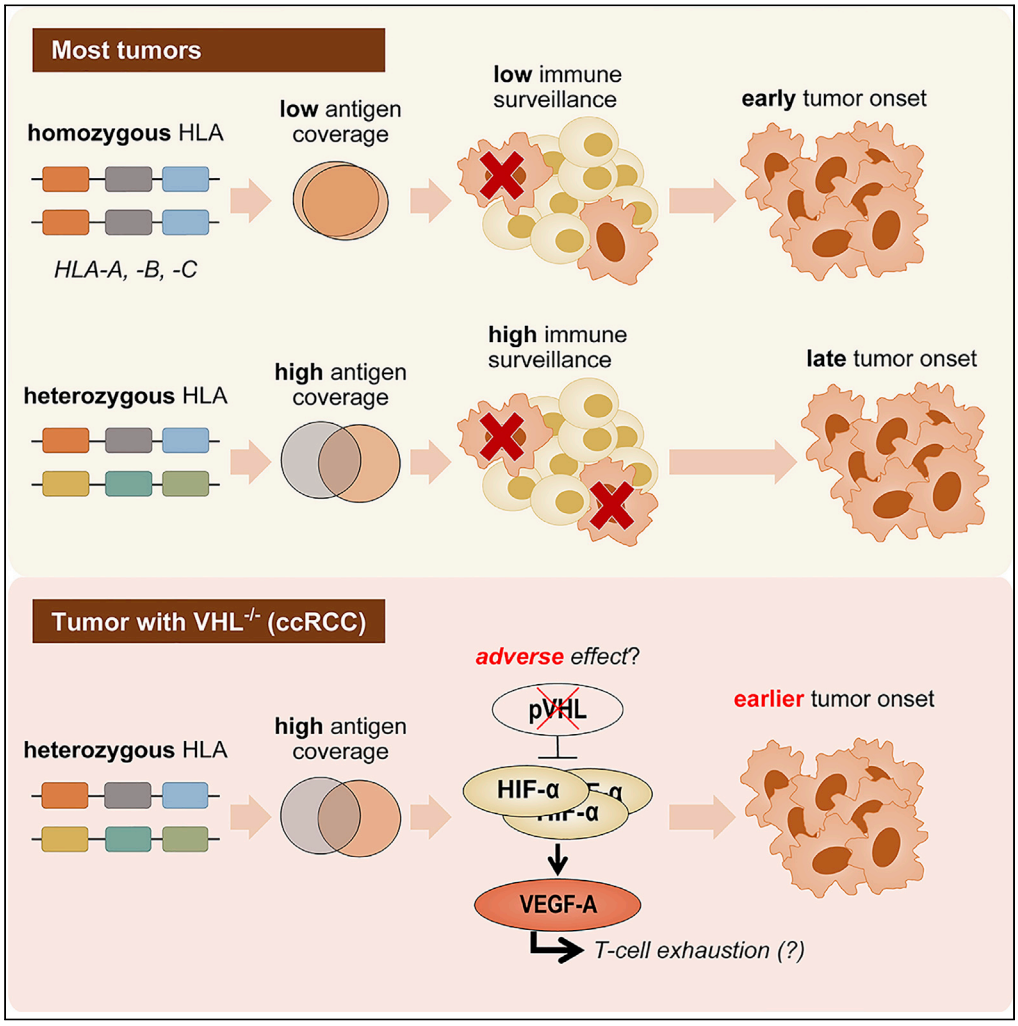
iScience

CellPress
OPEN ACCESS

Article

HLA-I-restricted CD8$^+$ T cell immunity may accelerate tumorigenesis in conjunction with VHL inactivation



BeumJin Park, Seok-Jae Heo, Yong Joon Lee, ..., Eui-Cheol Shin, Inkyung Jung, Sangwoo Kim

swkim@yuhs.ac

Highlights

HLA homozygosity reduces antigen coverage and is associated with earlier tumor onset

Tumors with VHL$^{-/-}$, such as ccRCC, have the opposite association

In VHL$^{-/-}$ tumors, CD8$^+$ T cell immunity may have adverse effects in imunosurveillance

Park et al., iScience 25, 104467
June 17, 2022
https://doi.org/10.1016/j.isci.2022.104467

iScience

CellPress
OPEN ACCESS

Article

HLA-I-restricted CD8$^+$ T cell immunity may accelerate tumorigenesis in conjunction with VHL inactivation

BeumJin Park,[1] Seok-Jae Heo,[2] Yong Joon Lee,[3] Mi-Kyoung Seo,[1] Jiyun Hong,[1] Eui-Cheol Shin,[3] Inkyung Jung,[2] and Sangwoo Kim[1,4,*]

SUMMARY

CD8$^+$ T cells recognize and kill tumor cells with HLA-I tumor antigens in early tumorigenesis, the efficiency of which differs according to antigen-recognition coverage, as shown in earlier tumor onset in HLA-I homozygosity. However, the universality of these associations remains unknown. Here, we assessed the tumor type and driver mutation specificity in the association between tumor onset age and HLA-I zygosity. Statistical analyses identified an unexpected negative relationship in tumors with VHL biallelic loss, wherein HLA-I heterozygosity was associated with earlier tumor onset, while all others showed either no or a positive association. Testing on an independent dataset reproduced the VHL-dependent acceleration of tumor onset in the HLA-I heterozygous group, confirming the association. Further speculation proposed VEGF-A-mediated T cell exhaustion under VHL inactivation as a potential mechanism. Our findings suggest that CD8$^+$ T cell immunity in early tumor suppression can be conditional to the genetic status of tumors and may even lead to adverse consequences.

INTRODUCTION

The dynamic interplay between the immune system and cancer cells, known as immunoediting, has been a long-standing theory in cancer immunology (Dunn et al., 2002). This interplay is believed to affect all phases of tumor development and progression, including killing of initially transformed tumor cells (the elimination phase or immunosurveillance), suppression of tumor growth (the equilibrium phase), and the forced modification of tumors and the consequential escape from the suppression (the escape phase) (O'Donnell et al., 2019). While direct monitoring of immunoediting in human cancer development is yet to be accomplished, recent genome-level studies have provided compelling evidence, such as the negative selection of random mutations in human leukocyte antigen class I (HLA-I) and T cells (Lakatos et al., 2020), and confinement of oncogenic mutation types to the HLA-I genotypes (Marty et al., 2017). In particular, the role of immune response in the elimination phase has only recently been revealed by the pan-cancer level association between delayed tumor onset time and the lower coverage of immunogenic epitopes that are measured by HLA-I genotype and zygosity (Marty Pyke et al., 2018); that is, homozygous HLA-I is unfavorable in cancer immunosurveillance due to the reduced range for epitope recognition. These results provide convincing and crucial effects of human T cell immunity on tumorigenesis and progression, including the removal of early transformed cells.

The immunological characteristics of cancer are highly heterogeneous and vary by its types and subtypes, as shown by their distinct immunogenicity (Lv et al., 2019), immune microenvironment (Thorsson et al., 2018), and responses to therapy (Chen and Mellman, 2017). In particular, the role of T cell immunity and its effects on tumor survival often deviates from expectations in specific types of cancers, including increased expression of HLA genes in cancer and shorter survival with T cell infiltration (Jager et al., 2020). These findings imply that T cell immunity can be compromised in, or even co-opted by, cancer cells for tumor progression under certain circumstances. Likewise, the size and direction of immunoediting by T cell immunity in tumor initiation would also be variable. We presumed that a more specialized statistical analysis would enable a closer look at the cancer-specific effects of immunosurveillance.

[1]Department of Biomedical Systems Informatics and Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, South Korea

[2]Division of Biostatistics, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea

[3]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[4]Lead contact

*Correspondence: swkim@yuhs.ac

https://doi.org/10.1016/j.isci.2022.104467

Table 1. HLA-I zygosity status in 11 TCGA cancer types

TCGA cancer type	Heterozygous group No.	Homozygous group No.	Total No.
THCA	203	61	264
LUAD	162	54	216
LUSC	148	48	196
ccRCC (KIRC)	142	59	201
BRCA	117	42	159
KIRP	107	30	137
LIHC	89	27	116
TGCT	72	28	100
UCEC	72	14	86
SKCM	52	11	63
STAD	24	11	35
Total	1,188	385	1,573

Here, we report our analysis of the relationships between HLA-I zygosity and tumor onset age in 1,573 patients of 11 cancer types that demonstrated unexpected associations in VHL-inactivated tumors. VHL inactivation is a major oncogenic event in clear cell renal cell carcinoma (ccRCC), which is known for its distinct immunogenic characteristics (Hsieh et al., 2017), including an increase in HLA expression compared to autologous normal tissues (Stickel et al., 2011), and association of T cell infiltration with shorter survival (Geissler et al., 2015). We confirmed the uniqueness and statistical significance of this association in an independent dataset. Finally, we speculate on the potential mechanism of this association.

RESUTLS

Cancer type-specific association of HLA-I genotype and tumor onset age

We examined the effects of HLA-I zygosity on immunosurveillance based on 1,573 patients with cancer from 11 cancer types. The 1,573 patients were classified into the homozygous group (n = 385, 24.5%) or the heterozygous group (n = 1,188, 75.5%) (Table 1), and applied to accelerated failure time model (AFT) regression with three covariates: race, gender, and tumor type (STAR Methods). The baseline distribution at regression was determined by Akaike information criterion (AIC) (Table S1). The time ratio (TR), which denotes the effect of the variable (here, HLA-I zygosity) on the time to tumor onset (*i.e.*, acceleration or deceleration), was evaluated for the 11 cancer types (Figure 1 and Table S2). At the pan-cancer level, we found that HLA-I heterozygosity insignificantly delayed tumor onset by 0.65 years (TR, 0.98; 95% CI 0.96–1.01; p = 0.36). Among the covariates, race (white *vs.* non-white) was the only significant factor, wherein white patients took a longer time to tumor occurrence by 2.52 years (TR, 1.04; 95% CI 1.01–1.07; p = 0.01).

Deeper assessment showed diversity in the effects of HLA-I zygosity depending on the tumor type. When the AFT model was applied on each individual cancer type, the acceleration of tumor onset by HLA-I homozygosity ranged from +9.68 years (later in the heterozygous group; in stomach adenocarcinoma, STAD) to −4.26 years (earlier in the heterozygous group; in clear cell renal cell carcinoma, ccRCC) (Table S3). STAD and ccRCC were the only tumor types of statistical significance in AFT regression (TR, 0.87 and 1.08; p = 0.007 and 0.01, respectively). We found that pan-cancer analysis of 10 tumor types, excluding ccRCC, TR reached statistical significance (TR, 0.97; 95% CI 0.94–0.99; p = 0.04) (Table S4). These results are consistent with those of a previous study (Marty Pyke et al., 2018) that low MHC-I coverage (corresponding to HLA-I homozygosity) is associated with earlier tumor onset, measured in 14 tumor types without ccRCC. Overall, the effect of immunosurveillance efficiency on tumorigenesis was reproduced in terms of HLA-I zygosity, while showing tumor type specificity. In particular, the potential adverse effect in ccRCC is noteworthy.

Furthermore, we evaluated tumor acceleration by HLA-I various criteria. The effects of HLA-I was analyzed by homozygosity in each of the three genes and the number of homozygous gene (Table S5). At the pan-cancer level, no pattern of tumor acceleration by zygosity of each gene was observed. On the other hand, in

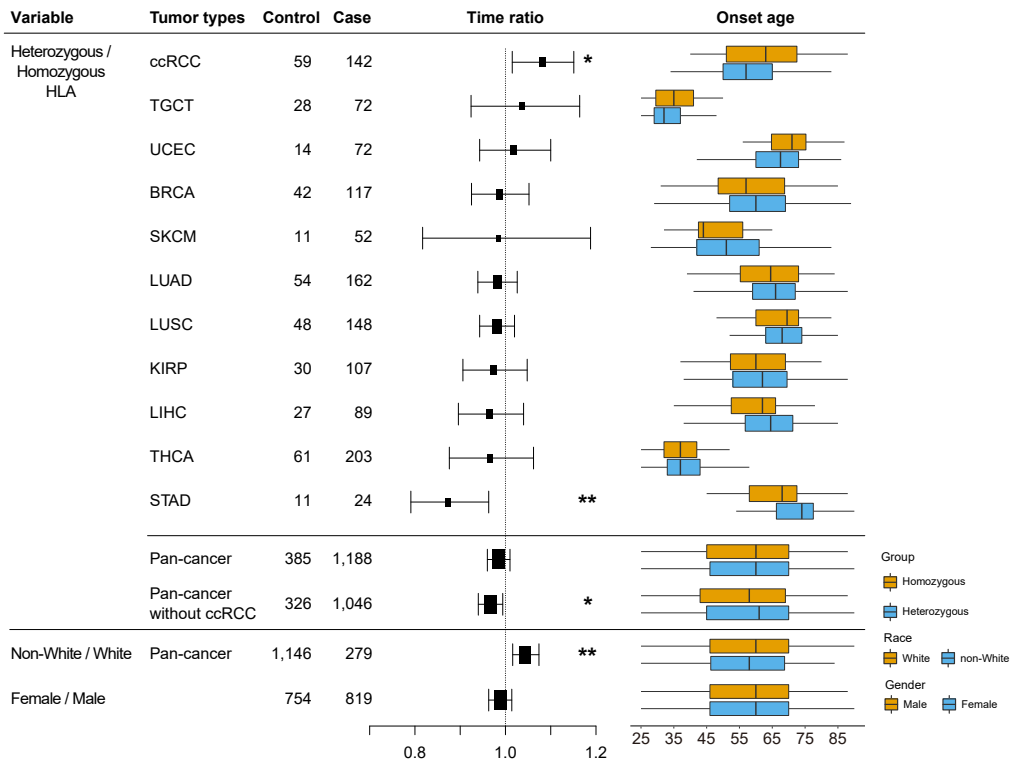


Figure 1. Summary of the Time ratio and onset age distribution

The figure shows the TR by HLA-I zygosity for each tumor type. The TRs for race and gender used as adjusted variables in the pan-cancer analysis are shown at the bottom of the figure. The p values are represented by AFT regression. *$p \leq 0.05$, **$p \leq 0.01$. Box plot on the right side of the figure shows the onset age distribution for each group. Data were presented as median +/− interquartile range (IQR).

ccRCC, homozygous HLA-A was significant for tumor delay (TR, 1.09; 95% CI 1.01–1.18; p = 0.03). In addition, an increase in the number of homozygous HLA-I genes in ccRCC was also associated with tumor delay (TR, 1.05; 95% CI 1.01–1.10; p = 0.03).

Next, we focus on effect of the particular HLA-I allele. In pan-cancer excluding ccRCC, A*29:02, A*33:01, and B*41:01 were significant alleles for tumor acceleration. Significant alleles for tumor delay were found to be B*07:05, C*07:01, C*04:01, C*05:01, C*16:01, C*14:02, and C*17:01. These results are consistent with alleles that are advantageous for tumor suppression in the immune checkpoint inhibitor therapy in previous studies (Manczinger et al., 2021). This finding suggests that differences in the effects of HLA alleles on tumor suppression may affect the immunosurveillance consistency of HLA zygosity.

VHL biallelic loss accelerates tumorigenesis with heterozygous HLA-I

We further investigated the genetic effect on immunosurveillance in ccRCC tumorigenesis. The most frequent genetic aberrations in ccRCC were mutations in VHL (46%), PBRM1 (43%), and 3p loss (79%), leading to major driver events (Figure S1). Among the drivers, 3p loss was the earliest event, followed by VHL and/or PBRM1 inactivation as a second hit (Batavia et al., 2019). As 3p harbors VHL and PBRM1, the final genetic status can be biallelic, where both copies of VHL or PBRM1 are lost (e.g., 3p loss + VHL/PBRM1 inactivation), or monoallelic (e.g., 3p loss only, or VHL/PBRM1 inactivation without 3p loss) (Figure 2A). A study of multi-regional sequencing analysis identified that approximately 15–30 years are required from the driver event to cancer diagnosis (Mitchell et al., 2018). A long-term oncogenic setting after genetic mutations suggests potential continuous effects on pre-malignant cells, such as the microenvironment or immunoediting.

We subdivided the 208 patients with ccRCC into five different categories based on the genetic status of VHL and PBRM1: (1) wild type in both genes (n = 20); (2) PBRM1 monoallelic loss ($PBRM1^{WT/-}$, n = 30); (3)

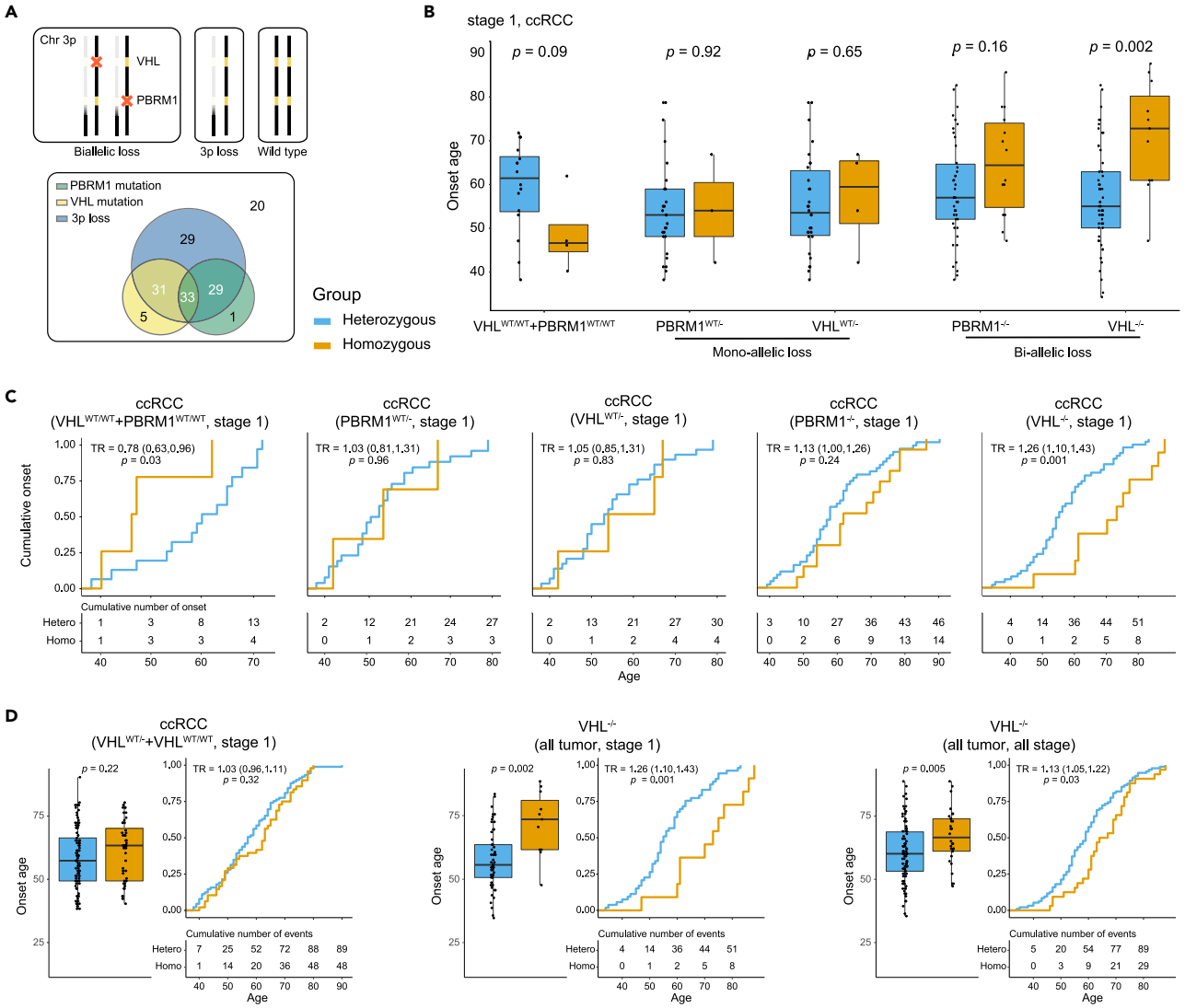


Figure 2. Tumor onset age distribution by HLA-I zygosity and inactivation of VHL and PBRM1 in TCGA cohort

(A) The classification of VHL and PBRM1 by allelic loss. The Venn diagram shows the sample size by the inactivation classification in stage 1 ccRCC.

(B) Age distribution of onset age by HLA-I zygosity with VHL and PBRM1 biallelic loss in stage 1 ccRCC.

(C) Cumulative onset comparison by HLA-I zygosity and by biallelic loss status in stage 1 ccRCC. The difference between the two groups was compared with a log rank test. TR denotes the effect of the variable on the time to tumor onset.

(D) Comparison of onset in each condition. TR denotes the effect of the variable on the time to tumor onset. Left, stage 1 ccRCC with VHL$^{WT/WT}$ or VHL$^{WT/-}$. Middle, all tumors in stage 1 with VHL biallelic loss. Right, all tumors at all stage with VHL biallelic loss. Data in boxplot were presented as median +/− IQR.

VHL monoallelic loss (VHL$^{WT/-}$, n = 34); (4) PBRM1 biallelic loss (PBRM1$^{-/-}$, n = 60); and (5) VHL biallelic loss (VHL $^{-/-}$, n = 64), and applied it to the same AFT model. Comparisons of onset age between the HLA-I zygosity groups revealed that earlier onset in HLA-I homozygous (as in other cancer types) was confined to the wild-type group (median onset age = 61.5 vs. 46.5 in hetero- and homozygous; p = 0.03 by log-rank test; TR, 0.78, 95% CI 0.63–0.96) (Figures 2B and 2C). In contrast, ccRCC with PBRM1 and VHL loss showed an earlier onset in heterozygous HLA-I. Specifically, the gap between tumor onset ages was greatest in the VHL biallelic loss group (median onset age = 55 vs. 73 in hetero- and homozygous) with statistical significance (p = 0.002 by Wilcoxon rank-sum test). The difference in VHL biallelic loss was maintained even in PRBM1 wild type (VHL$^{-/-}$PBRM1$^{WT/WT}$; p = 0.002 by Wilcoxon rank-sum test) (Figure S2), confirming that the major genetic effect on immunosurveillance in ccRCC is VHL inactivation, not PBRM1. Log-rank test on the same groups showed a similar gap of tumor onset in the VHL biallelic group with statistical significance (TR, 1.26; 95% CI 1.10–1.43; p = 0.001) (Figure 2C).

Finally, we extended the analyses to determine whether the origin of the association was ccRCC or VHL inactivation. We found that the previous difference in onset age between HLA-I homo- and heterozygous genotypes in ccRCC was diminished when VHL biallelic loss was excluded (median onset age = 57 vs. 63 in hetero- and homozygous; p = 0.22 by Wilcoxon rank-sum test; TR, 1.03, 95% CI 0.96-1.11, p = 0.37) (Figure 2D). However, the difference was maintained in all tumor types with VHL biallelic loss (median onset age = 55 vs. 73 in hetero- and homozygous genotypes, p = 0.002 by Wilcoxon rank-sum test; TR, 1.26, 95% CI 1.10–1.43, p = 0.001), or even extended to all tumor types at all stages (median onset age = 59 vs. 65.5, hetero- and homozygous, p = 0.005 by Wilcoxon rank-sum test; TR, 1.13, 95% CI 1.05–1.22, p = 0.002). These results confirm that the HLA-I type-dependent acceleration of tumorigenesis stems from VHL inactivation, while ccRCC is subordinate.

Validation in an independent cohort

We tested whether VHL inactivation was dependent on earlier tumor onset in HLA-I heterozygotes and this was apparent in an independent dataset. A total of 182 patients with stage I tumors with HLA-I genotype information were used for the analysis (STAR Methods). Among 182 tumors, 70 and 29 were ccRCC and VHL biallelic loss, respectively.

Statistical analysis clearly reproduced the association found in the discovery set (Figure 3). Thus, earlier tumor onset in HLA-I heterozygotes was statistically significant only in the VHL biallelic loss group (median onset age 59 vs. 67 in hetero- and homozygous, p = 0.03 by Wilcoxon rank-sum test) (Figure 3A). Likewise, the log-rank test and AFT analysis showed statistical significance (p = 0.016 by log-rank test; TR, 1.19, 95% CI 1.06–1.35) (Figure 3B). We again confirmed that the association was specific to VHL inactivation, not ccRCC (Figure 3C). These results indicate that VHL biallelic loss is strongly associated with tumor acceleration of heterozygous HLA-I.

DISCUSSION

A role for the immune system in recognition and elimination of early cancer cells has been a controversial concept for a long time. Today, while extensive supporting evidence strongly advocates immunosurveillance, long-term, gradual, and continuous effects on non-cancerous cells hinder direct observation, especially *in vivo*. Most of the evidences, therefore, are based on statistical analysis of the outcome, such as increased tumor growth or incidence in animal models with conditional immunodeficiency (Teng et al., 2013). In this study, we adopted the same approach, but with a different hypothesis, to discover diverse effects by applying more specified grouping and stringent criteria to minimize the effects of other factors.

In the current study, an unfavorable effect of HLA-I heterozygosity on antitumor immunity was observed in patients with ccRCC with VHL biallelic loss, indicating that VHL inactivation can negatively modulate antitumor $CD8^+$ T cell responses, as reported in a previous retrospective study (Zhang et al., 2020). Under normal conditions, pVHL, the protein product of the VHL gene and an E3 ubiquitin ligase, regulates transcription factor HIF by targeting the alpha-subunit of the heterodimer and inducing proteasomal degradation (Lendahl et al., 2009). However, pVHL inactivation by VHL biallelic loss, which is common in ccRCC tumors, induces accumulation of HIF-alpha and, consequently, hypoxia-responsive genes are transcriptionally upregulated, including the VEGFA gene, which encodes VEGF-A (Majmundar et al., 2010; Ramakrishnan et al., 2014; Semenza, 2013). Although VEGF-A is known to be a critical factor for tumor angiogenesis, it also acts as an immunosuppressive factor and directly influences $CD8^+$ T cells in the tumor microenvironment. VEGR2 expression, which can be induced by antigen stimulation, makes tumor-infiltrating $CD8^+$ T cells available to respond to VEGF-A. The VEGF-A/VEGFR2 interaction can enhance the expression of TOX, a master regulator of T cell exhaustion in $CD8^+$ T cells (Kim et al., 2019; Voron et al., 2015). As a result, tumor-infiltrating $CD8^+$ T cells undergo exaggerated exhaustion in a VEGF-A-rich environment resulting from VHL biallelic loss.

Here, we provide evidence of T cell exhaustion through RNA data of samples that have completed tumor development. Differential expression gene (DEG) and Gene Ontological terms (GO) enrichment analysis were performed in normal matched samples for VHL biallelic loss (STAR Methods). Various immune activation and proliferation were included in the significant GO terms, and genes associated with T cell activation were most enriched (Figure S3). Despite T cell activation gene enrichment,

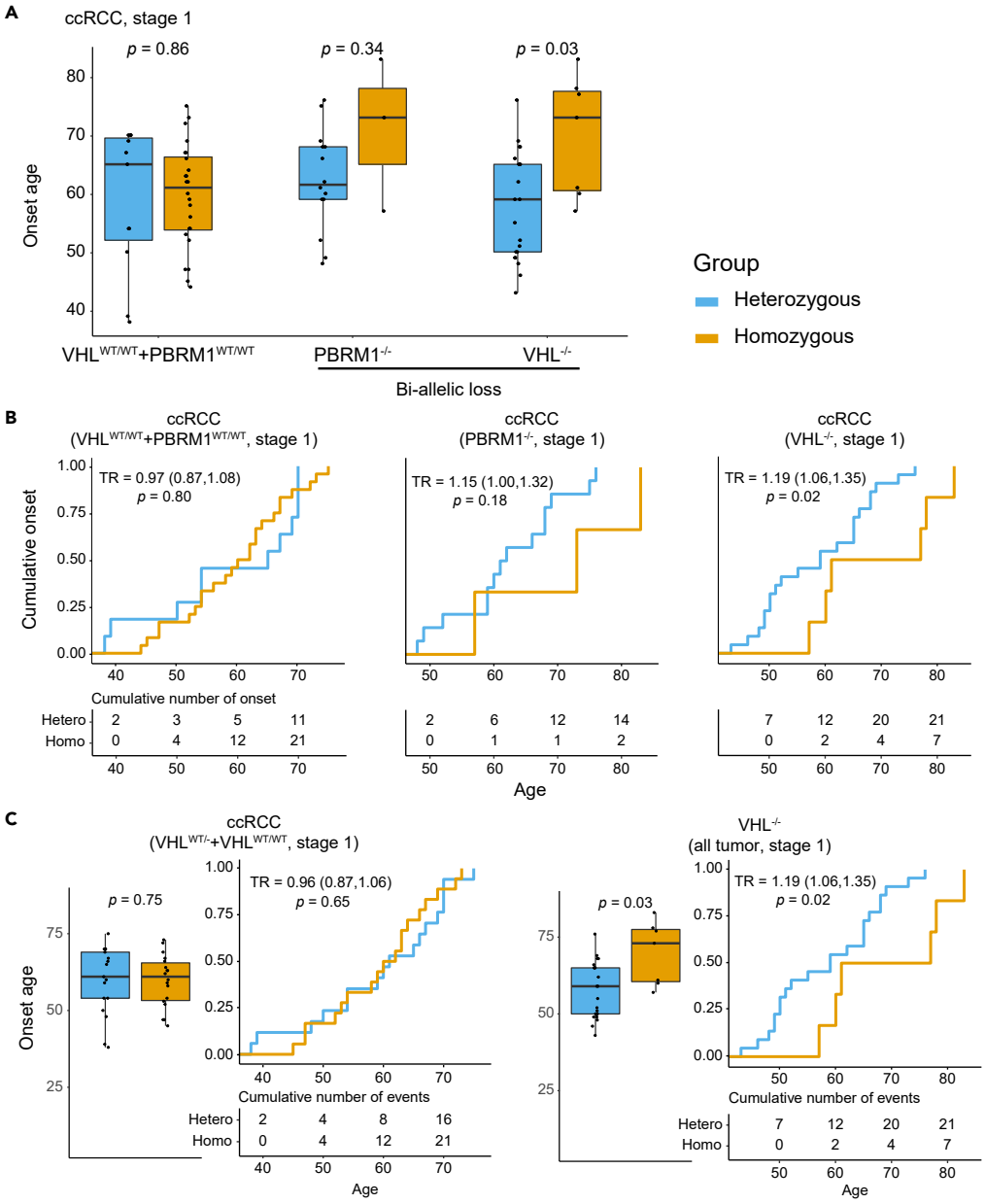


Figure 3. Tumor onset age distribution by HLA-I zygosity and inactivation of VHL and PBRM1 in ICGC cohort

(A) Age distribution of onset age by HLA-I zygosity with VHL and PBRM1 biallelic loss in stage 1 ccRCC.

(B) Cumulative onset comparison by HLA-I zygosity and by biallelic loss status in stage 1 ccRCC. The difference between the two groups was compared with a log rank test. TR denotes the effect of the variable on the time to tumor onset.

(C) Comparison of onset in each condition. TR denotes the effect of the variable on the time to tumor onset. Left, stage 1 ccRCC with $VHL^{WT/WT}$ or $VHL^{WT/-}$. Right, all tumors in stage 1 with VHL biallelic loss. Data in boxplot were presented as median +/− IQR.

CTLA4 and PD1 (log fold change = 3.99 and 3.53, respectively) as T cell exhaustion markers (Wherry and Kurachi, 2015) were overexpressed along with VEGFA (log fold change = 3.36) in tumor samples (Table S6).

Considering that HLA-I zygosity determines the antigen-recognition coverage of $CD8^+$ T cells, HLA-I heterozygosity in patients with VHL biallelic loss can enrich more exhausted $CD8^+$ T cells in the tumor microenvironment than HLA-I homozygosity.

However, the status of dysfunction is only limited to exhausted $CD8^+$ T cells themselves and cannot explain the paradoxical adverse effects of $CD8^+$ T cells on tumor development. Regarding this point, we hypothesize that exaggerated exhaustion of $CD8^+$ T cells in a VHL biallelic loss-associated VEGF-A-rich condition makes them not only dysfunctional cells but also immune-suppressive cells, for which their abundance in the tumor microenvironment can restrain antitumor immune activity. Further studies are required to unveil the mechanism by which exaggerated exhaustion of $CD8^+$ T cells leads to immune-suppressive functions in the tumor microenvironment. The results should be interpreted with caution.

Limitations of the study

This study focused on the onset of tumors based on the HLA-I genotype. However, it is challenging to determine the actual time of tumor occurrence, as the observation begins with clinical diagnosis. Although we expect that the time gap between the diagnosis with stage 1 and the actual tumor occurrence is close, patient-specific variations in the gap would be a confounding factor for our analysis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Acquisition of discovery dataset
  - Patient selection and downstream analyses
  - Covariates selection for AFT regression
  - Judgement of VHL genotypes
  - Differential expression analysis and enrichment analysis
  - Acquisition and analysis of validation dataset
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104467.

ACKNOWLEDGMENTS

The authors thank BiCU (Bioinformatics Collaboration Unit, Yonsei University College of Medicine) for assisting with the resources. This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI14C1324) and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2019R1A2C2008050).

AUTHOR CONTRIBUTIONS

B.J.P., S.J.H., and J.H. analyzed the data; B.J.P and M.-K.S. contributed to the data acquisition; I.J provided methodology advice; S.K. acquired funding and administered the project. B.J.P., S.J.H., Y.J.L., E.-C.S., and S.K. wrote the original draft; I. J., E.-C.S., and S.K. reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 25, 2021
Revised: March 28, 2022
Accepted: May 19, 2022
Published: June 17, 2022

REFERENCES

Akaike, H. (1981). Citation Classic - A New Look at the Statistical-Model Identification (Current Contents/Engineering Technology & Applied Sciences), p. 22.

Batavia, A.A., Schraml, P., and Moch, H. (2019). Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: a non-existent or new tumour entity? Histopathology *74*, 60–67. https://doi.org/10.1111/his.13749.

Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature *541*, 321–330. https://doi.org/10.1038/nature21349.

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. https://doi.org/10.1038/s41586-020-1969-6.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998. https://doi.org/10.1038/ni1102-991.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7. https://doi.org/10.1016/j.cels.2018.03.002.

Geissler, K., Fornara, P., Lautenschlager, C., Holzhausen, H.J., Seliger, B., and Riemann, D. (2015). Immune signature of tumor infiltrating immune cells in renal cancer. Oncoimmunology *4*, e985082. https://doi.org/10.4161/2162402X.2014.985082.

Hayashi, S., Yamaguchi, R., Mizuno, S., Komura, M., Miyano, S., Nakagawa, H., and Imoto, S. (2018). ALPHLARD: a Bayesian method for analyzing HLA genes from whole genome sequence data. BMC Genomics *19*, 790. https://doi.org/10.1186/s12864-018-5169-9.

Hsieh, J.J., Purdue, M.P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., Heng, D.Y., Larkin, J., and Ficarra, V. (2017). Renal cell carcinoma. Nat. Rev. Dis. Prim. *3*, 17009. https://doi.org/10.1038/nrdp.2017.9.

Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic germline variants in 10,389 adult cancers. Cell *173*, 355–370.e14. https://doi.org/10.1016/j.cell.2018.03.039.

Jager, M.J., Shields, C.L., Cebulla, C.M., Abdel-Rahman, M.H., Grossniklaus, H.E., Stern, M.H., Carvajal, R.D., Belfort, R.N., Jia, R., Shields, J.A., and Damato, B.E. (2020). Uveal melanoma. Nat. Rev. Dis. Prim. *6*, 24. https://doi.org/10.1038/s41572-020-0158-0.

Kahles, A., Lehmann, K.V., Toussaint, N.C., Huser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., Cancer Genome Atlas Research, N., et al. (2018). Comprehensive analysis of alternative splicing across tumors from 8,705 patients. Cancer Cell *34*, 211–224.e6. https://doi.org/10.1016/j.ccell.2018.07.001.

Kim, C.G., Jang, M., Kim, Y., Leem, G., Kim, K.H., Lee, H., Kim, T.S., Choi, S.J., Kim, H.D., Han, J.W., et al. (2019). VEGF-A drives TOX-dependent T cell exhaustion in anti-PD-1-resistant microsatellite stable colorectal cancers. Sci. Immunol. *4*, eaay0555. https://doi.org/10.1126/sciimmunol.aay0555.

Lakatos, E., Williams, M.J., Schenck, R.O., Cross, W.C.H., Househam, J., Zapata, L., Werner, B., Gatenbee, C., Robertson-Tessi, M., Barnes, C.P., et al. (2020). Evolutionary dynamics of neoantigens in growing tumors. Nat. Genet. *52*, 1057–1066. https://doi.org/10.1038/s41588-020-0687-1.

Lendahl, U., Lee, K.L., Yang, H., and Poellinger, L. (2009). Generating specificity and diversity in the transcriptional response to hypoxia. Nat. Rev. Genet. *10*, 821–832. https://doi.org/10.1038/nrg2665.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Lv, J.W., Zheng, Z.Q., Wang, Z.X., Zhou, G.Q., Chen, L., Mao, Y.P., Lin, A.H., Reiter, R.J., Ma, J., Chen, Y.P., and Sun, Y. (2019). Pan-cancer genomic analyses reveal prognostic and immunogenic features of the tumor melatonergic microenvironment across 14 solid cancer types. J. Pineal Res. *66*, e12557. https://doi.org/10.1111/jpi.12557.

Majmundar, A.J., Wong, W.J., and Simon, M.C. (2010). Hypoxia-inducible factors and the response to hypoxic stress. Mol. Cell *40*, 294–309. https://doi.org/10.1016/j.molcel.2010.09.022.

Manczinger, M., Koncz, B., Balogh, G.M., Papp, B.T., Asztalos, L., Kemeny, L., Papp, B., and Pal, C. (2021). Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity. Nat Cancer *2*, 950–961. https://doi.org/10.1038/s43018-021-00226-4.

Marty Pyke, R., Thompson, W.K., Salem, R.M., Font-Burgada, J., Zanetti, M., and Carter, H. (2018). Evolutionary pressure against MHC class II binding cancer mutations. Cell *175*, 416–428.e13. https://doi.org/10.1016/j.cell.2018.08.048.

Marty, R., Kaabinejadian, S., Rossell, D., Slifker, M.J., van de Haar, J., Engin, H.B., de Prisco, N., Ideker, T., Hildebrand, W.H., Font-Burgada, J., and Carter, H. (2017). MHC-I genotype restricts the oncogenic mutational landscape. Cell *171*, 1272–1283.e15. https://doi.org/10.1016/j.cell.2017.09.050.

Mitchell, T.J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J.H.R., O'Brien, T., Martincorena, I., Tarpey, P., Angelopoulos, N., Yates, L.R., et al. (2018). Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. Cell *173*, 611–623.e17. https://doi.org/10.1016/j.cell.2018.02.020.

O'Donnell, J.S., Teng, M.W.L., and Smyth, M.J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. Nat. Rev. Clin. Oncol. *16*, 151–167. https://doi.org/10.1038/s41571-018-0142-8.

Orenbuch, R., Filip, I., Comito, D., Shaman, J., Pe'er, I., and Rabadan, R. (2020). arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics *36*, 33–40. https://doi.org/10.1093/bioinformatics/btz474.

Ramakrishnan, S., Anand, V., and Roy, S. (2014). Vascular endothelial growth factor signaling in hypoxia and inflammation. J. Neuroimmune Pharmacol. *9*, 142–160. https://doi.org/10.1007/s11481-014-9531-7.

Semenza, G.L. (2013). HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. J. Clin. Invest. *123*, 3664–3671. https://doi.org/10.1172/JCI67230.

Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158. https://doi.org/10.1038/nbt.3344.

Stickel, J.S., Stickel, N., Hennenlotter, J., Klingel, K., Stenzl, A., Rammensee, H.G., and Stevanovic, S. (2011). Quantification of HLA class I molecules on renal cell carcinoma using Edman degradation. BMC Urol. *11*, 1. https://doi.org/10.1186/1471-2490-11-1.

Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316. https://doi.org/10.1093/bioinformatics/btu548.

Teng, Michele.W.L., Kershaw, M.H., and Smyth, M.J. (2013). Cancer Immunoediting: From Surveillance to Escape. In Cancer Immunotherapy. Immune Suppression and Tumor Growth., E.M.J. George and C. Prendergast, eds. (John G. Delinassios), pp. 85–99.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity *48*, 812–830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

Voron, T., Colussi, O., Marcheteau, E., Pernot, S., Nizard, M., Pointet, A.L., Latreche, S., Bergaya, S., Benhamouda, N., Tanchot, C., et al. (2015). VEGF-A modulates expression of inhibitory checkpoints on CD8+ T cells in tumors. J. Exp. Med. *212*, 139–148. https://doi.org/10.1084/jem.20140559.

Wei, L.J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Stat. Med. *11*, 1871–1879. https://doi.org/10.1002/sim.4780111409.

Wherry, E.J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. Nat. Rev. Immunol. *15*, 486–499. https://doi.org/10.1038/nri3862.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The international cancer genome Consortium data portal. Nat. Biotechnol. *37*, 367–369. https://doi.org/10.1038/s41587-019-0055-9.

Zhang, J., Yan, A., Cao, W., Shi, H., Cao, K., and Liu, X. (2020). Development and validation of a VHL-associated immune prognostic signature for clear cell renal cell carcinoma. Cancer Cell Int. *20*, 584. https://doi.org/10.1186/s12935-020-01670-5.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA clinical data resources	(Liu et al., 2018)	https://gdc.cancer.gov/about-data/publications/pancanatlas
TCGA HLA-I genotypes	(Kahles et al., 2018; Thorsson et al., 2018)	https://gdc.cancer.gov/about-data/publications/panimmune, https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Splicing-2018
TCGA Germ line genetic variants	(Huang et al., 2018)	https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Germline-AWG
TCGA viral infection status	(Thorsson et al., 2018)	https://gdc.cancer.gov/about-data/publications/panimmune
TCGA somatic mutation	(Ellrott et al., 2018)	https://gdc.cancer.gov/about-data/publications/mc3-2017
TCGA aneuploidy	Broad GDAC Firehose	https://gdac.broadinstitute.org/
TCGA mRNAseq	Broad GDAC Firehose	https://gdac.broadinstitute.org/
ICGC clinical data resources and raw RNA sequence	(Zhang et al., 2019)	https://dcc.icgc.org/
ICGC HLA-I genotypes, somatic mutation, aneuploidy	(The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020)	https://dcc.icgc.org/releases/PCAWG
Software and algorithms		
R (version 3.6.3)	R Foundation	https://cran.r-project.org/
Survreg	R library	https://cran.r-project.org/web/packages/survival/index.html
ggplot2	R library	https://ggplot2.tidyverse.org/
Optitype	(Szolek et al., 2014)	https://github.com/FRED-2/OptiType
arcasHLA	(Orenbuch et al., 2020)	https://github.com/RabadanLab/arcasHLA
DESeq2	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
clusterProfiler	(Yu et al., 2012)	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sangwoo Kim (swkim@yuhs.ac).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The TCGA and ICGC data reported in this study cannot be deposited in a public repository because data policy. To request access, contact the TCGA data portal (https://portal.gdc.cancer.gov) and ICGC Data Portal (https://dcc.icgc.org/), respectively. The accession numbers are also listed in the key resources table.
- All original code has been deposited at https://github.com/Yonsei-TGIL/HLA-classI-tumorigenesis and is publicly available as of the date of publication.
- Any additional information and data required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Acquisition of discovery dataset

We downloaded clinical data of 11,163 patients with 33 cancer types from The Cancer Genome Atlas (TCGA)-Clinical Data Resources (Liu et al., 2018) using the Pancancer Atlas publication

(https://gdc.cancer.gov/about-data/publications/pancanatlas). TCGA data were approved by Genomic Data Commons (accession number: phs000178). The clinical data included pathological stage, age at diagnosis, race, sex, and information on prior malignancy. HLA-I genotype information of the patients in the same cohort was acquired from two prior studies (Kahles et al., 2018; Thorsson et al., 2018), wherein the HLA-I genotype was predicted using PolySolver (Shukla et al., 2015) and OptiType (Szolek et al., 2014), respectively. Germ line genetic variations and viral infection status were assessed as predisposing factors for cancer. Pathogenic genetic variants of the patients were collected from a recent study (Huang et al., 2018). Viral infection status was determined by the normalized number of sequencing reads that mapped to the genomes of human papilloma virus, Epstein-Barr virus, and hepatitis B virus, following the procedures described in a previous study (Thorsson et al., 2018).

Patient selection and downstream analyses

We applied strict inclusion criteria that consisted of two-phase filtration (based on clinical and genetic information, respectively; Figure S4) of the 11,163 patients in the discovery (TCGA) cohort. The main purposes of the filtration were to (1) achieve the best approximation of the time of tumor onset from the time of diagnosis; and (2) remove any other that predisposes tumorigenesis, so the role of immunosurveillance can be fairly assessed.

In the first phase, 8,838 patients with (1) tumor pathological stage of >1 (n = 8,438), (2) prior malignancy (n = 314), (3) missing diagnosis age (n = 21) and (4) missing HLA-I genotype (n = 65) were excluded, to retain 2,325 patients. In the second phase, 544 patients with (5) predisposing germline mutations (n = 77), and (6) detectable oncogenic viral infection (n = 467) were excluded, leaving 1,781 patients for analyses.

Based on the genotyping of the three classical HLA-I genes (HLA-*A*, -*B*, and -*C*), the 1,781 patients were classified into two major groups: the homozygous group, wherein any of the three genes were homozygous, and otherwise the heterozygous group. Finally, tumor types with a homozygous group of less than 10 patients were excluded due to the lack of statistical insignificance, to retain 1,573 patients of 11 tumor types for further analyses.

Covariates selection for AFT regression

We obtained 186 variables for clinical and exposure information from GDC (https://portal.gdc.cancer.gov) to select variables to use as covariates for AFT regression. 137 variables for which all values were missing were removed preferentially. 16 variables including alcohol history and smoked history were excluded because more than 80% were missing values. We then filtered out 28 variables that were duplicated or not related to tumor development, such as patient ID, tumor staging method, and treatment information. The remaining variables were tumor type, ethnicity, gender, race, and prior malignancy.

Among them, prior malignancy affects tumor development, but it is difficult to analyze without additional information, so it was used for patient filtering. We analyzed the categorical correlation between race and ethnicity variables. As a result of the Chi-square test, the two variables were determined to have a correlation with a p-value of 4.9e-06. Finally, race, gender and tumor type were selected as covariates.

Judgement of VHL genotypes

Inactivating genetic mutations in VHL was judged based on somatic mutation status and copy number loss. Somatic mutation information was collected from a previous study (Ellrott et al., 2018). To select deleterious mutations, only nonsynonymous mutations with a variant allele frequency exceeding 0.1 were included. For copy number variation, previously analyzed data based on GISTIC2 were downloaded from Firehose (http://firebrowse.org/). Genes with log2 copy number ratio < -0.3 were determined as copy number loss. We classified allelic loss using two alterations: (1) bi-allelic loss, genes with both somatic mutations and copy number loss; (2) mono-allelic loss, genes with only one of the two alterations; and (3) wildtype, in which neither mutation was found.

Differential expression analysis and enrichment analysis

Read count data for ccRCC was downloaded the from Firehose (http://firebrowse.org/). We extracted a normal matched sample for VHL bi-allelic loss (n = 8). DEG was calculated by DESeq2 (Love et al., 2014) and the genes with a FDR adjusted p-value < 0.05 and log-fold change > 1 were considered DEGs. To

observe the biological functions of the DEGs, GO enrichment analysis was performed. Package clusterProfiler (Yu et al., 2012) was used to convert gene ID and profile the GO analysis.

Acquisition and analysis of validation dataset

For the validation dataset, clinical information of 12,790 patients from 61 cancer types was downloaded from the International Cancer Genome Consortium (ICGC) (Zhang et al., 2019). ICGC data were approved by ICGC Data Portal (https://dcc.icgc.org/). Datasets only deposited in the ICGC part of the data portal were used to prevent cross-contamination between the discovery and validation datasets. Procedures for patient selection and HLA-I genotyping were similarly conducted using the discovery set. RNA sequencing data were downloaded from the Cancer Genome Center Collaboratory (https://cancercollaboratory.org/) and used for HLA-I genotyping using OptiType and arcasHLA (Orenbuch et al., 2020). For tumors without available RNA sequencing data, HLA-I genotypes predicted from Pan-Cancer Analysis of Whole Genomes (PCAWG) were used (The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020), wherein a whole genome sequencing based HLA genotype (ALPHLARD) (Hayashi et al., 2018) was used. From 12,790 patients, 1,591 were annotated with pathological stage and HLA-I genotype, 182 of which were stage I tumors. Mutation information was imported from the PCAWG study (The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020) and used to identify VHL mutation status.

QUANTIFICATION AND STATISTICAL ANALYSIS

For statistical analyses, the AFT was applied, together with the conventional log-rank test. AFT is a parametric model similar to proportional hazard models, whereas AFT assumes the effect of a covariate to be the acceleration or deceleration of events, which more directly corresponds to our assumption. Briefly, the AFT model was specified as follows:

$$s(t|x) = s_0\left(exp(\beta' x)t\right) \text{ for } t \geq 0$$

where $s(t|x)$ and $s_0(exp(\beta' x)t)$ are the survival and the baseline survival functions at time t, with the acceleration factor $exp(\beta' x)$ (Wei, 1992). If $exp(\beta^{'})$ is larger than 1, the effect of the covariate is decelerated; if $exp(\beta^{'})$ is less than 1, the effect of the covariate is accelerated. In our analysis, t is the observed (approximated by the diagnosis time) time to tumor onset, whereby the HLA-I zygosity is presumed to function as the acceleration factor along with other covariates, race, gender and tumor type (if applicable). The effects and direction (acceleration or deceleration of tumor onset) of HLA-I zygosity are presented as the TR with a 95% confidence interval (95% CI) and p-values. In addition, to evaluate the effect of specific HLA, we built three AFT models for the zygosity of each gene (HLA-A, -B, -C), the number of homozygosity and HLA allele. In all three models, race, gender, and tumor type (if applicable) were used as covariates with each HLA variable. For the first model, HLA was used as a binary variable for zygosity for three genes. Next, the number of HLA zygosity was used as a continuous variable in the model. In the allele model, minor alleles (less than 1% of the TCGA population) were filtered, and the inclusion of each patient's allele was used as a binary variable.

For different distributions, that is, exponential, Weibull, lognormal, and log-logistic, were initially assumed as the baseline distributions of AFT, followed by selection with the AIC (Akaike, 1981):

$$AIC = -2\ln(\widehat{L}) + 2k$$

where $\widehat{L}$ is the log maximum likelihood of the regression model using each distribution, and k is the number of parameters. The optimal distribution that minimizes information loss was selected with the lowest AIC value. The cumulative incidence plots created using the Kaplan-Meier (K-M) estimator and the log-rank test were used to compare differences in cumulative incidence curves. The Wilcoxon rank-sum test was used to test the difference in onset age for HLA-I zygosity. All analyses were performed using R software (version 3.6.3; www.r-project.org; R Foundation for Statistical Computing, Vienna, Austria). The *Survreg* function in the R survival package was used to perform the AFT model and compute AIC. Data in boxplot were presented as median +/− IQR. All statistical tests were conducted on two sides, and the threshold of statistical significance was set at $p < 0.05$. A p-value is denoted by an asterisk, in which * $\leq$0.05, ** $\leq$0.01.