathogenic mechanism of KS. In this scenario, genetic polymorphisms inducing low expression of hIL-6 and/or high levels of sIL-6R would favor trans-signaling by vIL-6. Our study revealed the -174 G>C promoter polymorphism of *IL-6* (rs1800795) and the *IL-6R* SNP rs4129267, which is associated with a high level of serum level of sIL-6R ($p=10^{-57}$)²⁶, as gene interaction variants conferring excess risk for HIV-KS. In line with our hypothesis is a report of an experimental evidence for trans-signaling and expression of sgp130 in HIV-KS cell cultures.²⁷

The observation of several independent gene interactions comprising *CCND1* is appealing. First, this gene encodes the proto-oncogene cyclin D1, a powerful regulator of cell cycle progression extensively investigated in human cancer.²⁸ Second, polymorphism rs603965 (SNP 28) is the splice donor site G870A (Pro241Pro) that produces cyclin D1b, an alternate splice variant overexpressed in several cancer and associated with poor disease outcome.²⁹ Third, regulation by cyclin D1 is a late event in the mitogenic signal transduction cascade;

this suggests that dysregulation of cyclin D1 activity is a determinant and necessary event prior to engagement of the cell in the neoplastic process.

As outlined earlier, our study was not powered to detect higher order gene interactions; nevertheless, several identified gene interactions are predicted by system biology to take place in the same cascade of signaling events or in distinct cellular pathways connected by pluripotent molecules. The pluripotent IL-6 cytokine is important in both physiological and pathological angiogenesis and its role in VEGF production in some cancer have been reported^{30, 31}; thus IL-6 may bridge IFN-mediated cell cycle arrest and apoptosis (abrogated by vIL-6), and angiogenesis in KS lesions.

Endothelin-1 (ET-1), the product of *EDN1*, plays a crucial role in migration and metastasis of human cancer cells^{32, 33} and is likely to be up-regulated in multifocal angioproliferative conditions such as KS. Indeed, overexpression of this angiogenic factor and its receptor ETBR, which is encoded by *EDRNB*, is a hallmark of KS development.^{34, 35}

It is not quite clear how the other interaction models identified in this study relate to each other; our data suggest possible cross-talk between the identified signaling pathways, with the angiogenic ET-1 and IL-6 possibly as key molecules linking the pathogenic pathways. The identification of *IL-18*, *IL-6*, *CCL2* and *CXCR5* as potential candidate susceptibility genes suggests a skewed antiviral immune response towards the less effective T_H2-cell response as a possible mechanism of KS pathogenesis.

Implicit in this approach to selecting candidate genes for HIV-KS is the evaluation of a strategy, a proof of concept for investigating cancer induced not only by the viruses of the *herpesviridae* family but also by those of other families that have similarly captured and evolved cellular genes for immune evasion.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) with centers (Principal Investigators) at The Johns Hopkins University Bloomberg School of Public Health (Joseph B. Margolick, Lisa Jacobson), Howard Brown Health Center and Northwestern University Medical School (John Phair), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles Rinaldo). Website located at <http://www.statepi.jhsph.edu/mac/mac.html>.

Grant support

This work was supported by the National Institutes of Health administrative supplement (BA) to R01-CA106168 (RAK). The Multicenter AIDS Cohort Study (MACS) is funded by the National Institute of Allergy and Infectious Diseases, with additional supplementary funding from the National Cancer Institute and the National Heart, Lung and Blood Institute, U01-AI-35042, 5-MO1-RR-00722 (GCRC), U01-AI-35043, U01-AI-37984, U01-AI-35039, U01-AI-35040, U01-AI-37613, U01-AI-35041.

Abbreviations

KS Kaposi's sarcoma

HIV	Human Immunodeficiency Virus
KSHV	Kaposi's sarcoma herpes virus
HHV-8	human herpes virus 8
AIDS	Acquired ImmunoDeficiency Syndrom
HAART	Highly active antiretroviral therapy
MSM	men who have sex with men
MACS	Multicenter AIDS Cohort Study
HWE	Hardy Weinberg equilibrium
HLA	human leukocyte antigen
CCND1	cyclin D1 gene
IL-6	interleukin-6
CCL2	C-C chemokine ligand 2
CFLAR	CASP8 and FADD-like apoptosis regulator

References

1. Cathomas G. Kaposi's sarcoma-associated herpesvirus (KSHV)/human herpesvirus 8 (HHV-8) as a tumour virus. *Herpes*. 2003; 10:72–7. [PubMed: 14759339]
2. Cathomas G. Human herpes virus 8: a new virus discloses its face. *Virchows Arch*. 2000; 436:195–206. [PubMed: 10782877]
3. Hengge UR, Ruzicka T, Tyring SK, Stuschke M, Roggendorf M, Schwartz RA, Seeber S. Update on Kaposi's sarcoma and other HHV8 associated diseases. Part 1: epidemiology, environmental predispositions, clinical manifestations, and therapy [SEE PART II in LID Issue 6]. *Lancet Infect Dis*. 2002; 2:281–92. [PubMed: 12062994]
4. Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, Lee B, Doms RW, Margolick J, Buchbinder S, Goedert JJ, O'Brien TR, et al. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science*. 1998; 282:1907–11. [PubMed: 9836644]
5. Edelman DC. Human herpesvirus 8—a novel human pathogen. *Virology*. 2005; 2:78. [PubMed: 16138925]
6. Franceschi S, Geddes M. Epidemiology of classic Kaposi's sarcoma, with special reference to Mediterranean population. *Tumori*. 1995; 81:308–14. [PubMed: 8804445]
7. Goedert JJ. The epidemiology of acquired immunodeficiency syndrome malignancies. *Semin Oncol*. 2000; 27:390–401. [PubMed: 10950365]
8. International Agency of Research on Cancer. IARC Monographs on the evaluation of carcinogenic risks to humans. Lyon, France: IARC Press; 1996. Vol. 67: Human Immunodeficiency Viruses
9. Renwick N, Halaby T, Weverling GJ, Dukers NH, Simpson GR, Coutinho RA, Lange JM, Schulz TF, Goudsmit J. Seroconversion for human herpesvirus 8 during HIV infection is highly predictive of Kaposi's sarcoma. *AIDS*. 1998; 12:2481–8. [PubMed: 9875587]
10. Regamey N, Cathomas G. HHV-8: a newly recognized pathogen in transplantation. *Transplantation*. 2000; 69:1768–9. [PubMed: 10830205]
11. Frances C, Mouquet C, Marcelin AG, Barete S, Agher R, Charron D, Benalia H, Dupin N, Piette JC, Bitker MO, Calvez V. Outcome of kidney transplant recipients with previous human herpesvirus-8 infection. *Transplantation*. 2000; 69:1776–9. [PubMed: 10830210]

12. Foster CB, Lehrnbecher T, Samuels S, Stein S, Mol F, Metcalf JA, Wyvill K, Steinberg SM, Kovacs J, Blauvelt A, Yarchoan R, Chanock SJ. An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency virus. *Blood*. 2000; 96:2562–7. [PubMed: 11001912]
13. Lehrnbecher TL, Foster CB, Zhu S, Venzon D, Steinberg SM, Wyvill K, Metcalf JA, Cohen SS, Kovacs J, Yarchoan R, Blauvelt A, Chanock SJ. Variant genotypes of FcγRIIIA influence the development of Kaposi's sarcoma in HIV-infected men. *Blood*. 2000; 95:2386–90. [PubMed: 10733511]
14. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR Jr. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol*. 1987; 126:310–8. [PubMed: 3300281]
15. Cole SR, Napravnik S, Mugavero MJ, Lau B, Eron JJ Jr, Saag MS. Copy-years viremia as a measure of cumulative human immunodeficiency virus viral burden. *Am J Epidemiol*. 2010; 171:198–205. [PubMed: 20007202]
16. Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, Li MD. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet*. 2007; 80:1125–37. [PubMed: 17503330]
17. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*. 2001; 69:138–47. [PubMed: 11404819]
18. Zhang K, Sun F, Zhao H. HAPLORE: a program for haplotype reconstruction in general pedigrees without recombination. *Bioinformatics*. 2005; 21:90–103. [PubMed: 15231536]
19. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 2007; 35:W193–200. [PubMed: 17478515]
20. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–9. [PubMed: 10802651]
21. Robinson PN, Mundlos S. The human phenotype ontology. *Clin Genet*. 2010; 77:525–34. [PubMed: 20412080]
22. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010; 38:D355–60. [PubMed: 19880382]
23. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*. 2009; 37:D619–22. [PubMed: 18981052]
24. Kacser H, Burns JA. The molecular basis of dominance. *Genetics*. 1981; 97:639–66. [PubMed: 7297851]
25. Chatterjee M, Osborne J, Bestetti G, Chang Y, Moore PS. Viral IL-6-induced cell proliferation and immune evasion of interferon activity. *Science*. 2002; 298:1432–5. [PubMed: 12434062]
26. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, Lauretani F, Murray A, Gibbs JR, Paolisso G, Rafiq S, Simon-Sanchez J, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet*. 2008; 4:e1000072. [PubMed: 18464913]
27. Murakami-Mori K, Taga T, Kishimoto T, Nakamura S. The soluble form of the IL-6 receptor (sIL-6R α) is a potent growth factor for AIDS-associated Kaposi's sarcoma (KS) cells; the soluble form of gp130 is antagonistic for sIL-6R α-induced AIDS-KS cell growth. *Int Immunol*. 1996; 8:595–602. [PubMed: 8722639]
28. Knudsen KE, Diehl JA, Haiman CA, Knudsen ES. Cyclin D1: polymorphism, aberrant splicing and cancer risk. *Oncogene*. 2006; 25:1620–8. [PubMed: 16550162]
29. Sergeantanis TN, Economopoulos KP. Cyclin D1 G870A polymorphism and breast cancer risk: a meta-analysis comprising 9,911 cases and 11,171 controls. *Mol Biol Rep*. 2011; 38:4955–63. [PubMed: 21161398]

30. Su JL, Lai KP, Chen CA, Yang CY, Chen PS, Chang CC, Chou CH, Hu CL, Kuo ML, Hsieh CY, Wei LH. A novel peptide specifically binding to interleukin-6 receptor (gp80) inhibits angiogenesis and tumor growth. *Cancer Res.* 2005; 65:4827–35. [PubMed: 15930303]
31. Mahnke JL, Dawood MY, Huang JC. Vascular endothelial growth factor and interleukin-6 in peritoneal fluid of women with endometriosis. *Fertil Steril.* 2000; 73:166–70. [PubMed: 10632434]
32. Nelson J, Bagnato A, Battistini B, Nisen P. The endothelin axis: emerging role in cancer. *Nat Rev Cancer.* 2003; 3:110–6. [PubMed: 12563310]
33. Bagnato A, Tecce R, Moretti C, Di Castro V, Spergel D, Catt KJ. Autocrine actions of endothelin-1 as a growth factor in human ovarian carcinoma cells. *Clin Cancer Res.* 1995; 1:1059–66. [PubMed: 9816080]
34. Rosano L, Spinella F, Di Castro V, Nicotra MR, Albini A, Natali PG, Bagnato A. Endothelin receptor blockade inhibits molecular effectors of Kaposi's sarcoma cell invasion and tumor growth in vivo. *Am J Pathol.* 2003; 163:753–62. [PubMed: 12875994]
35. Bagnato A, Rosano L, Di Castro V, Albini A, Salani D, Varmi M, Nicotra MR, Natali PG. Endothelin receptor blockade inhibits proliferation of Kaposi's sarcoma cells. *Am J Pathol.* 2001; 158:841–7. [PubMed: 11238033]

Impact

Molecular mimicry is an important immune-evasion strategy used by members of the herpesviridae family to escape cellular control. Seroepidemiological studies have established human herpes virus 8 as the etiologic agent of all clinical forms of Kaposi's sarcoma. This study exploits molecular mimicry to identify host susceptibility genes in HIV-related Kaposi's sarcoma.

Table 1

Best three-way gene interaction models at risk for HIV-related Kaposi's sarcoma

model #	SNP #	SNP description ^b	best gene models ^c	OR ^a	Lower 95% CL	Upper 95% CL	Bonferroni-corrected p
1	100-20-27	(rs2043055)-rs220022-rs598493	<i>IL18-HIVEP1-TNFAIP3</i>	3.33	2.31	4.79	2.6 E-04
2	39-52-22	rs745191 -rs1047033-rs318477	<i>AKAP13-ID4-SERPINB9</i>	3.17	2.22	4.53	5.2 E-04
3	79-165-244	rs1794849-rs3181245-rs9504743	<i>EDN1-TDP2-F13A1</i>	3.06	2.15	4.35	1.2 E-03
4	28-36-66	rs603965-rs730775-rs1474347	<i>CCND1-NFKBIE-IL6</i>	3.11	2.17	4.45	1.4 E-03
5	28-36-86	rs603965-rs730775-(rs1800795)	<i>CCND1-NFKBIE-IL6</i>	2.99	2.10	4.25	2.8 E-03
6	3-125-238	<u>rs3922</u> - rs228212 -rs9457935	<i>CXCR5-HIVEP1-LPA</i>	3.07	2.14	4.42	3.1 E-03
7	28-36-107	rs603965-rs730775-rs2069833	<i>CCND1-NFKBIE-IL6</i>	2.94	2.07	4.18	4.0 E-03
8	28-73-215	rs603965-rs1629862-rs6918586	<i>CCND1-EDN1-HFE</i>	3.92	2.51	6.14	5.4 E-03
9	182-39-22	rs3818416- rs745191 -rs318477	<i>EDNRB-AKAP13-SERPINB9</i>	3.02	2.10	4.34	5.5 E-03
10	28-145-41	rs603965-rs2540489-(rs849877)	<i>CCND1-LTA4H-PRL</i>	2.98	2.07	4.29	9.9 E-03
11	28-40-284	rs603965-rs833069-rs4129267	<i>CCND1-VEGF-IL6R</i>	2.84	2.01	4.03	9.9 E-03

^a Estimates of the odds ratio (OR) and confidence limits (CL) from conditional logistic models adjusted for the effects of age at the time of seroconversion to HIV-1 and KSHV infections and of immunosuppression (area under the curve and slope of CD4+ counts) from the time of dual seroconversion to the time of diagnosis of Kaposi's sarcoma in the matched cases and back imputation in the controls.

^b Non-synonymous SNPs (in bold), SNPs in gene promoters (in parentheses) and SNPs in 3' UTR (underlined).

^c Cellular genes acquired by KSHV (in bold). *IL18*, interleukin 18 (interferon-gamma-inducing factor); *HIVEP1*, human immunodeficiency virus type I enhancer binding protein 1; *TNFAIP3*, tumor necrosis factor, alpha-induced protein 3; *AKAP13*, AKAP13 A kinase (PRKA) anchor protein 13; *ID4*, inhibitor of DNA binding 4, dominant negative helix-loop-helix protein; *SERPINB9*, serpin peptidase inhibitor; *EDN1*, endothelin 1; *TDP2*, tyrosyl-DNA phosphodiesterase 2; *F13A1*, coagulation factor XIII, A1 polypeptide; *CCND1*, cyclin D1; *NFKBIE*, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon; *IL6*, interleukin 6 (interferon, beta 2); *CXCR5*, chemokine (C-X-C motif) receptor 5; *LPA*, lipoprotein; *HFE*, hemochromatosis; *EDNRB*, endothelin receptor type B; *LTA4H*, leukotriene A4 hydrolase; *PRL*, prolactin; *VEGF*, vascular endothelial growth factor A; *IL6R*, interleukin 6 receptor.

Table 2
Gene ontology for the genes in the most significant three-way gene interaction models

Obs	p-value	T	Q	Q&T	Q&T/Q	Q&T/T	ID	name	gene list
1	4.4 E-05	647	20	<u>8</u>	0.4	0.012	GO:0033993	response to lipid	IL-18, TNFAIP3, AKAP13, SERPINB9, EDN1, CCND1, IL6, IL6R
2	5.3 E-05	131	19	<u>5</u>	0.263	0.038	GO:0010951	negative regulation of endopeptidase activity	TNFAIP3, SERPINB9, IL6, LPA, VEGF
3	9.0 E-06	87	20	<u>5</u>	0.25	0.057	GO:0048660	regulation of smooth muscle cell proliferation	TNFAIP3, EDN1, IL6, VEGF, IL6R
4	5.6 E-05	125	20	<u>5</u>	0.25	0.04	GO:0030595	leukocyte chemotaxis	EDN1, IL6, EDNRB, VEGF, IL6R
5	9.8 E-06	781	20	<u>9</u>	0.45	0.012	GO:0009725	response to hormone stimulus	IL18, AKAP13, SERPINB9, EDN1, CCND1, IL6, LTA4H, PRL, IL6R
6	8.9 E-05	709	20	<u>8</u>	0.4	0.011	GO:0008284	positive regulation of cell proliferation	IL18, TNFAIP3, ID4, EDN1, CCND1, IL6, VEGF, IL6R

Gene ontology databases Gene Ontology (GO), Human Phenotype Ontology (HPO) and the biological pathway databases KEGG and Reactome were queried via g:Profiler, a web server for functional profiling, using as input gene query (Q) the set of 20 genes identified in the eleven most significant (Bonferroni-adjusted $p < 10^{-2}$) gene interaction models shown in Table 1. Only gene ontology pathways predicted at FDR-adjusted $p < 10^{-4}$ are shown. T denotes the number of genes associated to GO (KEGG, HPO...etc) terms. Q&T is the number of genes shared between Q and T (listed in the last column). The most important parameter is Q&T/T which represents the percentage of query genes found in a given GO pathway. Note that only pathway #3 (regulation of smooth muscle cell proliferation) remains significant when a more stringent cut-off ($T=100$) for the number of functional categories was used.

Table 3
Gene ontology for the most frequently occurring genes in gene interactions at risk for HIV-KS

Obs	p-value	T	Q	Q&T	Q&T/Q	Q&T/T	ID	name	gene list
1	4.3 E-06	87	25	<u>6</u>	0.24	0.069	GO:0048660	regulation of smooth muscle cell proliferation	IL6, IL6R, VEGF, EDN1, TNFAIP3, IFNG
2	2.6 E-06	55	16	<u>5</u>	0.312	0.091	GO:0002690	positive regulation of leukocyte chemotaxis	IL6, IL6R, VEGF, EDN1, CCL2
3	3.8 E-07	52	28	<u>6</u>	0.214	0.115	GO:0046427	positive regulation of JAK-STAT cascade	IL6, IL6R, PRL, TNFRSF1A, IFNG, CD40
4	4.1 E-07	52	28	<u>6</u>	0.214	0.115	GO:0046427	positive regulation of JAK-STAT cascade	IL6, IL6R, PRL, TNFRSF1A, IFNG, CD40
5	9.9 E-06	87	28	<u>6</u>	0.214	0.069	GO:0048660	regulation of smooth muscle cell proliferation	IL6, IL6R, VEGF, EDN1, TNFAIP3, IFNG

Gene ontology databases Gene Ontology (GO), Human Phenotype Ontology (HPO) and the biological pathway databases KEGG and Reactome were queried via g:Profiler, a web server for functional profiling, using as input gene query (Q) the set of 28 genes (Supplementary Table S2) that occurred more than once in 60 significantly associated (Bonferroni-adjusted $p < 0.05$) gene interaction models. In querying g:Profiler, input genes were entered either as an ordered list of genes with decreasing order of frequency (pathways 1, 2 and 3) or unordered (pathways 4 and 5). Only gene ontology pathways predicted at a FDR-adjusted $p < 10^{-5}$ are shown. T denotes the number of genes associated to GO (KEGG, HPO...etc) terms. Q&T is the number of genes shared between Q and T (listed in the last column). The most important parameter is Q&T/T which represents the percentage of query genes found in a given GO pathway.