



Computing Skin Cutaneous Melanoma Outcome From the HLA-Alleles and Clinical Characteristics

Anjali Dhall¹, Sumeet Patiyal¹, Harpreet Kaur^{1,2}, Sherry Bhalla¹, Chakit Arora¹ and Gajendra P. S. Raghava^{1*}

¹ Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India, ² Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India

OPEN ACCESS

Edited by:

Yusuf Akhter,
Babasaheb Bhimrao Ambedkar
University, India

Reviewed by:

Navchetan Kaur,
University of California,
San Francisco, United States
Gustavo Fioravanti Vieira,
Universidade La Salle, Brazil

*Correspondence:

Gajendra P. S. Raghava
raghava@iiitd.ac.in

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 26 November 2019

Accepted: 25 February 2020

Published: 26 March 2020

Citation:

Dhall A, Patiyal S, Kaur H,
Bhalla S, Arora C and Raghava GPS
(2020) Computing Skin Cutaneous
Melanoma Outcome From
the HLA-Alleles and Clinical
Characteristics. *Front. Genet.* 11:221.
doi: 10.3389/fgene.2020.00221

Human leukocyte antigen (HLA) are essential components of the immune system that stimulate immune cells to provide protection and defense against cancer. Thousands of HLA alleles have been reported in the literature, but only a specific set of HLA alleles are present in an individual. The capability of the immune system to recognize cancer-associated mutations depends on the presence of a particular set of alleles, which elicit an immune response to fight against cancer. Therefore, the occurrence of specific HLA alleles affects the survival outcome of cancer patients. In the current study, prediction models were developed, using 401 cutaneous melanoma patients, to predict the overall survival (OS) of patients using their clinical data and HLA alleles. We observed that the presence of certain favorable superalleles like HLA-B*55 (HR = 0.15, 95% CI 0.034–0.67), HLA-A*01 (HR = 0.5, 95% CI 0.3–0.8), is responsible for the improved OS. In contrast, the presence of certain unfavorable superalleles such as HLA-B*50 (HR = 2.76, 95% CI 1.284–5.941), HLA-DRB1*12 (HR = 3.44, 95% CI 1.64–7.2) is responsible for the poor survival. We developed prediction models using key 14 HLA superalleles, demographic, and clinical characteristics for predicting high-risk cutaneous melanoma patients and achieved HR = 4.52 (95% CI 3.088–6.609, p -value = 8.01E-15). Eventually, we also provide a web-based service to the community for predicting the risk status in cutaneous melanoma patients (<https://webs.iiitd.edu.in/raghava/skcmhrp/>).

Keywords: cutaneous melanoma, survival analysis, HLA, superalleles, Hazard ratio, regression, machine learning, prognosis

INTRODUCTION

The HLA complex is the highly polymorphic genetic region located on chromosome 6, precisely in the 6p21.3 region (Beck and Trowsdale, 2000; Choo, 2007). Major histocompatibility complex (MHC) encodes more than 200 immune-related genes, from which approximately 40 genes are associated with the development of leukocyte antigen, i.e., class I and class II HLA genes (Bonamigo et al., 2012). Class I and II regions are categorized into classical and non-classical, where, classical HLA-class I comprises of HLA-A, HLA-B, and HLA-C, and class II HLA gene loci are HLA-DR, HLA-DP, and HLA-DQ (Shiina et al., 2009). Out of which, class I genes encode proteins which

present endogenous antigen to CD8 + T lymphocytes, while, class II genes encode proteins which present exogenous antigens to CD4 + T cells (Watts, 2004; Traherne, 2008; Cruz-Tapias et al., 2013). The class I complex is generally located on all nucleated cell surfaces, and class II genes are expressed on the specific antigen-presenting cells (APCs), B lymphocytes and activated T cells (Choo, 2007). By the cross-presentation process, certain APCs present exogenous antigens on HLA class I molecules for the activation of cytotoxic CD8 + T cells responses (Bevan, 2006; Joffre et al., 2012).

HLA molecules play a significant role in the induction and regulation of immune responses. The role of HLA class I molecules has been implied in tumor resistance to apoptosis (Sabapathy and Nam, 2008). Recent findings suggest that the altered expression of HLA molecules is associated with metastatic progression and poor prognosis in the tumor (Aptsiauri et al., 2007; Mendez et al., 2009; Johansen et al., 2016). The modification of surface molecules, lack of co-stimulatory molecules, production of immunosuppressive cytokines, and alterations in HLA molecules are some of the primary escape mechanisms used by tumor cells to evade the immune response (Garrido et al., 2010), which can directly distress the survival of an individual. Previous studies reveal that cutaneous melanoma is one of the most threatening and fatal form of skin cancer and scrutinized multi-omics signatures for the progression of malignancy (Li et al., 2015; Ossio et al., 2017; Bhalla et al., 2019). Further, in the past, it has been shown that if melanoma is detected at an early stage, the OS rate is 95%; but, once it is metastasized (lesion thickness > 4 mm); they are tough to cure, and the survival rate is reduced to less than 50% (Büttner et al., 1995; Bristow et al., 2010). Therefore, tumor staging is crucial to provide fundamental prognostic information to clinicians. To this end, the American Joint Committee on Cancer (AJCC), and the Melanoma Staging Committee, provides information related to Tumor-Nodes-Metastasis (TNM) classification and tumor stage grouping (Gershenwald et al., 2017). Primary tumors (stage I and II), are categorized into T1, T2, T3, and T4 with a corresponding tumor thickness such as ≥ 1.00 mm, 1.01 – 2.0 mm, 2.01 – 4.0 mm and >4.0 mm, respectively. Regional Lymph Nodes (stage III) are classified into N0, N1, N2, and N3, which represent the number of metastatic tumor nodes (0, 1, 2–3, 4+), respectively. Distant metastasis (stage IV) is divided into four categories M0 (No distant metastases), M1a (metastasis to distant skin, subcutaneous tissues, and/or lymph nodes), M1b (metastasis to the lungs), and M1c (metastasis to any non-pulmonary visceral site) (Gershenwald et al., 1998; Dickson and Gershenwald, 2011). Earlier, it has been observed that melanoma tumor cells escape the immune checkpoints and proliferate at a higher rate than normal tissue cells (Khair et al., 2019). Further, it is categorized as an immunogenic tumor as its lesions have been found to have signatures of several immune escape mechanisms such as the downregulated expression of HLA molecules, secretion of cytokines like IL-10, and loss of tumor-specific antigens (Nestle et al., 1997). For instance, the downregulation of class I antigen have been associated with the poor prognosis and inadequate treatment in melanoma cases (Sun and Schuchter, 2001; Cabrera et al., 2007;

Carretero et al., 2008). Moreover, recent studies demonstrate the importance of HLA alleles in the prognosis of melanoma. For example, the loss of heterozygosity in HLA class I allele (HLA-B*15:01) has been shown to be related with poor survival outcome. In addition, HLA-C alleles and the HLA-B44 supertype has been shown to enhance OS (Campillo et al., 2006; Gogas et al., 2010; Chowell et al., 2018), thereby showing that these molecules could be considered prognostic markers for melanoma. Thus, it is vital to ascertain the role of class I and II antigen in the survival of melanoma patients. With the knowledge of accurate HLA typing, one can design immunotherapy-based prognostic biomarkers and personalized vaccines against cancer.

In the current study, we have attempted to understand the role of HLA (class I and II) alleles and superalleles in the survival of cutaneous melanoma patients using The Cancer Genome Atlas (TCGA-SKCM) dataset. Here, first, we have performed the HLA typing of patients for class I and II alleles, followed by their assignment to superallele (i.e., low-resolution HLA alleles) groups. Subsequently, we categorized the HLA superalleles into survival favorable and unfavorable groups based on the significant impact of their presence on the survival of patients. Further, we have developed survival prediction models employing key HLA superalleles, and demographic and clinical features of patients by using different machine learning techniques. In service of the scientific community, we have also developed a webserver “SKCMhrp” to predict low-risk and high-risk patient groups based on HLA superalleles, and clinical and demographic features.

MATERIALS AND METHODS

Study Design and Dataset Collection

The complete pipeline of the study is illustrated in **Figure 1**. The description of each step is given below.

Dataset Collection

Collection of Whole-Exome Dataset

We obtained the TCGA-SKCM controlled access dataset from the Genome Data Commons (GDC) data portal. Specifically, the whole-exome sequencing (WXS) BAM files of individual melanoma patients were downloaded [under the approval of dbGap (Project No. 17674)] according to the GDC protocols (Grossman et al., 2016) with the help of an in-house high-performance computing (HPC) facility and scripts. Clinical information for 470 patients was also obtained, that included age, gender, stage, tumor status, treatment status, Breslow depth, vital status, OS, etc. using TCGA assembler 2 (Zhu et al., 2014; Liu et al., 2018). We were able to extract the HLA typing information for 415 out of 470 TCGA-SKCM patients only, after removing irrelevant errors in the BAM files. Out of 415 samples, 14 patients lacked OS information. In summary, we used 401 cutaneous melanoma patients for which complete survival information with exome sequencing data was available. Clinical information like the type of melanoma, tumor stage, tumor site, Breslow depth, treatment etc., of the patients is shown in **Supplementary Table S1**.

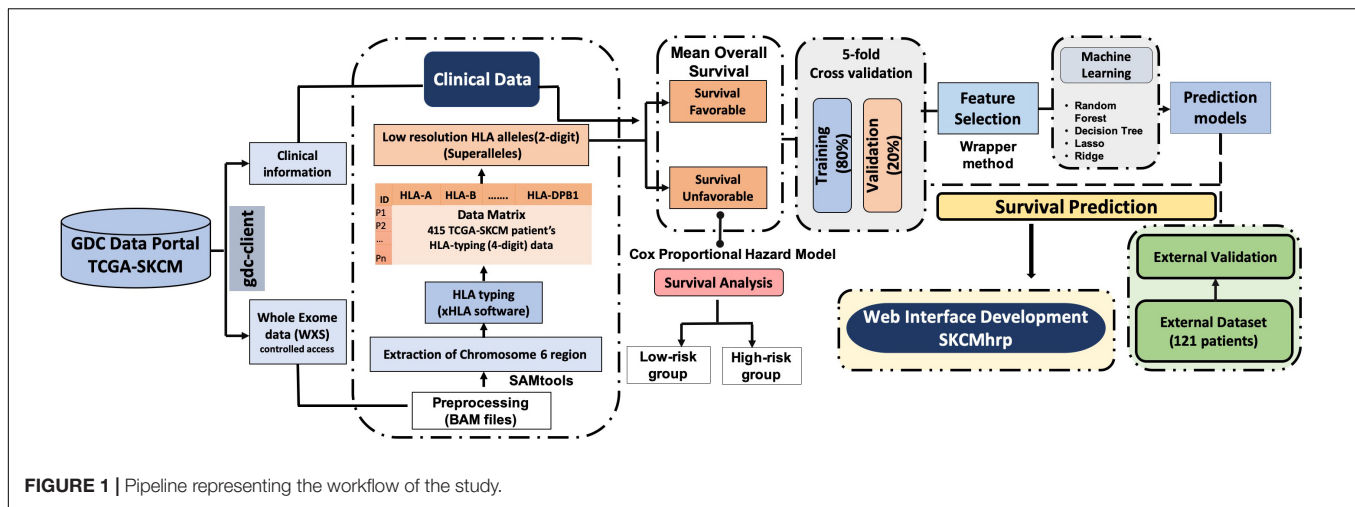


FIGURE 1 | Pipeline representing the workflow of the study.

Dataset for Prediction Models

We used the TCGA-SKCM dataset to train our prediction models and assessed the performance of our models using a particular set of features, which included HLA alleles and clinical characteristics. Eventually, the performance was evaluated on the external dataset. For the external validation dataset, we collected data from 121 cutaneous melanoma patients from various studies (Snyder et al., 2014; Van Allen et al., 2015; Hugo et al., 2016; Riaz et al., 2017), which incorporated 145 unique class I and II HLA alleles with two demographics (age and gender) and one clinical feature (tumor stage). We trained our machine learning model on the TCGA-SKCM dataset and evaluated it on the external dataset with a similar set of features.

HLA Typing

After downloading the whole exome BAM files of cutaneous patients from TCGA, chromosome 6 was extracted from these BAM files using the SAMtools package (Li et al., 2009). Subsequently, we used xHLA software (Xie et al., 2017) for HLA typing from the chromosome 6 region. In this study, four-digit HLA typing was performed for each patient for the assignment of both class I (-A, -B, -C) and class II (-DP, -DQ, -DR) HLA alleles, which are represented in **Supplementary Table S2**.

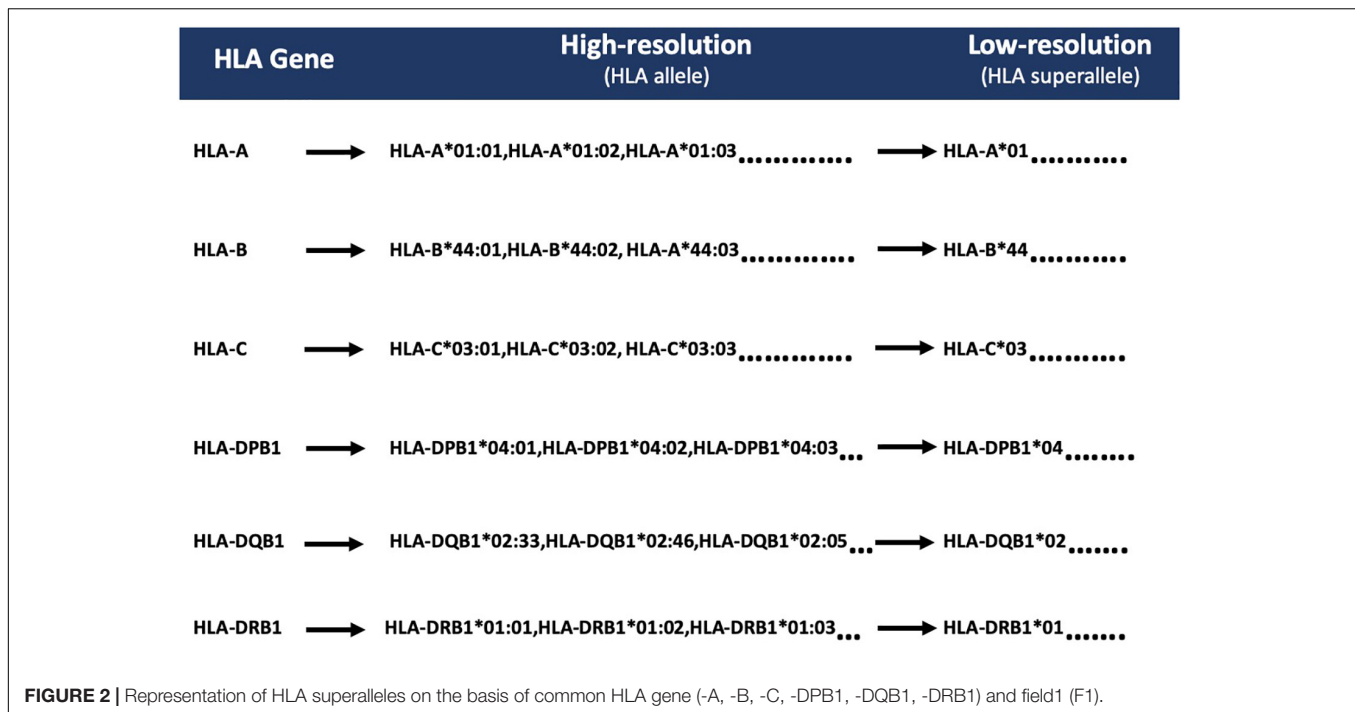
HLA Superallele

According to IMGT/HLA nomenclature, each HLA-allele is assigned to a unique name, followed by the asterisk (*) and separated by colons (Marsh, 2003; Robinson et al., 2016). The first two-digits represent an allele group (field1) (Listgarten et al., 2008); the third and fourth digit corresponds to the specific HLA protein (field2). Due to low-frequency distribution of high-resolution HLA alleles among patients, we combined the HLA alleles on the basis of field1 [which correspond to the historical serological antigen group (or allele family)] (Giannopoulos and Kriebardis, 2017) to form low-resolution HLA alleles. In this study, we used the term “superallele” for the first time for low-resolution HLA

alleles, in which we assigned a high resolution (i.e., four-digit typing) to low-resolution (i.e., two-digit typing) HLA allele. For example, HLA-A*01:01/A*01:02/A*01:03 alleles are assigned to the HLA-A*01 superallele. The complete representation of the superallele is shown in **Figure 2**, and **Supplementary Table S2**.

Categorization of HLA Superalleles

Here, we categorized HLA superalleles into favorable and unfavorable groups based on the impact of their presence on the survival of patients, i.e., whether the presence of the superallele improves or deteriorates the survival rate. First, all patients were divided into two groups, i.e., patients with a particular HLA-allele and patients lacking that particular HLA-allele; subsequently, the mean survival of patients was computed in each group. Further, an allele was assigned as a survival favorable allele if the mean survival of the patients with this allele was significantly (p -value < 0.05) higher than the mean survival of patients without this allele. Similarly, an allele is assigned as an unfavorable allele, if the mean survival of patients with this allele is lower than the mean survival of patients without this allele. It has been observed that an individual allele is only present in a limited number of patients; thus, grouping based on the occurrence of alleles will be skewed. Therefore, we analyzed the presence and absence of HLA superalleles in patients and assigned them to survival favorable (SF) and survival unfavorable (SU) superallele groups. Here, we applied a two-sample t -test to check the statistical significance (p -value < 0.05) of these superalleles. Notably, we considered only those superallele, that must be present in at least 10 samples before assigning it to any of these groups. Further, to study the overall impact of the presence of SF and SU superalleles, we combined SF and SU superalleles and prepared a matrix; where, we assigned a score of +1 if an unfavorable superallele was present, and a score of -1 if a favorable superallele was present in an SKCM patient, otherwise 0. Eventually, all the scores were cumulatively added to generate a single score called risk score (RS). Subsequently, threshold-based methods have been developed using these superalleles as features. Finally, we



assigned a patient as high-risk if the score was more than the threshold of RS, otherwise the patient was classified as low-risk.

Survival Analysis

In the current study, “univariate” and “multivariate” survival analyses were performed by cox proportional hazard (Cox PH) models and implemented by the ‘survival’ package in R (V.3.5.1). Univariate analysis was performed to understand the impact of each variable like age, tumor stage, tumor status, gender, class I, II HLA alleles, HLA superalleles, and RS in the prognosis of cutaneous melanoma patients. Further, multivariate survival analysis was performed to understand the independent clinical impact of these HLA superalleles in the presence of other multiple factors such as age, tumor stage, tumor status, gender, and class I, II HLA superalleles (Bradburn et al., 2003). The log-rank test was used to estimate the significant survival distributions between high-risk and low-risk groups in terms of the *p*-value. Kaplan-Meier (KM) survival curves were used for the graphical representation of high-risk and low-risk groups (Kishore et al., 2010).

Development of Prediction Models

Models Based on Machine Learning Techniques

In the current study, various machine learning techniques were implemented to develop regression models for OS time prediction in cutaneous melanoma patients. These machine learning techniques include a random forest (RF), ridge, lasso, and a decision tree (DT). These techniques were implemented using python-library scikit-learn (Pedregosa et al., 2011). Regression using the decision tree method, results in the supervised machine learning model, which predicts the response variable by learning the decision rules from the predictor

variables. It is a tree-based approach in which a decision tree is constructed using the recursive partitioning approach in a top-down manner (Pedregosa et al., 2011). The random forest is a supervised machine learning method that implements ensemble learning. It proceeds by assembling a number of decision trees at the time of training of a model and predicts the response variable as the average prediction of the individual trees (Geurts et al., 2006). Least absolute shrinkage and selection operator or LASSO, is a type of a linear regression method that employs the shrinkage approach. It performs the L1 regularization, which leads to the model with coefficients for predictor variables, which aids in the prediction of the response variable. On the other hand, ridge regression performs the L2 regularization to calculate the coefficients (Friedman et al., 2010). To develop prediction models, we used a wide range of features that include HLA superalleles, and clinical and demographic characteristics of the patients like age, gender, stage, tumor status, Breslow depth, and their combination.

Wrapper Based Feature Selection Method

Here, a recursive feature selection model was developed by adding HLA superalleles to the clinical and demographic features one-by-one. Then, survival time was predicted and followed by the computation of the hazard ratio (HR) for each combination. Briefly, every time input matrix was updated by adding a new column with a HLA superallele, which had the HR just higher than that of the previous input matrix. We repeated this process until there was no further improvement in the HR. Finally, we were left with the matrix which attained the highest HR. Subsequently, this matrix was used to build the final prediction model for the estimation of OS time.

Evaluation of Models

Five-Fold Cross-Validation

In order to avoid the over-optimization in the training of models, we used standard 5-fold cross-validation (Patiyal et al., 2019). In brief, all instances are randomly divided into five sets; where, four sets are used for the training and the remaining fifth set for testing. This process is repeated five times so that each set is used for testing at least once. The final performance is calculated by averaging the performance on all five sets.

Parameters for Measuring Performance

The major challenge in these types of studies is to use appropriate parameters to evaluate the performance of models. In this study, we used the standard parameter HR to measure the performance of the models. HR is a measure of the effect of an intervention on an outcome of interest over time. Our regression models segregate patients into high-risk and low-risk groups by taking a median cut-off. In order to evaluate our model, we compute HR from the predicted OS time for the group of patients (high-risk or low-risk patients). Additionally, we also measured the confidence interval (CI) with HR and reported the HR at 95% CI. In order to measure the significance of prediction, we also calculated the *p*-value using the log-rank test. These parameters were implemented previously in similar kinds of studies (Schemper, 1993; Chen et al., 2012).

RESULTS

Distribution of HLA Alleles

We extracted a total of 4,711 HLA alleles from 415 TCGA-SKCM patients by performing the HLA typing using xHLA software (Xie et al., 2017), out of which 367 HLA alleles were unique. Among them, 367 alleles, 237 belong to the HLA class I gene (-A, -B, -C), and 130 alleles correspond to class II genes (-DPB1, -DQB1, -DRB1). We computed the frequency distribution of different alleles in patients. Due to heterogeneity in HLA genes, all alleles were not found all individuals, so the frequency of alleles vary from patient to patient (Williams, 2001). Out of 415 patients, only 357 patients had all the six alleles of the HLA class I genes. In the case of HLA class II gene, only 264 patients had all six alleles. The complete frequency distribution of class I and class II HLA alleles in the TCGA-SKCM patients is provided in **Supplementary Table S3**. Among them, the most abundant (present in more than 20% of the population) class-I and class-II HLA alleles include HLA-A*02:01, HLA-A*01:01, HLA-C*07:02, HLA-C*07:01, HLA-B*07:02, HLA-A*03:01, HLA-DPB1*04:01, HLA-DQB1*03:01, HLA-DQB1*02:01, HLA-DPB1*02:01, HLA-DRB1*07:01, HLA-DQB1*05:01, HLA-DRB1*15:01, respectively, as shown in **Supplementary Table S3**.

Categorization of Superalleles Into Favorable and Unfavorable Groups

To understand whether an allele is favorable for the survival of the patient or not, we computed the difference in mean overall survival (MOS) of patients. HLA allele is assigned as favorable

if the difference in MOS is positive, otherwise it is classified as unfavorable. For instance, the class I allele HLA-A*01:01 was present in 110 patients with a MOS of 72.21 months; while MOS was reduced to 55.25 months in 291 patients that lack the class I allele. Therefore, we conclude that, HLA-A*01:01 is a favorable allele as its presence enhances the MOS. Similarly, the class I allele HLA-A*24:02 is present in 72 patients with a MOS of 45.73 months, and it is absent in 329 patients with a MOS of 63 months. This is an unfavorable allele as its presence decreases the MOS of patients, as represented in **Supplementary Table S3**. These alleles can be used to predict the risk of survival; unfortunately, this statistic could be biased as the number of patients with a particular allele is very small for most of the alleles. This prompted us to create the HLA superalleles (low-resolution HLA alleles) from the high-resolution HLA alleles on the basis of field1 (F1). Here, 367 alleles were further categorized into 121 superalleles. Out of 121 superalleles, 60 and 61 belong to class I and II, respectively. HLA-A*01/02, HLA-B*07, HLA-C*07, HLA-B*44, HLA-DPB1*04/02, HLA-DQB1*02/03/06/05, HLA-DRB1*07/15 are the most frequent class I and class II HLA superalleles in TCGA-SKCM patients, as shown in **Supplementary Figure S1**. Distribution of superalleles which are present in at least ten patients, is shown in **supplementary Figure S2**. The abundance of all superalleles is given in **Supplementary Table S4**. Further, the HLA superalleles are categorized into two groups, i.e., SF and SU, on the basis of the difference in MOS between patients with a specific HLA superallele and patients without that specific HLA superallele. Among the 24 superalleles, 9 were SF (HLA-B*55, HLA-DPB1*01, HLA-DPB1*10, HLA-B*08, HLA-B*49, HLA-A*01, HLA-DRB1*03, HLA-C*05, HLA-C*07) and 15 were SU (HLA-B*14, HLA-A*24, HLA-DPB1*05, HLA-A*31, HLA-DPB1*11, HLA-DRB1*07, HLA-DPB1*06, HLA-C*14, HLA-B*18, HLA-C*01, HLA-B*13, HLA-A*30, HLA-DRB1*16, HLA-B*50, HLA-DRB1*12) with their MOS and frequency represented in **Table 1**.

Univariate Survival Analysis HLA Superalleles

It is clear from the above analysis that certain alleles/superalleles are essential for the survival of cutaneous melanoma patients. The next challenge is to utilize this information to predict high-risk cancer patients based on the presence of key alleles or superalleles. Here, we used HLA superalleles to predict high-risk patients, employing the univariate survival analysis due to the poor distribution of alleles in patients. We observed that HLA-B*50, which is responsible for poor survival in patients, assigns patients as a high risk if this superallele is present and obtained a HR of 2.77 (95% CI 1.284 to 5.941) with a *p*-value of 0.009. Similarly, HLA-DRB1*12 achieved the maximum performance of HR 3.13 (95% CI 1.687–5.826) with a *p*-value < 0.001. The combined effect of the presence of HLA-B*50 and HLA-DRB1*12 was also used to predict high-risk patients and obtained HR 3.15, 95% (CI 1.906–5.194) with a *p*-value less than 0.001, as shown in **Supplementary Table S5**.

TABLE 1 | Classification of HLA-superalleles in to SF and SU on the basis of mean OS difference.

HLA superalleles	#No. of Samples		#Mean OS		Mean Diff OS (P-A)	P-value	Risk status
	Present (P)	Absent (A)	Present (P)	Absent (A)			
HLA-B*55	16	385	94.58	58.46	36.12	0.002	Low Risk
HLA-DPB1*01	34	367	87.51	57.34	30.17	6.82E-07	
HLA-B*08	80	321	81.09	54.62	26.47	6.36E-14	
HLA-DRB1*03	85	316	80.14	54.46	25.69	2.29E-14	
HLA-B*49	11	390	77.87	59.39	18.48	0.037	
HLA-A*01	115	286	72.88	54.68	18.2	1.24E-17	
HLA-C*05	61	340	72.74	57.6	15.15	1.82E-12	
HLA-DPB1*10	16	385	72.87	59.36	13.51	0.0004	
HLA-C*07	217	184	66.01	52.7	13.31	3.65E-31	
HLA-B*14	27	374	48.34	60.74	-12.39	2.20E-05	
HLA-A*24	81	320	48.59	62.77	-14.18	5.61E-13	
HLA-DPB1*05	17	384	46.26	60.51	-14.25	0.001	
HLA-A*31	26	375	46.34	60.84	-14.5	1.76E-05	
HLA-DPB1*11	10	391	45.32	60.27	-14.95	0.003	
HLA-DRB1*07	103	298	48.37	63.89	-15.51	4.31E-14	
HLA-DPB1*06	12	389	43.68	60.4	-16.72	0.014	
HLA-C*14	10	391	43.44	60.32	-16.88	0.003	
HLA-B*18	39	362	44.41	61.57	-17.16	1.07E-08	
HLA-C*01	42	359	44.35	61.72	-17.37	9.08E-07	
HLA-B*13	19	382	41.94	60.79	-18.86	0.03	
HLA-A*30	26	375	42.14	61.13	-19	5.22E-06	
HLA-DRB1*16	23	378	29.53	61.75	-32.22	7.00E-06	
HLA-B*50	12	389	25.03	60.98	-35.95	6.33E-05	
HLA-DRB1*12	19	382	23.46	61.71	-38.26	9.43E-05	

#Samples (P): No of SKCM-patients in which HLA-superallele is present; # Samples (A): No of SKCM-patients in which HLA-superallele is absent; #Mean OS (P): Average OS in which HLA-superallele is present; # Mean OS (A): Average OS in which HLA-superallele is absent; Mean Diff OS: Mean difference in mean OS between patients with a specific HLA typing and patients without it; Risk status Survival Favorable (SF) or Survival Unfavorable (SU) HLA-superallele, SF considered as low-risk and SU taken as high-risk groups.

Risk Score (RS)

From the above univariate analysis, we have identified key HLA superalleles, which exhibit a significant role in the prognosis of melanoma patients as a single feature. We next aimed to use them as features for the development of prediction methods. Therefore, we developed a threshold-based method using RS, which was derived from multiple HLA superalleles. To understand how well RS based on multiple superalleles stratified risk-groups of cutaneous melanoma patients, a survival analysis was performed using this RS as an input feature. For instance, if the threshold value is ≥ 2 , then the patients are significantly divided into high-risk and low-risk groups with HR 2.18 (95% CI 1.441–3.297) and a *p*-value of 0.000223, as given in **Table 2**. Conclusively, we found that RS thresholds act as a prognostic indicator to stratify melanoma patients into high-risk and low-risk groups, as shown in **Table 2**. Additionally, KM survival plots represent the segregation of risk groups of melanoma patients based on different threshold values of RS (shown in **Figure 3**).

Clinical and Demographic Characteristics

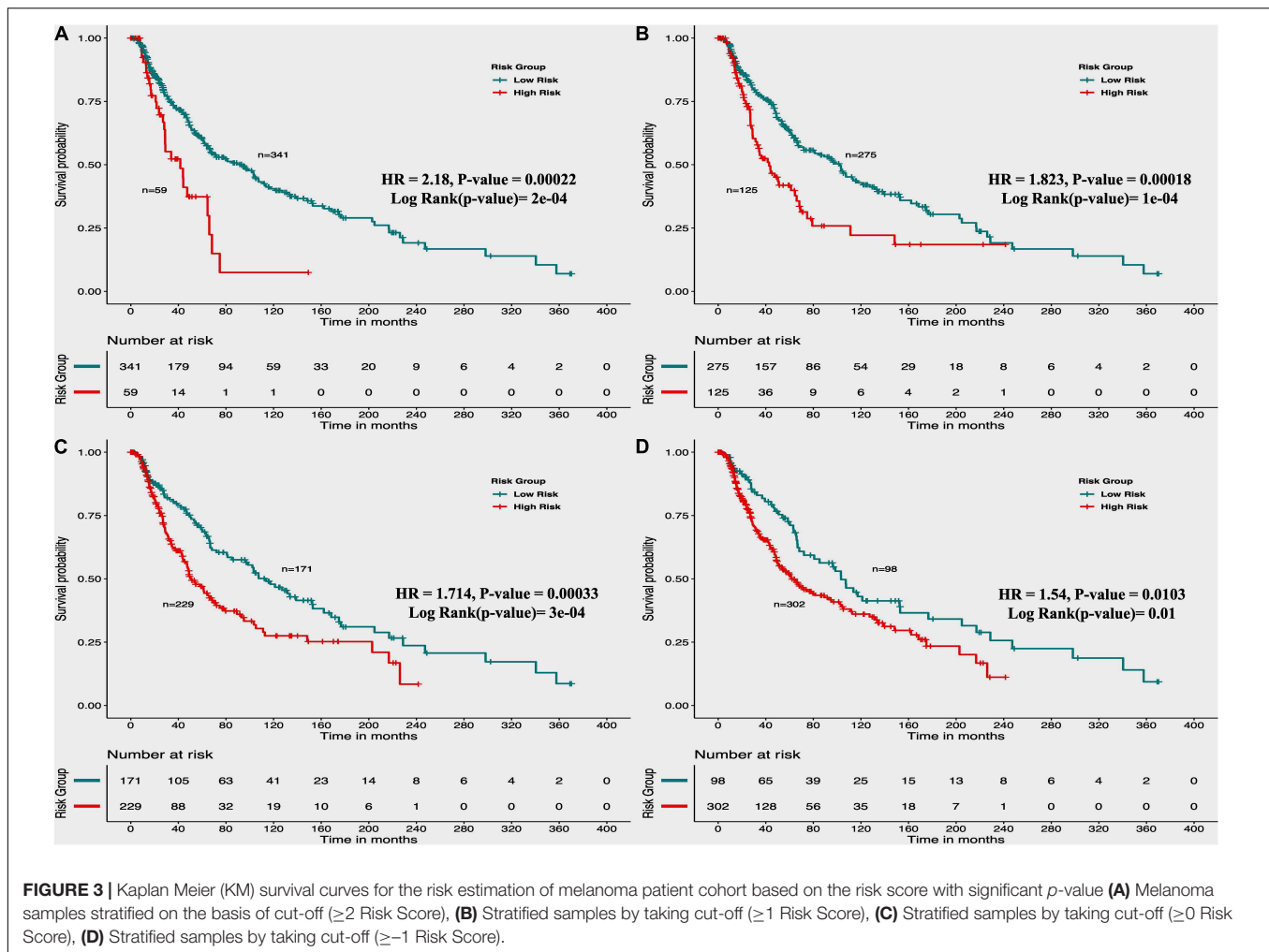
In the past, clinical and demographic features like age, gender, tumor stage, tumor status, and Breslow depth, have shown a significant effect on skin cancer incidence and a bias toward a

particular group (Zhang and Zhang, 2017). For instance, even in the current study, male incidences are higher than that of females, as shown in **Supplementary Table S1**. This prompted us to analyze the association between these clinical features and the survival of patients. Thus, we performed a univariate survival analysis using these clinical and demographic features. This analysis indicates that the tumor status is a major significant prognostic factor in the prediction of survival time of melanoma

TABLE 2 | Survival analysis based on Risk score to discriminate low-risk and high-risk samples.

Threshold (Risk Score)	#G1	#G2	HR	95% CI	P-value
≥ 3	375	26	1.84	0.97–3.51	6.35E-02
≥ 2	341	60	2.18	1.44–3.30	2.23E-04
≥ 1	275	126	1.82	1.33–2.50	1.83E-04
≥ 0	171	230	1.71	1.28–2.30	3.35E-04
≥ -1	98	303	1.55	1.11–2.16	1.03E-02
≥ -2	61	340	1.26	0.87–1.82	0.23
≥ -3	37	364	1.55	0.98–2.46	0.06

#G1: No of SKCM-patients representing low-risk group; #G2: No of SKCM-patients denoting high-risk group; HR: Hazard Ratio; 95% CI: 95% confidence interval.



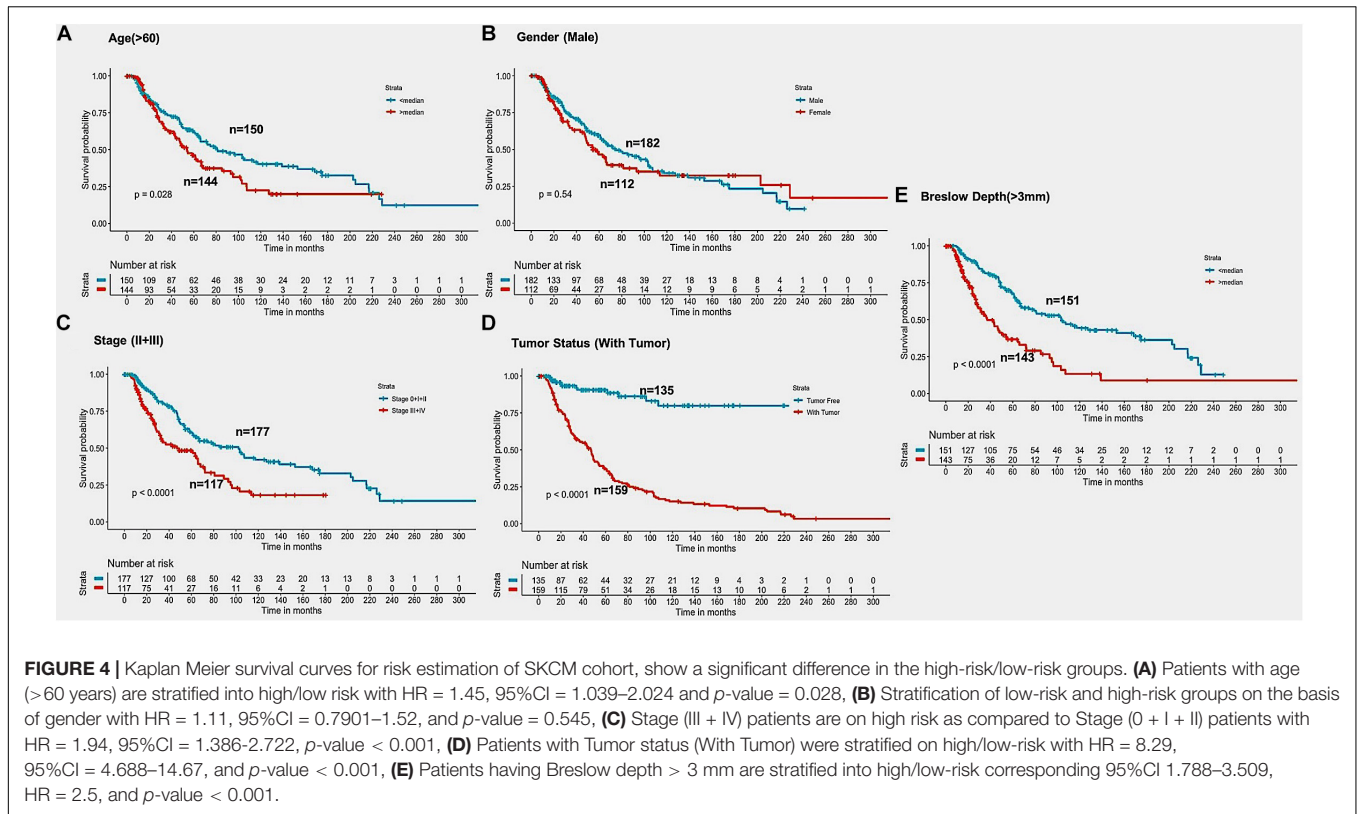
patients. We predict patients to be at high-risk if the score is more than zero and obtained HR 8.293 (95% CI 4.688–14.67) with a *p*-value of less than 0.0001 (Supplementary Table S6). Age, tumor stage, and Breslow depth are other features that are significantly associated with the prognosis of patients, as shown in Figure 4. However, samples are unable to stratified into high-risk and low-risk groups based on the gender (Figure 4).

Prediction Models

Machine Learning Based Prediction Models

It is clear from the above results that HLA superalleles, clinical and demographic features (such as age, gender, tumor stage, tumor status, and Breslow depth) are essential to identify high-risk patients. The threshold-based method, however, is simple, but not very efficient when multiple features were used. Thus, to further improve the performance, we implemented a wide range of machine learning techniques (such as, lasso, RF, ridge, DT) to develop prediction models. First, we considered all 121 superalleles to develop machine learning models. The RF model is the top performing model and achieves a maximum HR of 3.264, and a *p*-value of 1.03E-10, as represented in

Supplementary Table S7. Subsequently, a prediction model was developed by considering clinical, demographic, and 24 HLA superalleles as input features. In addition, Lasso and RF based models were also developed using only clinical and demographic features (Feature Set-1) and obtained a maximum performance with HR 3.17 (*p*-value 3.50E-11), and HR 3.09 (*p*-value 2.87E-11), respectively, as shown in Table 3. Further, we developed models by eliminating two factors, i.e., tumor status and tumor stage. Although the tumor stage is an important clinical factor, this information is sometimes available only for a few patients. So, the prediction model was developed without considering these clinical factors and a maximum HR of 2.99 (with *p*-value 9.37E-12) was achieved by the RF model. To further improve the performance of the machine learning based models, we used all clinical and demographic features with the key 24 HLA superalleles (Feature Set-2). Models based on the lasso regressor achieved the maximum performance with a HR 4.05 and a significant *p*-value of 4.01E-13. However, RF prediction models also performed reasonably well, but had a lower HR than that of the lasso models. The complete results of the survival prediction models are represented in Table 3. It has been reported in the literature that class I



alleles are important in tumor cell elimination (Garcia-Lora et al., 2003; Chang et al., 2005). Therefore, we also tried to build prediction models employing class-I alleles only. Here, the prediction model was developed using 15 class I superalleles, two demographic, and three clinical features. RF performs best among other machine learning models with a HR of 2.91 and a *p*-value of 1.79E-07. The complete results are shown in **Supplementary Table S8**.

Machine Learning Prediction Models Based on Wrapper Method

It is important to have a minimum number of features to avoid over-optimization and for practical implementation in real life. Therefore, a further wrapper method was used to decrease the number of features recursively. Finally, prediction models were developed using five clinical and demographic characteristics (age, gender, tumor stage, tumor status, and Breslow depth) and various HLA superalleles, by implementing different machine learning techniques. Similar to the above analysis, the lasso method, based on five clinical features and 14 superalleles, was the top performer with a HR of 4.52 and a *p*-value of 8.01E-15, as given in **Table 4**. The KM plot represents the stratification of high-risk and low-risk patients based on the estimated OS using the lasso recursive regression model, as shown in **Figure 5**.

Performance on the External Validation Dataset

In order to evaluate performance on the external validation dataset, we considered only 27 available input features, which included 24 superalleles, two demographic (age and gender),

and one clinical feature (tumor stage) for both training and validation, represented as ‘Set-A.’ Here, training was done on the TCGA-SKCM dataset and validated on the external dataset. The prediction model based on lasso, with 24 HLA superalleles, stratified the risk groups with a HR of 2.66 (*p*-value 6.94E-08) and 2.24 (*p*-value 0.000778) for training and validation datasets, respectively. Whereas, after applying the wrapper method we got 14 HLA superalleles and then, we trained and validated our model on these 14 HLA superalleles along with two demographic and one clinical feature, which resulted in 17 input features, represented as ‘Set-B.’ The performance of the lasso model using features of ‘Set-came out’ with a HR of 2.81 (*p*-value 1.06E-08) and a HR of 2.11 (*p*-value 0.0018) for the training and validation dataset, respectively, as shown in **Table 5**.

Multivariate Survival Analysis for SF and SU HLA Superalleles

Further, to understand the independent impact of the different variables, like SF and SU HLA-superalleles, RS, and clinical and demographic features in the presence of all the factors, on the survival of the patients, we performed a multivariate survival analysis using the cox proportional hazard model (Bradburn et al., 2003). This analysis revealed that RS is a significant independent factor associated with the survival of patients. Results (shown in **Supplementary Figure S2**) indicate that the presence of SU superalleles reduces the survival of melanoma samples. The SU patients group is at approximately two times higher risk as compared to the SF patients group as indicated

by a HR of 2.44 (95% CI 1.68–3.5) with a p-value less than 3.02E-06 (shown in **Table 6**). Both multivariate and univariate analysis revealed that age (> 60), stage (III and IV), Breslow depth (>3 mm), and RS (>0) are associated with poor survival in melanoma patients, as represented in **Supplementary Figure S3**.

Further, to scrutinize which specific superalleles out of SF and SU superallele groups, are significantly associated with good and poor outcomes in patients, multivariate analysis was performed using each of the SF and SU superalleles with the clinical and demographic characteristics. Results from this analysis show that the presence of HLA-B*55 and HLA-A*01 superalleles is significantly associated with a good outcome; while, HLA-DRB1*12, HLA-B*50, HLA-B*13, HLA-DPB1*06, HLA-A*31, HLA-A*24 is significantly associated with a poor outcome of a melanoma cohort in terms of their survival time, as given in **Supplementary Tables S9, S10** and **Supplementary Figures S4, S5**.

Web Server for Risk Prediction in SKCM Patients: SKCMhrp

To serve the scientific community, we developed a web server, “SKCMhrp” <https://webs.iiitd.edu.in/raghava/skcmhrp/>. SKCMhrp is designed to predict risk using clinical, demographic features and HLA superalleles. It has two modules; one is based on clinical features and the second is based on superalleles.

TABLE 3 | Performance of the survival prediction models based on Clinical Characteristics and 24 HLA-Class I, II Superalleles implemented using various regression techniques.

Method	Feature Set-1		Feature Set-2	
	HR	P-value	HR	P-value
	All Features			
LASSO	3.17	3.50E-11	4.05	4.01E-13
RIDGE	3.01	1.76E-10	3.80	2.30E-12
RF	3.09	2.87E-11	3.77	8.15E-12
DT	2.25	6.93E-07	2.00	5.29E-05
	Clinical features without tumor status			
LASSO	3.50	3.93E-13	3.46	1.54E-11
RIDGE	3.49	3.93E-13	2.97	2.89E-09
RF	3.74	3.01E-14	2.96	8.23E-10
DT	2.15	2.24E-06	1.83	3.12E-04
	Clinical features without tumor stage			
LASSO	2.80	9.96E-10	3.51	1.32E-11
RIDGE	2.43	4.68E-08	3.55	7.56E-12
RF	2.81	2.05E-10	3.18	2.14E-10
DT	2.50	1.64E-08	2.76	2.38E-09
	Clinical features without tumor stage and tumor status			
LASSO	2.40	4.41E-08	3.11	5.60E-10
RIDGE	2.40	4.41E-08	2.57	5.81E-08
RF	2.99	9.37E-12	2.59	1.55E-08
DT	2.54	1.06E-08	2.65	7.37E-09

#HR: Hazard Ratio; RF: Random Forest; DT: Decision Tree; Feature Set-1: Clinical and demographic features, Feature Set-2: Clinical, demographic and 24 HLA superalleles features.

TABLE 4 | Performance of the recursive prediction models based on selected features (clinical features and superalleles) implemented using various regression technique.

Method	Attribute	HR	P-value
LASSO	Clinical + 14 HLA-superalleles (A*31_A*24_DPB1*10_B*08_DRB1*03_DRB1*07_B*18_B*55_A*01_C*05_DRB1*16_DRB1*12_B*49_DPB1*11)	4.52	8.01E-15
RIDGE	Clinical + 19 HLA-superalleles (DPB1*10_B*50_C*07_B*49_B*55_B*08_C*01_C*14_DPB1*06_C*05_DRB1*03_A*30_DRB1*07_A*31_B*14_DRB1*16_B*13_DPB1*01_A*01)	3.85	3.35E-12
RF	Clinical + 3 HLA-superalleles (DPB1*11_C*05_B*08)	3.53	2.84E-11
DT	Clinical + 2 HLA-Superalleles (A*01_DPB1*01)	2.59	6.92E-08

#HR: Hazard Ratio; RF: Random Forest; DT: Decision Tree; Attribute: Clinical features and selected HLA-superalleles.

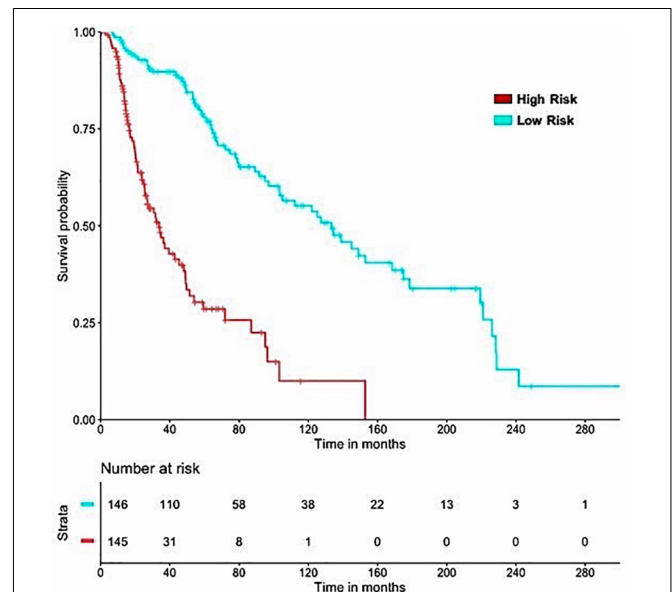


FIGURE 5 | SKCM-patients were stratified based on predicted OS by using Lasso recursive regression model after applying fivefold cross validation. Samples with predicted OS < median (predicted OS) were at fourfold higher risk as compared to the patients predicted OS > median (predicted OS) (HR = 4.52, 95% CI = 3.088 to 6.609, p-value = 8.01E-15).

The first module predicts the risk status of melanoma patients based on their clinical and demographic characteristics, i.e., age, gender, tumor stage, tumor status, Breslow depth. Here, a user can predict the survival time (in months) of the individual sample, even by choosing a single clinical feature. Input values are given to a regression model to estimate the risk status. The second module predicts the risk status of melanoma patients using all 121 superalleles and 14 superalleles with five clinical and demographic features. The webserver “SKCMhrp” was built

TABLE 5 | Performance of the prediction models based on lasso method using selected features (2 demographic and 1 clinical features, 24 and 14 HLA superalleles) on training and external validation dataset.

Dataset	Set-A		Set-B	
	HR	P-value	HR	P-value
Training data (TCGA-SKCM)	2.66	6.94E-08	2.81	1.06E-08
External validation data	2.24	0.000778	2.11	0.0018

#HR: Hazard Ratio; Set-A: 1 Clinical, 2 demographic features and selected 24 HLA-superalleles; Set-B: 1 Clinical, 2 demographic features and selected 14 HLA-superalleles.

using HTML, PHP 5.2.9, and JAVA scripts. To make the website compatible with mobiles and tablets, we used the HTML5 web template. The abovementioned technologies that have been implemented are open source and platform-independent.

DISCUSSION

Skin cutaneous melanoma is a lethal malignancy as indicated by the rise in incidence of melanoma (Siegel et al., 2020). The FDA (Food and Drug Administration) has approved several therapies and strategies to curb melanoma in the past few years. Choosing a treatment from the available options, however, requires information about the tumor such as its location, stage, etc. Accurate tumor stage identification with high precision is itself a challenging task in different malignancies (Jagga and Gupta, 2014; Bhalla et al., 2017, 2018; Saghapour et al., 2017; Kaur et al., 2019; Yu et al., 2019). Recent findings have suggested that antigenic repertoire variability is a crucial factor in tumor progression and immunosurveillance (Dunn et al., 2004). For instance, HLA-class I and II proteins have been shown to have a significant role in the progression of melanoma (Gogas et al., 2010; Bonamigo et al., 2012; Kandilarova et al., 2016). Thus, it is important to understand which specific HLA alleles from class-I and class-II could affect the survival of the patients. To this end, the current study is a systematic attempt to understand the prognostic roles of class-I/II alleles in the survival of melanoma patients.

In this study, 367 unique HLA alleles were identified for 415 cutaneous melanoma patients using the xHLA software. The low-frequency distribution of these 367 alleles among patients

(as shown in **Supplementary Table S3**) made it difficult to delineate any reliable conclusion regarding any of the alleles from the analysis. This propelled us for their assignment into low-resolution HLA typing, i.e., 121 HLA superalleles. Thereafter, these superalleles were categorized into SF and SU groups based on the impact of their presence on the survival of the patients, i.e., higher MOS or lower MOS of the patients with their occurrence, respectively. Here, among the 24 superalleles, nine were SF including HLA-B*55, HLA-DPB1*01, HLA-DPB1*10, HLA-B*08, HLA-B*49, HLA-A*01, HLA-DRB1*03, HLA-C*05, HLA-C*07; while, 15 were SU that include HLA-B*14, HLA-A*24, HLA-DPB1*05, HLA-A*31, HLA-DPB1*11, HLA-DRB1*07, HLA-DPB1*06, HLA-C*14, HLA-B*18, HLA-C*01, HLA-B*13, HLA-A*30, HLA-DRB1*16, HLA-B*50, HLA-DRB1*12. In the literature, HLA-A*01, HLA-C*05, and HLA-C*07 have been shown to be positively associated with survival of melanoma patients (Paschen et al., 2005; Campillo et al., 2006; Zhu et al., 2012), whereas HLA-B*14, HLA-A*24, HLA-A*31, HLA-C*14, and HLA-B*13 are negatively associated with survival of melanoma patients (Marincola et al., 1995; Kawakami et al., 2000; Akiyama et al., 2005; Kandilarova et al., 2016; Rogel et al., 2019). Apart from these, HLA-DRB1*07 has been shown to be negatively associated with patient survival in other cancers such as lung cancer, cervical cancer, and breast cancer (Ferreiro-Iglesias et al., 2018; Hu et al., 2018; Spraggs et al., 2018). Further, in the current study, a parameter RS was computed to evaluate the cumulative effect of the presence of SF and SU superalleles in patients. After that, 24 HLA superalleles, and clinical features like tumor status, Breslow depth, and tumor stage were identified which can significantly stratify high-risk and low-risk survival groups, by employing univariate survival analysis and a log rank test (**Supplementary Tables S5, S6**). Furthermore, prediction models were developed based these 24 superalleles, and clinical and demographic features, which stratified the risk groups with HR 4.05 (p -value 4.01E-13). In the past, the role of class I alleles was reported to be crucial for the defense against a tumor. Therefore, prediction models were developed by employing 15 class I superalleles with clinical and demographic features only. The RF model attained the maximum HR of 2.9 (p -value 1.79E-07). This indicates that not only class I, but rather both class I and II superalleles are important in the stratification of survival risk groups as performance decreases on the exclusion of class-II superalleles. Subsequently, the performance of the model based

TABLE 6 | Comparison of univariate and multivariate analysis.

Covariate	Univariate survival analysis			Multivariate survival analysis		
	HR	95% CI	P-value	HR	95% CI	P-value
Age (>60 years)	1.45	1.04–2.02	0.029	1.45	1.03–2.00	3.20E-02
Gender (Female)	1.11	0.79–1.52	0.545	0.98	0.69–1.40	0.896
Tumor Stage (III + IV)	1.94	1.39–2.72	0.001	1.89	1.33–2.70	4.00E-04
Tumor Status (With Tumor)	8.29	4.69–14.67	<0.001	9.24	5.21–2.80	2.76E-14
Breslow Depth(>3 mm)	2.50	1.79–3.51	<0.001	1.96	1.38–2.80	1.70E-04
Risk Score (>0)	1.82	1.33–2.50	<0.001	2.44	1.68–3.50	3.02E-06

HR: Hazard Ratio, 95% CI: 95% Confidence Interval.

on 24 superalleles with age, gender, and tumor stage was also evaluated on the external validation dataset. This model stratified survival risk groups of the external dataset with a HR of 2.24 (p -value 0.000778). Recently, Chen et al. (2019) also reported that the higher expression of HLA-class II genes enhances the survival of melanoma patients. Our analysis also revealed that, the higher expression of HLA (-A, -B, -C, -DPB1, -DQB1, -DRB1) genes, are associated with poor survival, as shown in **Supplementary Table S11**. Besides, our study also indicates that stage is a major prognostic factor for melanoma patients. It significantly stratified high-risk and low-risk patients in both univariate and multivariate analyses with a p -value of 0.001 and 4.00E-04, respectively (**Table 6**). It corroborates with previous literature as well; where, it has been reported that the tumor stage of melanoma patients drastically affects their prognosis. For instance, the OS of stage-Ia patients is quite good, i.e., a 10-year survival of 95%, while it is only 30% in the case of stage-IV patients. A moderate survival rate has been reported in the case of stage 1b, II, and III patients, which is shown to vary between 85 and 40% (Balch et al., 2009; Yélamos and Gerami, 2015).

Multivariate survival analysis was performed to better understand the prognostic impact of the association of SF and SU superalleles with the clinical and demographic features on the survival of patients. This analysis revealed that SF and SU superalleles also act as independent prognostic indicators. For instance, the presence of HLA-class I superalleles, such as HLA-B*55 (HR = 0.15, p -value = 0.013) and HLA-A*01 (HR = 0.54, p -value = 0.011) is significantly associated with the good outcome (**Supplementary Table S9** and **Supplementary Figure S4**). On the other hand, the presence of superalleles such as HLA-B*50 (HR = 3.1 and p -value = 0.03), HLA-DRB1*12 (HR = 3.77 and p -value < 0.001), HLA-DRB1*16 (HR = 2.18, p -value = 0.04), HLA-B*13 (HR = 2.49, p -value = 0.046), HLA-DPB1*06 (HR = 3.53, p -value = 0.006), HLA-A*31 (HR = 2.09, p -value = 0.04), and HLA-A*24 (HR = 1.79, p -value = 0.006) is associated with a poor survival outcome (**Supplementary Table S10** and **Supplementary Figure S5**). In addition to this, RS, tumor status, tumor stage, Breslow depth, and age were also revealed as major independent prognostic factors for melanoma patients.

Furthermore, with an aim to estimate the survival time of melanoma patients, various regression models were developed based on survival-associated superalleles, clinical and demographic features, and their combination. For this, we implemented diverse machine learning regressors like lasso, RF, DT, and ridge regressor. The predicted OS from these models was further employed for the stratification of high-risk and low-risk survival groups. Although the prediction, based on five clinical and demographic factors, attained a consistent performance (HR = 3.17). The accurate determination of the stage and tumor status remains a difficult task. Therefore, prediction models were also developed after the exclusion of these two factors. The performance of our ML models substantially decreased to a HR of 2.99. Thereafter, prediction models were developed based on HLA-superalleles and conveniently available demographic and clinical factors like age, gender, and Breslow depth. The performance improved considerably from a HR of 2.99 to 3.11. Lasso and RF recursive regression models were found to be

among the top performers for the prediction of the survival of melanoma samples. In particular, predicted OS obtained from the lasso recursive model, based on clinical and demographic characteristics and 14 superalleles, significantly (p -value = 8.01E-15) stratified the high-risk and low-risk survival groups of the cutaneous melanoma patients with a HR = 4.52. Although the RF-based models performed reasonably well in the estimation of OS, however, they achieved a lower HR of 3.53 than that of lasso models. The performance of this model was further evaluated on an external validation dataset considering 14 superalleles with age, gender, and tumor stage; attained an HR 2.11 (p -value 0.0018), as represented in **Table 5**.

CONCLUSION

Altogether, our findings show that the presence of HLA-class I and II alleles influence the OS of TCGA-SKCM patients both favorably and unfavorably. Eventually, survival analysis and recursive machine learning regression models revealed the prognostic potential of 14 superalleles, clinical and demographic features in the stratification of high-risk and low-risk survival groups and the estimation of OS time. Further, these HLA-based signatures could be considered in the design of personalized vaccines in several clinical cohorts. For clinical utility, this needs to be further confirmed by exploring the role of these superalleles in other cohorts. Finally, to provide a service to the scientific community to predict high-risk patients based on their clinical features, demographic features, 14 and 121 HLA-superalleles, we designed a webserver “SKCMhrp.”

LIMITATION OF THE STUDY

In the current study, the prognostic potential of 14 superalleles (low-resolution HLA allele), demographic and clinical features were revealed to estimate the survival of cutaneous melanoma patients. One of the limitations of these features/markers is that they are derived from low-resolution HLA alleles. Researchers can, however, implement a similar strategy if sufficient data is available for high-resolution HLA alleles. Further, we have not considered the ethnicity of the patients for the analysis or development of the prediction models in this study.

DATA AVAILABILITY STATEMENT

All the datasets generated for this study are either included in this article/**Supplementary Material** or available at the “SKCMhrp” webserver <https://webs.iitd.edu.in/raghava/skcmhrp/data.php>, as mentioned in the “Materials and Methods” section.

AUTHOR CONTRIBUTIONS

AD, HK, and SB collected and processed the datasets. AD, HK, and SP implemented the algorithms. AD and SP

created the back-end of the web server and front-end user interface. AD and SP developed the prediction models. AD, HK, and GR analyzed the results. AD, HK, SP, CA, and GR penned the manuscript. GR conceived and coordinated the project and provided overall supervision to the project. All authors have read and approved the final manuscript.

FUNDING

This research was funded by the J. C. Bose National Fellowship (with Grant No. SRP076), Department of Science and Technology (DST), India.

REFERENCES

- Akiyama, Y., Tanosaki, R., Inoue, N., Shimada, M., Hotate, Y., Yamamoto, A., et al. (2005). Clinical response in Japanese metastatic melanoma patients treated with peptide cocktail-pulsed dendritic cells. *J. Transl. Med.* 3:4. doi: 10.1186/1479-5876-3-4
- Aptsiauri, N., Cabrera, T., Mendez, R., Garcia-Lora, A., Ruiz-Cabello, F., and Garrido, F. (2007). Role of altered expression of HLA class I molecules in cancer progression. *Adv. Exp. Med. Biol.* 601, 123–131. doi: 10.1007/978-0-387-72005-0-13
- Balch, C. M., Gershenwald, J. E., Soong, S. J., Thompson, J. F., Atkins, M. B., Byrd, D. R., et al. (2009). Final version of 2009 AJCC melanoma staging and classification. *J. Clin. Oncol.* 27, 6199–6206. doi: 10.1200/JCO.2009.23.4799
- Beck, S., and Trowsdale, J. (2000). The human major histocompatibility complex: lessons from the DNA sequence. *Annu. Rev. Genomics Hum. Genet.* 1, 117–137. doi: 10.1146/annurev.genom.1.1.117
- Bevan, M. J. (2006). Cross-priming. *Nat. Immunol.* 7, 363–365. doi: 10.1038/nri0406-363
- Bhalla, S., Chaudhary, K., Kumar, R., Sehgal, M., Kaur, H., Sharma, S., et al. (2017). Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. *Sci. Rep.* 7:44997. doi: 10.1038/srep44997
- Bhalla, S., Kaur, H., Dhall, A., and Raghava, G. P. S. (2019). Prediction and analysis of skin cancer progression using genomics profiles of patients. *Sci. Rep.* 9:15790. doi: 10.1038/s41598-019-52134-4
- Bhalla, S., Kaur, H., Kaur, R., Sharma, S., and Raghava, G. P. S. (2018). Expression based biomarkers and models to classify early and late stage samples of papillary thyroid carcinoma. *bioRxiv* [Preprint]. doi: 10.1101/393975
- Bonamico, R. R., Carvalho, A., de Sebastiani, V. R. Z., Silva, C. M., and Pinto, A. C. (2012). HLA and skin cancer. *An. Bras. Dermatol.* 87, 9–16. doi: 10.1590/s0365-05962012000100001
- Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003). Survival analysis part II: multivariate data analysis— an introduction to concepts and methods. *Br. J. Cancer* 89, 431–436. doi: 10.1038/sj.bjc.6601119
- Bristow, I. R., de Berker, D. A., Acland, K. M., Turner, R. J., and Bowling, J. (2010). Clinical guidelines for the recognition of melanoma of the foot and nail unit. *J. Foot Ankle Res.* 3:25
- Büttner, P., Garbe, C., Bertz, J., Burg, G., d'Hoedt, B., Drepper, H., et al. (1995). Primary cutaneous melanoma. optimized cutoff points of tumor thickness and importance of Clark's level for prognostic classification. *Cancer* 75, 2499–2506. doi: 10.1002/1097-0142(19950515)75
- Cabrera, T., Lara, E., Romero, J. M., Maleno, I., Real, L. M., Ruiz-Cabello, F., et al. (2007). HLA class I expression in metastatic melanoma correlates with tumor development during autologous vaccination. *Cancer Immunol. Immunother.* 56, 709–717. doi: 10.1007/s00262-006-0226-7
- Campillo, J. A., Martínez-Escribano, J. A., Muro, M., Moya-Quiles, R., Marín, L. A., Montes-Ares, O., et al. (2006). HLA class I and class II frequencies in patients with cutaneous malignant melanoma from southeastern Spain: the role of HLA-C in disease prognosis. *Immunogenetics* 57, 926–933. doi: 10.1007/s00251-005-0065-2

ACKNOWLEDGMENTS

All the authors acknowledge funding agencies J. C. Bose National Fellowship (DST). AD, SP, HK, and SB are thankful to DST INSPIRE, DBT, CSIR, and ICMR for providing fellowships, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00221/full#supplementary-material>

- Carretero, R., Romero, J. M., Ruiz-Cabello, F., Maleno, I., Rodriguez, F., Camacho, F. M., et al. (2008). Analysis of HLA class I expression in progressing and regressing metastatic melanoma lesions after immunotherapy. *Immunogenetics* 60, 439–447. doi: 10.1007/s00251-008-0303-5
- Chang, C. C., Campoli, M., and Ferrone, S. (2005). Classical and nonclassical HLA class I antigen and NK cell-activating ligand changes in malignant cells: current challenges and future directions. *Adv. Cancer Res.* 93, 189–234. doi: 10.1016/S0065-230X(05)93006-6
- Chen, H.-C., Kodell, R. L., Cheng, K. F., and Chen, J. J. (2012). Assessment of performance of survival prediction models for cancer prognosis. *BMC Med. Res. Methodol.* 12:102. doi: 10.1186/1471-2288-12-102
- Chen, Y.-Y., Chang, W.-A., Lin, E.-S., Chen, Y.-J., and Kuo, P.-L. (2019). Expressions of HLA class II genes in cutaneous melanoma were associated with clinical outcome: bioinformatics approaches and systematic analysis of public microarray and RNA-Seq Datasets. *Diagnostics* 9:59. doi: 10.3390/diagnostics9020059
- Choo, S. Y. (2007). The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* 48, 11–23. doi: 10.3349/ymj.2007.48.1.11
- Chowell, D., Morris, L. G. T., Grigg, C. M., Weber, J. K., Samstein, R. M., Makarov, V., et al. (2018). Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 359, 582–587. doi: 10.1126/science.aao4572
- Cruz-Tapias, P., Castiblanco, J., and Juan-Manuel, A. (2013). “Major histocompatibility complex: antigen processing and presentation,” in *Autoimmunity: From Bench to Bedside*, eds J. M. Anaya, Y. Shoenfeld, and A. Rojas-Villarraga, (Bogota: El Rosario University Press), 169–184.
- Dickson, P. V., and Gershenwald, J. E. (2011). Staging and prognosis of cutaneous melanoma. *Surg. Oncol. Clin. N. Am.* 20, 1–17. doi: 10.1016/j.soc.2010.09.007
- Dunn, G. P., Old, L. J., and Schreiber, R. D. (2004). The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 21, 137–148. doi: 10.1016/j.immuni.2004.07.017
- Ferreiro-Iglesias, A., Lesseur, C., McKay, J., Hung, R. J., Han, Y., Zong, X., et al. (2018). Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nat. Commun.* 9:3927. doi: 10.1038/s41467-018-05890-2
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.
- García-Lora, A., Algarra, I., and Garrido, F. (2003). MHC class I antigens, immune surveillance, and tumor immune escape. *J. Cell. Physiol.* 195, 346–355. doi: 10.1002/jcp.10290
- Garrido, F., Cabrera, T., and Aptsiauri, N. (2010). “Hard” and “soft” lesions underlying the HLA class I alterations in cancer cells: implications for immunotherapy. *Int. J. Cancer* 127, 249–256. doi: 10.1002/ijc.25270
- Gershenwald, J. E., Buzaid, A. C., and Ross, M. I. (1998). Classification and staging of melanoma. *Hematol. Oncol. Clin. North Am.* 12, 737–765. doi: 10.1016/S0889-8588(05)70021-6
- Gershenwald, J. E., Scolyer, R. A., Hess, K. R., Sondak, V. K., Long, G. V., Ross, M. I., et al. (2017). Melanoma staging: evidence-based changes in the american joint committee on cancer eighth edition cancer staging manual. *CA. Cancer J. Clin.* 67, 472–492. doi: 10.3322/caac.21409

- Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1
- Giannopoulos, A., and Kriebardis, A. G. (2017). “Future perspectives in HLA typing technologies,” in *Umbilical Cord Blood Banking for Clinical Application and Regenerative Medicine*, ed. A. C. Mauricio, (London: IntechOpen), doi: 10.5772/64850
- Gogas, H., Kirkwood, J. M., Falk, C. S., Dafni, U., Sondak, V. K., Tsoutsos, D., et al. (2010). Correlation of molecular human leukocyte antigen typing and outcome in high-risk melanoma patients receiving adjuvant interferon. *Cancer* 116, 4326–4333. doi: 10.1002/cncr.25211
- Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., et al. (2016). Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* 375, 1109–1112. doi: 10.1056/NEJMp1607591
- Hu, J. M., Liang, W. H., Qi, C. H., Wang, X. L., Pan, X. L., Qi, L. W., et al. (2018). HLA-DQB1*03 and DRB1*07 alleles increase the risk of cervical cancer among Uighur and Han women in Xinjiang, China. *Future Oncol.* 14, 2005–2011. doi: 10.2217/fo-2018-0048
- Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., et al. (2016). Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell* 165, 35–44. doi: 10.1016/j.cell.2016.02.065
- Jagga, Z., and Gupta, D. (2014). Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms. *BMC Proc.* 8:S2. doi: 10.1186/1753-6561-8-S6-S2
- Joffre, O. P., Segura, E., Savina, A., and Amigorena, S. (2012). Cross-presentation by dendritic cells. *Nat. Rev. Immunol.* 12, 557–569. doi: 10.1038/nri3254
- Johansen, L. L., Lock-Andersen, J., and Hviid, T. V. F. (2016). The pathophysiological impact of HLA class Ia and HLA-G expression and regulatory T cells in malignant melanoma: a review. *J. Immunol. Res.* 2016:6829283. doi: 10.1155/2016/6829283
- Kandilarova, S. M., Paschen, A., Mihaylova, A., Ivanova, M., Schadendorf, D., and Naumova, E. (2016). The influence of HLA and KIR genes on malignant melanoma development and progression. *Arch. Immunol. Ther. Exp.* 64, 73–81. doi: 10.1007/s00005-016-0437-3
- Kaur, H., Bhalla, S., and Raghava, G. P. S. (2019). Classification of early and late stage liver hepatocellular carcinoma patients from their genomics and epigenomics profiles. *PLoS One* 14:e0221476. doi: 10.1371/journal.pone.0221476
- Kawakami, Y., Dang, N., Wang, X., Tupesis, J., Robbins, P. F., Wang, R. F., et al. (2000). Recognition of shared melanoma antigens in association with major HLA-A alleles by tumor infiltrating T lymphocytes from 123 patients with melanoma. *J. Immunother.* 23, 17–27. doi: 10.1097/00002371-200001000-00004
- Khair, D. O., Bax, H. J., Mele, S., Crescioli, S., Pellizzari, G., Khiabany, A., et al. (2019). Combining immune checkpoint inhibitors: established and emerging targets and strategies to improve outcomes in melanoma. *Front. Immunol.* 10:453. doi: 10.3389/fimmu.2019.00453
- Kishore, J., Goel, M., and Khanna, P. (2010). Understanding survival analysis: kaplan-meier estimate. *Int. J. Ayurveda Res.* 1:274. doi: 10.4103/0974-7788.76794
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, Y., Krahn, J. M., Flake, G. P., Umbach, D. M., and Li, L. (2015). Toward predicting metastatic progression of melanoma based on gene expression data. *Pigment Cell Melanoma Res.* 28, 453–463. doi: 10.1111/pcmr.12374
- Listgarten, J., Brumme, Z., Kadie, C., Xiaojiang, G., Walker, B., Carrington, M., et al. (2008). Statistical resolution of ambiguous HLA typing data. *PLoS Comput. Biol.* 4:e1000016. doi: 10.1371/journal.pcbi.1000016
- Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173, 400–416.e18. doi: 10.1016/j.cell.2018.02.052
- Marincola, F. M., Shamamian, P., Rivoltini, L., Salgaller, M., Cormier, J., Restifo, N. P., et al. (1995). HLA associations in the antitumor response against malignant melanoma. *J. Immunother. Emphasis Tumor Immunol.* 18, 242–252. doi: 10.1097/00002371-199511000-00005
- Marsh, S. G. E. (2003). HLA nomenclature and the IMGT/HLA sequence database. *Novartis Found. Symp.* 254, 165–173.
- Mendez, R., Aptsiauri, N., Del Campo, A., Maleno, I., Cabrera, T., Ruiz-Cabello, F., et al. (2009). HLA and melanoma: multiple alterations in HLA class I and II expression in human melanoma cell lines from ESTDAB cell bank. *Cancer Immunol. Immunother.* 58, 1507–1515. doi: 10.1007/s00262-009-0701-z
- Nestle, F. O., Burg, G., Föh, J., Wrone-Smith, T., and Nickoloff, B. J. (1997). Human sunlight-induced basal-cell-carcinoma-associated dendritic cells are deficient in T cell co-stimulatory molecules and are impaired as antigen-presenting cells. *Am. J. Pathol.* 150, 641–651.
- Ossio, R., Roldán-Marín, R., Martínez-Said, H., Adams, D. J., and Robles-Espinoza, C. D. (2017). Melanoma: a global perspective. *Nat. Rev. Cancer* 17, 393–394. doi: 10.1038/nrc.2017.43
- Paschen, A., Jing, W., Drexler, I., Klemm, M., Song, M., Müller-Berghaus, J., et al. (2005). Melanoma patients respond to a new HLA-A*01-presented antigenic ligand derived from a multi-epitope region of melanoma antigen TRP-2. *Int. J. Cancer* 116, 944–948. doi: 10.1002/ijc.21132
- Patiyal, S., Agrawal, P., Kumar, V., Dhall, A., Kumar, R., Mishra, G., et al. (2019). NAGbinder: an approach for identifying N-acetylglucosamine interacting residues of a protein from its primary sequence. *Protein Sci.* 29, 201–210. doi: 10.1002/pro.3761
- Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., et al. (2011). *Scikit-learn: Machine Learning in Python* Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. Available at: <http://scikit-learn.sourceforge.net> (accessed September 27, 2019).
- Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* 171, 934–949.e16. doi: 10.1016/j.cell.2017.09.028
- Robinson, J., Soormally, A. R., Hayhurst, J. D., and Marsh, S. G. E. (2016). The IPD-IMGT/HLA database - new developments in reporting HLA variation. *Hum. Immunol.* 77, 233–237. doi: 10.1016/j.humimm.2016.01.020
- Rogel, C. S., Souza-Santana, F. C., de Marcos, E. V. C., Ogawa, M. M., Basso, G., and Tomimori, J. (2019). HLA alleles in renal transplant recipients with nonmelanoma skin cancer in southeastern Brazil. *An. Bras. Dermatol.* 94, 287–292. doi: 10.1590/abd1806-4841.20197322
- Sabapathy, K., and Nam, S. Y. (2008). Defective MHC class I antigen surface expression promotes cellular survival through elevated ER stress and modulation of p53 function. *Cell Death Differ.* 15, 1364–1374. doi: 10.1038/cdd.2008.55
- Saghapour, E., Kermani, S., and Sehhati, M. (2017). A novel feature ranking method for prediction of cancer stages using proteomics data. *PLoS One* 12:e0184203. doi: 10.1371/journal.pone.0184203
- Schemper, M. (1993). The relative importance of prognostic factors in studies of survival. *Stat. Med.* 12, 2377–2382. doi: 10.1002/sim.4780122413
- Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* 54, 15–39. doi: 10.1038/jhg.2008.5
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA. Cancer J. Clin.* 70, 7–30. doi: 10.3322/caac.21590
- Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199. doi: 10.1056/NEJMoa1406498
- Spraggs, C. F., Parham, L. R., Briley, L. P., Warren, L., Williams, L. S., Fraser, D. J., et al. (2018). Characterisation of the HLA-DRB1*07:01 biomarker for lapatinib-induced liver toxicity during treatment of early-stage breast cancer patients with lapatinib in combination with trastuzumab and/or taxanes. *Pharmacogenomics J.* 18, 480–486. doi: 10.1038/tpj.2017.39
- Sun, W., and Schuchter, L. M. (2001). Metastatic melanoma. *Curr. Treat. Options Oncol.* 2, 193–202.
- Traherne, J. A. (2008). Human MHC architecture and evolution: Implications for disease association studies. *Int. J. Immunogenet.* 35, 179–192. doi: 10.1111/j.1744-313X.2008.00765.x
- Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211. doi: 10.1126/science.1240095

- Watts, C. (2004). The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. *Nat. Immunol.* 5, 685–692. doi: 10.1038/ni1088
- Williams, T. M. (2001). Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *J. Mol. Diagn.* 3, 98–104. doi: 10.1016/S1525-1578(10)60658-7
- Xie, C., Yeo, Z. X., Wong, M., Piper, J., Long, T., Kirkness, E. F., et al. (2017). Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8059–8064. doi: 10.1073/pnas.1707945114
- Yélamos, O., and Gerami, P. (2015). Predicting the outcome of melanoma: can we tell the future of a patient's melanoma? *Melanoma Manag.* 2, 217–224. doi: 10.2217/MMT.15.15
- Yu, L., Tao, G., Zhu, L., Wang, G., Li, Z., Ye, J., et al. (2019). Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. *BMC Cancer* 19:464. doi: 10.1186/s12885-019-5646-9
- Zhang, M., and Zhang, N. (2017). Clinical and prognostic factors in 98 patients with malignant melanoma in China. *J. Int. Med. Res.* 45, 1369–1377. doi: 10.1177/0300060517708922
- Zhu, S., Van den Eynde, B. J., Coulie, P. G., Li, Y. F., El-Gamil, M., Rosenberg, S. A., et al. (2012). Characterization of T-cell receptors directed against HLA-A*01-restricted, and C*07-restricted epitopes of MAGE-A3, and MAGE-A12. *J. Immunother.* 35, 680–688. doi: 10.1097/cji.0b013e31827338ea
- Zhu, Y., Qiu, P., and Ji, Y. (2014). TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* 11, 599–600. doi: 10.1038/nmeth.2956

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Dhall, Patiyal, Kaur, Bhalla, Arora and Raghava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.