Aus dem
Lehrstuhl für Biometrie und Bioinformatik, IBE
Ludwig-Maximilians-Universität München

Diagnosing and predicting clinical outcomes based on computational methods for immune microenvironment patterns: two examples

Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Shangming Du

aus
Guangdong China

Jahr
2023

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten:	Prof. Dr. Ulrich Mansmann
Zweites Gutachten:	Prof. Dr. Michael Ingrisch
Drittes Gutachten:	PD Dr. André Jakob
Viertes Gutachten:	Prof. Dr. Sibylle Ziegler

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 22.11.2023

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Affidavit

DU, Shangming

Surname, first name

Street

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Diagnosing and predicting clinical outcomes based on computational methods for immune microenvironment patterns: two examples

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Guangdong, China, 10.01.2024 — Shangming Du

place, date — Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

DU, Shangming

Surname, first name

Street

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Diagnosing and predicting clinical outcomes based on computational methods for immune microenvironment patterns: two examples

is congruent with the printed version both in content and format.

Guangdong, China, 10.01.2024

place, date

Shangming Du

Signature doctoral candidate

Table of contents

List of figures

List of tables

List of abbreviations

AUC	Area under the curve
BM	Basement membrane
CAA	Coronary artery aneurysm
CIBERSORT	Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts
CTL	Cytotoxic T lymphocyte
DAMP	Damage-associated molecular pattern
DC	Dendritic cell
DCA	Decision curve analysis
DEG	Differentially expressed gene
DIS	Diagnostic immune score
ECM	Extracellular matrix
FC	Febrile control
FPKM	Fragments-per-kilobase-per-million
GEO	Gene expression omnibus
GSEA	Gene set enrichment analysis
HC	Healthy control
HLA	Human leukocyte antigen
ICD	Immunogenic cell death
IFN	Interferon
IL	Interleukin
IPF-LASSO	Integrative LASSO with penalty factors
IQR	Interquartile range
IVIG	Intravenous immunoglobulin
KD	Kawasaki disease
KEGG	Kyoto Encyclopedia of Genes and Genomes
K-M curve	Kaplan–Meier curve
LASSO	Least absolute shrinkage and selection operator
MFS	Metastasis-free survival
MSigDB	The Molecular Signatures Database
NK	Natural killer cell
OS	Overall survival
PCA	Principal component analysis
ROC curve	Receiver-operating characteristic curve
TAM	Tumor-associated macrophage
TIL	Tumor infiltration lymphocyte
TME	Tumor microenvironment
TNF	Tumor necrosis factor
TPM	Transcripts-per-kilobase-million
UVM	Uveal melanoma

Abstract

The immune system is a critical component of the delicate balance between human health and disease, particularly as immunotherapy gains popularity. In order to identify patients at an early stage and develop individualized disease prevention strategies, it is important to systematically and accurately describe the immune environment before disease onset. However, the immune environment is a complex system consisting of immune cells, antibodies, complement, and cytokines. Traditional methods of monitoring immune patterns in clinical settings are limited in accuracy and reliability, which have created a pressing need for more advanced technologies. Therefore, bioinformatics has become an important tool in the field of disease immunology research. In recent years, computational methods and approaches have been developed, which specifically enumerate the immune microenvironment and allow for the further quantification of the complex immune system.

The objective of this project is to explore the role of bioinformatics in the prediction of the diagnosis or prognosis of immune-related diseases. The focus of this dissertation is on two types of diseases, each with a different prediction model, which are described separately in two independent chapters due to their respective specificities. In Chapter 2, the immune cell compositions of blood samples from patients with Kawasaki disease (KD)—an immune-mediated inflammation in children—were enumerated. A novel algorithm was developed for predicting KD diagnosis based on this enumeration. Using the model, patients with KD and febrile controls could be well distinguished in the test set, with an AUC of 0.80. In Chapter 3, a study concerning a tumor disease, uveal melanoma (UVM), was conducted. The study integrates the patterns of basement membrane and immunogenic cell death to investigate the immune microenvironment patterns in UVM patients. On this basis, three models using different algorithms were constructed and the optimal one was selected after comparing them in validation set, which was the model generated by the IPF-LASSO algorithm, with an AUC of 0.740, 0.841 and 0.835 for 1-, 3-, and 5-year overall survival, respectively. Furthermore, we assessed its performance on the test sets with different survival outcomes and preliminary investigated its association with the response to UVM immunotherapy.

As such, this dissertation highlights how the integration of machine learning and high-throughput data can improve the characterization of the disease-associated immune microenvironment and the development of better prediction models. The study results demonstrate that the bioinformatics-based approach presented in this project holds great potential for predicting the diagnosis or prognosis of immune-related diseases, and could ultimately improve patient outcomes.

1. Introduction

The immune system performs a vital function in the balance between health and disease especially in the current environment where immunotherapy is becoming increasingly popular. Its over-activation can lead to autoimmune or inflammatory diseases, while a state of immunosuppression increases the risk of infection and abnormal expression of tumor cells. If the immune environment can be systematically and accurately described before disease onset, patients can be identified early enough to develop more precise individualized disease prevention strategies or immunotherapy regimens, thereby improving their health quality [1, 2]. However, the immune environment is made up of a complex variety of immune cells, antibodies, complement and cytokine systems. Traditional methods for monitoring immune patterns in clinical settings have limitations in accuracy and reliability, leading to an urgent need for more advanced technologies [3, 4]. To overcome these limitations, bioinformatics has become an important tool in disease immunology research.

Since the 21st century, bioinformatics has made significant progress in various aspects [5]. On one hand, the emergence and improvement of high-throughput sequencing technology have generated vast amount of genomic data, which provides a good foundation to conduct biomedical big data analysis; on the other hand, major breakthroughs in machine learning algorithms have enabled us to comprehensively and efficiently interpret these data and leverage them to develop better biomarkers for disease diagnosis or prognosis prediction. Machine learning algorithms based on biomedical big data have been applied to a variety of diseases including infectious diseases and cancer [6]. In addition, in recent years, researchers have developed a series of computational methods-based approaches specifically for the enumeration of immune microenvironment, helping further quantify the intricate immune system [7, 8].

1.1 Kawasaki disease (KD)

1.1.1 Epidemiology and pathophysiology

Kawasaki disease (KD), which was initially identified by Tomisaku Kawasaki in Japan in 1967 [9], is an acute severe febrile inflammatory illness of unknown etiology that primarily affects young children and infants. For children living in developed countries, it causes the majority of heart disease acquired over time [10-12]. Despite the fact that

it occurs at any age in children, KD is most commonly seen in children under the age of five, and the risk is 1.5 times higher for boys than for girls [13]. KD is often characterized by a high fever (over 38.5°C) lasting for more than five days, associated with at least four of the following clinical manifestations: rash, conjunctivitis, cervical lymphadenopathy, oral changes (i.e., cracking, erythema, and/or strawberry tongue), and peripheral extremity changes (i.e., edema, erythema, and/or peeling) [10, 14]. KD principally attacks blood vessels throughout the body. If left untreated appropriately, it can result in fatal complications such as coronary artery aneurysm (CAA), myocardial infarction, and sudden death, by initially targeting the coronary arteries [15-17]. With regards to standard treatment, high doses of oral aspirin and intravenous immunoglobulin (IVIG) therapy are recommended according to the current guidelines. In patients without response to IVIG, glucocorticoids are administered as preemptive therapy [18, 19].

To date, the exact mechanism underlying KD pathogenesis has not yet been fully understood despite extensive researches. The pathophysiology of KD is hypothesized to begin with an abnormal immune response to infection, most likely to be viral or *Yersinia*, which triggers a T-helper cell-mediated immune response, generating a range of pro-inflammatory cytokines, including interleukin (IL)-6 and tumor necrosis factor-α (TNF-α). As a result, inflammation and damage occur in the endothelium of the coronary arteries [20-22]. Recent clinical and laboratory findings have also reported the involvement of innate immune cells (i.e., neutrophils and macrophages), in the pathogenesis of KD, which is similar to those febrile illness caused by bacterial infection [23, 24]. Besides, activated γδT cells appear to be the predominant T cell subset [25]. These innate immune cells release effector molecules such as matrix metalloproteinases (MMPs), which can impair the arterial wall formed from elastic lamina [26]. The inflammasome is considered to be a crucial component of the innate immune system. Researchers have also found the up-regulation of IL-1 signaling pathway in KD, which indicates that the inflammasome is involved in the progression of vasculitis in KD [27]. Notably, the NLRP3 inflammasome seems to serve as one of the inflammatory signaling factors that contribute to KD vasculitis [28]. Furthermore, several studies on the blood samples from KD patients have revealed that the IVIG treatment may benefit from binding multiple pattern recognition molecules, which are indispensable to innate immunity [29].

1.1.2 Current diagnosis

Until now, KD diagnosis relies primarily on clinical criteria, but additional diagnostic tools such as laboratory testing and imaging techniques are also important. These can include auscultation, chest X-ray, electrocardiograph (ECG), and echocardiography [15, 30]. Standardized diagnostic criteria have been developed internationally, but variability in clinical interpretation can lead to missed or misdiagnosed cases, ultimately leading to delayed diagnosis [15]. Recent studies suggest that assessing peripheral inflammatory factors, such as CRP, neutrophils, albumin, and hemoglobin, during the acute stage of KD is essential for early diagnosis [31]. Elevated blood immunoglobulin and platelet levels during recovery from KD may also reflect the systemic inflammation present during the acute stage [32]. By utilizing these complementary diagnostic tools and monitoring systemic inflammatory parameters, physicians can improve their ability to diagnose KD in a timely manner. KD is generally divided into an acute, sub-acute and convalescent phase. The acute febrile phase, characterized by high fever, lasts 7 to 14 days. The sub-acute phase follows, during which the patient may experience finger peeling, joint pain, and abnormal laboratory findings. This period carries the greatest risk of cardiac sequelae (i.e., CAAs) and lasts for about 4 weeks. The convalescent phase is typically asymptomatic and lasts for 4 to 8 weeks after the onset of the disease. However, the risk of aneurysm development remains during this phase. It's important to note that individual clinical manifestations may not present at the same time, and careful questioning and examination may be necessary to further assist in the diagnosis of KD [33]. Echocardiography should be considered in children with any major clinical features with prolonged unexplained fever. However, there is no clear test to distinguish KD from scarlet fever, hand-foot-and-mouth-disease, roseola infantum, or juvenile idiopathic arthritis, which cause similar symptoms [34, 35]. The results of all these studies suggest that, we need to improve the early and accurate identification of KD patients for early IVIG treatment in order to minimize the number of undiagnosed cases.

With the introduction of artificial intelligence algorithms, researchers can convert massive amounts of clinical documentation data into models of treatment methods to guide clinical practice. For example, Wang et al [36] retrospectively retrieved clinical electronic case information, then successfully used convolutional neural networks to identify KD patients. Subsequently, they again successfully construct a predictive model for identifying IVIG-insensitive KD patients using a new machine-learning algorithm. In addition, the development of high-throughput techniques has also facilitated individualization of the diagnosis of KD at the genetic level, for instance, Jaggi et al. [37] and Wright et al. [35] have successfully constructed diagnostic signatures consisting of different sets of genes that could be used to discriminate KD

from other inflammatory illness, based on transcriptomic profiling data, respectively. With the rapid development of machine learning algorithms, it is believed that in the near future, artificial intelligence will be able to provide clinical practitioners with more accurate decision support analysis on all aspects of the diagnosis, differential diagnosis and treatment, and even the long-term management of KD patients.

1.2 Uveal melanoma (UVM)

1.2.1 Epidemiology and pathophysiology

Despite its rarity, uveal melanoma (UVM) is a significant primary intraocular malignant neoplasm, which originates from melanocytes located in the uveal tract [38]. It accounts for around 5% of all melanoma types [39]. The incidence of UVM is correlated with geographical region, race, and gender. In Western countries, incidence is estimated to be approximately 5 cases per million people per year, with a slightly higher incidence in men than women [40, 41]. There is also evidence that individuals with light-colored eyes may be at an increased risk of developing UVM [42]. Most people with UVM experience only mild vision problems or even no symptoms, and the disease is therefore usually diagnosed during a routine eye exam [43]. Valuable tools for diagnosis include slit-lamp, gonioscopy, ultrasonography (USG), and optical coherence tomography (OCT). The occurrence of complications (i.e., hyphemia, extraocular extension, cataracts, and secondary glaucoma) is also a significant sign to be considered [40, 44]. However, ultimate diagnosis needs to be confirmed by the gold standard – fine needle aspiration biopsy (FNAB) [44].

The pathophysiology of UVM usually involves genetic mutations or/and epigenetic alterations. These abnormal changes lead to highly malignant proliferation of melanocytes and even metastasis. Available researches [45-47] have shown that BAP1, GNAQ, GNA11, and SF3B1 seem to be the main sites of these mutation. Among them, the BAP1, located on chromosome 3 and found to be mutated in nearly 50% of all UVM patients, is involved in DNA repair and chromatin remodeling [48]. Other genes are usually enriched in RNA splicing and cell signaling pathways, specifically the G-protein signaling [45, 49-51]. On the other hand, epigenetic alterations are proved to play a preeminent role in the UVM pathophysiology as well. DNA methylation can alter gene expression patterns and contribute to tumorigenesis. PRAME, or the preferentially expressed antigen in melanoma, is being studied as a novel epigenetic biomarker for UVM metastasis [52]. In addition, the immune system is closely associated with the progression and metastasis of various tumors, and UVM is

no exception. Available studies [24, 53, 54] have indicated that the immune microenvironment in KD is characterized by an overactive innate immune system, with increased infiltration levels of monocyte, macrophage and neutrophil. Besides, several investigations in UVM [55-57] have observed multiple signals of inflammatory microenvironment, which is initiated by inflammatory factors, such as HLAs, cytokines, and chemokines, which lead in recruitment of inflammation-related cells and promotion of angiogenesis.

1.2.2 Current prognostic prediction approaches

Uveal melanoma is an extremely aggressive form of cancer, with a poor prognosis and a strong tendency toward metastasis. This metastasis most commonly occurs through the bloodstream, with up to 80% of cases resulting in liver metastases [58-60]. Unfortunately, UVM patients with metastasis have a median survival of only 2 to 8 months [40, 47, 59]. Identifying potential prognostic indicators for metastatic UVM, therefore, can be crucial in understanding its heterogeneity, improving prognostic predictions, and guiding individualized treatments. Numerous studies have found that UVM patients of advanced age experienced a poorer prognosis [61]. In a retrospective study of 99 patients with metastatic UVM, Lorenzo et al. [58] discovered that the median survival for patients with age ≤ 65 years was 11 months, compared to 8 months for those over the age of 65. The relationship between gender and prognosis in UVM remains controversial right now. Some studies suggest that hormonal factors may result in a lower metastatic rate for female patients [62], whereas more researchers have not found a statistical difference in the effect of gender on prognosis [58, 63, 64]. The stage classification system is a crucial prognostic factor for UVM. An investigation of 7,731 UVM patients [65] revealed that the 10-year metastatic rates for UVM with T1, T2, T3, and T4 stages were 15%, 25%, 49%, and 63%, respectively.

As modern molecular biology techniques have advanced, researchers have increasingly turned their attention to the development of prognostic biomarkers at the molecular level. As mentioned in **Section 1.2.1**, genetic traits are highly associated with the progression and metastasis, making them crucial for prognosis of UVM. Monosomy 3 is the aberration of chromosome 3. It is linked to histopathologic risk factors that lead to metastasis [66]. Additionally, higher rate of chromosome 8q-gain or 6p-loss also indicates a worse UVM outcome. In contrast, chromosome 6p-gain usually predicts a good prognosis [67]. In terms of bioinformatics approaches, Xu et al. [68] identified 103 potential candidate prognostic factors for UVM by analyzing differentially expressed genes between patients with primary and metastatic UVM. Next, they further identified a six-gene signature using the Kaplan–Meier method. By providing additional

clues for prognosis, this study enhanced our understanding of the process of tumor metastasis occurrence and progression in UVM. However, multifactorial analysis was not included to confirm the independent prognostic role of candidate genes, and its diagnostic value remains unclear. Despite this, there is still no doubt that high-throughput techniques could facilitate the identification of ideal prognostic biomarkers and improve risk stratification for patients with UVM.

1.2.3 Traditional treatment and immunotherapy

Traditional UVM treatment includes surgical excision, chemotherapy drugs and radiotherapy. For primary tumor, radiotherapy is one of the most common and effective treatments, which has been proved to be effective in controlling local tumor growth in many patients [44]. Accompanying complications can occur locally, however [69]. Moreover, although various radiotherapeutic approaches for primary UVM have improved over time, treatments for metastasis remain disappointing [67, 70]. For metastatic UVM, chemotherapeutic agents are usually applied, either alone or in combination with other treatments. However, clinical trials have shown that they induced a response in a small subset of UVM patients, but their overall efficacies were limited, with no statistically significant survival improvement found [71-73].

Thanks to the era of immunotherapy, immunotherapeutic agents have been widely tested and shown promise in treating UVM [72, 74, 75]. CTLA-4 monoclonal antibodies were approved in the United States and Europe in 2011 as first-line agents therapies for the treatment of advanced melanoma [76]. Nevertheless, the response rate for Ipilimumab in metastatic UVM has been modest, ranging from 5%-10% [77]. Another immune checkpoint inhibitor, Pembrolizumab against the PD-1 receptor, has also been approved for advanced melanoma. However, it exhibited a response rate of only 3.6% in a large retrospective analysis [78]. These non-responders may be associated with low mutation load and new immune checkpoint acquisition [79], which are possible to be identified by the changes of immune landscape. Therefore, a novel predictive algorithm based on the characterization of tumor immune microenvironment may be potential to serve as a predictor for UVM prognosis and immunotherapeutic response.

1.3 Immune microenvironment

Tumor cells are not isolated entities as they exist within an active microenvironment that plays a crucial role in tumor development. The cell infiltration within the tumor microenvironment is critical in both tumor killing and immune escape. It comprises

intrinsic non-tumor cells (i.e., endothelial cells and fibroblasts) and extracellular matrices, all of which contribute to tumor formation and progression. Moreover, the tumor microenvironment contains cells from the immune system and non-cellular molecules such as growth factors and proteases, in addition to stromal components [80, 81]. In the advanced stages of the disease, the immune system portrays duality, as immune cells exert an anti-tumor effect in the early stages of tumor invasion, but later turn into pro-tumor phenotypes, assisting in tumor immune escape, leading to the formation of inflammatory microenvironments. There is a great deal of diversity and complexity in the of immune cell components, including both adaptive immune system and innate immune system, including dendritic cells (DCs), natural killer cells (NKs) and macrophages that play a role in antigen presentation [56, 81]. Tumor infiltration lymphocytes (TILs) infiltrate in the local tumor lesions. In interactions with tumor cells, components of the tumor immune microenvironment, represented by tumor-associated macrophages (TAMs) and regulatory T cells (Tregs), contribute to immunosuppression and tumor growth [82-84]. Numerous studies have demonstrated that components of the immune microenvironment are associated with the occurrence, development, and therapeutic efficacy of a wide range of tumors [85-88]. Additionally, it has been found that recurrent tumors of the same size exhibit greater resistance to treatment than primary tumors, and this phenomenon was accompanied by the differences of immune microenvironments, although tumor cells from both tumors show no heterogeneity [89]. Apart from tumors, the immune microenvironment also plays a crucial role in autoimmune and inflammatory diseases [90-92].

1.4 Basement membrane

As specialized structures of the extracellular matrices (ECMs), basement membranes (BMs) provide mechanical support and regulate cellular behaviors. They are dense and thin sheets (50–100 nm) beneath epithelial and endothelial cells. Their primary components are proteoglycans and glycoproteins, including type IV collagens, laminins, nidogens, and perlecans, while other components, such as type XV and XVIII collagens and osteopontins, are present in smaller amounts. These components act synergistically with each other to influence both morphology and behavior of cells. In turn, cells adapt to different ECMs by adjusting the expression of surface receptors. During metastatic process, tumor cells degrade the BM by up-regulating expression of MMPs and heparanase. The damaged BM not only ceases to be a barrier to tumor cells, but may even become an accomplice to tumor metastasis [93]. BM proteins are key pathogenic factors and targets of autoantibodies in various diseases [94].

Additionally, several studies have also found the correlation between poor prognosis and the proteins relevant to BM components or stiffness in different types of tumors [95, 96]. Recently, Jayadev et al. [97] summarized 224 BM-related genes for the first time, bringing fresh perspectives on the role of BM in diseases and providing a more comprehensive understanding and prediction of tumor progression and metastasis.

1.5 Immunogenic cell death

In the last years, investigations have demonstrated that stimulating tumor cells with various therapeutic agents not only kills them, but also triggers the expression of antigenic signaling molecules on their surface, converting non-immunogenic cells into immunogenic cells. This phenomenon of immune-related anti-tumor effect in the body is named immunogenic cell death (ICD) [98, 99]. In vivo, ICD inducers recruit initial immune effector cells to the local cancer site, leading to inhibitions of tumor growth. This inhibitory effect is dependent (at least partially) on the immune system [100]. Inflammatory mediators, precursor inflammation cytokines and danger associated molecular patterns (DAMPs), are released and/or expressed during the process [101]. In the tumor ICD process, various DAMPs are released to regulate different immune responses with "find me" or "eat me" signals [102]. The patterns of predominant DAMPs expressed include: calreticulin (CRT) exposure, heat shock protein (HSP) translocation, interferon I (IFN I) secretion, and release of high mobility group box protein B1 (HMGB1) and/or adenosine triphosphate (ATP). Subsequently, Immature DCs transform into mature DCs by binding to these factors via pattern recognition receptors, and further phagocytize the apoptotic tumor cells. In the meanwhile, they present these cells to $CD8^+$ T cells together with co-stimulatory molecules (CD83/86) and major histocompatibility complex I (MHC I), leading to the activation of cytotoxic T lymphocyte (CTL). In addition, tumor cell-secreted cytokines recruit NKs to participate in the regulation of adaptive immune response. Moreover, different studies [103-105] have shown that ICD-inducing chemotherapy enhanced the efficacy of immune checkpoint blockades. For melanoma, Choi et al [106] substantiated in experiments conducted in vitro and in vivo that the complex agent of Oxaliplatin and *N*α-deoxycholyl-L-lysyl-methylester (DCK) effectively induced the expression of CRT and HMGB1 in tumor cells, resulting in ICD and effective suppression of tumor. Overall, due to its ability to drive immunogenicity and release tumor antigens within tumors, ICD is anticipated to generate novel insights and strategies for immunotherapy.

1.6 Bioinformatic methods for model development

1.6.1 Regression

1.6.1.1 Logistic regression

Logistic regression, which falls under the category of generalized linear models (GLMs), is a commonly employed statistical modeling technique. It establishes a relationship between outcomes and predictor variables, assuming that outcome variables are distributed from the exponential family. For binary outcomes, it is assumed to follow a Bernoulli distribution. [107]. In biomedical research, logistic regression has numerous applications, including diagnostic prediction, treatment efficacy, and patient outcome prediction. The most popular application is the prediction of disease diagnosis. By analyzing multiple predictor variables, including demographic information, lifestyle habits, medical history, and various omics data, logistic regression models can estimate the probability of a patient developing a particular disease. This information can enable healthcare providers to intervene early and personalize treatment plans for patients. In this project, we combine it and LASSO regression to develop a predictive model for KD diagnosis using the “glmnet” R package [108].

1.6.1.2 Cox regression

In biomedical research, survival data are analyzed using Cox regression, which is additionally referred to as proportional hazards regression [109]. This method models the relationship between predictor variables and the time to an event of interest. It is commonly used in medical research to analyze right-censored survival data. Cox regression estimates hazard ratios, which represent how much more or less likely an individual is to experience the event of interest based on their level of each predictor variable. This helps identify prognostic predictors linked with disease outcomes and establish predictive models for patient survival. By determining the relative impact of different predictors on survival time, Cox regression allows for more accurate risk assessments and informed decision-making in healthcare. In this project, we implement it to develop a predictive model for UVM prognosis using the "survival" R package. Kaplan–Meier (K-M) curves and log-rank tests using the “survminer” R package were utilized to visualize and compare the survival probability, respectively.

1.6.1.3 LASSO regression

The least absolute shrinkage and selection operator (LASSO) regression, proposed by Robert Tibshirani [110], is an algorithm used for constructing interpretable models. It achieves both estimation of unknown parameters and variable screening by penalizing the absolute value of the regression coefficients in the least squares estimation, hence its name L_1-penalty. The penalty parameter is λ and it takes values in the range (0, +∞). The penalty function represents a trade-off between goodness-of-fit and model complexity, which is influenced by λ. As λ approaches 0, the model contains more predictors and fits better, but at the cost of predictive performance and interpretability. In contrast, as the value of λ increases, the regression coefficients of some independent variables gradually compress to 0, and the predictors in the model become less. It follows that choosing an appropriate λ is essential for model's interpretability and performance. The cross-validation method is the most commonly used method to achieve this goal, where the *n* samples in a dataset are divided into K subsets to produce K-fold cross-validation ($1<K<n$). In practice, 10-fold cross-validation is generally used [111]. The LASSO method can be used for general linear regression models, as well as Logistic and Cox regression models, and is an essential tool used to solve the problem of model sparsity. Since its introduction, the LASSO method has received widespread attention and has been cited by a large number of researchers, becoming one of the classical statistical methods for high-dimensional variable screening. In this project, we conducted the LASSO using the "glmnet" R package [108] and every penalty parameter (λ) was determined via 10-fold cross-validation.

1.6.1.4 IPF-LASSO

Integrative LASSO with penalty factors (IPF-LASSO) is an extension of the standard LASSO algorithm that has the ability to accommodate diverse sources of data, as proposed by Boulesteix et al. [112]. It is specially designed to handle situations where multiple types of biomarkers (e.g., multi-omics) are available for the same patient [113]. The method globally assigns the penalty parameters λ to each block of data, which can be determined according to practical considerations, identify the contributions of these blocks. The "ipflasso" R package developed by Boulesteix et al. [112] can implement this IPF-LASSO algorithm and its variant that can process continuous, binary, and survival outcomes. In this project, we leveraged the IPF-LASSO as one of the three algorithms to construct the model for UVM prognostic prediction. It was applied on transcriptomic datasets with two different gene sets that could be treated as two blocks in the "ipflasso" R package.

1.6.1.5 ROC and time-dependent ROC curve

In a binary classifier system, the receiver operating characteristic (ROC) curve graphs its performance, such as a diagnostic test [114]. By varying the recognition threshold, the ROC curve depicts the values of sensitivity (true-positive rates, TPRs) and values of 1-specificity (false-positive rates, FPRs), which can visually display the accuracy of a diagnostic test. A high-performing test yields an area under the curve (AUC) close to 1, while a useless test has an AUC of 0.5. ROC curve has several advantages over other measures of accuracy. For example, they provide a comprehensive evaluation of the trade-off between specificity and sensitivity for different thresholds. In addition, they convey less subjectivity about changes in prevalence than other measures of diagnostic accuracy. Finally, their concise visual presentation provides clarity and ease of interpretation for researchers and clinicians.

In standard ROC curves, we assume that the predictor works equally well at all time points. For survival data, however, many outcomes are time dependent, therefore a ROC curve that varies over time may be a better solution. Considering this, time-dependent ROC curve was developed as an extension of the standard ROC curve to account for the predictive accuracy of a test with censored survival time outcomes (i.e., survival/decease). For prognostic prediction, the Cumulative/Dynamic ROC curve proposed by Heagerty et al. [115] has been popular to use. This approach defines a time-dependent version of sensitivity and specificity: time-dependent sensitivity refers to the likelihood that a subject who experiences the event before time t will be correctly classified as positive by the predictor at threshold c, while time-dependent specificity refers to the likelihood that a subject without the event before time t will have a predicted probability less than or equal to c, including those who have not yet experienced the event.

In this project, we generated ROC curves and their AUCs by using “pROC” R package [116] for the diagnostic model (in **Chapter 2**), and the AUCs of time-dependent ROC curves by using “timeROC” R package for the different prognostic models (in **Chapter 3**) to evaluate their performances.

1.6.2 Consensus clustering

Unsupervised clustering is a statistical method that can divide a set of objects into distinct groups based on similarities in their features, without prior knowledge of their labels [117]. This process can be carried out using different algorithms. One such algorithm, consensus clustering, is widely used in molecular subtyping of patients with

tumors to determine the optimal number of clusters due to its capability of analyzing biomedical data with complex structures. This algorithm involves resampling the original dataset to obtain different subsets, clustering each subset separately and then combining the clusters to generate the consensus index (range [0,1]). The larger the value of the consensus index, the more stable the clusters are. Either of the following criteria is used to determine the optimal number of clusters: (1) comparing the consensus matrices at different clustering numbers, setting the number of clusters that achieves the comprehensive trade-off between the highest possible consensus index for samples within subgroups and the lowest possible consensus index for samples between subgroups; (2) selecting the number of clusters where the middle segment of the cumulative distribution function (CDF) curve of the consensus index is the flattest, resulting in the fewest sample pairs with ambiguous clustering. In this project, we conduct the consensus clustering using the “ConsensusClusterPlus” R package [118]. The K-means algorithm was employed on a random selection of 80% of the given samples with 1,000 iterations. We chose the first criterion to identify the optimal number of clusters.

1.6.3 Differential gene expression analysis

Differential gene expression analysis is a statistical technique commonly used in biomedical research to identify differences between two or more groups of samples in gene expression levels. It helps researchers gain insight into the underlying biological or functional mechanisms of diseases and identify potential therapeutic targets by identifying [119] differentially expressed genes (DEGs). To fit models to this data format, several R packages can be leveraged: 1) “limma” fits a linear model; 2) “DESeq2” and “edgeR” fit generalized linear models, specifically negative binomial distribution models. All of these packages implement an empirical Bayesian approach to borrow information across features. In brief, “limma” is useful for continuous data such as microarray data, while “edgeR” and “DESeq2” are useful for count data like high-throughput sequencing read counts. In this project, both the GEO microarray data and TCGA normalized data (in the form of FPKM or TPM) are continuous data. Therefore, we performed the analysis employing the “limma” R package [120] here.

1.6.4 Functional and pathway enrichment analysis

Functional enrichment analysis is a computational biological technique that can be applied in biomedical studies to highlight significant biological pathways and processes involved in diseases [121, 122]. The approach involves identifying the specific gene

sets where the DEGs are enriched and retrieving corresponding functional terms from an assigned bioinformatics database, in order to identify the up-regulated functional pathways or processes. These results significantly shed light on the underlying molecular mechanisms of diseases and its potential therapeutic targets. Traditional enrichment analysis (based on Fisher's exact test) encounters two common problems: firstly, certain remarkable genes may be excluded because they do not reach the uniform threshold for differentially expressed genes, and secondly, it is difficult to define the overall regulation of a functional pathway with both up- and down-regulated genes – whether it is suppressed or activated? To address these issues, Gene Set Enrichment Analysis (GSEA), a non-parametric algorithm variant of functional enrichment analysis, was proposed [123]. Its basic idea is to utilize a predefined gene sets (usually derived from annotations in databases or previous lab results), and subsequently sorts them according to their expression levels in two groups of samples, then test whether the gene set is enriched in one of the groups. GSEA detects changes of functional enrichment in gene sets rather than individual genes, and therefore can provide more desirable results compared to traditional approaches. In this project, we conducted the GSEA function in the ‘‘clusterProfiler’’ R package [124]. Predefined gene sets were downloaded from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) [125].

1.6.5 Bioinformatics methods for immune microenvironment

1.6.5.1 CIBERSORT

The Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) is a machine-learning algorithm specifically designed for identifying human immune cell phenotypes [7]. It combines support vector regression techniques with a priori knowledge of expression profiles from purified immune cells, enabling highly accurate identification of 22 different cell types. Furthermore, the algorithm utilizes deconvolution algorithm based on Monte Carlo sampling to generate *p*-values for each sample, and the root mean squared error (RMSE) is utilized for assessing the accuracy of the predicted fractions. Notably, Newman et al. [7] and Zhong et al. [92] have demonstrated its ability to enumerate patterns of immune cell infiltration in human lung cancer and systemic lupus erythematosus, respectively, highlighting its tremendous value in scrutinizing the immune microenvironment in tumors and autoimmune disorders, as well as its potential for immune scoring. In this project, we performed this algorithm using the “CIBERSORT” R package [7].

1.6.5.2 ESTIMATEscore

The ESTIMATEscore is a bioinformatics tool used to estimate the immune and stromal content in tumors. It is calculated by summing the scores of immune (ImmuneScore) and stroma (StromalScore), which represent the abundance of respective components. A higher ESTIMATEscore corresponds to a greater presence of immune and stromal cells within the tissue, indicating a lower tumor purity. The technique infers the presence of these cells in tumor samples via analyzing their gene expression data. Through ESTIMATEscores, researchers can gain insight into the tumor microenvironment and interplay amid the neoplasm and the immunological and stromal cells present. It has been widely used in cancer research to prognosticate patient survival outcomes and their susceptibility to immunotherapy, making it a valuable tool for clinicians and researchers alike. In this project, we calculate the ESTIMATEscore, ImmuneScore, and StromalScore by using the “ESTIMATE” R package [8].

1.7 Objective of this project

In this Ph.D. project, our aim is to explore the role of bioinformatics in the prediction of diagnosis or prognosis of immune-related diseases. Specifically, we focus on the combination of gene expression data and machine-learning methods to characterize the immune microenvironment in diseases and develop more accurate and reliable biomarkers to guide clinical decision-making and improve patient outcomes.

Here, two types of diseases with different prediction models are the focuses of this dissertation. Due to their respective specificities, these two studies are described separately in two independent chapters. The study in **Chapter 2** enumerates the immune cell compositions of the blood samples from patients with Kawasaki disease—a severe immune-mediated inflammatory disease in children—and develops a novel algorithm for predicting Kawasaki disease diagnosis based on this. This study has been published in a pediatric journal [126]. The other study in **Chapter 3** is devoted to a tumor disease. This study integrates the patterns of basement membrane and immunogenic cell death to investigate the immune microenvironment status in patients with uveal melanoma, then further develops a biomarker that can prognosticate the clinical outcomes of these patients.

In conclusion, the investigation presented in this dissertation highlights how the integration of machine learning and high-throughput data improves the characterization of immune microenvironment and the development of better prediction models.

2. A Novel Algorithm for Predicting the Diagnosis of Kawasaki Disease

2.1 Introduction

Kawasaki disease (KD) is an immunological disorder characterized by acute febrile inflammation in the pediatric population. It is the predominant etiology acquired heart disease in children, specifically in developed nations, however, there is no known cause for it [10-12]. Some of its complications may be associated with fatal risks, including thrombotic occlusions, coronary aneurysms, and even myocardial infarctions [15]. At present, KD cannot be diagnosed with a specific test, as it is a clinical diagnosis. The distinction between KD and other febrile illnesses can sometimes be tricky, specifically infectious etiologies that present with comparable clinical features or occur concurrently [127], resulting in delayed treatment and complications [128, 129].

In recent years, several biomarkers have been found for diagnosing KD, however they either lack adequate specificity or verification in sufficient numbers of patients [16, 35, 130-133]. Due to inflammatory nature of KD, immunological biomarkers may hold promise [130]. Nevertheless, these biomarkers are identified by flow cytometry or immunohistochemistry, making them difficult to standardize or implement in ordinary clinical practice. These limitations have led researchers to search for new diagnostic strategies, including high-throughput screening methods, but the role of immune cells in whole blood with respect to these approaches remains unclear [35, 37, 132].

In the identification of immune cells, CIBERSORT has been reported to exhibit a high degree of accuracy because it is based on transcriptomic profiling. The use of CIBERSORT to differentiate KD from febrile illness in children remains to be seen. The objective of this study was to infer the CIBERSORT-derived fractions of immune cells in children with KD, and compare it with those in febrile controls (FCs), in order to establish a potential biomarker for KD diagnosis.

2.2 Material and methods

2.2.1 Data sources and preprocessing

To obtain data from the Gene Expression Omnibus (GEO) database, we employed the following search strategy: ((("Expression profiling by array" [DataSet Type] OR "expression profiling by high throughput sequencing" [DataSet Type]) AND "Homo sapiens" [Organism] AND ("mucocutaneous lymph node syndrome" [MeSH Terms] OR Kawasaki disease [Title]) AND (("diagnosis" [MeSH Terms] OR diagnostic [Title]) OR signature [Title])) AND blood [Sample Source]) AND ("infection" [All Fields] OR "infectious" [All Fields]). Initially, there were 5 items recognized (until 2021). Among them, GSE73464 was the SuperSeries composed of three sub-series: GSE73461, GSE73462, and GSE73463, so we excluded it. Next, we checked in each dataset whether their blood samples were collected according to the following criteria: 1) KD cases were diagnosed according to the American Heart Association (AHA) criteria [15]; 2) FC cases manifested symptoms of fever and were only diagnosed infection of definite bacterial, definite viral, or "uncertain" pathogen. Ultimately, four datasets were obtained. **Table 2.1** provides an overview of all datasets used here. Microarray gene expression data on Illumina® platform were processed and normalized using the "lumi" R package [134]. In detecting a gene by multiple probes, only the one with the highest level was selected.

Table 2.1 Summary of KD datasets

Cohort	Dataset	Platform	No. of KDs	No. of FCs	No. of HCs
Discovery: Training & Test set (7:3)	GSE73461	Illumina® HumanHT-12 V4.0	78	242	55
	GSE73462	Illumina® HumanHT-12 V3.0	1	130	16
	GSE73463	Illumina® HumanHT-12 V4.0	146	0	87
	GSE68004 (2 batches)	Illumina® HumanHT-12 V4.0	57	25	19
			19	11	31
Validation	GSE15297	SMD Print_853	23	13	0

Note: *This table was adapted from Du et al. [126].*

2.2.2 Composition estimation of immune cells

We used the LM22 gene signature and the CIBERSORT algorithm with 1,000 permutations to compute the fractions of immune cells in different samples [7]. See **Section 1.6.5.1** for more details.

2.2.3 Pathway and functional enrichment analysis

Functional enrichment analyses were conducted by running the GSEA function [123] in ‘‘clusterProfiler’’ R package [124]. The predefined gene sets for GSEA were downloaded from the MSigDB. See **Section 1.6.4** for more details.

2.2.4 Other statistical analysis

The discovery cohort, which was the merger of GSE73461, GSE73462, GSE73463, and GSE68004 datasets, was randomly allocated into two distinct sets in a 7:3 ratio, namely a training and a held-out test set. The GSE15297 dataset was used for external validation. We employed the LASSO algorithm for binary outcomes, as described in **Section 1.6.1.1** and **Section 1.6.1.3**. The performance of the DIS was assessed using ROC curves and the optimal cut-off point was identified through reaching the largest Youden's index in the “OptimalCutpoints” R package [135]. Group comparisons were performed using the Wilcoxon’s test and Fisher’s exact test for continuous variables and categorical variables, respectively. Kruskal-Wallis tests were used in multiple group comparisons. Spearman’s rank correlation coefficient was used to analyze correlations between the DIS and the expression levels of the selected genes. All statistical tests were two-sided. Statistical significance was all set at $p < 0.05$.

2.3 Results

2.3.1 Characteristics of participants

The strategy for the overall work can be found in **Figure 2.1**. Based on the filter criteria, a discovery cohort of 708 children (300 KDs and 408 FCs) was obtained, which was then randomized as described in **Section 2.2.4**. To verify our results externally, we obtained another independent dataset (GSE15297), which contained 23 KDs and 13 FCs. In **Table 2.2**, we list the baseline clinical and demographic characteristics of the patients.

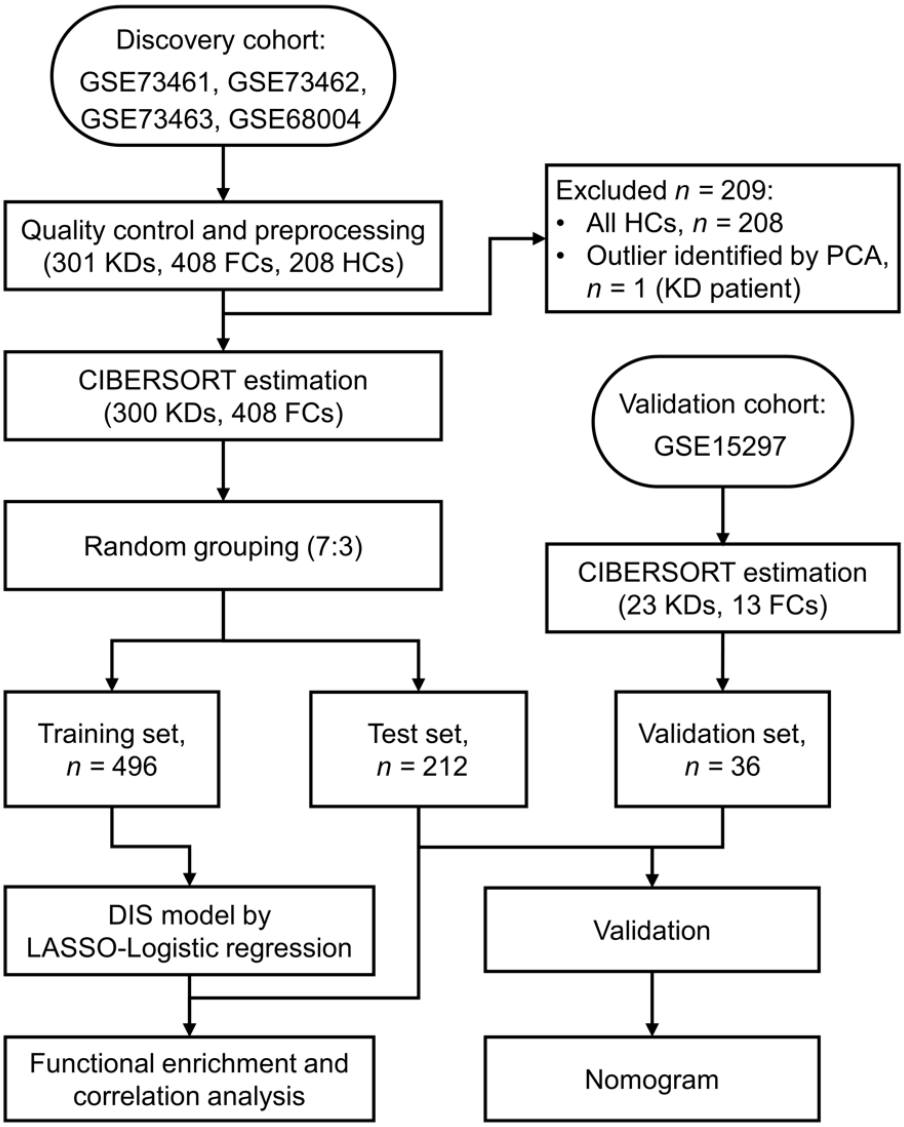


Figure 2.1 Flow chart of this study. *This figure was adapted from Du et al. [126]. DIS, diagnostic immune score.*

Table 2.2 Clinical and demographic characteristics in all KD datasets.

	Discovery*			Validation
	Entire set	**Training set**	**Test set**	**Validation set**
No. of patients	708	496	212	36
Age, months	27.0 [9.0 to 59.0]	27.0 [9.0 to 54.0]	29.0 [8.0 to 66.0]	NA
Sex				
Male	417 (58.9)	291 (58.7)	126 (59.4)	23 (63.9)
Female	291 (41.1)	205 (41.3)	86 (40.6)	13 (36.1)
Outcome				
Kawasaki disease	300 (42.4)	210 (42.3)	90 (42.5)	23 (63.9)
Febrile controls	408 (57.6)	286 (57.7)	122 (57.5)	13 (36.1)
Pathogens in febrile controls				
Definite bacterial	92 (22.5)	70 (24.5)	22 (18.0)	8 (61.5)
Definite viral	141 (34.6)	95 (33.2)	46 (37.7)	5 (38.5)
Uncertain	175 (42.9)	121 (42.3)	54 (44.3)	0

Note: *There are two types of data presented: median [IQR] or n (%).* Children's data did not show any statistically significant differences between the Test and the Training sets. This table was adapted from Du et al. [126]. NA, not applicable.*

2.3.2 Gene expression data preprocessing

We prepossessed the gene expression profile data before model construction, containing quantile normalization, log2-transformation and batch-effect removal. **Figure 2.2** indicates that the samples from GEO were well normalized, guaranteeing the majority of the differences subsequently found between study groups were most probably caused by different environments [136].

After log2-transforming these normalized data, we performed the principal component analysis (PCA) to check the batch effect. According to PCA, a batch effect between datasets were apparently observed and needed to be removed. Here, we used the "ComBat" function in the "SVA" R package, which has been widely used for batch-effect correction by fitting a linear model for each gene [137]. Subsequent PCA

plot improves that this batch-effect was eliminated. Additionally, we identified an outlier from GSE73461 and excluded it (**Figure 2.3A, B**).

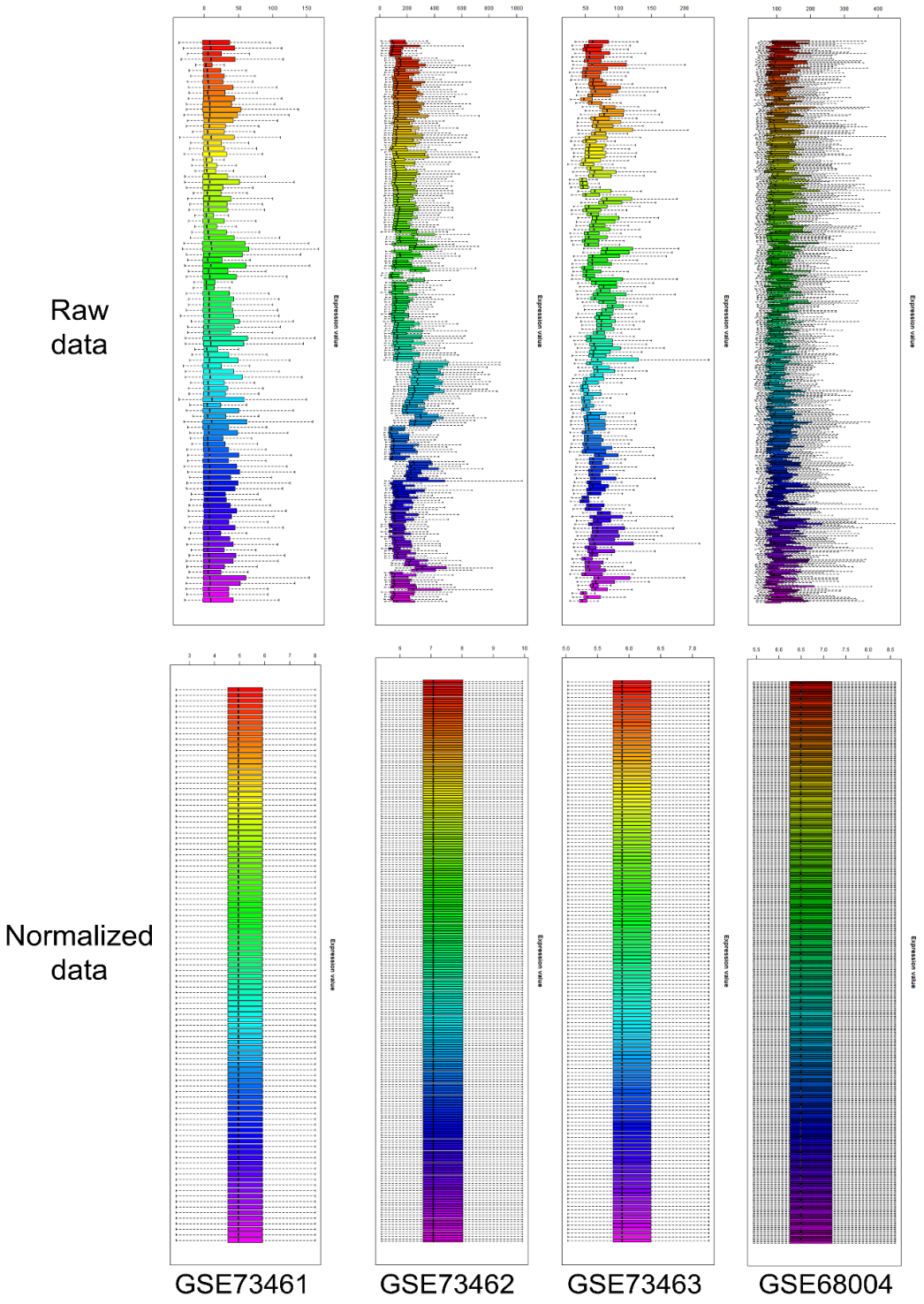


Figure 2.2 Box plots for expression intensity of each sample before and after normalization. *Vertical axis represents individual samples.*

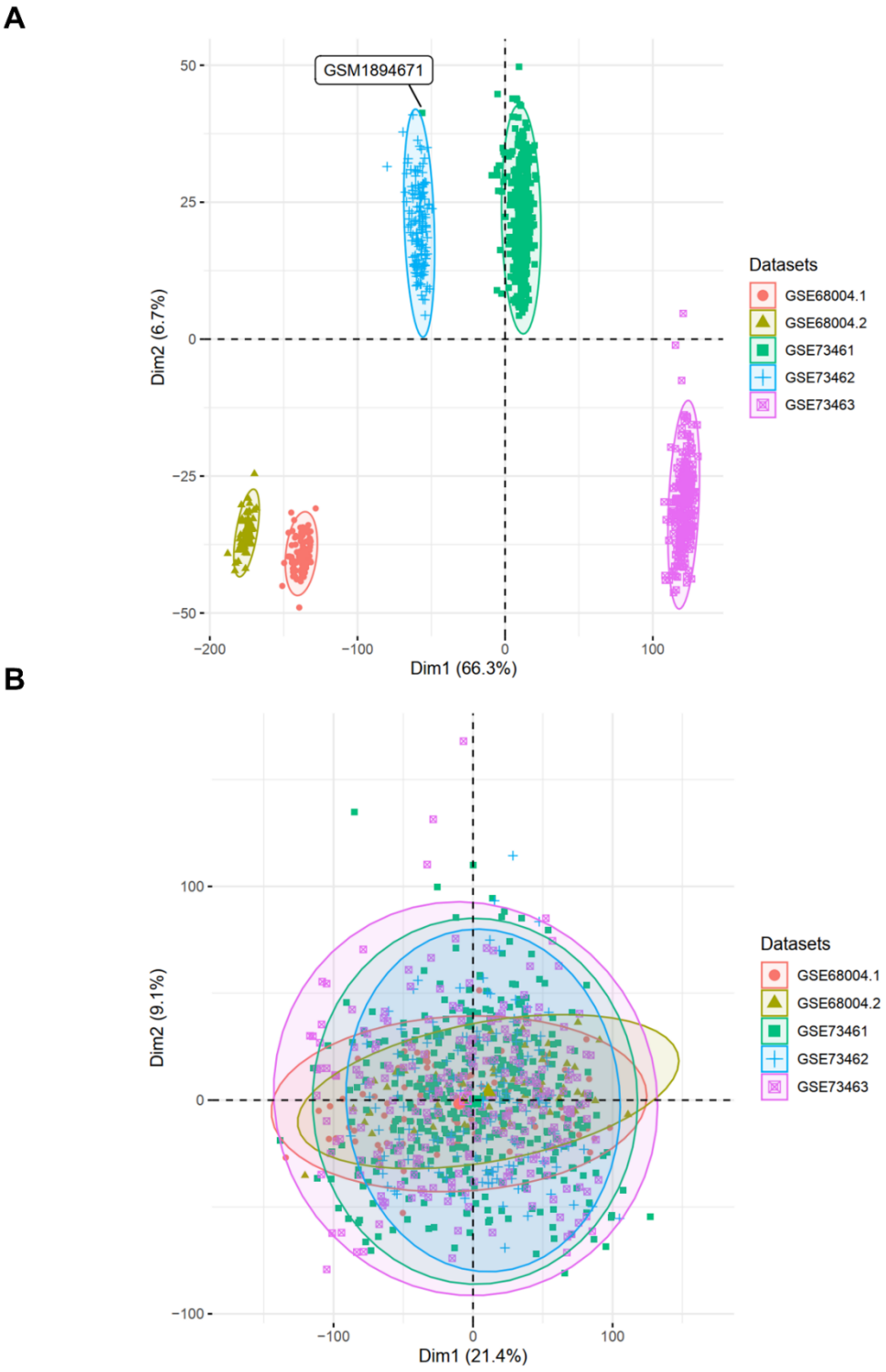


Figure 2.3 Principal component analysis on the discovery cohort before (A) and after (B) batch-effect removal. *This figure was adapted from Du et al. [126].*

2.3.3 Immune cell fractions in different patient groups

The CIBERSORT algorithm was employed in each individual sample to enumerate its immune cell composition. As compared to FC samples, KD samples showed significantly greater proportions of activated mast cells ($p < 0.001$), monocytes ($p = 0.035$), neutrophils ($p < 0.001$), and M0 macrophages ($p < 0.001$). The fractions of M1 and M2 macrophages, $CD8^+$ and $CD4^+$-naïve T cells, γδ T cells, resting mast cells and plasma cells, on the other hand, were lower in KD samples (all p values were less than 0.001; **Figure 2.4**). Generally, blood samples from KD patients tended to contain mostly five cell types making up more than three quarters of the total: $CD4^+$-naïve and $CD8^+$ T cells, neutrophils, monocytes, and M0 macrophages.

The $CD4^+/CD8^+$ T cell ratio has been reported as a potential biomarker to distinguish KD from children with infectious febrile illness [53, 138]. Taking this into account, we additionally measured this ratio. The ratios of $CD4^+/CD8^+$ T cells in KD patients were significantly higher than those in FC patients ($p < 0.001$, **Table 2.3**). Furthermore, **Figure 2.5** shows a significant difference between the two groups in terms of immune cell composition landscape.

In addition, A healthy control group (HC) was also compared with them. In the majority of cell types, higher fractions of these cells observed when comparing the HC group with the KD group were also observed when comparing the FC group with the KD group, and vice versa. In a particular instance, the fraction of plasma cells was reduced when comparing the HC group with the KD group, yet conversely elevated when comparing the FC group with the KD group (**Figure 2.6**).

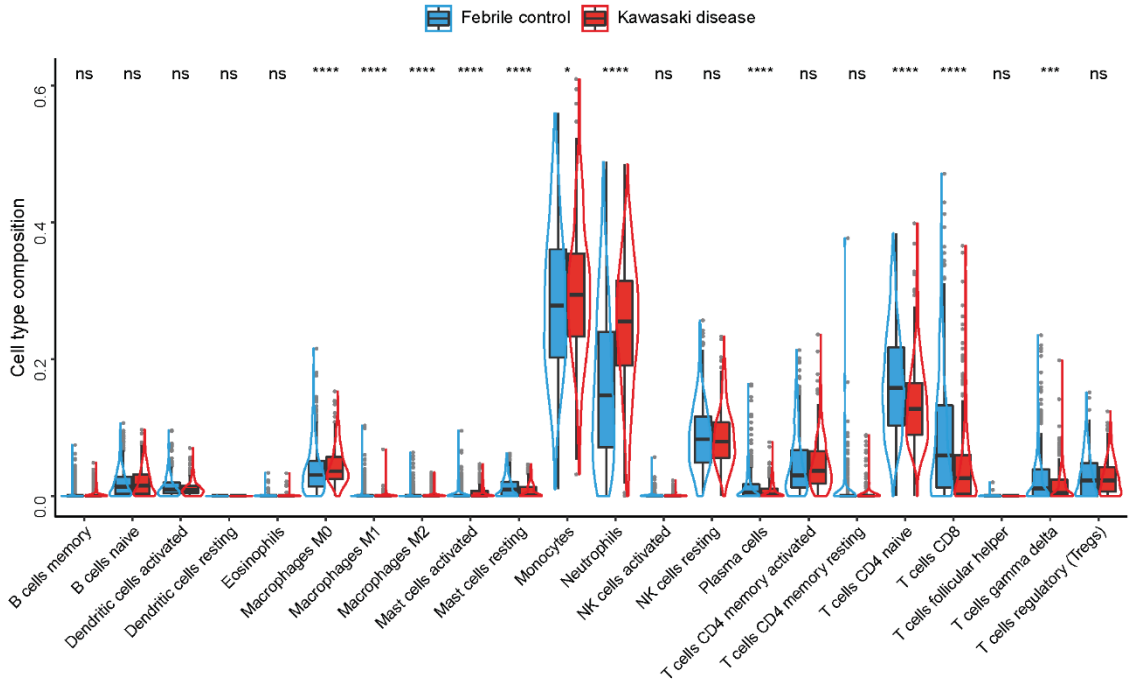


Figure 2.4 CIBERSORT-inferred compositions of various immune cells in the discovery cohort. *Wilcoxon-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, not significant. This figure was adapted from Du et al. [126].*

Table 2.3 Comparison of the percentages of the T cell subsets between the KD and the FC group

Group	$CD4^+$	$CD8^+$	$CD4^+/CD8^+$
KD group	0.187 [0.149-0.240]	0.038 [0.018-0.075]	4.95 [2.49-9.94]
FC group	0.222 [0.163-0.287]	0.081 [0.031-0.142]	2.82 [1.56-6.72]
p value	6.6×10^{-6}	6.2×10^{-10}	5.2×10^{-7}

Note: *Data were shown as median [IQR]. This table was adapted from Du et al. [126].*

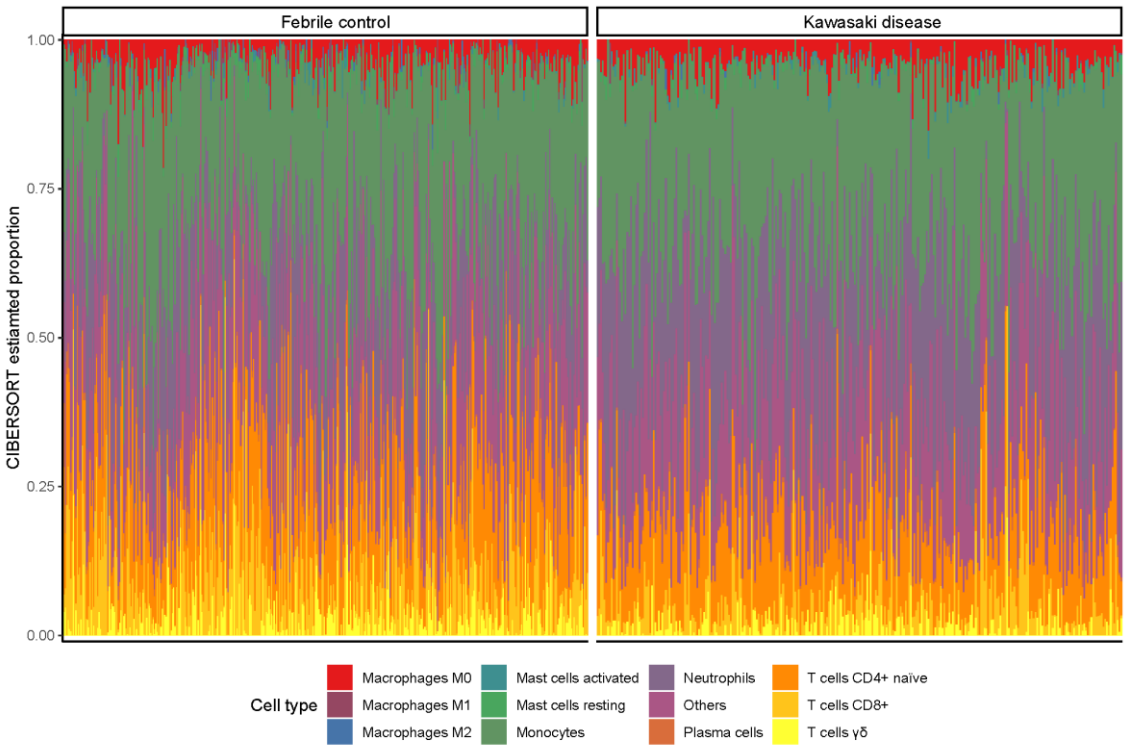


Figure 2.5 Estimated proportions of immune cell types in the discovery cohort presented as Stacked bar plots. *Only eleven significantly differently expressed cell types (KD vs. FC) are shown. This figure was adapted from Du et al. [126].*

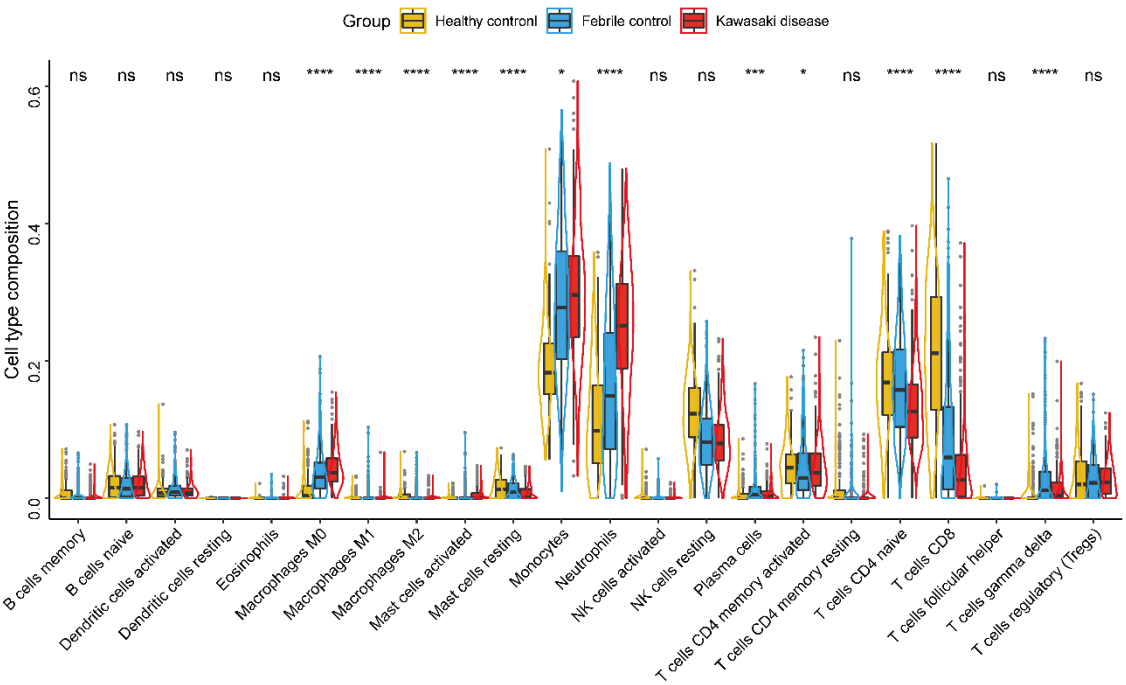


Figure 2.6 The immune cell fractions in the HCs (yellow), FCs (blue), and KDs (red) in the discovery cohort. Kruskal-Wallis test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, not significant. *This figure was adapted from Du et al. [126].*

2.3.4 Diagnostic immune score (DIS) derivation and validation

To begin, we constructed our diagnostic model, which we named “DIS” (diagnostic immune score), by extracting highly significantly different immune cells ($p < 0.01$). For model development, we used LASSO regression for binary outcomes and determined the penalty regularization λ based on 10-fold cross-validation and the 1-Se rule [139]. The candidate cell types were then narrowed down to eight (**Figure 2.7, Figure 2.8**).

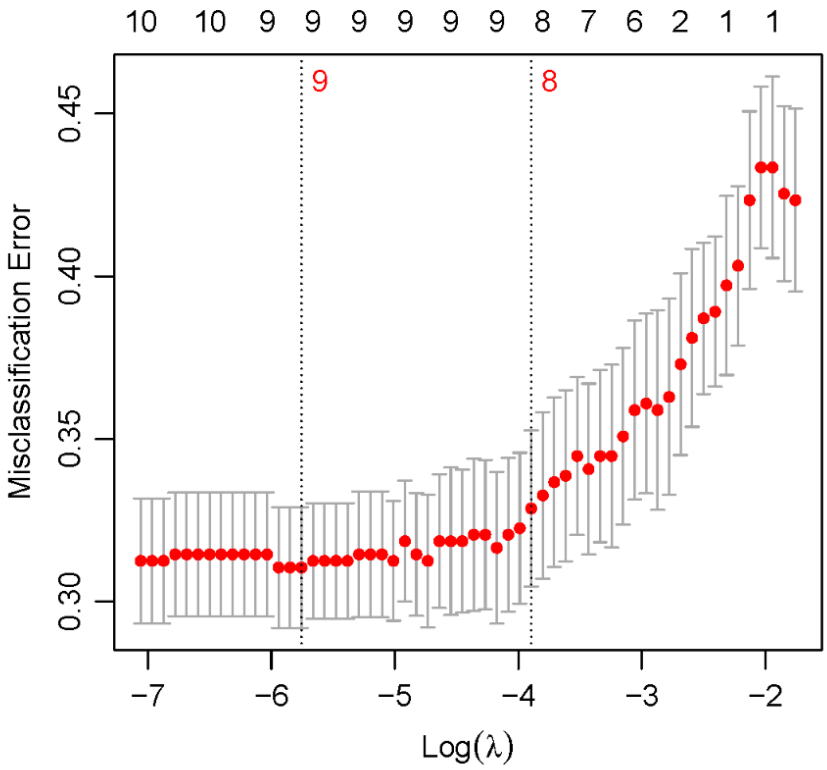


Figure 2.7 The construction of the DIS model. *Error bars illustrate the standard errors of the necessarily biased cross-validation estimates. This figure was adapted from Du et al. [126].*

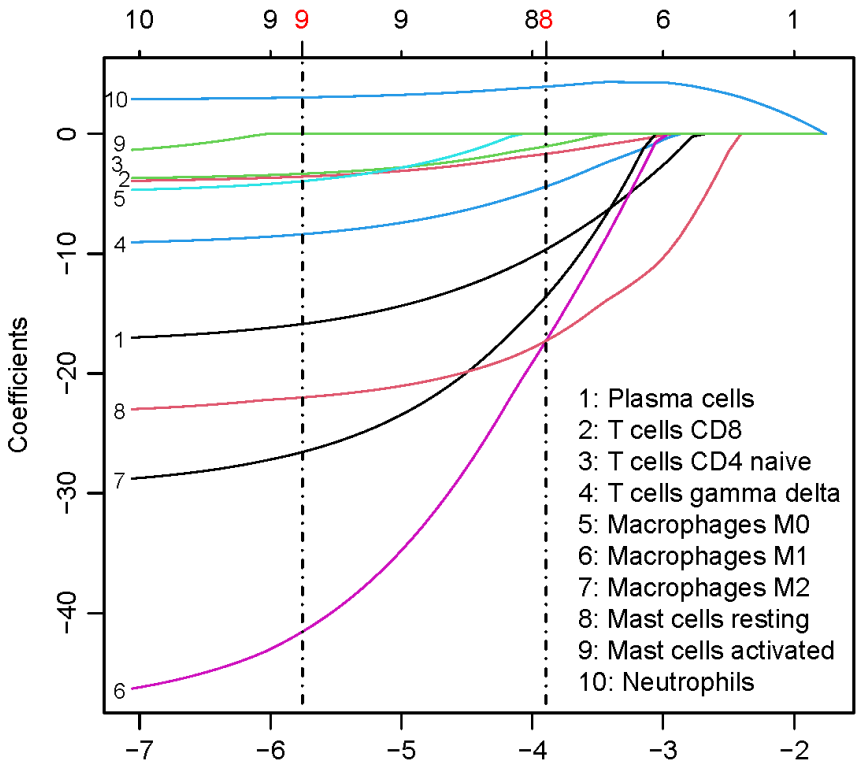


Figure 2.8 Fractions of candidate immune cell types with different paths of LASSO coefficient. *The left and the right vertical lines are the optimal values according to the minimum and the 1-Se rule, respectively. This figure was adapted from Du et al. [126].*

Table 2.4 Estimated coefficients of the diagnostic prediction model for KD diagnosis. *This figure was adapted from Du et al. [126].*

Immune cell marker	Coefficient
Intercept of the model	0.60
Neutrophils	3.23
M1 macrophages	-44.59
M2 macrophages	-27.99
Mast cells resting	-27.73
Plasma cells	-20.00
Gamma delta T cells	-10.43
CD4+ naïve T cells	-3.98
CD8+ T cells	-2.60

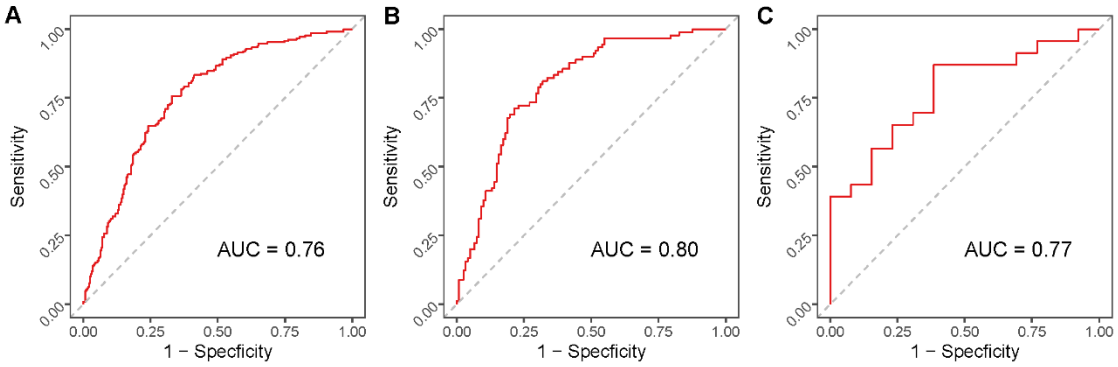


Figure 2.9 Evaluation of the performance of the DIS model in the training (A), held-out test (B), and validation (C) set*. This figure was adapted from Du et al. [126].*

Incorporating these eight variables, a diagnostic prediction model was constructed by using logistic regression. **Table 2.4** shows the estimated coefficients of the model. The overall performance of the model was evaluated using ROC curves. Held-out test results showed an AUC of 0.80 (**Figure 2.9B**). In order to evaluate the capability of the DIS in external settings, we examined its performance on GSE15297, an independent dataset. Again, ROC curve presented a similar result (AUC = 0.77; **Figure 2.9C**).

2.3.5 Evaluation of model performance

We compared the DIS distribution across the different datasets studied, in order to assess its capability to discriminate KD from infection-caused febrile conditions (**Figure 2.10**). In every dataset, KD children had higher DIS values than FC children according to the violin plots. Although the DIS distributions in the training and test set were highly comparable, they differed in the validation set. This could be attributed to the diverse microarray platforms underlying different cohorts, which may result in different inputs in the CIBERSORT program. Nevertheless, the DIS was still valid since in each of the datasets, the DIS values of KD cases were significantly higher, and moreover, the cut-off point effectively distinguished the two conditions. Additionally, our exploration of the relationship between the DIS and gender observed no significant differences between males and females (p = 0.57; **Figure 2.11**).

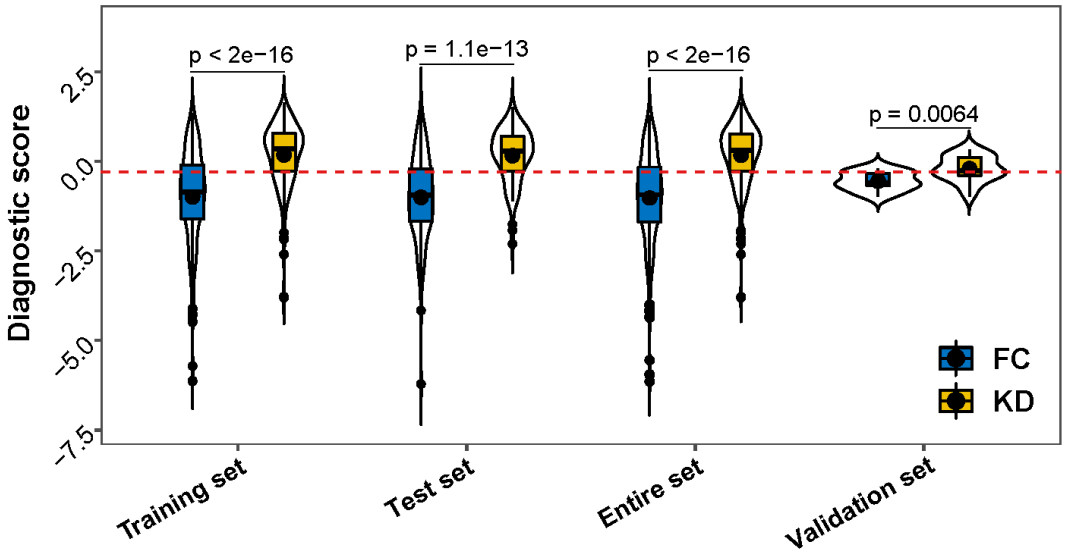


Figure 2.10 The distribution of DIS across the different datasets studied. *Box plots indicate the interquartile ranges and median values. Red dashed lines indicate the cut-off point based on the largest Youden's index. This figure was adapted from Du et al. [126].*

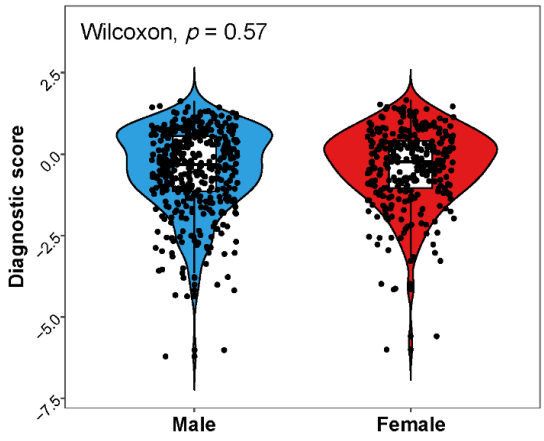


Figure 2.11 Distribution of DIS values in male (blue) and female children (red). *This figure was adapted from Du et al. [126].*

2.3.6 Nomogram construction and evaluation

In order to create a more practical tool for use in clinical settings, we established a nomogram that integrated these diagnostic factors based on immune cell type fractions (**Figure 2.12**). Additionally, we generated calibration and decision curves separately in

all datasets we studied. These curves indicated that the nomogram has great potential for clinical application due to the strong correspondence between observed and predicted outcomes (**Figure 2.13**).

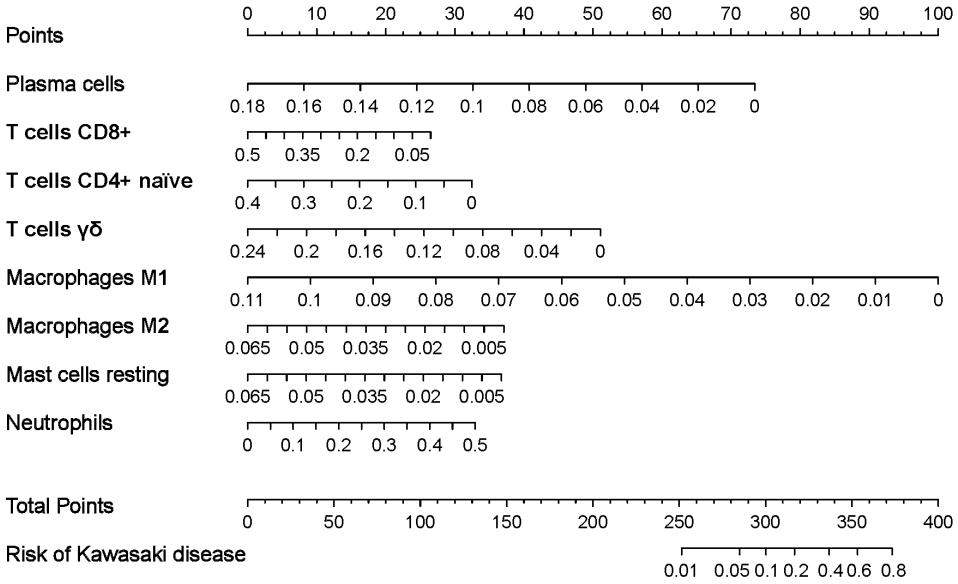


Figure 2.12 Nomogram for predicting the risk for KD using the training set. *This figure was adapted from Du et al. [126].*

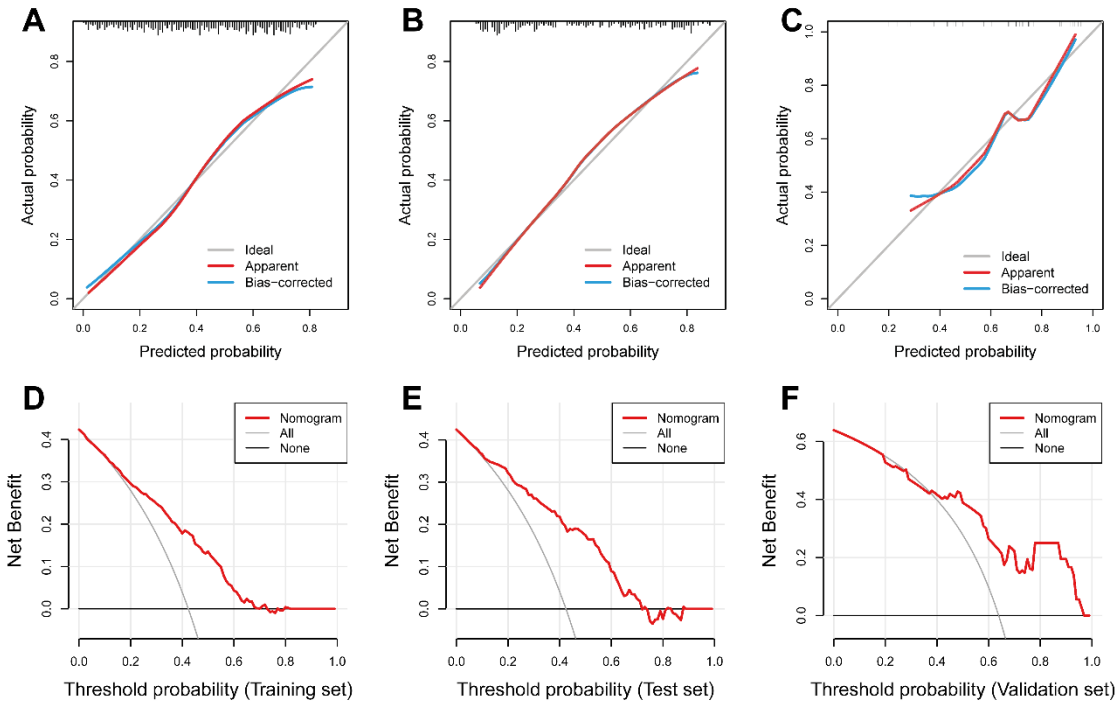


Figure 2.13 Evaluations of the established nomogram. *(**A-C**) Calibration curves for the concordance of nomogram between the predicted and the actual outcomes in training (**A**), test (**B**) and external validation set (**C**). The apparent and bias-corrected performance (Bootstrap repetition, B = 1,000) lines are shown in red and blue. (**D-F**) Decision curves in training (**D**), test (**E**) and external validation set (**F**).*

2.3.7 KD-related genes associated with the DIS

In addition, the Spearman's rank correlation analysis mirrored a significant positive correlation between the DIS value and the expression level of majority factors that were known to be involved in KD [140, 141] (**Figure 2.14**).

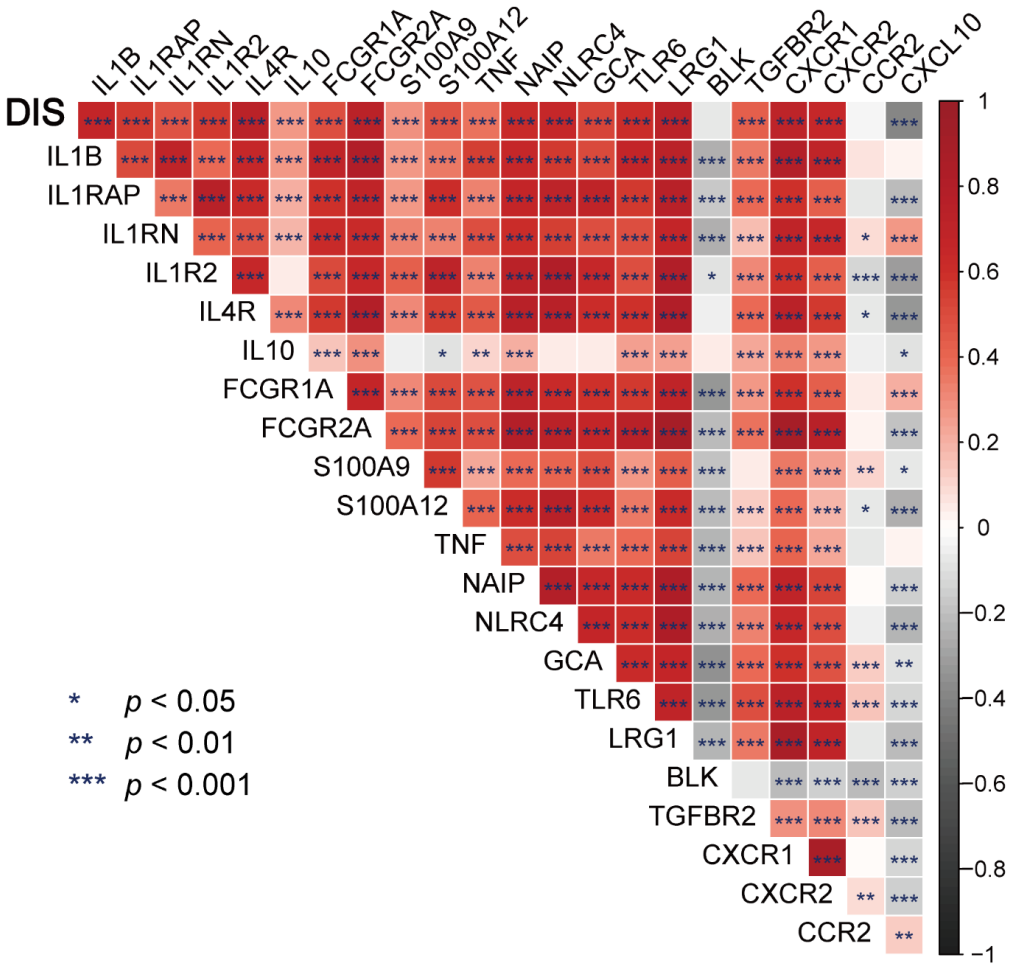


Figure 2.14 Heat map illustrating the correlation between the DIS and the expression level of KD-related genes previously reported. *This figure was adapted from Du et al. [126].*

2.3.8 Functional enrichments associated with the DIS

We also conducted the functional enrichment analysis using GSEA approach to evaluate biological plausibility of the model, by focusing on three gene sets of a series of well-defined biological processes and canonical pathways, which were the Hallmark, KEGG and BioCarta. All GSEA results for three gene sets mirrored that differentially expressed genes in the high-DIS group were predominantly enriched in inflammatory and immune-related pathways (e.g., IL-6/JAK/STAT3 signaling, TNFα signaling,

reactive oxygen species, complement and coagulation cascades, and β-arrestin/ Src pathways), compared to the low-DIS group (**Figure 2.15**).

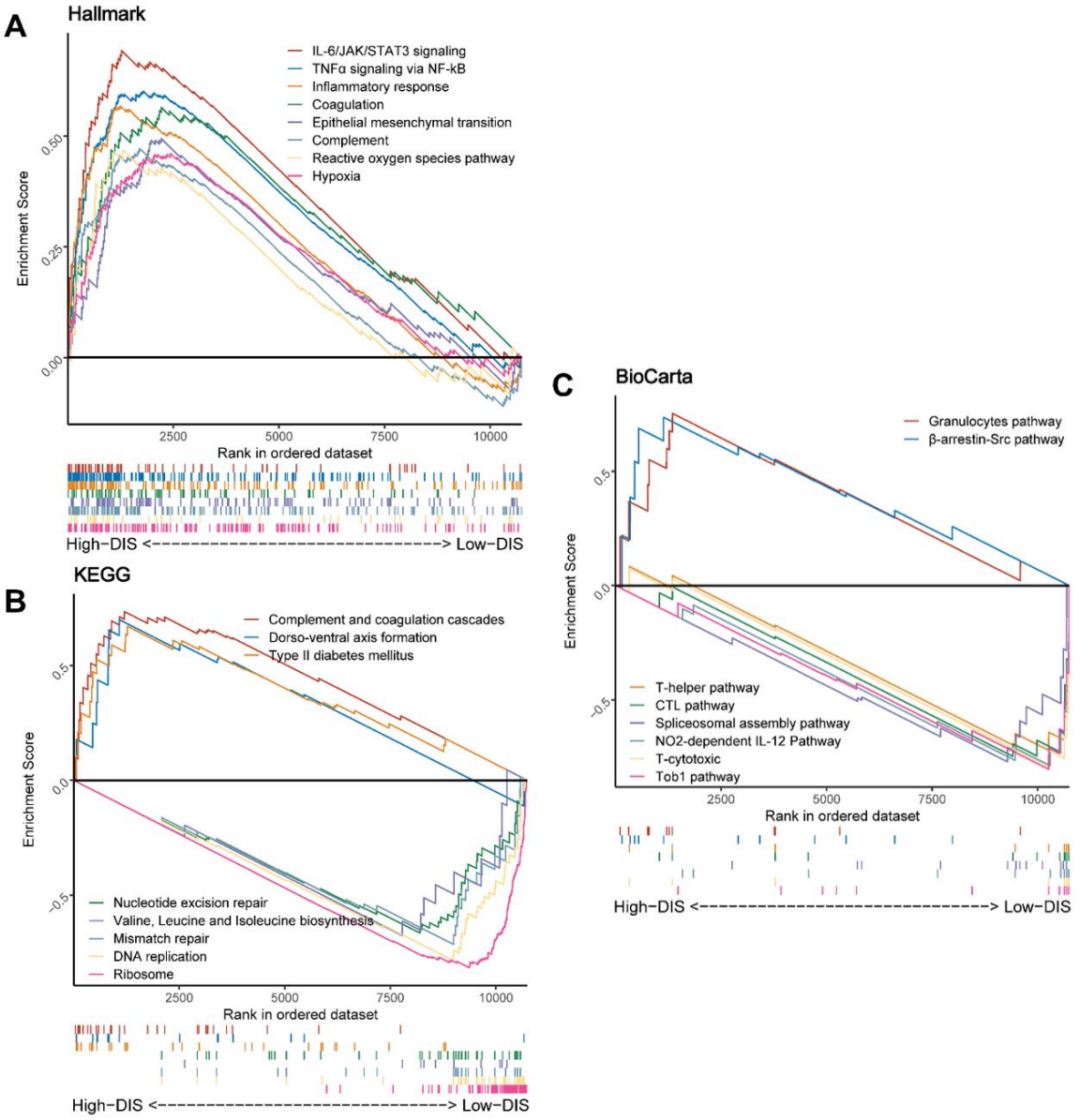


Figure 2.15 Biological processes and pathways in KD children with different DIS. *Gene sets were obtained from (**A**) Hallmark, (**B**) KEGG and (**C**) BioCarta on MSigDB. Top eight results are displayed in different colors based on their enrichment scores in each figure. This figure was adapted from Du et al. [126].*

2.4 Discussion

In this study, a DIS model for KD was developed based on a combination of machine-learning techniques and the inferred fractions of eight immune cells. Using the DIS, KD

and FC children could be distinguished well in the test set (AUC = 0.80). In addition, we scrutinized the functional contributions of the immune-related molecules (i.e., interleukins) and pathways (i.e., IL-6/JAK/STAT3 pathways) in KD. It is important to note that because all the analysed data were obtained from GEO database, its practicability and generalizability might be limited to a certain extent. Our results should, therefore, be ideally prospectively validated.

To date, numerous investigations have well-reported the influence of the immune microenvironment on KD pathogenesis, and there is a consensus among them that the innate immune system is crucial during KD acute phase. In particular, monocyte/macrophage and neutrophil levels were markedly increased in KD children [24, 53, 54]. According to our results, these high-throughput studies were also in agreement [24, 54]. Our DIS focused heavily on macrophages and neutrophils. Since experimental validation was lacking, we had to rely on computational analyses for biological procedures. These analyses, however, were speculative necessarily. The results of this study suggested that innate immune-related genes (i.e., certain chemokines and S100 proteins) and signaling pathways (i.e., TNFα signaling and IL-6/JAK/STAT3 pathways) could contribute to a high DIS. Moreover, researchers have noticed that KD and FC differed due to imbalances in T cell subsets in the body. Moreover, the $CD4^+/CD8^+$ T ratio in KD patients is significantly higher than that in FC patients, according to the studies of Furukawa et al. [138] and Ding et al. [53]. To a certain extent, our results supported this finding. It is nevertheless likely that B cell subsets are involved in KD, based on Xu et al.'s findings [142]. In contrast, in our diagnostic score, plasma cells negatively impacted the probability for KD and both groups exhibited similar proportions of naïve and memory B cells. One reason for this may be that Xu et al. had made comparison with healthy controls rather than febrile controls, and another reason may be that the presence of an infection could be indicated by certain B cell subsets.

Currently, the diagnosis of KD is based on clinical criteria along with additional non-specific laboratory testing in certain cases. [15, 129, 132]. Due to their mutual clinical manifestations, nevertheless, it is possible to confuse KD with other febrile illnesses including infections, leading to delays in treatment [128, 129, 143]. Conversely, underdiagnosed cases may be overtreated with intravenous immunoglobulin and/or other immunosuppressants [35]. The development of a precise biomarker that can accurately differentiate between KD and infectious febrile diseases would be a significant breakthrough, as it would decrease improper treatment and enable prompt intravenous immunoglobulin therapy for true cases. Our current diagnostic model was created as a potential biomarker by utilizing machine-learning

techniques on high-throughput data, instead of solely relying on clinical features. The ROC curves demonstrated that in either internal or external prediction settings, the DIS model exhibits an acceptable level of performance. Nonetheless, this DIS model's performance still needs to be compared with that of the current clinical KD-diagnostic algorithm, which necessitates further studies.

As an immune-related disease, immunological methods may potentially be used to identify KD. However, immunohistochemistry or flow cytometry may not be practical for frequent application [130]. High-throughput methods, on the other hand, may be suitable for comparing the characterization of the immune landscape in KD and febrile disease in children. Although some studies have identified certain transcriptomic biomarkers for KD diagnosis [35, 37, 132], they did not fully characterize the immune landscape. By integrating high-throughput data with enumeration of immune cell types, CIBERSORT could potentially offer a solution for this challenge.

There are some limitations to this study, which we acknowledge. Firstly, the DIS was derived from public datasets, where complete clinical information for each patient remained difficult to obtain. However, characteristics like ethnicity may have an impact on the variability of gene expression patterns. Secondly, batch effects may occur when we combined data from various microarray datasets. However, the results of PCA indicated that the majority of batch effects were corrected by the "ComBat" algorithm. Thirdly, mislabeling for some patients with KD or febrile disease by the provider of datasets we used can introduce bias into our analysis. It is necessary for prospective studies, therefore, to have a larger sample size and data collected in a more standardized process. Lastly, a bias may be introduced when comparing patients with different lengths of time from disease onset, and unfortunately, the datasets used in our study did not provide complete information on this aspect. Consequently, in the data collection of future studies, "days of illness" should be considered a significant factor.

2.5 Conclusion

In summary, we developed a potential predictive algorithm, namely DIS, for early diagnosis of KD by using CIBERSORT-derived fractions of immune cell types. However, future studies are required to further validate the DIS and ideally assimilate it into a novel clinical diagnostic model to improve the accuracy of KD diagnosis.

3. Integration of Basement Membrane and Immunogenic Cell Death Patterns Predicts the Prognosis in Uveal Melanoma

3.1 Introduction

Uveal melanoma (UVM) is an infrequent but formidable type of malignancy that originates from melanocytes within the ocular region [38]. Despite the recent treatment improvements, up to 50% of patients experience metastasis within a decade, rendering it a highly lethal disease [144]. In the recent decade, Immunotherapies have been approved for melanoma treatment, which offers hope for UVM patients. However, only under 5% of patients show a clinical response [145]. Therefore, there has been an urgent requirement for a novel algorithm for predicting the response to UVM immunotherapy.

The tumor microenvironment (TME) is comprised of diverse non-tumor cellular populations, such as immune cells and stromal cellular elements including extracellular matrices [81, 87, 146]. Researches over the past decade have proved the pivotal function of TME in the development and metastasis of UVM [56, 147]. Specifically, as two predominant components in UVM, the tumor-associated macrophages (TAMs) facilitate metastasis of tumor cells [148], while tumor-infiltrating lymphocytes (TILs) are strongly related to poor prognosis and immunotherapeutic response [149, 150]. In order to foster new approaches for prognostic and/or therapeutic prediction, a comprehensive understanding of the TME in UVM is imperative.

Basement membranes (BMs) are specialized extracellular matrices that surround tissues and cells, and alterations in their composition and structure can promote tumor invasion and metastasis [95, 151, 152]. A recent investigation has revealed BM stiffness is a crucial factor in the metastasis formation, and the expression level of netrin-4, a type of BM protein, was significantly associated with melanoma outcomes [95]. As a type of programmed cell death, the immunogenic cell death (ICD) generates long-term immunological memory, making it a potential target for immunotherapy against UVM [99, 153]. There is accumulating literature suggest that in the context of ICD, a variety of damage-associated molecular patterns (DAMPs) are released, which stimulate the maturation and recruitment of antigen-presenting cells to initiate adaptive

immunity, and have potential predictive value for immunotherapy efficacy [153, 154]. Although BM and ICD are crucial in the progression and metastasis of UVM, there remains a paucity of studies on their potential for developing predictive models.

The objective of this study was to identify the roles of the BM and ICD patterns in UVM progression, then exploit them to develop a predictive algorithm for UVM prognosis and immunotherapeutic response strategies.

3.2 Material and methods

3.2.1 Data sources and preprocessing

Table 3.1 summarizes the detailed information of all datasets involved in this study. For the training set, we used the TCGA-UVM cohort, which contains fragments-per-kilobase-per-million (FPKM)-normalized RNA-seq data with clinical information and were downloaded from the UCSC Xena browser (https://xenabrowser.net/datapages/). Next, these FPKM values were transformed into transcripts-per-kilobase-million (TPM) values, whose distribution is more similar to microarrays profile and more comparable between samples [155].

To obtain data from the GEO database, we used the following search strategy: (("Expression profiling by array" [DataSet Type] OR "expression profiling by high throughput sequencing" [DataSet Type]) AND ("Homo sapiens" [Organism]) AND ((Uveal Neoplasms [MeSH Terms]) OR ((((((((((Uveal neoplasm [Title]) OR Uveal cancer [Title]) OR Uveal tumor[Title]) OR Uveal melanoma [Title]) OR intraocular neoplasm [Title]) OR intraocular cancer [Title]) OR intraocular tumor [Title]) OR intraocular melanoma [Title]) OR eye neoplasm [Title]) OR eye cancer [Title]) OR eye tumor [Title]) OR eye melanoma [Title])). Initially, there were 29 items recognized (until 2022). Next, we selected only those subject to the following 1) prognostic information provided; 2) sample collected from tissues in UVM patient; 3) not a SuperSeries or repeated series. After filtering, we obtained three series with survival data: GSE22138, GSE44295, and GSE39717, which would subsequently be used as the validation and test cohort. Microarray gene expression data on Illumina® platform were processed with the "lumi" R package [134], while data on Affymetrix® were processed with the "Affy" and “limma” R package [120]. In detecting a gene by multiple probes, only the one with the highest level was selected.

Table 3.1 Information of UVM datasets

Cohort	Datasets	Platform	No. of patients	Outcome
Training	TCGA-UVM	Illumina® HiSeq	80	OS
Validation	GSE22138	GPL570 Affymetrix® Human Genome U133 Plus 2.0 Array	63	OS
Test	GSE44295	GPL6883 Illumina® HumanRef-8 v3.0 expression beadchip	59*	OS
	GSE39717	GPL6098 Illumina® humanRef-8 v1.0 expression beadchip	30	MFS

Note: *, Two of these patients lacked survival data and were excluded. OS, overall survival; MFS, metastasis-free survival.*

3.2.2 Molecular subtype identification

The unsupervised consensus clustering method for analysis of a certain gene set was employed to discover discrete subtypes based on Ward's linkage and Euclidean distance [156]. See **Section 1.6.2** for more details.

3.2.3 DEG analysis and functional enrichment analysis

DEGs between subtypes were determined with significance criteria as false discovery rate (FDR) < 0.05 and | log2 (Fold change) | > 1.0 using the "limma" R package [120]. See **Section 1.6.3** for more details. Functional enrichment analyses were implemented as described in **Section 2.2.3**.

3.2.4 Immune microenvironment characterization

CIBERSORT was implemented as described in **Section 2.2.2**. The computation of the Estimate-Stromal-Immune scores for each UVM sample was performed using the "ESTIMATE" R package [8]. See **Section 1.6.5** for more details.

3.2.5 Survival prediction model construction and validation

In order to identify a biomarker that better integrates the basement membrane (BM) and immunogenic cell death (ICD) patterns, we have attempted three options of algorithms based on the identified subtypes of BM and ICD in the training set. These algorithms were:

- Model I, based on the Separate-LASSO algorithm:

$$\boldsymbol{BMscore} = \sum_{i=1}^{10} \boldsymbol{Coef}(\boldsymbol{i}) \times \boldsymbol{expr}(\boldsymbol{i});\ \ \boldsymbol{ICDscore} = \sum_{i=1}^{8} \boldsymbol{Coef}(\boldsymbol{i}) \times \boldsymbol{expr}(\boldsymbol{i})$$
$$\boldsymbol{BMICD\ classifier}: \text{High } \boldsymbol{BM} - \text{High } \boldsymbol{ICD},\ \text{Low } \boldsymbol{BM} - \text{Low } \boldsymbol{ICD},\ \text{Mixed}$$

- Model II, based on the IPF-LASSO algorithm:

$$\boldsymbol{BMscore} = \sum_{i=1}^{10} \boldsymbol{Coef}(\boldsymbol{i}) \times \boldsymbol{expr}(\boldsymbol{i});\ \ \boldsymbol{ICDscore} = \sum_{i=1}^{7} \boldsymbol{Coef}(\boldsymbol{i}) \times \boldsymbol{expr}(\boldsymbol{i})$$
$$\boldsymbol{BMICDscore} = \boldsymbol{BMscore} + \boldsymbol{ICDscore}$$

- Model III, based on the PCA algorithm:

$$\boldsymbol{BMscore} = (\sum \boldsymbol{PC1}_{\boldsymbol{BM.P}} + \sum \boldsymbol{PC2}_{\boldsymbol{BM.P}}) - (\sum \boldsymbol{PC1}_{\boldsymbol{BM.N}} + \sum \boldsymbol{PC2}_{\boldsymbol{BM.N}})$$
$$\boldsymbol{ICDscore} = (\sum \boldsymbol{PC1}_{\boldsymbol{ICD.P}} + \sum \boldsymbol{PC2}_{\boldsymbol{ICD.P}}) - (\sum \boldsymbol{PC1}_{\boldsymbol{ICD.N}} + \sum \boldsymbol{PC2}_{\boldsymbol{ICD.N}})$$
$$\boldsymbol{BMICD\ classifier}: \text{High } \boldsymbol{BM} - \text{High } \boldsymbol{ICD},\ \boldsymbol{Low BM} - \text{Low } \boldsymbol{ICD},\ \text{Mixed}$$

  where P is the expression of gene whose Cox coefficient is positive, and N is the expression of gene whose Cox coefficient is negative.

Here, “Separate-LASSO” refers to the following procedure: 1) apply standard LASSO separately to each of the two gene sets; 2) divide the patients into respective risk groups according to the results of two gene-set-based LASSO models separately; 3) combine the primarily grouping results to develop the classifier. See **Section 1.6.1** for more details of modeling algorithms and their metrics of performance (i.e., time-dependent ROC curves).

Subsequently, an independent validation cohort was used to compare the capability of the three models and identity the most optimal algorithm as our risk scoring model. Furthermore, we used two independent GEO datasets as the test cohorts to assess the generalization ability of the selected model. Independent prognostic factors were determined using Cox regression models and visualized with the “forestplot” R package. Moreover, we incorporated clinical features to generated a nomogram and its calibration curves of 3- and 5-year survival, using the “rms” and “regplot” R package. Decision curve analysis (DCA) was conducted to determine the suitability of our established nomogram for clinical use [157]. The detailed study design is shown in **Figure 3.1**.

3.2.6 Other statistical analysis

Group comparisons were performed using the Wilcoxon's test and Fisher's exact test for continuous variables and categorical variables, respectively. Kruskal-Wallis tests were used multiple group comparisons. Spearman's correlation coefficient was used to analyze correlations between Estimate-Stromal-Immune scores and the BMICDscore. The Boruta dimensionality reduction was performed using "Boruta" R package [158]. All statistical tests were two-sided. Statistical significances were all set at $p < 0.05$.

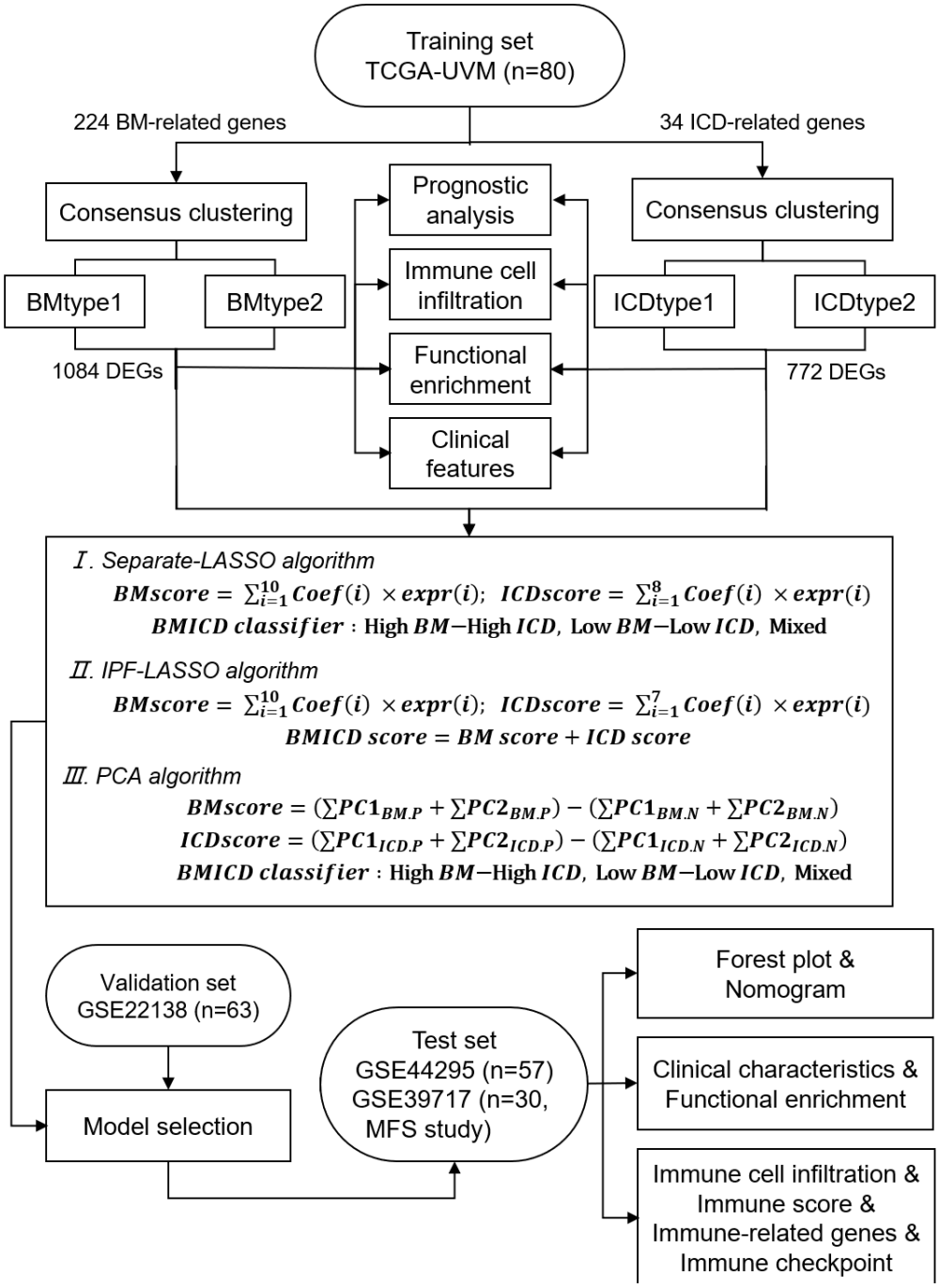


Figure 3.1 Overview of the study design. *Abbreviations: BM, basement membrane; ICD, immunogenic cell death; DEGs, differentially expressed genes; MFS, metastasis-free survival.*

3.3 Results

3.3.1 Characteristics of participants

According to the screening strategy outlined in **Section 3.2.1**, we identified a TCGA dataset with 80 samples containing survival data, as well as three GEO datasets (GSE22138, GSE44295, and GSE39717) with 63, 57, and 30 patients with survival data, respectively. The TCGA-UVM cohort, which served as the training set, consisted of 57 alive patients and 23 deceased patients, while the GSE22138 (validation set) and GSE44295 (test set) had alive: deceased ratios of 28:35 and 33:24, respectively. For assessing the predictive power to other survival outcomes, we selected the GSE39717 dataset with metastasis as the endpoint outcome, which was comprised of eight metastasis cases and 22 non-metastasis cases. In **Table 3.2**, we compared the baseline data and clinical characteristics of the study cohorts.

Table 3.2 Clinical and demographic characteristics in all UVM datasets.

	Training	**Validation**	**Test**	
	TCGA-UVM	**GSE22138**	**GSE44295**	**GSE39717**
Number of patients	80	63	57	30
Age, months	61.5 [51.0, 74.2]	62.1 [52.1, 69.1]	NA	62.5 [54.2, 67.8]
Sex				
Male	45 (56.2)	39 (61.9)	32 (56.1)	23 (76.7)
Female	35 (43.8)	24 (38.1)	25 (43.9)	7 (23.3)
Survival				
Alive	57 (71.2)	28 (44.4)	33 (57.9)	NA
Deceased	23 (28.7)	35 (55.6)	24 (42.1)	NA
Metastasis				
No	73 (96.1)	NA	NA	22 (73.3)
Yes	3 (3.9)	NA	NA	8 (26.7)
Survival time	26.1 [14.9, 39.3]	32.1 [14.9, 65.7]	39.5 [26.4, 52.8]	22.1 [8.4, 37.3]
Clinical stage				
II	36 (45.0)	—	—	—
III	40 (50.0)	—	—	—
IV	4 (5.0)	—	—	—
T stage				
T2	4 (5.0)	—	—	—

T3	36 (45.0)	—	—	—
T4	39 (48.8)	—	—	—
NA	1 (1.3)	—	—	—
M stage				
M0	76 (95.0)	—	—	—
M1	4 (5.0)	—	—	—

Note: *There are two types of data presented: median [IQR] or n (%). NA*, *not applicable.*

3.3.2 Identification of BM and ICD subtypes in UVM patients

We used the unsupervised-consensus-clustering approach introduced in **Section 1.6.2** and **Section 3.3.2** to identify the BM and ICD subtypes of UVM. Jayadev et al. [97] and Garg et al. [159] have summarized 224 major BM-related genes and 34 major ICD-related genes, respectively, through large-scale studies, which we used as gene sets for UVM subtyping. After 1,000 iterations, it was clear from the consensus clustering results that the highest variability interpretation was observed in both BM (**Figure 3.2**) and ICD (**Figure 3.3**) subtypes when patients were divided into two clusters. We named the two subtypes based on each gene set as BMtype1/BMtype2 and ICDtype1/ICDtype2, respectively. Survival analysis using K-M curves revealed that UVM patients' overall survival (OS) probability was statistically significantly different between subtypes (both in BMtypes and ICDtypes; log-rank test, both *p* values are less than 0.0001; **Figure 3.4**). BMtype1 and ICDtype1 showed the prominent survival advantage compared to BMtype2 and ICDtype2, respectively. Besides, we employed heat maps utilizing hierarchical clustering to further depict the intrinsic distinctions between BM-related or ICD-related genes and clinical features among these subtypes. Results indicated that there were a higher proportion of death and increased severities of various tumor stages in BMtype2 patients, where most BM-related genes were significantly up-regulated (**Figure 3.5**), and we observed a similar trend in ICDtype2, where most ICD-related genes were significantly up-regulated (**Figure 3.6**).

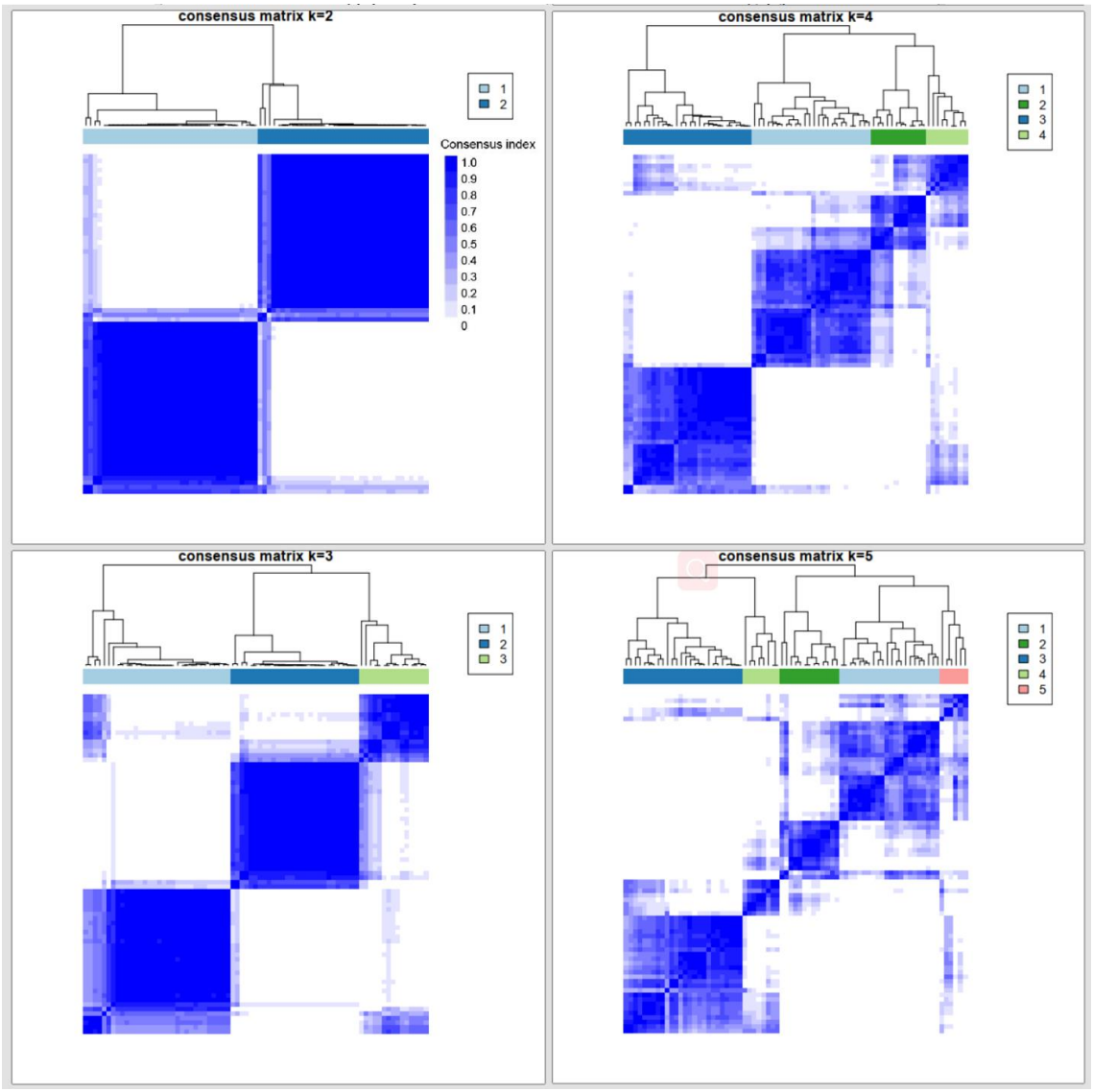


Figure 3.2 Tumor BM-based subtyping of patients from TCGA-UVM cohort. *Consensus matrix heat maps for each k (the number of selected clusters).*

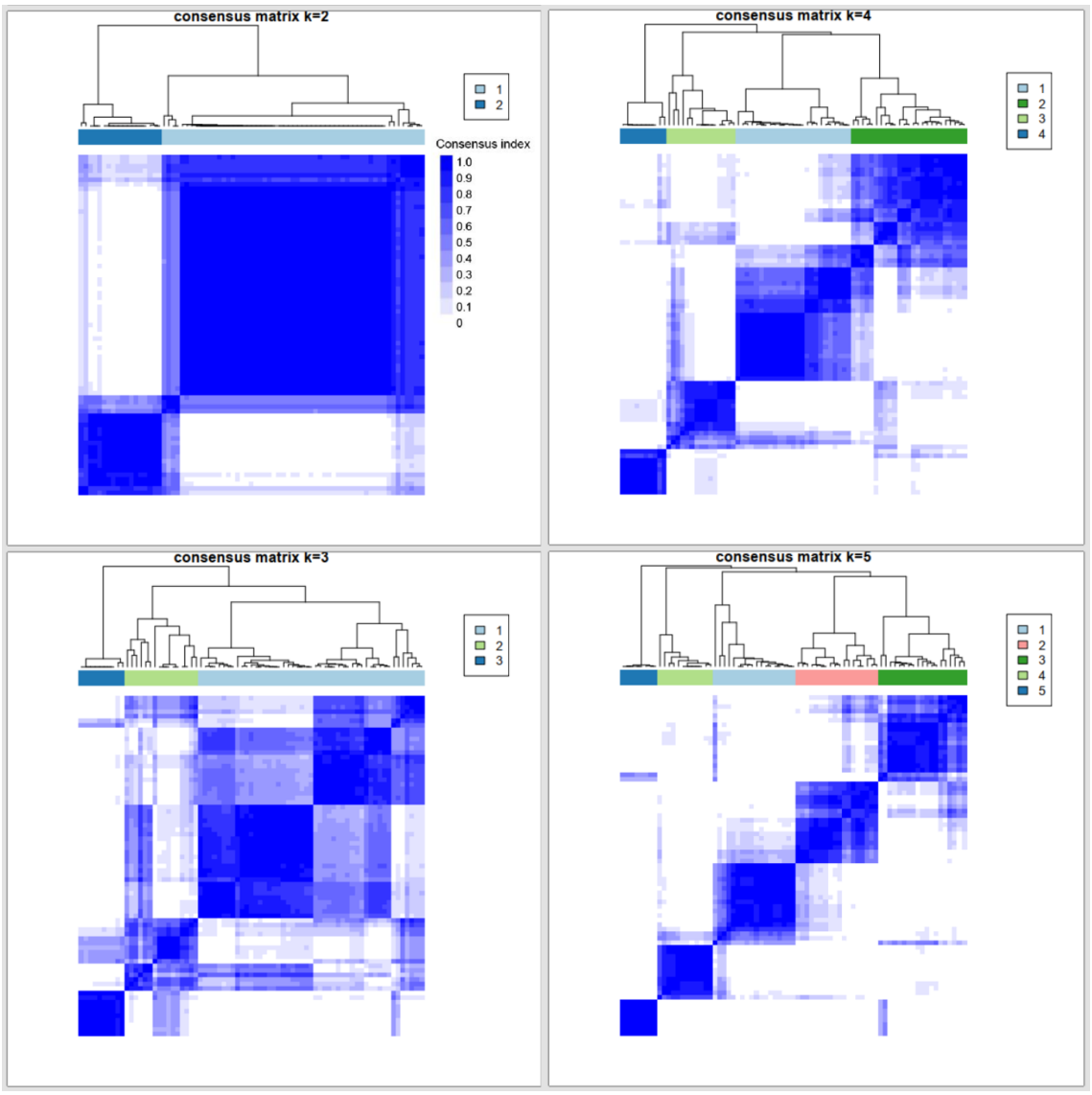


Figure 3.3 Tumor ICD-based subtyping of patients from TCGA-UVM cohort. *Consensus matrix heat maps for each k (the number of selected clusters).*

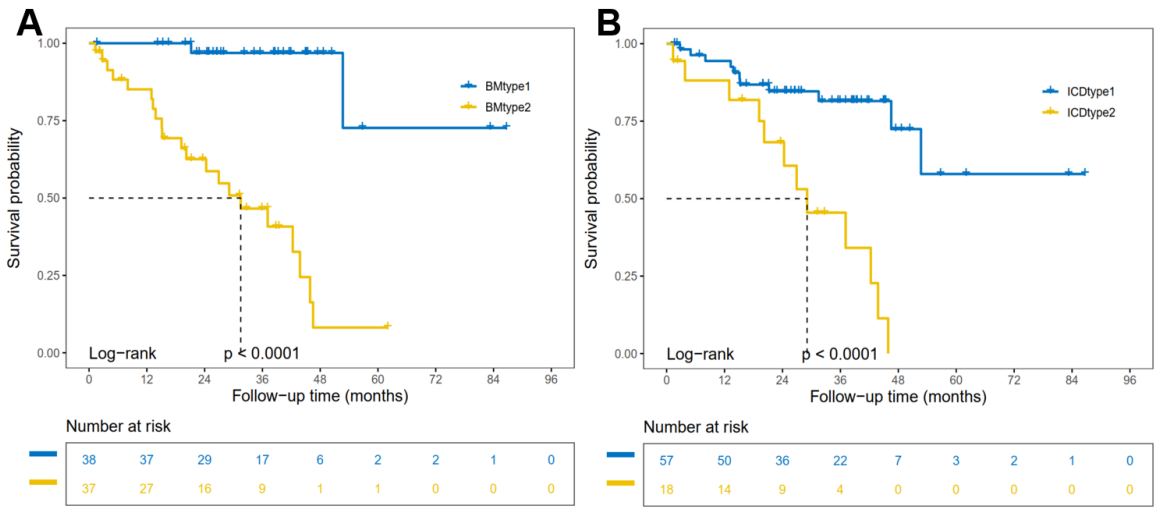


Figure 3.4 Kaplan–Meier (K-M) curves for the OS of patients in different BMtypes (A) and ICDtypes (B) in the TCGA-UVM cohort.

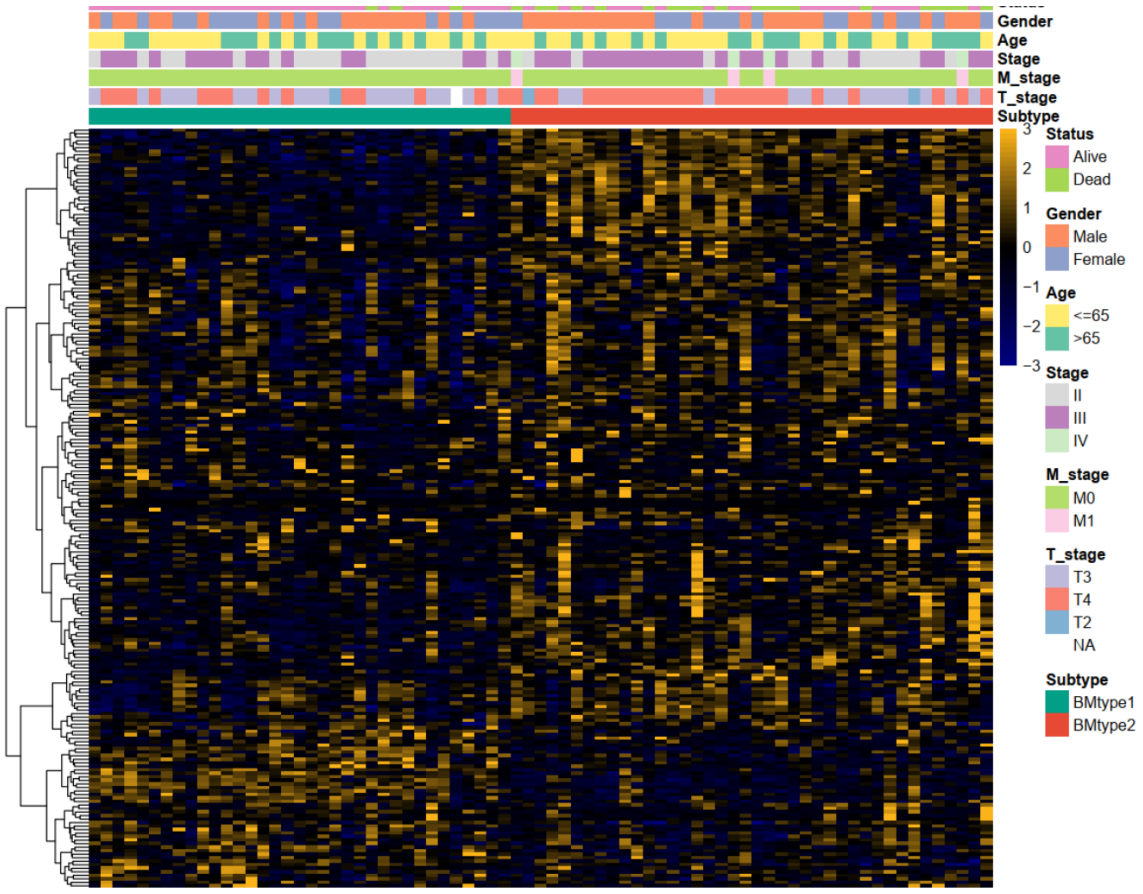


Figure 3.5 Expression of BM—related genes in TCGA-UVM cohort presented as a heat map. *Unsupervised clustering was performed with Hierarchical clustering. Columns represent UVM patients, while rows represent BM-related genes.*

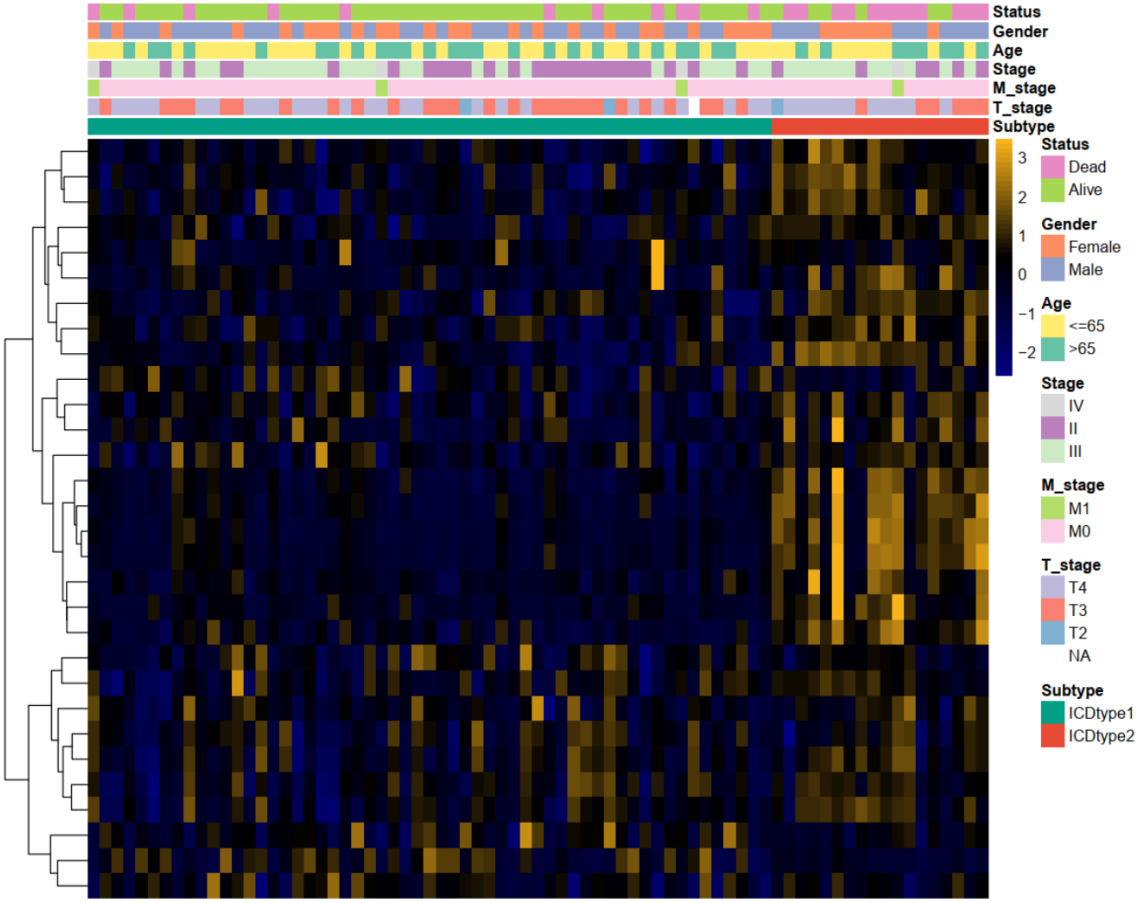


Figure 3.6 Expression of ICD—related genes in TCGA-UVM cohort presented as a heat map. *Unsupervised clustering was performed with Hierarchical clustering. Columns represent UVM patients, while rows represent ICD-related genes.*

3.3.3 Immune landscape in different subtypes

Next, by using the CIBERSORT algorithm, we characterized the tumor microenvironment (TME) in different UVM subtypes. BMtype2 was characterized by elevated infiltration of M1 macrophages, $CD8^+$ and follicular helper T cells but depressed infiltration of monocytes, naïve B cells, and $CD4^+$ memory resting T cells, compared to BMtype1 (**Figure 3.7A**). ICDtype2 exhibited high infiltration levels of M1 macrophages, and activated $CD4^+$ memory, $CD8^+$, follicular helper, and γδ T cells compared to ICDtype1, and its cell composition at low infiltration levels was the same as BMtype2 (**Figure 3.7B**).

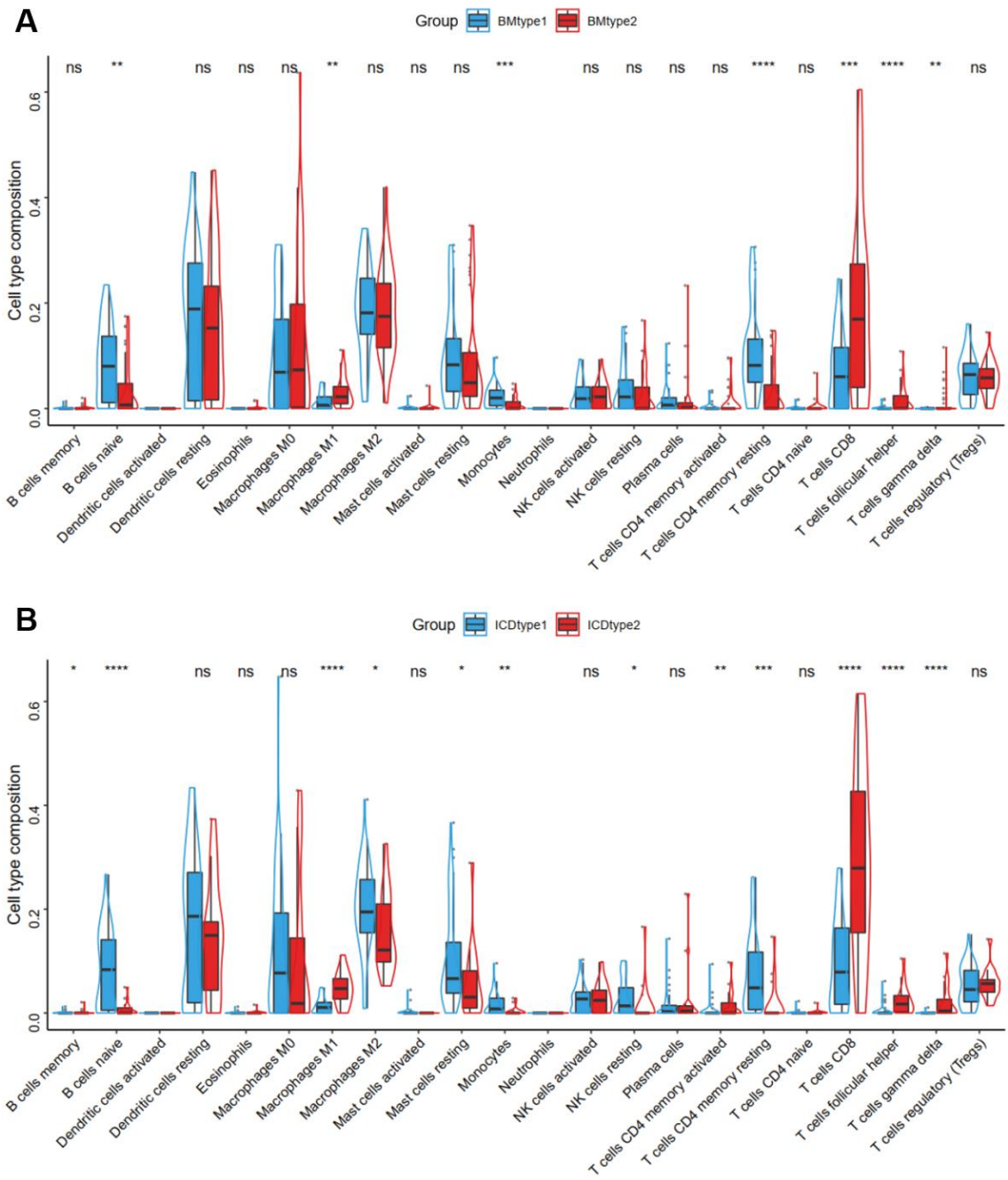


Figure 3.7 Compositions of TME immune cell types expressed in BMtypes (A) and ICDtypes (B). **, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001; ns, not significant.*

3.3.4 Biological processes and clinical features associated with different subtypes

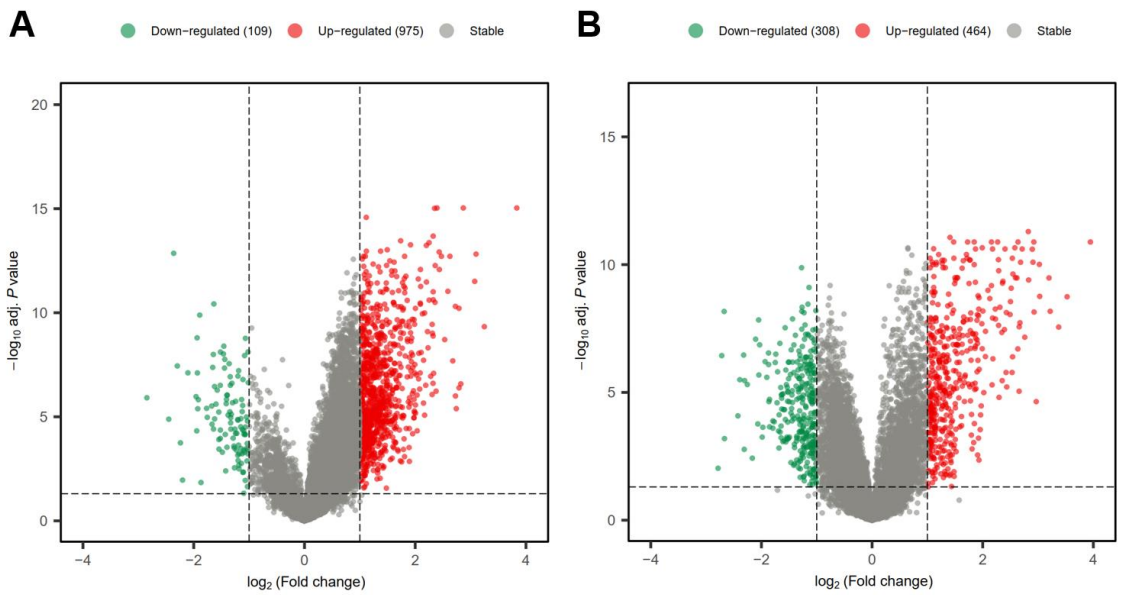


Figure 3.8 Volcano plots for DEGs between BMtype1 and BMtype2 (A), and between ICDtype1 and ICDtype2 (B) in TCGA-UVM cohort.

We identified the DEGs between BMtype1 and BMtype2, and between ICDtype1 and ICDtype2, respectively, to scrutinize the potential molecular mechanism associated with them, and provide the features for the construction of predictive model. According to **Figure 3.8**, a total of 1,084 DEGs between BMtypes and 772 DEGs between ICDtypes were identified, which were subsequently exploited for model construction.

To better understand the underlying biological processes involved in BMtypes and ICDtypes, we analyzed the enriched “Hallmark” and “KEGG” pathways using GSEA approach. In BMtype2, both “Hallmark” and “KEGG” GSEA plots revealed principal up-regulation of pathways of immune response, cytokine and chemokine signaling, cell cycle and proliferation, and cell adhesion, which are crucial for tumor progression, invasion and metastasis (**Figure 3.9A,B**). Again, a similar pathway enriched phenomenon was observed in ICDtype2 (**Figure 3.9C,D**).

The next investigation was to compare the differences of clinical characteristics between different subtypes, as shown in those stacked bar charts in **Figure 3.10**. In both BMtypes and ICDtypes, we observed the significant differences of survival rates between different subtypes, or, more precisely, alive individuals accounting for a major

percentage of either BMtype2 or ICDtype2. The compositional difference of clinical stages was seen in BMtypes, whereas no statistical difference was seen in ICDtypes. Moreover, there were no statistical difference presented in other clinical traits, even though the proportion of T stages exhibited a few trends. Future work with a larger sample size would be required to confirm these trends.

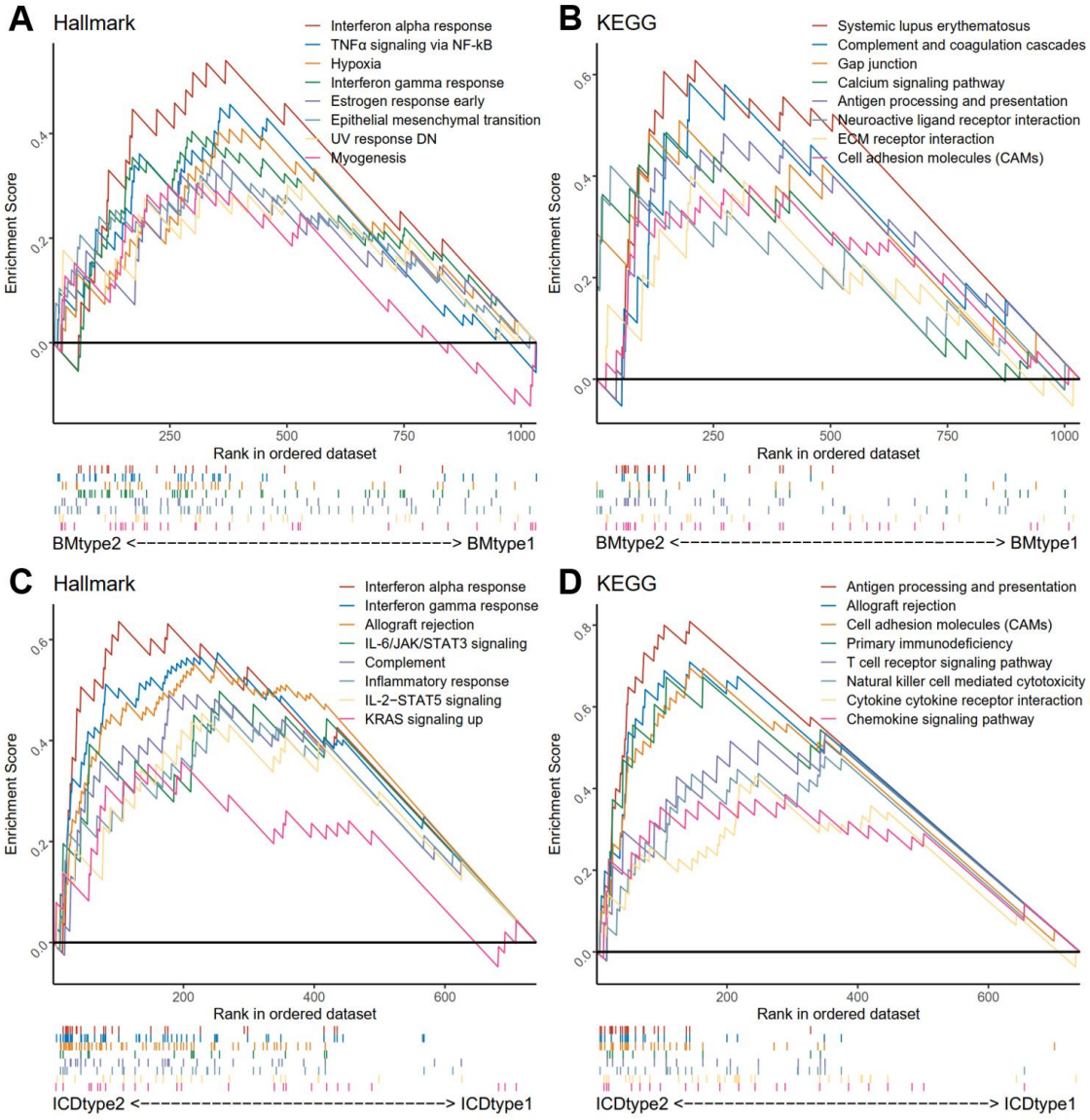


Figure 3.9 GSEA showing signaling pathways and biological processes in different BMtypes (A,B) and ICDtypes (C,D). *Gene sets were obtained from "Hallmark" (**A,C**) and "KEGG" (**B,D**) on MSigDB. The top eight results are displayed in different colors based on their enrichment scores in each figure.*

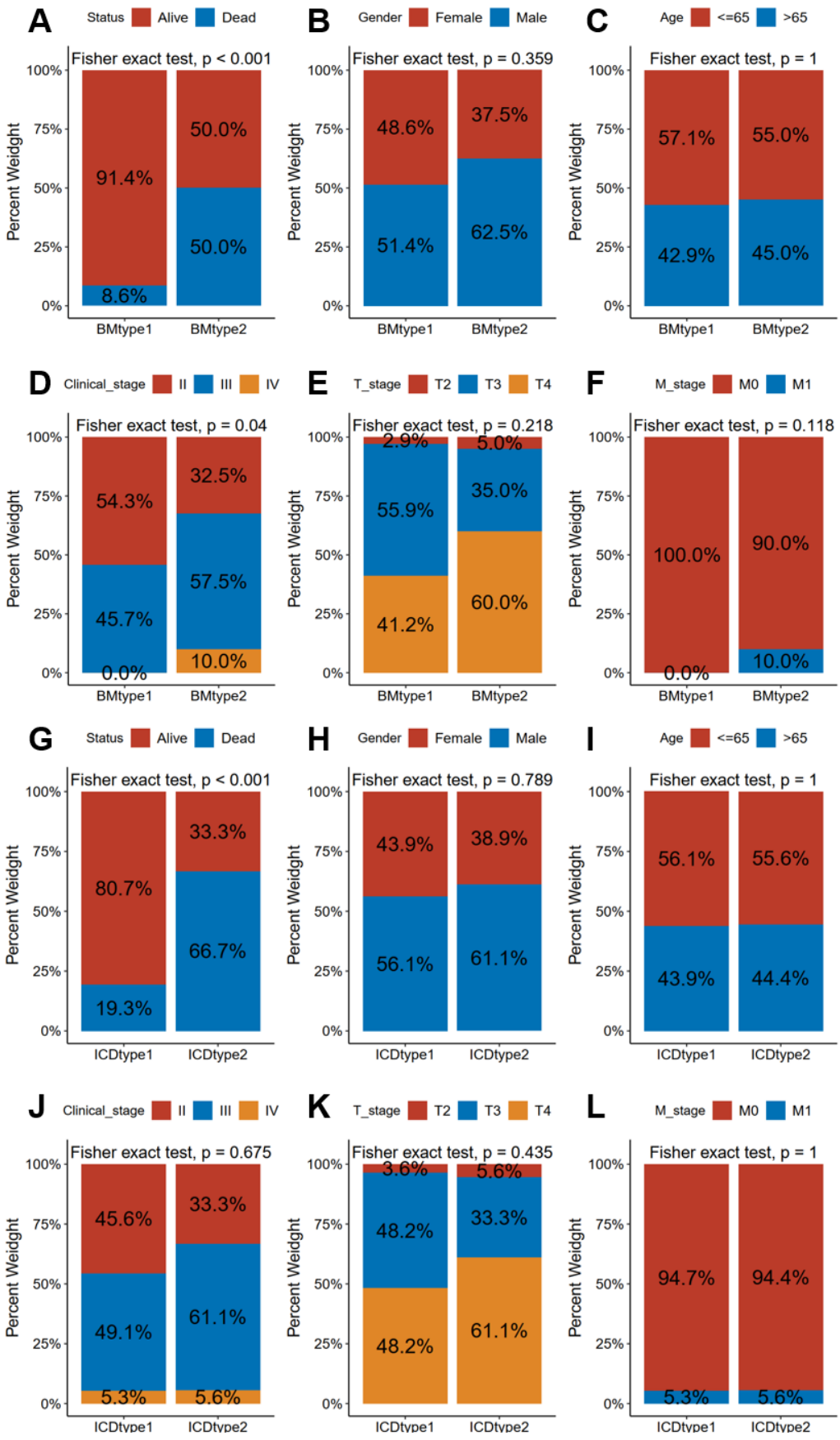


Figure 3.10 Proportions of patients with different survival statuses (A,G), genders (B,H), ages (C,I), clinical stages (D,J), T stages (E,K), and M stages (F,L) between different BMtypes (A-F) and ICDtypes (G-L) in the TCGA-UVM cohort.

3.3.5 Risk score calculation and classifier construction using Separate-LASSO algorithm

We built and compared three predictive models with different algorithms to find the optimal one. To develop the first model, we used Separate-LASSO algorithm for survival outcomes, as described in **Section 3.2.5**. In the training set, based on the ideal value of penalty regularization λ with 10-fold cross-validation [139], the BMscore including ten genes and the ICDscore including eight genes were ultimately found (**Figure 3.11A-D**). Detailed parameters with coefficients were listed in **Table 3.3**. K-M curves exhibited statistically significant differences in OS probability between the high- and low-risk group in both BMscore and ICDscore (both two log-rank p<0.001; **Figure 3.11E,F**). Both high BMscore and high ICDscore were associated with the unfavorable survival outcome. The time-dependent ROC curves were used to estimate the performance of two predictive scores, which indicated the good power of both two models for predicting the OS with different time points for UVM patients (BMscore: 1-, 3-, and 5-year OS AUC = 0.922, 0.970 and 0.948, respectively; ICDscore: 1-, 3-, and 5-year OS AUC = 0.896, 0.969, and 0.948, respectively; **Figure 3.12**).

Considering the combination of two risk scores simultaneously could be preferable to characterize the tumor progression, we then constructed a BMICD classifier according to the risk groups based on two gene sets (BMscore: “High-BMscore” group and “Low-BMscore” group; ICDscore: “High-ICDscore” group and “Low-ICDscore” group), consisting of three groups: “High BM**–**High ICD”, “Mixed”, and “Low-BM**–**Low ICD”. According to the OS curves, “High BM**–**High ICD” patients had the worst prognostic outcomes; between the other two groups, however, no significant difference was found (**Figure 3.13A**). We believe it might be due to the limited sample size of the “Mixed” group (only 2 patients). Additionally, AUCs of 1-, 3-, and 5-year OS prediction were 0.796, 0.908, and 0.762, respectively (**Figure 3.13B**). Notably, the K-M curves for the “Low-BM**–**Low ICD” group and “Mixed” group were overlapped until the time point of 36 months, which indicated that there was no statistically significant difference between the two groups. Therefore, there was only one cutoff point in their corresponding ROC curve (of either 1- or 3-year OS). We speculated the reason may be that the sample size of the “Mixed” group was insufficient, with only 2 cases.

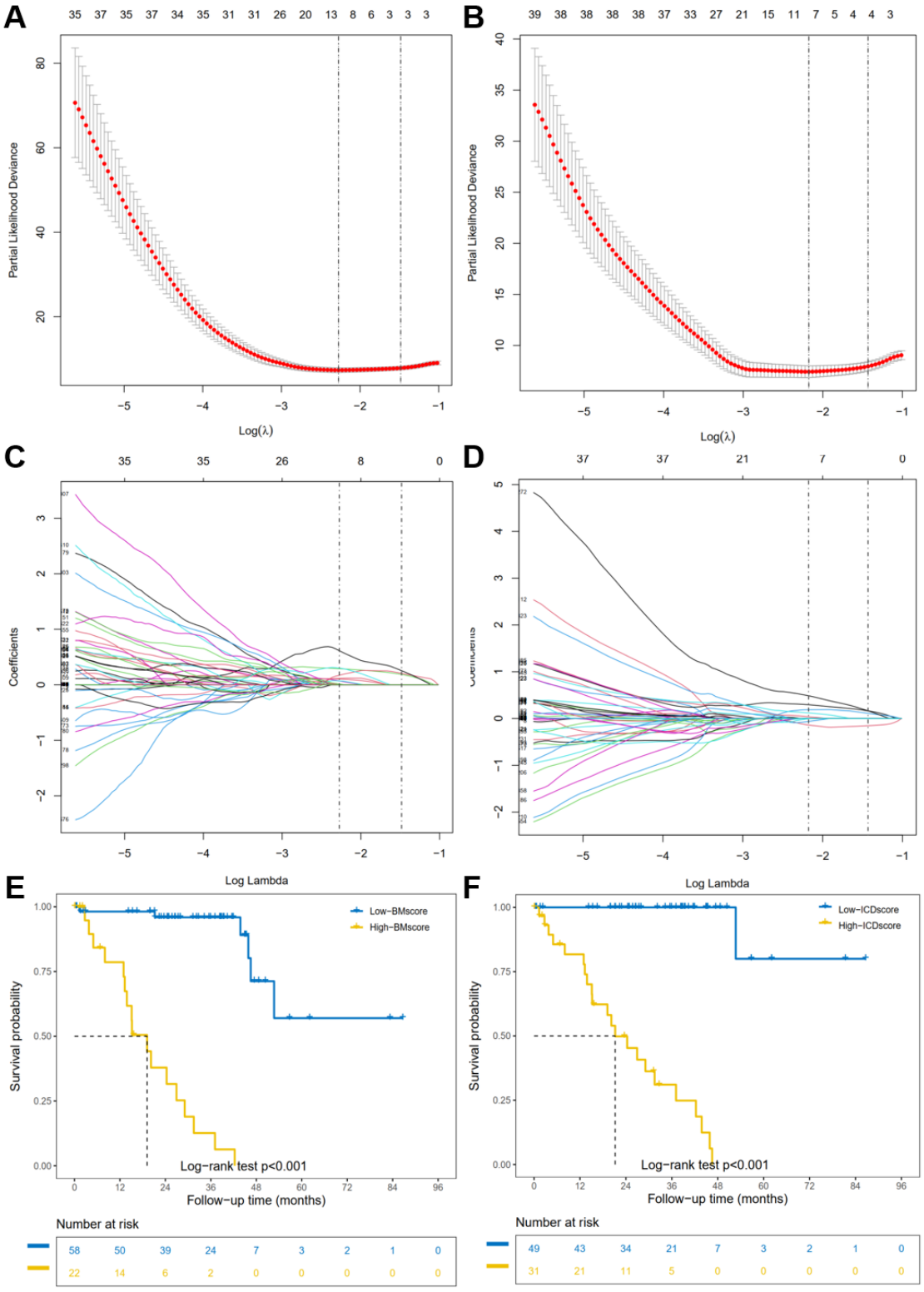


Figure 3.11 Performing Separate-LASSO algorithm to calculate the BMscore (A,C,E) and the ICDscore (B,D,F) in the training set. *(**A, B**) Tuning the model using a 10-fold cross-validation. (**C, D**) Fractions of candidate genes with different paths of LASSO coefficient. (**E, F**) K-M curves for the OS of UVM patients.*

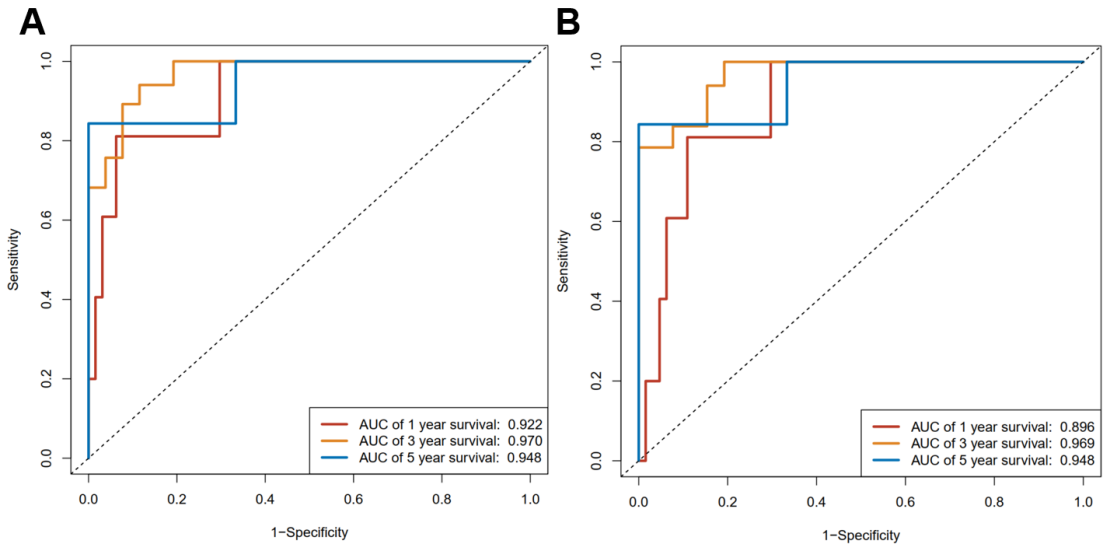


Figure 3.12 ROC curves for the BMscore (A) and ICDscore (B) depicting the AUCs of 1-, 3-, and 5-year OS in the TCGA-UVM cohort (training set).

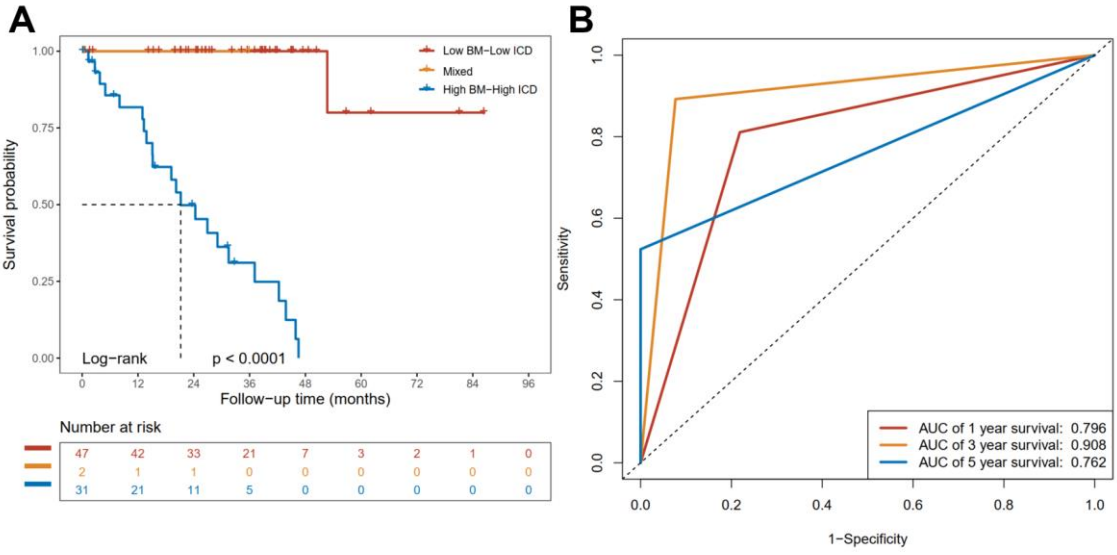


Figure 3.13 Construction and evaluation of the BMICD classifier in the TCGA-UVM cohort (training set). *(**A**) K-M curves for the OS of three BMICD group. (**B**) ROC curves depicting the AUCs of 1-, 3-, and 5-year OS among three BMICD group.*

Table 3.3 Coefficients of selected genes by Separate-LASSO algorithm

Gene	Coefficient	Gene set
NECAB2	0.005	BM
CA12	0.289	BM
ARX	0.071	BM
KDELR3	0.171	BM
ADAM12	0.071	BM
PARP8	0.614	BM
MMP9	0.105	BM
POLA1	0.115	BM
SLC45A2	0.016	BM
S100A13	0.301	BM
CA12	0.295	ICD
KDELR3	0.187	ICD
PARP8	0.486	ICD
MMP9	0.119	ICD
SLCO5A1	0.038	ICD
TJP2	-0.139	ICD
CHRNB2	-0.063	ICD
SH3D19	-0.006	ICD

Abbreviations: *BM, basement membrane; ICD, immunogenic cell death.*

3.3.6 BMscore, ICDscore and BMICDscore calculation using IPF-LASSO algorithm

For the second model, we tried to introduce the “modality” to the same gene set and separate the whole dataset into two blocks based on two modalities, so that we can directly calculate the incorporated risk score simultaneously containing two gene sets. We assigned penalty factors to each block and chose the best model according to the cross-validated prediction error. Finally, we got the BMICDscore including ten BM genes (also served as BMscore) and seven ICD genes (also served as ICDscore). Detailed parameters with coefficients were listed in **Table 3.4**. K-M curves depicted the statistically significant differences in OS between high- and low-risk group in all risk scores (all log-rank p values are less than 0.001; **Figure 3.14**). The time-dependent ROC curves showcased predictive performance of the OS with different time points (BMscore: 1-, 3-, and 5-year OS AUC = 0.924, 0.962 and 0.948, respectively; ICDscore: 1-, 3-, and 5-year OS AUC = 0.843, 0.885, and 0.818, respectively; BMICDscore: 1-, 3-, and 5-year OS AUC = 0.921, 0.966, and 0.948, respectively) , suggesting the good power of all models for UVM prognostic prediction.

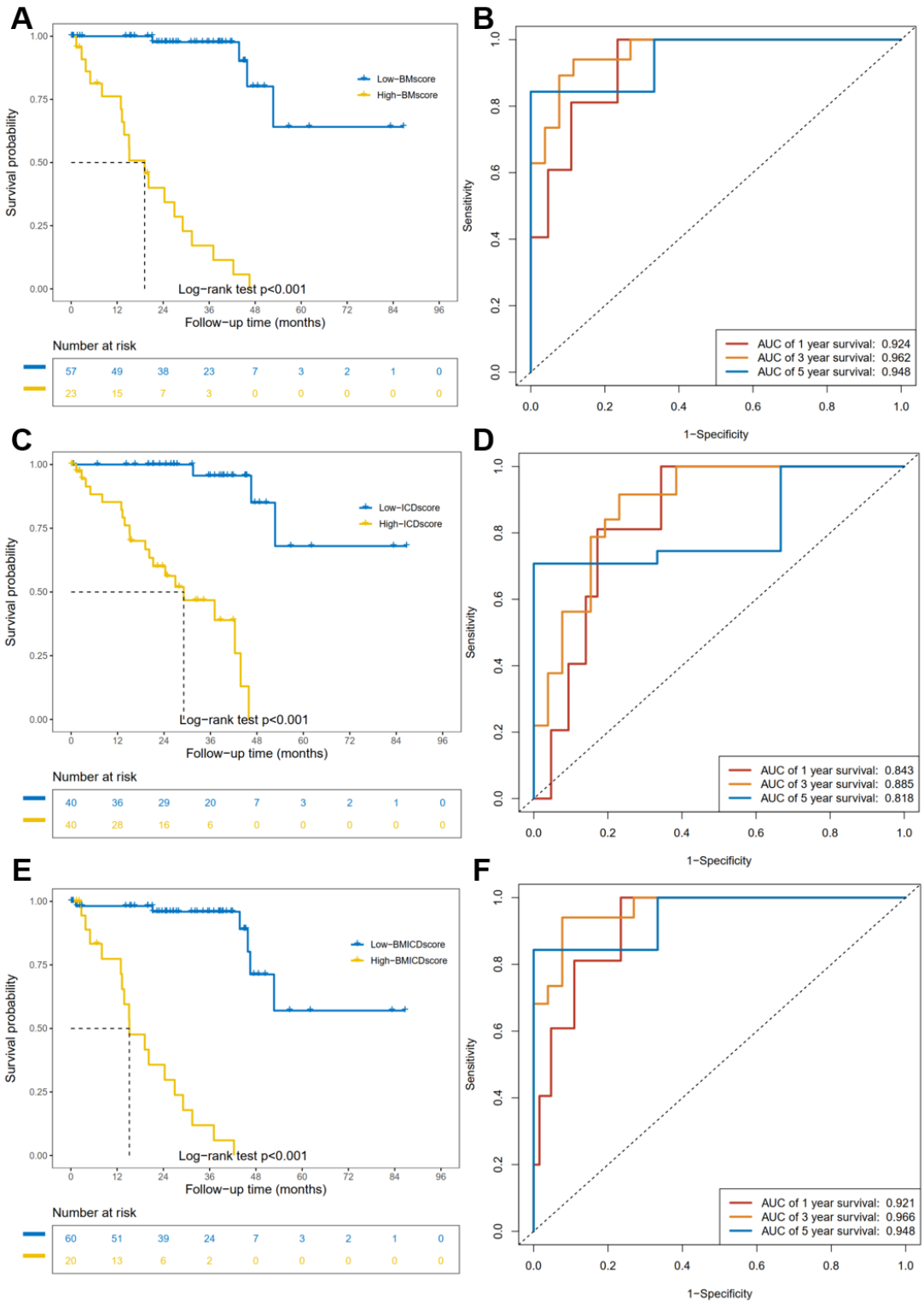


Figure 3.14 Calculation and evaluation of BMscore (A, B), ICDscore (C,D) and BMICDscore (E, F) based on IPF-LASSO algorithm in the TCGA-UVM cohort (training set). *(**A, C, E**) K-M curves for the OS of BMscore (**A**), ICDscore (**C**) and BMICDscore (**E**). (**B, D, F**) ROC curves depicting the AUCs of 1-, 3-, and 5-year OS for BMscore (**B**), ICDscore (**D**) and BMICDscore (**F**).*

Table 3.4 Coefficients of selected genes by IPF-LASSO algorithm

Gene	Coefficient	Gene set
S100A4	0.118	BM
NECAB2	0.140	BM
CA12	0.326	BM
ARX	0.125	BM
MATK	0.008	BM
MMP9	0.070	BM
ANKRD30B	0.079	BM
PCP4	-0.053	BM
MSC	-0.048	BM
GSTA3	-0.185	BM
IGJ	0.006	ICD
CA12	0.015	ICD
MMP9	0.023	ICD
MSC	-0.010	ICD
GSTA3	-0.004	ICD
HMCN1	-0.124	ICD
PDE6G	-0.050	ICD

Abbreviations: *BM, basement membrane; ICD, immunogenic cell death.*

3.3.7 Feature selection for the PCA algorithm

Before performing PCA algorithm, we attempted to reduce the noise or redundant genes in the DEGs of both BMtypes and ICDtypes. As a dimensionality reduction approach based on the random forest algorithm (Kursa, 2010), “Boruta” algorithm was conducted here. **Table 3.5** and **Table 3.6** list all included genes and excluded genes in the DEGs of both BMtypes and ICDtypes, respectively.

Table 3.5 BM-gene selection by Boruta algorithm

Decision	Gene

Confirmed (n=126)	HTR2B, CHAC1, COL9A3, ECM1, HTRA1, MYEOV, PLN, GEM, CARD11, PPM1K, PTP4A3, HTRA3, SDC2, DLL4, PRKCDBP, RAB31, NQO1, HES6, MRC2, ARMC9, IGFBP2, KIAA0196, SULF2, FABP3, CA12, TGFBI, MAP1A, ARC, PRKDC, CALHM2, FN1, FKBP11, TLCD1, JAG2, FKBP10, C4A, AMN, ITGA5, MAPK12, SPHK1, ASB9, COL18A1, ATP8B2, SGSM2, TRIB1, GREB1, TSC22D1, PAPSS2, ELFN1, OAF, SPARC, CSPG4, ZNF467, TTYH3, GPC4, SERPINE1, COL5A1, BASP1, LAPTM4B, SOCS2, TCIRG1, ECE1, BCL3, MGST2, DECR1, PDGFRB, CYC1, CDH24, ADCY3, PDLIM4, FOXD1, MFSD3, GAS8, DENND3, HM13, APLNR, ACADVL, DHRS7B, GPR162, PREX1, COL5A3, FAM105A, G6PC3, YWHAZ, DUSP14, NSMF, LIG3, VPS53, VOPP1, PVRL1, PLEKHG2, LAYN, KIAA1462, RFTN1, PLXDC1, GPAA1, CAMK1, TNFRSF1A, LTBR, CLPB, TRIL, DDRGK1, SHC1, AFAP1L2, SDHA, NDUFB9, TMEM208, BAX, GJA4, POR, GPR143, SLC35F6, DLST, SHPK, ACAD10, ALG1, GBA, FOXRED1, ITM2C, AIFM2, TM4SF1, SKIV2L, ZHX3, CTF1, ZNF835, LY6G5C
Tentative (n=56)	SPON2, IRX1, FERMT3, GPR56, CADM1, TMEM119, ASTN2, KLHL38, RCOR2, PRKCA, BAI1, COX7A1, ADAM11, PAM, GGT1, KLHL30, KDELR3, COL4A2, ASAP1, SLC16 A8, STEAP3, P2RX6, ITFG1, HOGA1, FAM132B, DERL1, TCEB1, PTPLA, THY1, BAI2, ADCK5, C8orf33, PPAP2B, ANPEP, SORL1, ODF3L1, P4HA2, KDR, PPIC, CKS2, MAF, MROH1, ATP8B3, VASH1, FADS2, QDPR, C8orf82, FOXS1, MYO1G, ERVMER34-1, CXorf65, RASL11B, RENBP, EMC7
Rejected (n=902)	VGF, SLC38A5, CYSLTR2, EEF1A2, PTGER4, TNFRSF19, WARS, KIT, HSPB8, CELF2, VTN, AHNAK2, S100A4, ISM1, GRID1, TRPV2, NT5E, SSX5, UCHL1, APOL1, HLA-DRB5, TMEM200B, LINGO1, GDF11, NECAB2, SLC16A6, TFPI2, RARRES2, C1S, NKG7, SPARCL1, C1R, B2M, SEL1L3, FCGR3A, HLA-DRA, S100A1, CCL5, HLA-B, HLA-DRB1, ADAM23, ADCY1, CDH1, CITED1, GALNT18, CLIP3, FIBCD1, CCNO, IRF1, HLA-DQB1, C1orf95, HLA-DQA1, TTR, TSPYL5, GBP1, ARX, C4B, LAG3, TAP1, ITPR2, PHLDA2, ADAMTS2, MMP2, THBS2, KCTD17, LY96, PSMB9, C1orf116, LYZ, FABP5, SIPA1L2, C1QB, CST7, CD74, DYSF, HTATIP2, IFI27, LOXL4, PFKP, CPLX1, HLA-DPA1, CYGB, CPXM1, PANX2, TMEM255A, PLEKHG4B, NFATC4, CD3D, CEBPD, COL1A1, IQGAP1, CPVL, KIF17, PCDHGA12, CDC25B, C1QC, C2, SECTM1, SQLE, FAM26F, CD8A, ME1, ATP1B2, FZD7, PIK3C2A, CXCL9, CDKN1A, GBP2, GZMA, RRM2B, ATP1B1, LHFP, FRZB, MAL2, ISG20, CRABP2, SGK1, CXCL10, IGFBP7, SCN1B, SSX1, BAG2, GJA1, MATK, MGLL, PXDN, TYMP, CD2, MT1E, SLC1A4, SULT1A1, CD209, MDFI, GDF15, CD79B, COL4A1, IMPA1, C1QA, COL3A1, RAB2A, CECR1, MAN1C1, PROS1, STAT1, CHI3L1, SLCO5A1, RTN4RL1, DNAJC3, MGST1, ARHGDIG, SOAT1, HLA-DOA, TNFRSF1B, APOL3, NRG3, LAPTM5, ALDH1L2, CD3E, JPH1, DNAJC13, RIMS2, HLA-C, NT5DC3, MTDH, NOV, LYN, FEZ1, AQP1, CD27, FKBP5, TAF2, SLC22A17, SCCPDH, SEPT2, KCNK1, SPESP1, HCN2, GPM6B, ITGAX, RPS6KA2, ADAM12, MRPS28, LGALS3, ARFGEF1, P4HA1, PRF1, GZMK, CPE, IL18BP, DKK3, STK32A, S100A6, GBP4, FMN1, RDX, FADS1, FBXL7, NLRC5, HLA-F, FOS, RAP1GAP, GPR176, SLC1A1, HLA-A, IL2RG, RARRES3, CD70, SPIDR, CTSO, TAPBPL, ADRA2A, TWIST2, IL32, SLC2A3, TRAM1, BST2, KIAA1644, STOM, BTN3A3, APOL6, PAEP, PARP8, NDUFA4L2, ABI3, CHSY3, OXCT1, TNFAIP2, NOX5, IFITM1, MDK, MMP9, EPDR1, WWTR1, VLDLR, ACSL1, ASS1, ATHL1, PRIMA1, LGALS1, ANKRD30B, FHDC1, ALG5, MAP1B, COL1A2, TGS1, YTHDF3, RMDN1, UBR5, KLF4, AK4, AGT, SLC29A4, IDH2, CNTNAP1, HLA-DPB1, PSMB8, RNASE6, PCOLCE, SYT11, TPPP, PTGDS, NCALD, COTL1, COL22A1, PGM1, TSC22D3, TMEM70, JAG1, MCM4, LMAN1, FJX1, HLA-DQB2, SAP30, SNX25, REEP1, ZNF704, MRPL13, UBE2V2, CXCR4, TNFSF9, LYPLA1, PTPN13, HCLS1, MAP3K12, DOCK10, OLFM1, NSMCE4A, EFR3A, POLA1, GABRA5, ISG15, BAIAP2L2, FGF9, SFXN3, ASIP, IFI16, LCP1, ENO3, CCL4, FAM49B, COPS5, FLNC, ST3GAL1, SRD5A3, PHACTR1, PTPRU, SLC45A2, PTGS1, HLA-DQA2, OLIG1, AIF1, NSMAF, MPEG1, TCF12, HLA-DMA, NIPAL2, PLXNC1, HDAC4, SLC38A6, SMOC2, TRPM2, NPTX1, WIPI1, MRPL15, EMC2, MTFR1, ADM2, TMEM173, OMG, FAM173B, CD109, RASD2, VAV2, SOX1, BMPR1B, IDUA, GALNT14, CXXC5, POMC, SH3PXD2A, MDGA1, GNG4, CD53, SLC2A10, ETV7, UBD, FFAR3, MT1A, TSPAN10, SDCBP, GIMAP4, HRSP12, BCAT1, ATP6V1H, SCD5, WSB1, FGL2, MAFB, RASA3, SOCS3, CORO1A, SGCD, ANO6, RAC2, SCG3, EMID1, FOSL1, C3, CTSW, RTF1, ABHD3, NAAA, RGS1, SLC11A2, STK39, PIK3R3, ABCC4, MS4A7, MAPT, DUSP9, VSTM2L, CPQ, CSF2RA, SLC25A32, MR1, LRP1, B3GNT7, SMARCA5, TUBB3, GABRB3, FIG4, GADD45G, NBN,

B4GALNT3, CXCR3, LFNG, C1QL1, MET, THEM6, CRIP2, LAP3, RECQL, SPOCK1, TNFRSF21, NID1, FGFR1, IGFBP5, RAB15, TMEM176A, NOTCH3, HLA-DMB, GOLM1, GIMAP7, GLIS2, ALK, PTN, VCPIP1, APOD, SEPHS2, TBC1D1, TYROBP, COLGALT2, SLC38A3, IMPAD1, CDC27, HMOX1, ANKDD1A, CD7, PTPN6, RAPGEF4, LDB3, FMNL1, EHD2, HECTD1, LHB, ANXA4, UBXN2B, FAM167B, VWCE, DOK5, ATP6V1C1, KIAA0020, SIGLEC10, PCDHGC3, TMEM176B, LRRC39, ITGA6, C16orf45, OLFML3, ABCB7, SCX, SPON1, IGF2, WDR34, CD300A, SLAMF8, PSME4, NCOA2, TNFRSF12A, CHD7, FAM91A1, RNF139, TMEM64, PGM2, SLC16A2, PRDX4, SNCG, ANO9, UBE2W, SUMF2, CPM, MYOF, KCND3, SH3PXD2B, PDLIM1, MGAT5, GZMH, TCEA1, NSF, PECAM1, CORO6, EXOSC4, SUSD2, ACSBG1, SPTBN1, EMILIN3, SYNE2, OLFML2B, CLIC2, DPYSL3, ADAM10, PAG1, DCAF13, CD163L1, ULK4, HPSE, GFPT1, SQRDL, OXR1, CLCNKA, RNF19A, TMEM55A, TMEM68, TMSB4X, ITGAL, PITPNC1, UCP2, LDLRAD3, ID3, FAS, LCK, SLC9A3, B4GALNT4, LACTB2, CCDC140, TPR, HIP1, RAMP1, IL1R1, PCSK1, HLA-E, TRPM4, KIF21A, GMFG, SLC38A2, MS4A6A, MTERF3, HERC2, SLC38A8, XPOT, ALDH4A1, TSPAN5, LAMC1, RGS5, PHYHIP, COBL, SLC17A9, BTBD19, COQ2, ANXA1, THEMIS2, SASH3, OSGIN2, CTSC, CCDC74A, CA2, KATNAL2, PSRC1, EVA1B, PAQR8, MICAL1, ELP2, SGCB, SYK, IFI6, ACACB, STAT3, CD48, CORO2B, PBX3, SEC14L2, NUP88, IL10RA, KIF20A, FCER1G, CAPN3, CD8B, TNFAIP8L2, SPIRE1, CCDC64, GPC1, SRGAP2, ASPH, PLEKHO1, SLC52A2, PLA2G4C, CNTF, CD52, PCP4, LAMA5, COMT, EBI3, AUTS2, LST1, HEXB, CD93, OLFM2, SLITRK2, DDIT4L, FOXC1, COX6C, FPR3, IARS, RASSF2, MT2A, MXRA8, TERF1, RNF145, ILKAP, ZMYND8, NACC2, CALU, APP, NCS1, AZIN1, ABCA8, PPFIBP1, GZMB, ABCB10, RTTN, MDM2, ARMC1, MAN1A1, CLIC4, PTK2, MAP2, PKM, SYNJ2, PPAPDC1B, TRPV4, MAD1L1, PSTPIP1, VPS13B, RAD21, PDCD1, CASP1, FAM134B, NRD1, BATF, PCDHGB5, CTNND2, GLIPR2, P2RX4, IRF8, MC1R, GPR124, IL2RB, SIT1, SLC22A18, CCL18, SLC17A5, DHRS3, REEP2, INPP4B, TP53INP1, ZNF385A, CXCL12, KIAA1429, FSTL4, SLC6A17, LHFPL2, HCK, TPP2, CPNE7, NDUFAF1, LAMA1, SH2B3, LRRCC1, ANGPTL7, CCL4L1, SH2D5, SPATA18, EIF4G3, TMEM206, SH3RF3, STAM, POLR2K, ADAM33, A2M, CACNA1H, ZDHHC7, GPX8, TRIM47, IL12RB1, SLIT1, CBR3, HEXIM1, CD9, COX4I2, TFRC, C10orf54, WFDC1, ZNF469, CCDC109B, TRAK2, SIRPG, CNTFR, CPNE3, CA8, IDO1, ATP8B1, C8orf76, DNASE1L3, CDCA7, NOP58, GJA3, SEC24D, IAH1, BGN, CCRN4L, SEMA5A, NBL1, PKD2, NCF4, KIAA0513, C1orf21, DIP2A, PSMA3, HSPA5, CHST11, SORD, NKX2-4, DUSP5, S100A13, MEIS3, NRSN2, ABTB2, GIMAP6, LSP1, MUT, TMEM87B, SOS1, SFXN5, JKAMP, AP1S3, GALNT3, RBFOX2, RRM2, CDK5R1, HPN, C2CD2, CA14, VAT1L, GABRD, FKBP7, NES, RCAN2, AUH, RDH10, MAN2A1, MT1G, HLCS, GNPTAB, ASIC1, F2R, TAGLN2, VSIG4, PANX1, CTSS, IGFBP4, CD14, SNX22, PCED1B, EDA2R, VIMP, NUSAP1, MANBA, METTL7B, FAM98A, FGD5, GPER1, PHTF1, COL9A2, MBP, BTBD1, JUNB, CP, TMEM87A, IDE, SLC4A11, DCSTAMP, RGS16, ATP9A, CALM2, TGFB1, WAS, PLEKHF2, NR2F1, CORO1C, KIRREL, PEX2, TBCE, GSDMD, NELL2, DPY19L4, EVI2B, TMCO1, MTSS1, EMP3, SEPT11, ATP2A2, ATAD1, ADAM9, CCL3, CTSH, CTPS1, SELPLG, MSC, GSTA3, SYNPR, HHATL, AZGP1, CLEC11A, COX6A2, GPR27, LIMS2, ENPP2, EFS, FBXO17, SPP1, BCHE, MLIP, LNP1, IL12RB2, ROPN1B, RNF208, CHL1, MTUS1, MPZ, SLC44A3, CAMSAP3, GATA4, PALMD, PDE3A, RAPGEF3, DLL3, EDNRB, SLC25A38, LGI4, ACSF2, ANG, SSUH2, KLC3, PPARG, BEX1, MEGF10, RNF43, WNK4, SNCA, ZNF415, PPP1R3C, LMCD1, SATB1, KLK13, ZNF391, BAP1, ROPN1, ANKRD65, KIAA1045, TFAP2A, RBP7, GIPC3, RORC, BEX5, MUC7, HNMT, CPS1, GSTO2, SERPINB9, ALDH1L1, SHC4, FBP2, PLEKHG4, DNALI1, AHCYL2, ATAD3C, NR6A1, KCNK2, RPSAP58, ROBO1, MANEAL, HRASLS, NEDD9, COL11A1, PLXNB1, EFCAB1, C4orf19, IRF6, DUSP15, EXTL1, ZNF667, DCT, MAMSTR, PDE4B, SCIN, SORBS2, ZNF560, FAM71E1, PPP1R3B, L1CAM, MID1, RPL32, TUBB4A, NFIA, TFF3, PRRT4, TSGA10IP, YJEFN3, HFE, SH3D19, MYH14

Table 3.6 ICD-gene selection by Boruta algorithm

Decision	Gene

Confirmed (n=100)	HTR2B, CYSLTR2, VGF, CHAC1, LY96, HLA-DRB5, WARS, LAG3, CD3D, NKG7, CCL5, TAP1, CD74, CD8A, CD2, C1QB, GZMK, C1QC, HLA-DQB1, CD3E, IRF1, CD27, FAM26F, HLA-A, APOL6, HLA-E, CXCR3, HTATIP2, HLA-DOA, FGL2, CCL4, NKX2-4, CD8B, CLIC2, FABP3, ARMC9, CD209, CD53, SIT1, LCK, ADAMDEC1, SIRPG, BTN3A3, ITGAL, GIMAP7, RRM2B, PDCD1, TNFRSF1B, HLA-DMB, PTPRC, SIGLEC10, LAP3, IRF8, GZMB, CD163L1, RAP1GAP, CD300A, AP1S3, PTPN7, BTBD19, SPN, SLC2A10, TBC1D10C, CD247, SH2D1A, MDM2, ZAP70, BIN2, SLA2, APOL4, TIGIT, SEP15, EIF4G3, LAMA1, TNFRSF9, SAMSN1, CLCNKA, MLIP, WNK4, PPP1R3C, LMCD1, JUP, IFT122, EPHX2, RPL32, KIF7, NMNAT3, SLC7A2, ZNF853, SCGB3A1, ZBTB47, ZNF395, TCTN1, FAM86B1, MSRA, C8orf58, SORBS3, FAM181B, PCBP4, RPL29
Tentative (n=41)	CXCL9, CXCL10, GZMA, HSPB8, HLA-C, HLA-DQA1, ISG15, IL18BP, ADCY1, RNASE6, KIAA0196, ITGAX, MS4A7, EVI2B, MS4A6A, ABI3, MTDH, CAPZA1, CHSY3, FKBP10, TNFSF9, KDELR3, TRPM2, PARP8, TGS1, TMEM70, GREB1, CXCR6, CD86, SLC38A6, RNF208, ACSF2, GSTO2, H1FX, GPD1L, CENPV, FAM127C, GLTSCR2, SF3A2
Rejected (n=631)	HLA-B, ECM1, EEF1A2, APOL1, GBP1, HLA-DRA, LYZ, IFI27, BST2, HLA-DRB1, PSMB9, B2M, FCGR3A, S100A4, CST7, RARRES3, GBP2, MYEOV, TRPV2, HLA-F, HLA-DPA1, ISM1, C1QA, TNFRSF19, UBD, STAT1, COL22A1, IL2RG, GEM, HTRA1, LAPTM5, SLC16A6, APOL3, LGALS3, IFI6, ETV7, PTGER4, LCP1, ADAM23, CTSS, IGJ, SECTM1, PRF1, CELF2, GBP4, PTP4A3, PSMB8, SDC2, SPON2, TYMP, NQO1, NT5E, IFITM1, DLL4, RIMS2, HLA-DQA2, HLA-DQB2, UCHL1, C1S, CTSW, HLA-DPB1, CD52, CD48, CORO1A, SEL1L3, PAEP, ITPR2, IL32, PPM1K, AIF1, SLAMF8, TYROBP, GBP5, IDO1, SGK1, CITED1, PIK3C2A, C1R, NLRC5, RAB31, BATF, CADM1, RAC2, SASH3, GZMH, HLA-DMA, GIMAP4, MPEG1, FCER1G, KIT, BTN3A1, GGT1, CECR1, ISG20, TRIM22, TMEM119, SULF2, IMPA1, CCR5, EOMES, HCLS1, RGS1, PLN, GDF15, FERMT3, CD5L, KLHL38, NOV, PLEK, IL2RB, PHLDA2, CASP1, AHNAK2, ADAMTS2, TMSB4X, SLC38A5, C2, TMEM200B, FMN1, MATK, CDKN1A, SSX5, ARC, CARD11, IL12RB1, GPR56, MAP1A, CCL18, KIF17, STOM, MRC2, SLC6A17, CD7, IL10RA, THBS2, PRKDC, QPCT, MAL2, BAG2, PTPN6, FIBCD1, SNX25, TRIM69, GBP3, TAPBPL, LTB, RAB2A, IL1R1, SNX10, FAS, COL9A3, CCL3, CD70, VSIG4, ME1, IQGAP1, ARFGEF1, UCP2, DOK2, UBE2L6, S100A1, DCSTAMP, CA12, HTRA3, FAM91A1, C4B, CYBB, ANO9, CCL4L1, IGFBP7, S100A6, CXCL13, SSX1, LOXL4, KCND3, IGSF6, LCP2, SLAMF7, LYPLA1, TNFAIP2, C4A, TRAM1, EBI3, PRKCDBP, TGFBI, TAF2, GMFG, NBN, SLAMF6, CPVL, YTHDF3, GRID1, FKBP11, TMEM255A, FGF9, CD79B, FABP5, SELPLG, FAM49B, GALNT18, MRPS28, SDCBP, WAS, TNFAIP8L2, SLC1A1, ASTN2, LST1, NCF4, PANX2, FPR3, EFR3A, SIGLEC8, HCK, SOAT1, GDF11, PARP9, HRSP12, RASAL3, MRPL13, IGFBP2, EPSTI1, NIPAL2, PHACTR1, CYGB, SCPEP1, EMC2, ABCC2, CD14, PRKCA, HAVCR2, LAPTM4B, CD5, PLEKHG4B, BTN3A2, DERL1, ATP6V1C1, SPHK1, CRABP2, TSPYL5, WIPI1, CLIP3, SPI1, NPTX1, C1orf116, OMG, PFKP, TAP2, WWTR1, OAS2, CD163, MTFR1, PSRC1, CDC25B, ASS1, DOCK10, SOCS2, ITGB2, MAN1C1, OXR1, UBR5, LRRC39, VCPIP1, IKZF3, NFATC4, PLA2G7, ASAP1, TRIB1, TMEM64, TNFRSF21, MMP9, ABHD3, MRPL15, UBE2V2, NCKAP1L, UBE2W, MT1E, CEBPD, FAM167B, RMDN1, IFI16, APOL2, ALDH1L2, TCEB1, ASIP, SQRDL, CCNO, ARHGAP30, FASLG, ACSBG1, RNF139, GPM6B, ARMC1, DNAJC13, THEMIS2, TLCD1, LRRCC1, CTSO, LGALS1, CCL24, COPS5, FTL, CD109, CYC1, BCL3, SLCO5A1, MAFB, CLIC4, LGALS9, SQLE, SLIT1, PAM, GIMAP6, CALHM2, FZD7, CNTF, P4HA1, PCED1B, MTERF3, RDX, LACTB2, MGST1, DDX60, HMOX1, LYN, FOSL1, CSF2RA, TNFSF13B, SMARCA5, TMEM68, KCTD17, RB1CC1, DNAJC3, MS4A4A, TP53INP1, NSMAF, ATP1B1, MAP3K12, IFI44, PTCHD4, LHFPL3, AK4, PARP14, SPARCL1, FBXO32, AOAH, SLC38A2, NT5DC3, AGT, RNF145, PGM1, FCGR1A, TMEM51, XCL2, RECQL, PROS1, CD300LF, ARHGAP9, HHATL, AZGP1, MSC, COX6A2, ENPP2, CTF1, SYNPR, EFS, LIMS2, GSTA3, HPGD, SLC44A3, CAMSAP3, BEX1, CLEC11A, FBXO17, KLC3, ALDH1L1, KCNK2, GPR27, CHADL, CHL1, BCHE, EDNRB, SEMA6A, BEX5, SEMA3B, LNP1, PDE3A, COL11A1, SNCA, SLC25A38, GATA4, MPZ, KIAA1045, MTUS1, PRRT4, CPS1, PDK4, ERBB3, RAPGEF3, CHGB, BAP1, ROPN1B, KCNA5, PPARG, RAB34, FAM129A, PERP, DUSP15, SOX15, NEDD9, GIPC3, SATB1, ZNF835, DLC1, TFAP2A, GPR37, PLXNB1, HNMT, SERPINB9, PPP1R3B, LGI4, MYH14, RBPMS, OSBP2, RBP7, ACCSL, MANEAL, PLEKHG4, MATN2, MARC2, PALMD, KCNE4, DLL3, NDN, ATAD3C, FDFT1, ZNF391,

SHC4, RPSAP58, MAMSTR, GJB1, MID1, CDCA7L, CDC42EP1, TUBB4A, GJC3, ID4, SOBP, AFF3, IL12RB2, ABCA3, APOM, HRASLS, RFTN2, DNALI1, ARSE, PHYHD1, FAM71E1, KREMEN2, ME3, H2AFY2, EMCN, TJP2, KCNG2, FMOD, HMCN1, HSD17B8, ANKRD65, PHLDA1, ANG, POMGNT2, TIMP3, CA4, FHL2, PLCD1, OVOL1, MCF2, NR6A1, KCNQ5, ABHD14B, MEGF10, ROBO1, NUDT18, HDAC11, NFIA, CITED4, PALM, EFEMP2, ADRA2C, FBP2, COL11A2, ITGB4, RAMP2, SSUH2, ASB10, ABLIM1, YJEFN3, SULT1C4, RNF43, ZNF581, ZNF608, ALG1L, ZSCAN18, LYRM4, LTA4H, DUSP8, NRTN, RPL14, PDGFD, WIPF3, GYG2, RAB11FIP1, GRIK3, SLC6A15, MPPED2, PCOLCE2, SORBS2, EEFSEC, C10orf2, STEAP1, CH507-42P11.8, PDE4B, TDRD3, BTG1, NMRK2, CAND2, RPL35A, KIAA1217, TGFB1I1, C6orf48, CHRNB2, CTNNB1, BOC, SH3D19, GAS1, MARCH9, NPM2, MUC7, ZNF415, C16orf86, RXRG, HOXD1, ASRGL1, TMEM8B, RHCG, RADIL, DVL3, RORC, TMEM97, CELF5, PDE6G, DDAH2, DPYSL2, ZBTB12, RPL15, DUOX1, DPP6, NIPSNAP3B, SLC35F2, ASPG, DALRD3, RPSA, FAM69B, DCT, SLC2A4, ZDHHC2, EFNB3, FAM212A, DDX11, SEMA3C, ASPHD1, RPL22L1, NOXA1, C6orf226, HYAL1, TIPARP, PCSK4, LTBP1, POU3F1, CDO1, EIF1B, ZNF135, NYNRIN, ZNF560, GNAI1, SLC27A5, IRF6, MARCKS, LRRN1, KIAA1549L, ROPN1, SLC41A3, ZNF229, TFF3, ENKD1, ID2, NHSL1, IGHMBP2, AMOTL2, PPP1R13L, ADCY6, MZT2A, SYN2, BEX2, OSBPL2, FHIT, KIF6, CASKIN1, KLF15, NPW, NCKAP5, SCARA3, ZFPM1, OGDHL, SCAP, SCT, SOSTDC1, ADPRHL1, SPAG16, RPL24, TSGA10IP, MYO7B, CDK2, POSTN, BAMBI, ARHGEF37, C2orf72, ADAMTSL4, EFCAB1, TCF7L1, TSPAN7, SUGCT

3.3.8 Risk score calculation and classifier construction using PCA algorithm

Based on the results of Boruta dimensionality reduction, we chose 126 BM-related genes and 100 ICD-related genes for modeling. We initially computed the Cox coefficients of these genes separately using univariate Cox regression, extracted their positive and negative signs, as summarized in **Table 3.7**. Then, we applied the PCA algorithm on them and calculate the BMscore and ICDscore according to the algorithms described in **Section 3.2.5**. Next, the training set was classified into high- and low-BMscore groups, and high- and low-ICDscore groups, respectively. Using the same approach as that in **Section 3.3.5,** we developed another BMICD classifier comprised of “High BM**–**High ICD”, “Mixed”, and “Low BM**–**Low ICD” groups. We visualized their performance in **Figure 3.15**. The K-M curves showed that high-BMscore and high-ICDscore were both associated with death (log-rank *p* values are all less than 0.0001). Patients in the “Low BM**–**Low ICD” group had the best prognosis, followed by the “Mixed” group, and the “High BM**–**High ICD” group had the worst outcome. The time-dependent ROC curves showed predictive performance of the OS with different time points (BMscore: 1-, 3-, and 5-year OS AUC = 0.804, 0.838, and 0.889, respectively; ICDscore: 1-, 3-, and 5-year OS AUC = 0.759, 0.833, and 0.854, respectively; BMICD classifier: 1-, 3-, and 5-year OS AUC = 0.852, 0.845, and 0.796, respectively).

Table 3.7 The selected genes and their signs of cox coefficients used for PCA

Gene set – Cox sign	Gene
BM – positive	HTR2B, CHAC1, COL9A3, ECM1, HTRA1, MYEOV, PLN, GEM, CARD11, PPM1K, PTP4A3, HTRA3, SDC2, DLL4, PRKCDBP, RAB31, NQO1, HES6, MRC2, ARMC9, IGFBP2, KIAA0196, SULF2, FABP3, CA12, TGFBI, MAP1A, ARC, PRKDC, CALHM2, FN1, FKBP11, TLCD1, JAG2, FKBP10, C4A, AMN, ITGA5, MAPK12, SPHK1, ASB9, COL18A1, ATP8B2, SGSM2, TRIB1, GREB1, TSC22D1, PAPSS2, ELFN1, OAF, SPARC, CSPG4, ZNF467, TTYH3, GPC4, SERPINE1, COL5A1, BASP1, LAPTM4B, SOCS2, TCIRG1, ECE1, BCL3, MGST2, DECR1, PDGFRB, CYC1, CDH24, ADCY3, PDLIM4, FOXD1, MFSD3, GAS8, DENND3, HM13, APLNR, ACADVL, DHRS7B, GPR162, PREX1, COL5A3, FAM105A, G6PC3, YWHAZ, DUSP14, NSMF, LIG3, VPS53, VOPP1, PVRL1, PLEKHG2, LAYN, KIAA1462, RFTN1, PLXDC1, GPAA1, CAMK1, TNFRSF1A, LTBR, CLPB, TRIL, DDRGK1, SHC1, AFAP1L2, SDHA, NDUFB9, TMEM208, BAX, GJA4, POR, GPR143, SLC35F6, DLST, SHPK, ACAD10, ALG1, GBA, FOXRED1, ITM2C, AIFM2, TM4SF1, SKIV2L, ZHX3
BM – negative	CTF1, ZNF835, LY6G5C
ICD – positive	HTR2B, CYSLTR2, VGF, CHAC1, LY96, HLA-DRB5, WARS, LAG3, CD3D, NKG7, CCL5, TAP1, CD74, CD8A, CD2, C1QB, GZMK, C1QC, HLA-DQB1, CD3E, IRF1, CD27, FAM26F, HLA-A, APOL6, HLA-E, CXCR3, HTATIP2, HLA-DOA, CCL4, NKX2-4, CD8B, CLIC2, FABP3, ARMC9, CD209, CD53, SIT1, LCK, ADAMDEC1, SIRPG, BTN3A3, ITGAL, GIMAP7, RRM2B, PDCD1, TNFRSF1B, HLA-DMB, PTPRC, SIGLEC10, LAP3, IRF8, GZMB, CD163L1, RAP1GAP, CD300A, AP1S3, PTPN7, BTBD19, SPN, SLC2A10, TBC1D10C, CD247, SH2D1A, MDM2, ZAP70, BIN2, SLA2, APOL4, TIGIT, EIF4G3, LAMA1, TNFRSF9, SAMSN1, CLCNKA
ICD – negative	MLIP, WNK4, PPP1R3C, LMCD1, JUP, IFT122, EPHX2, RPL32, KIF7, NMNAT3, SLC7A2, ZNF853, ZBTB47, ZNF395, TCTN1, FAM86B1, C8orf58, FAM181B, PCBP4, RPL29

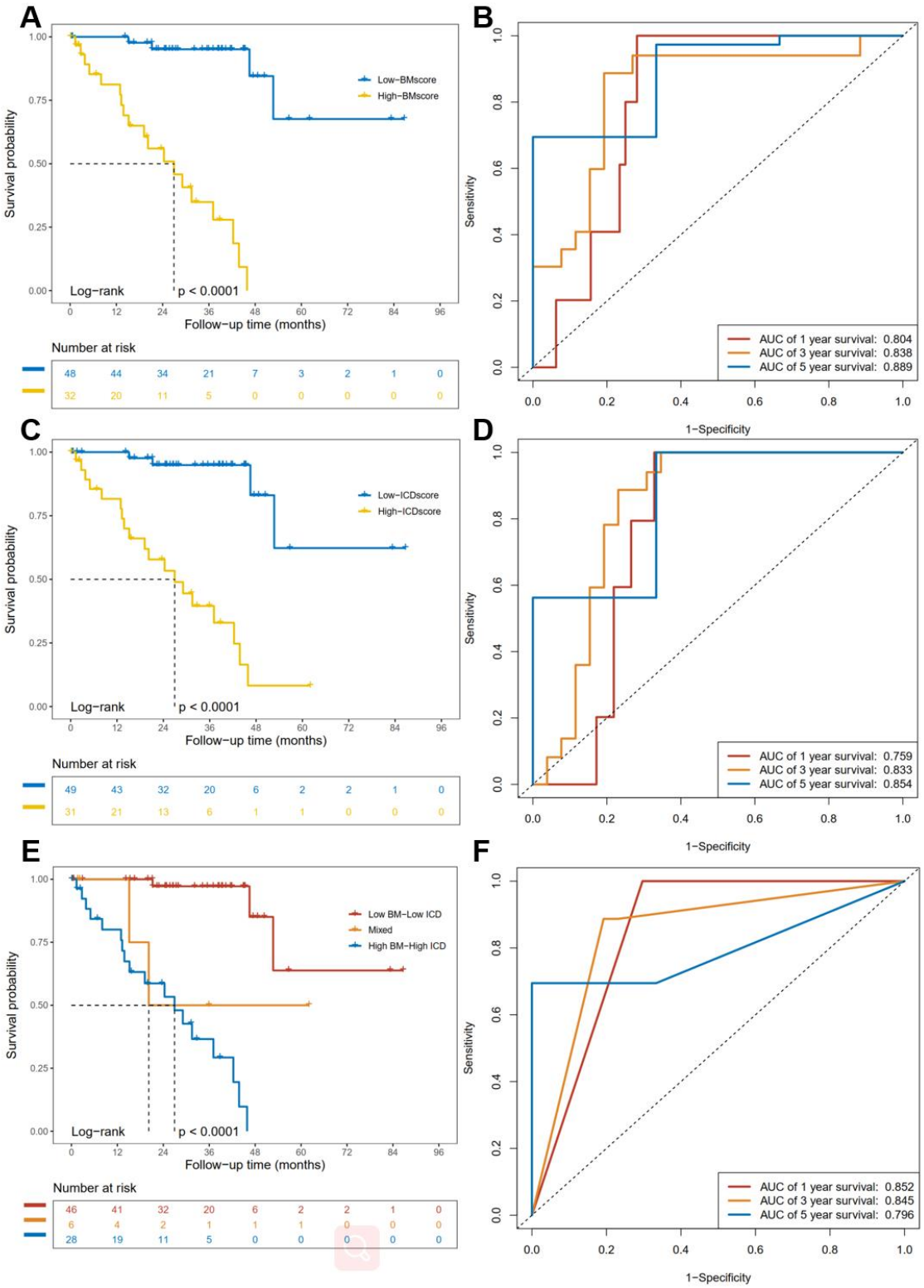


Figure 3.15 Calculation and evaluation of BMscore (A, B), ICDscore (C,D), and BMICD classifier (E, F) based on PCA algorithm in the TCGA-UVM cohort (training set). *(**A, C, E**) K-M curves for the OS of BMscore (**A**), ICDscore (**C**) and BMICD classifier (**E**). (**B, D, F**) ROC curves depicting the AUCs of 1-, 3-, and 5-year OS for BMscore (**B**), ICDscore (**D**) and BMICD classifier (**F**).*

3.3.9 Model selection in the validation set

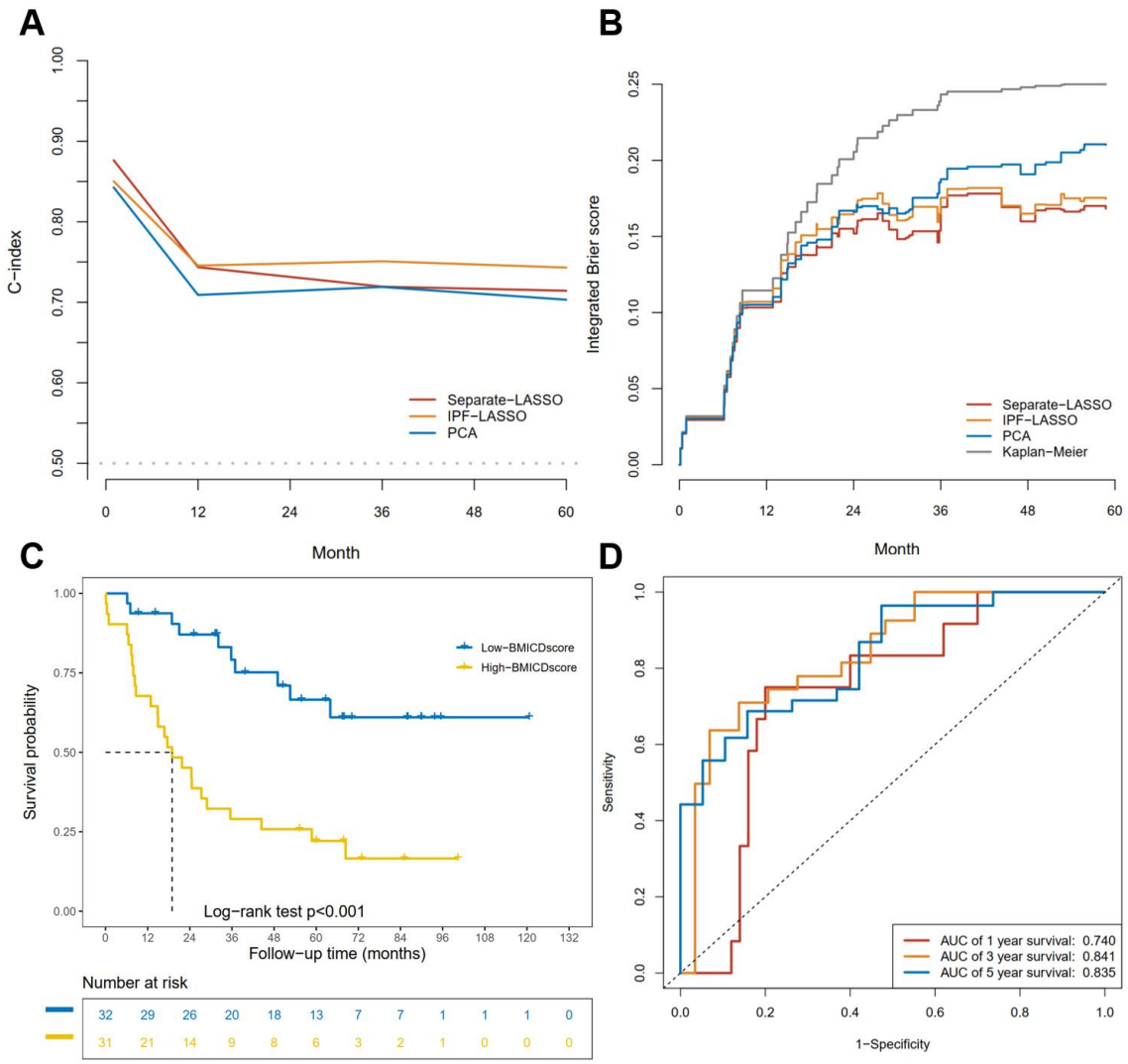


Figure 3.16 Model selection in the validation set (GSE22138). (A) *Curves of C-index up to 60 months for three predictive models.* ***(B)*** *Prediction error curves up to 60 months for three predictive models. Red, yellow and blue line represents the Separate-LASSO algorithm, IPF-LASSO algorithm, and PCA algorithm, respectively, and the black line represents the only K-M curve (null model).* ***(C)*** *K-M curves for the OS of BMICDscore in GSE22138.* ***(D)*** *ROC curves of 1-, 3-, and 5-year OS for BMICDscore in GSE22138.*

After completing three survival predictive models, we assessed and compared them using C-index and integrated Brier score in GSE22138 (validation set). Within one year, the OS predictive performance of Separate-LASSO model was higher than the other two, but the IPF-LASSO model obviously outperformed it after one year (**Figure 3.16A**).

In terms of predictive error, the integrated Brier scores of three models were all markedly lower than the null model. Separate-LASSO seemed to be the one with least prediction error, but the gaps between it and the IPF-LASSO was not significantly clear (**Figure 3.16B**). Given two points aforementioned, we considered the model obtained by IPF-LASSO algorithm with high accuracy and interpretability as the optimal one for our risk score. We further evaluated the predictive power of IPF-LASSO model in GSE22138. Obviously, patients in different risk score group have experienced different survival probabilities (log-rank $p < 0.001$; **Figure 3.16C**), and the BMICDscore predicted the prognosis fairly accurately according to the ROC analysis (1-, 3-, and 5-year OS AUC = 0.740, 0.841, and 0.835, respectively; **Figure 3.16D**).

3.3.10 Testing on independent GEO cohorts

For external validation of the BMICDscore generated by IPF-LASSO algorithm, we tested them on two independent GEO datasets: GSE44295 and GSE39717. In GSE44295, patients with high-BMICDscore experienced poorer survival outcomes (log-rank $p = 0.007$) and there was a similar trend in GSE39717. Moreover, the survival outcome of UVM patients in GSE39717 is metastasis-free survival, which implies that our model could be applied on researches with other survival outcomes beyond the OS (log-rank $p = 0.012$; **Figure 3.17A,C**). ROC curves provided another evidence for the satisfactory performance of our model in both GSE44295 (1-, 3-, and 5-year OS AUC = 0.900, 0.664, and 0.624, respectively) and GSE39717 (1-, 3-, and 5-year MFS AUC = 0.778, 0.799, and 0.871, respectively; **Figure 3.17B,D**).

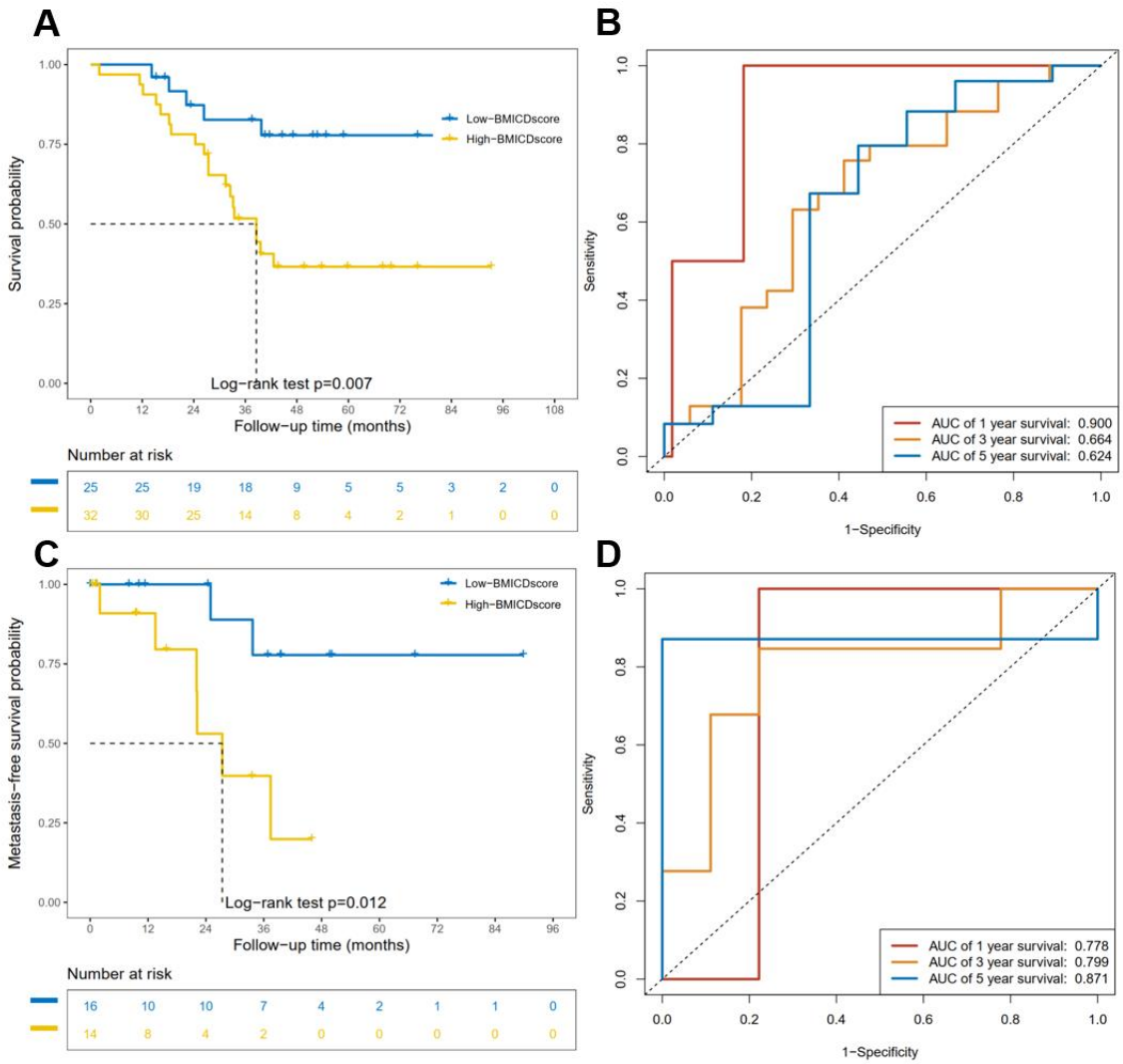


Figure 3.17 External validation of IPF-LASSO model in two test sets. (A) *K-M curves for the OS of BMICDscore in GSE44295.* ***(B)*** *ROC curves of 1-, 3-, and 5-year OS for BMICDscore in GSE44295.* ***(C)*** *K-M curves for the OS of BMICDscore in GSE39717.* ***(D)*** *ROC curves of 1-, 3-, and 5-year OS for BMICDscore in GSE39717.*

3.3.11 Relationship between the BMICDscore and clinical characteristics

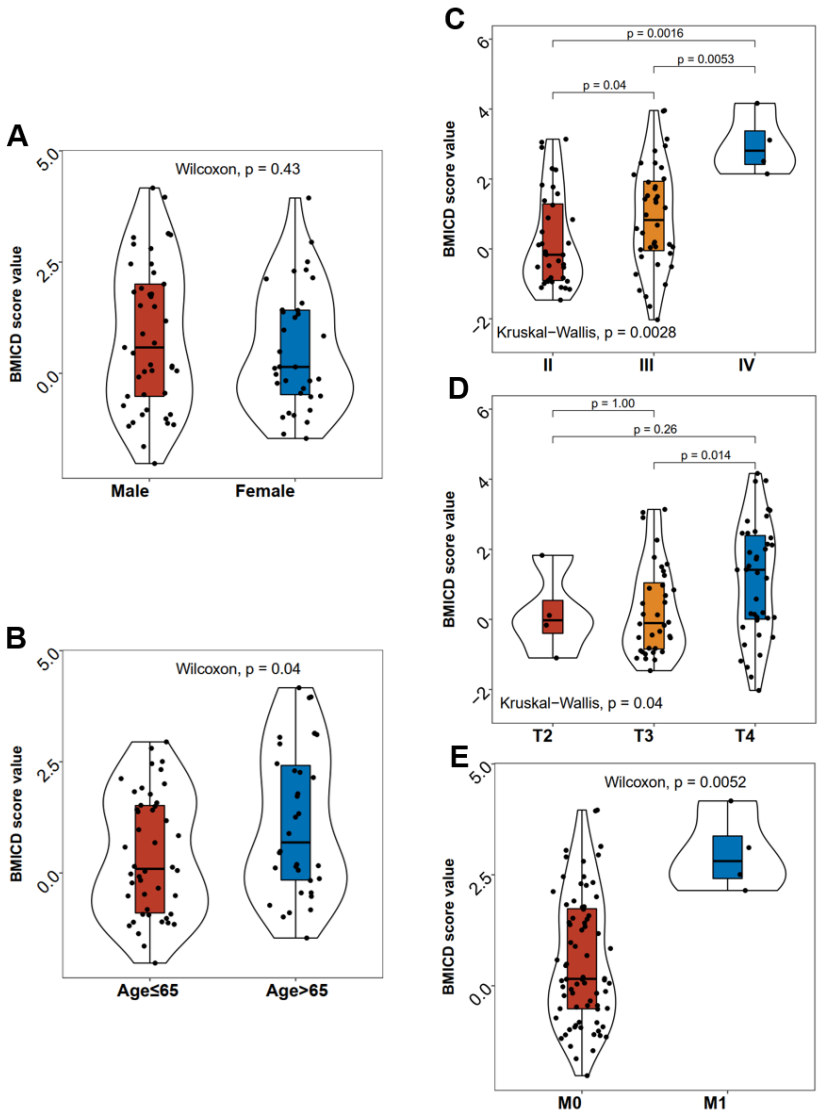


Figure 3.18 The distribution of BMICDscore values with different genders (A), ages (B), clinical stages (C), T stages (D), and M stages (E) in the TCGA-UVM cohort. *Box plots inside violin plots indicate interquartile ranges and median values.*

With violin plots, we portrayed the distributions of the BMICDscore values with different clinical traits in the TCGA-UVM cohort to compare the differences of the BMICDscore in different clinical subgroups. There were no statistically significant differences between males and females (Wilcoxon p = 0.43; **Figure 3.18A**). Nevertheless, greater ages (Wilcoxon p = 0.04), clinical stages (Kruskal-Wallis p = 0.0028), T stages

(Kruskal-Wallis p = 0.04), and M stages (Wilcoxon p = 0.0052) were all linked to a higher BMICDscore (**Figure 3.18B–E**), which indicates that the association between our model and these clinical features became stronger when we ungraded the BMtypes and ICDtypes to the ultimate BMICDscore.

3.3.12 Functional enrichments associated with the BMICDscore

Using GSEA method, we explored the potential biological processes and pathways contributing to the high-risk group. From both “Hallmark” and “KEGG” GSEA plots, we considered those enriched signaling pathways were related to immune response, cytokine and chemokine signaling, cell cycle and proliferation, and cell adhesion. This finding is similar to those results from the GSEA on BMtypes and ICDtypes (**Figure 3.19**).

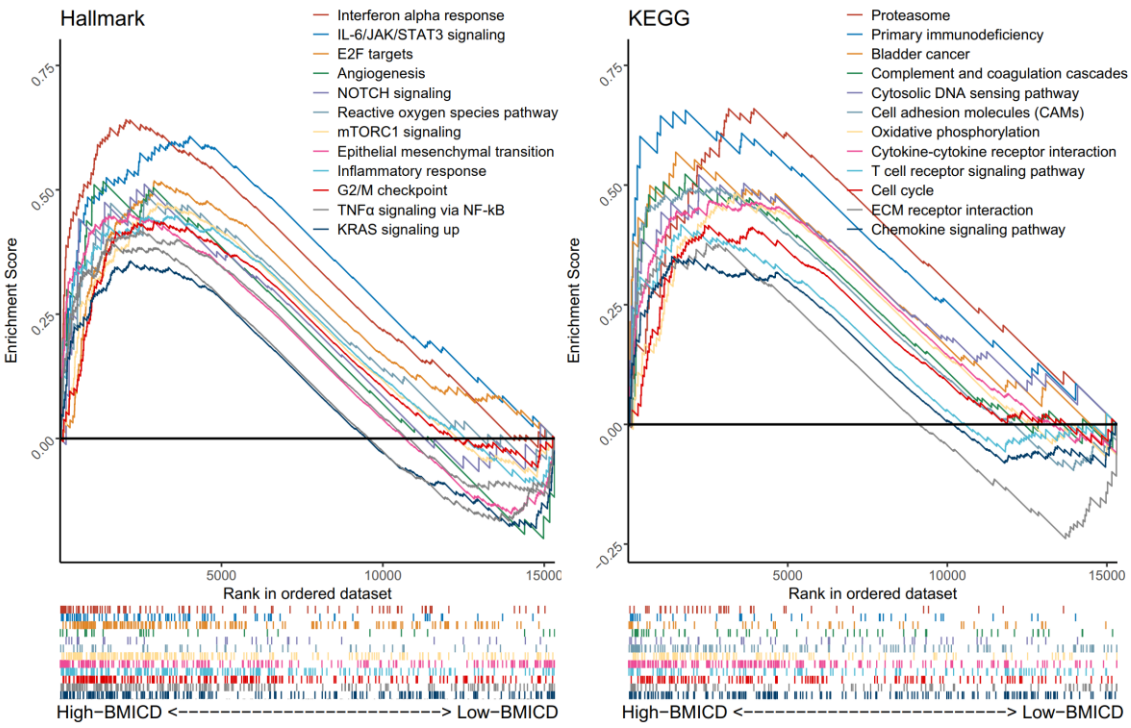


Figure 3.19 GSEA showing signaling pathways and biological processes in TCGA-UVM cohort. *Gene sets were obtained from "Hallmark" and "KEGG" on MSigDB. The top twelve results are displayed in different colors based on their enrichment scores in each figure.*

3.3.13 Association between the BMICDscore and immune landscape in UVM

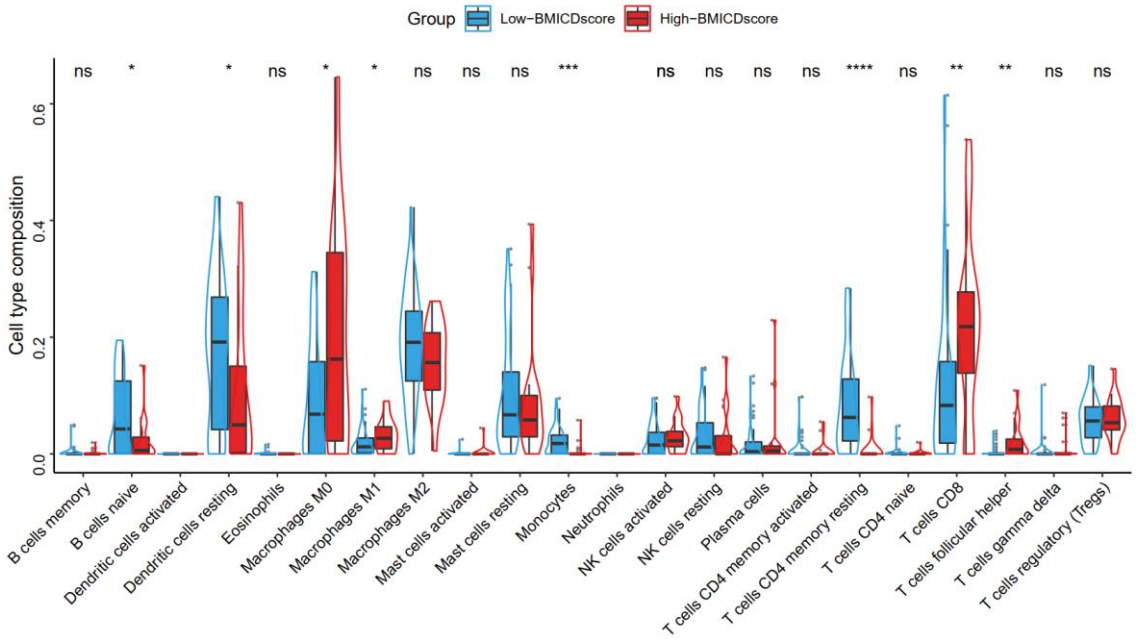


Figure 3.20 Fractions of 22 immune cell types in the different risk groups in the TCGA-UVM cohort. **, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001; ns, not significant.*

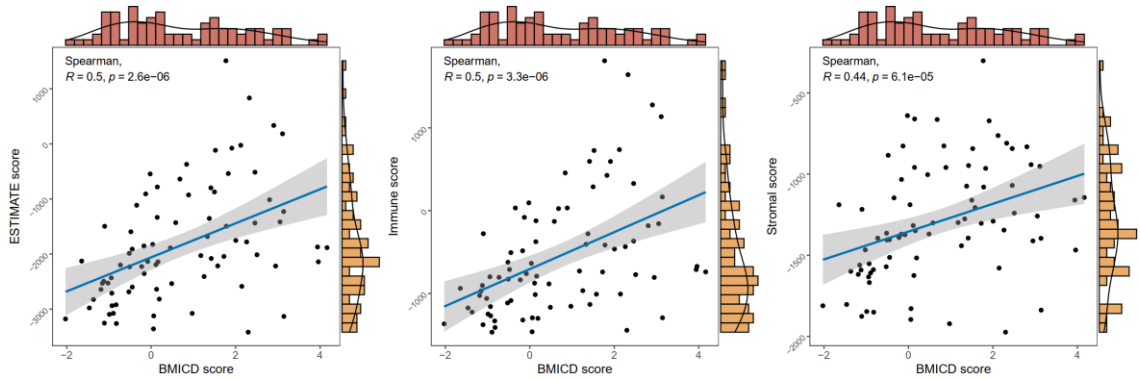


Figure 3.21 Scatter plots for correlations between the BMICDscore and the three immune scores in the TCGA-UVM cohort. *Histograms for distributions of the BMICDscore and corresponding scores are attached on the top and on the right, respectively.*

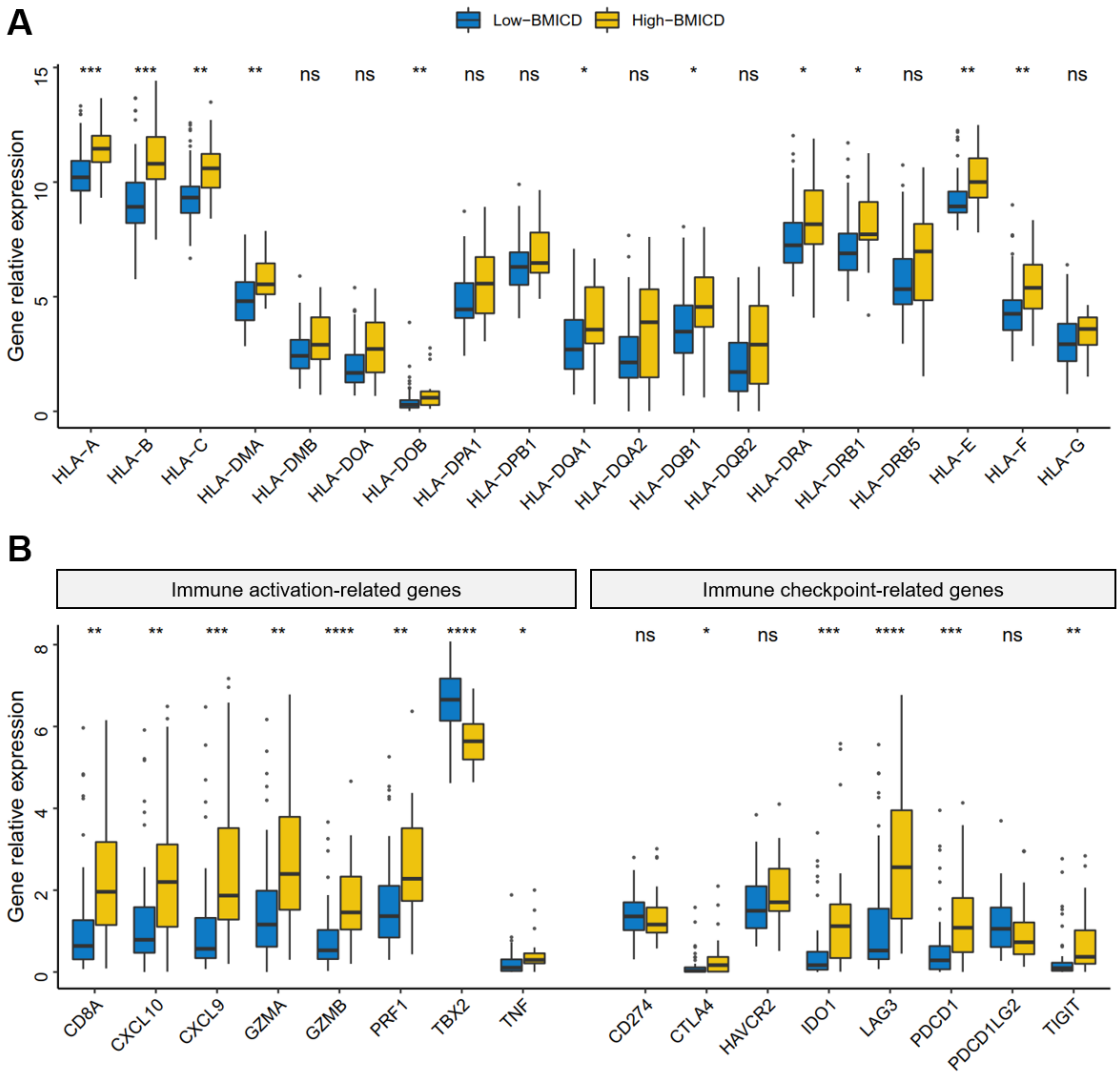


Figure 3.22 Box plots for the expression levels of HLA genes (A) and immune activation—related genes and immune checkpoints (B) in the TCGA-UVM cohort. **, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001; ns, not significant.*

To better investigate the role of the BMICDscore in UVM immune landscape, we performed several analyses. Using the CIBERSORT algorithm, we firstly compared twenty-two immune cell compositions between high- and low-BMICDscore group. Compared to the patients in low-BMICDscore group, those in the high-BMICDscore group exhibited higher infiltration of M0 and M1 macrophages, $CD8^+$ and follicular helper T cells, but lower infiltration of resting dendritic cells, monocytes, naïve B cells, and $CD4^+$ memory resting T cells (**Figure 3.20**).

Next, we exploited the ESTIMATE algorithm to generate three TME-specific scores and analyzed the correlations between the BMICDscore and them. All correlation analyses indicated the positive correlation between the BMICDscore and ESTIMATE score (Spearman r = 0.5, p = 2.6e-6), Immune score (Spearman r = 0.5, p = 3.3e-6) and Stromal score (Spearman r = 0.44, p = 6.1e-5; **Figure 3.21**), which provided evidence for a close association between our model and the previously reported TME scoring system.

Furthermore, human leukocyte antigens (HLAs), immune activation-related factors and immune checkpoints fulfil essential functions in the immune reaction to tumors in the microenvironment. Therefore, we also scrutinized the expression levels of the genes relevant to these molecules in different risk groups. Surprisingly, the results displayed a remarkable consistency among three gene types: the majority of genes in each gene functional type were significantly up-regulated in the high-BMICDscore group. These findings collectively suggest that the BMICDscore is highly correlated with the response to immunotherapy (**Figure 3.22**).

3.3.14 Nomogram establishment and evaluation

Based on the association of BMICDscore and the progression of UVM, we were also interested in the application of BMICDscore in clinical practice. We firstly performed univariate and multivariate Cox regression analyses involving the BMICDscore and other clinicopathological features as the covariates. Our results indicated that the age over 65 (HR (95%CI) = 2.311 (1.008–5.297), p = 0.048), T stage (HR (95%CI) = 2.862 (1.278–6.408), p = 0.011), and the BMICDscore ((HR (95%CI) = 6.426 (3.369–12.257), p <0.001) were the potential prognostic factors for the OS prediction (**Figure 3.23 A,B)**.

By integrating the prognostic factors, we established survival nomograms for 3- and 5-year OS, in order to provide clinically practical quantitative tools for survival probability predictions (**Figure 3.24 A**). Additionally, we evaluated the established nomogram using calibration curves and decisions curve analyses. The calibration curves revealed a good correspondence between predicted and observed outcome values for predicting both 3- and 5-year overall survival probability (**Figure 3.24 B**). Decision curves also indicated that either at 3- or 5-year risk, the benefit of prediction using the nomogram is greater than those of other single factors and their combinations (**Figure 3.24 C,D**)

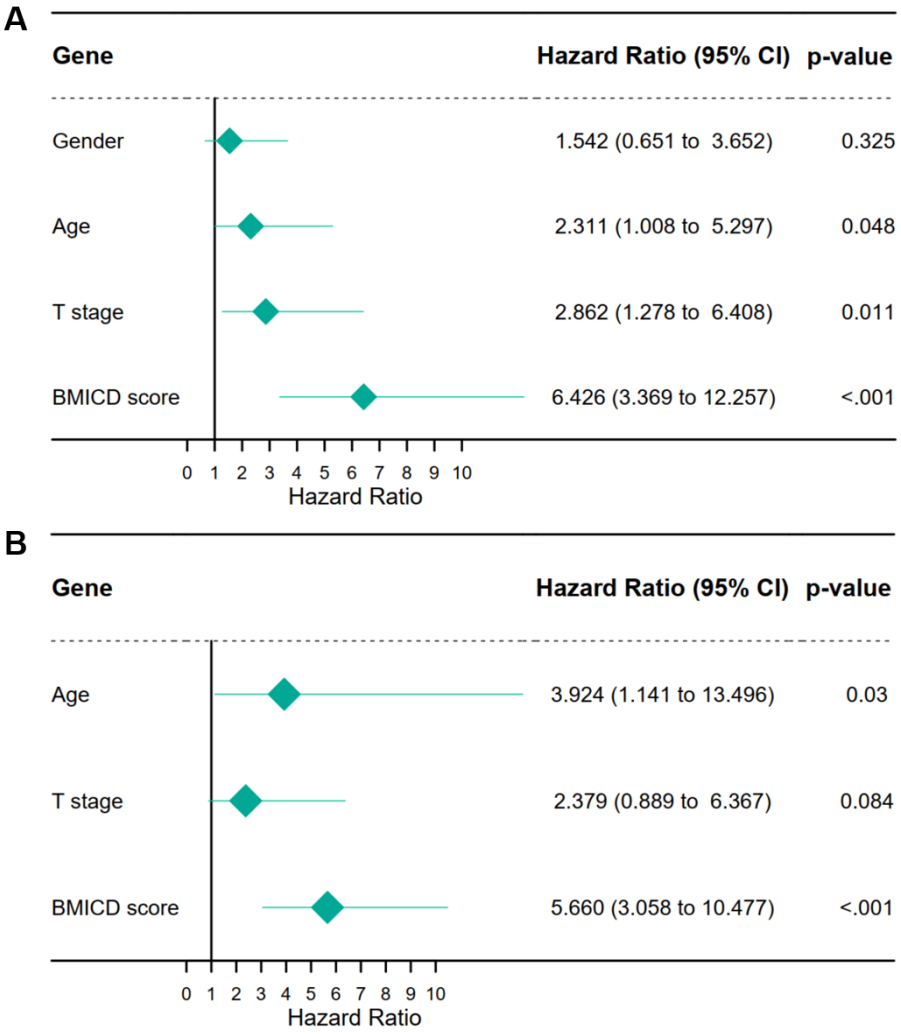
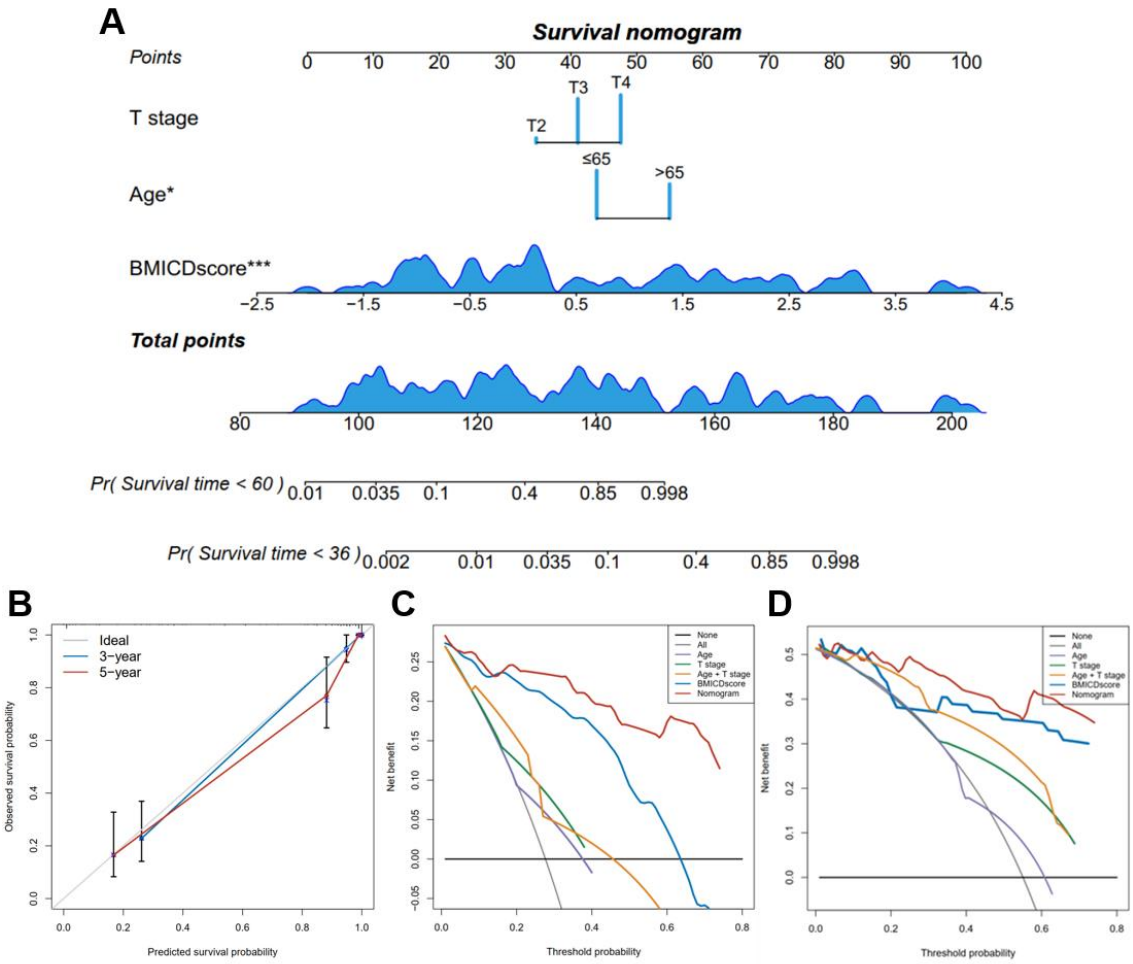


Figure 3.23 Forest plots summarizing the univariate (A) and multivariate (B) Cox regression models of the BMICDscore and clinicopathological characteristics. *The green rhombuses with transverse lines represent the hazard ratios (HRs) and their 95% confidence intervals (95%CIs). There are also detailed information of 95%CI and p value for each prognostic factors on the right side.*

Figure 3.24 Establishment and evaluation of the survival nomograms based on three prognostic factors (on the next page). *(**A**) Nomograms for predicting UVM death probability at 5- or 3-year (presented in 60- or 36-month survival time); ∗, p < 0.05; ∗∗∗, p < 0.001. (**B**) Calibration curves of the corresponding nomograms (Bootstrap repetition, B = 1,000). (**C, D**) Decision curve analyses of the corresponding nomograms based on various combinations of prognosis factors for predicting 3- (**C**) and 5-year (**D**) OS probability.*



3.4 Discussion

Uveal melanoma (UVM) remains an aggressive disease with poor prognostic outcomes [38, 72]. However, predicting the prognosis of UVM individually and accurately remains challenging due to the complexity of its influencing factors [67]. In recent years, cancer research has focused on the tumor microenvironment (TME), which encompasses the surrounding of oncocytes, including infiltrating immune cells and extracellular matrices [81, 87]. Lei and Zhang have discovered the involvement of the TME in the pathogenesis of UVM [86]. Numerous studies have demonstrated the importance of both basement membranes (BMs) [95, 151, 152] and immunogenic cell death (ICD) [99, 153] in tumor invasion and metastasis, as well as their strong interaction with TME. In the present study, we investigated the role of the BM-related and ICD-related genes in UVM, followed by integrating them with machine-learning methods to develop a scoring model for UVM prognosis prediction.

The implementation of unsupervised clustering has identified different UVM subtypes based on BM and ICD patterns. These subtypes varied in prognostic

outcome, TME characteristics, and biological processes. BMtype2 and ICDtype2, where most of the BM-related and ICD-related genes were up-regulated, respectively, exhibited a worse prognosis and a greater enrichment of pathways of cell proliferation, cytokine and chemokine signaling, and cell adhesion. Several reports have shown that in melanoma tumors, molecules of cytokine, chemokine and cell adhesion may play a dual role as both pro- and anti-tumor agents [160-162]. Their contribution to the tumor invasion and migration, together with cell cycle and proliferation pathways, which are essential for tumorigenesis, may be linked to poor clinical outcomes. In contrast to other types of cancer, some investigators have considered the infiltration of $CD8^+$ T cells as an unfavorable prognostic factor in UVM [163-165]. Interestingly, our results from CIBERSORT analysis were in accordance with them. We assumed a possible justification could be that the infiltrating $CD8^+$ T cell subset might be impaired in the TME in UVM and therefore exert an inhibitory effect.

Given the significant interplay between the extracellular matrix and immunogenic cell death in TME, we combined the BMtypes and ICDtypes to create a scoring system for more comprehensive prognosis prediction. Using the differentially expressed genes between these subtypes, we have constructed three models based on different algorithms and selected the optimal one named BMICDscore after comparing them in validation set, which was the BMICDscore generated by IPF-LASSO algorithm (1-, 3-, and 5-year survival AUC = 0.740, 0.841, and 0.835, respectively). This may be benefit from the enablement of integrating BM and ICD patterns by IPF-LASSO through assigning different penalty factors to each gene set [112]. In addition, the BMICDscore performed well in two independent test sets. Moreover, the survival outcome of UVM patients in GSE39717 is metastasis-free survival (MFS), which implies that our model could be applied on studies with other survival outcomes beyond the OS, and its prognostic capability is likely unaffected by gender according to the comparison of the distribution of its values. We consider it a novel prognostic algorithm since it has not been previously reported. Incorporating clinical features including age and different types of tumor stage, we additionally established nomograms to furnish a clinically applicable quantitative tool for survival probability predictions and results have shown its outperformance of these independent clinical prognostic factors as well as the BMICDscore alone.

The results of functional enrichment analysis and TME characterization were consistent with those obtained from the same analyses conducted on BMtypes and ICDtypes. Various in-silico approaches to TME analysis not only provide evidence for the close association between BMICDscore and TME, but also suggest that the BMICDscore may be correlated with the response to immunotherapy. Over the past

decade, there has been a significant increase in attention and investment towards the development of immunotherapies, though clinical research has revealed suboptimal efficacy in UVM treatment [81, 166]. The issue of immunotherapy resistance is multifaceted and extensive research has shed light on the crucial role TME plays in it [167]. In light of the complicacy of TME, we conducted a comprehensive study on the correlation between BMICDscore and various aspects of immune-related functions, beyond just infiltrated immune cells, including Estimate-Stromal-Immune score and the expression of genes associated with HLA, immune checkpoint and immune activation. Results signified the predictive value of BMICDscore for the response to UVM immunotherapy and we speculate that patients with high-BMICDscore would benefit from immunotherapies.

It is still important to note some limitations here. Firstly, the limited sample size may have affected the generalizability of our predictive model despite applying external validations. Future studies with larger sample sizes would be needed for better training and validating the model. Secondly, we have only performed analyses of immune-related functions in an in-silico way. However, bioinformatics approaches may not yield enough conclusive evidence, and therefore, these results require experimental verification. Thirdly, we failed to obtain clinical prognostic parameters previously reported, such as radiological data and chromosomal abnormality status, which could be incorporate to our model and improve its power and robustness. Fourthly, due to the lack of another dataset with the clinical features used in our nomogram, we have only assessed it in the training set. Consequently, subsequent researches with sufficient clinical information and sample sizes remain necessary. Last but not least, although we have preliminarily explored the relationship between the BMICDscore and those immunotherapeutic markers, the lack of a dataset with information on patients who have received UVM immunotherapy prevents us from making predictions and evaluations of response to immunotherapy. Hence, we plan to take the next step to conduct the investigation which take these aspects into account.

3.5 Conclusion

In summary, this study characterized the patterns of BM and ICD in UVM separately, and subsequently combined them to generate the BMICDscore that can be served for prognostic prediction and stratified clinical management of UVM patients. It may also help physicians to predict immunotherapeutic response and potentially, develop novel immunotherapies. However, future studies with large sample size and multiple information are still necessary to further optimize and validate the model.

4. Overall Discussion

The aim of this Ph.D. project was to explore the development of diagnostic and prognostic approaches for immune-related disease based on bioinformatical methods. To this end, we characterized the immune microenvironment and established two independent predictive models for Kawasaki disease (KD) and uveal melanoma (UVM), separately. We initially enumerated the compositions of immune cell types using CIBERSORT algorithm in both two diseases. Then, for KD, a diagnostic score named DIS was generated utilizing the combination of CIBERSORT-derived immune cell fractions and machine-learning techniques; for UVM, a prognostic model named BMICDscore was constructed by applying IPF-LASSO algorithm on the incorporation of BM-related and ICD-related genes.

KD is an acute inflammatory illness while UVM is a melanogenic eye tumor, and both are highly immune-related. As a crucial role in immune-related diseases, the immune microenvironment significantly influences their disease progression and treatment outcomes. Moreover, understanding the dynamic relationship between tumor cells and the tumor immune microenvironment will be valuable in devising new therapies and in understanding the pathophysiological mechanisms of tumor growth and immune evasion [168]. Advanced technologies, such as machine learning, have been used to study the microenvironment and its impact on diseases [7, 169, 170]. One such technology is CIBERSORT, a computational approach that can estimate cellular components of complex tissues. By using CIBERSORT, we can characterize and even quantify the immune microenvironment with fractions of specific immune cell subsets [7]. According to existing reports, the immune microenvironment in KD is characterized by an overactive innate immune system, with increased infiltration levels of monocyte, macrophage and neutrophil [24, 53, 54]. On the other hand, UVM exhibits multiple signals in an inflammatory microenvironment initiated by inflammatory factors, such as HLAs, cytokines, and chemokines, which lead in recruitment of inflammation-related cells and promotion of angiogenesis [55-57], and this pattern has also been found in hepatic metastasis lesion [171]. Results in both our studies corroborated the findings of these work.

Based on these observations, we attempted to develop two predictive models. In Chapter 2, we leveraged these immune cell fractions to construct the DIS for KD diagnostic prediction, while in the modeling in Chapter 3, we also took both tumor invasiveness and response to immunotherapy into account, so for the BMICDscore, we additionally incorporated the patterns of BM and ICD-related genes. In Chapter 2, we

merged multiple datasets as the entire discovery cohort before splitting it into a training set and a held-out test set, and utilized an additional GEO dataset for external validation. ROC curves mirrored the robust performance of DIS, with AUC values of 0.80 for the held-out test set and 0.77 for the external validation set. In Chapter 3, we tested and compared three modeling algorithms to optimally leverage the two gene sets. By applying the three fitted models to the validation set, the BMICDscore generated by the IPF-LASSO algorithm was ultimately determined as the best-performing algorithm (1-, 3-, and 5-year survival AUC = 0.740, 0.841, and 0.835, respectively). In addition, we analyzed two additional GEO datasets with different outcomes, namely overall survival and metastasis-free survival, to assess our model's capability for generalization and application on other prognostic outcomes. ROC analyses in two datasets also supported the strong predictive power of BMICDscore, with not only overall survival but also metastasis-free survival. Furthermore, we also successfully established two nomograms based on our predictive scores, separately, to further facilitate their application in clinical practice. With regards to the respond to UVM immunotherapy, we preliminarily investigated the association between its previously reported markers and our BMICDscore. According to the multiple results of immunotherapy-related scrutiny, we speculate that patients with high-BMICDscore would benefit from immunotherapies. Future prospective studies on patients who have received UVM immunotherapy, however, are still needed to further improve and validate our model. Last but not least, considering that KD and UVM are closely immune-related inflammatory diseases and tumors respectively, we believe it is worthwhile to explore the possibility of extending these two models to other autoimmune and cancerous conditions with similar characteristics.

As discussed in **Section 2.4** and **Section 3.4,** we acknowledge some limitations in this doctoral project. Apart from those limitations of disease-specific data that have been mentioned in the corresponding chapters, here, we highlight some common limitations we have met when modeling. First of all, the generalizability of our predictive models may have been influenced by the limited sample size of available datasets that we can obtain publicly. Thus, it would be important to conduct further investigations with larger sample sizes to optimize and validate our models. Secondly, it is possible for batch effects to arise during the combination of various datasets. To reduce this danger, we should try our best to find datasets in similar platforms, and implement approaches of batch-effect correction such as the "ComBat" algorithm. Thirdly, demographic and clinical information for every participant is challenging to obtain and usually incomplete on publicly available databases. However, incorporating these features, especially the response to immunotherapy, may effectively improve the power

and robustness of our models. Finally, we conducted all analyses about gene expressions and microenvironment by only using in-silico techniques that are difficult to reflect the exact and real biological situation. Therefore, laboratory experiments are required to be involved in our next studies for biological verification.

5. Conclusion and Outlook

Taken together, focusing on the intersection between biomedicine and informatics, this doctoral project has investigated the characterization of immune microenvironment in Kawasaki disease (KD) and uveal melanoma (UVM), and the development of innovative predictive algorithms for KD diagnosis and UVM prognosis based on their compositions of immune cell subsets, respectively. These models may offer significant contributions to the field of biomedical research and clinical management. On the one hand, they can help us comprehensively understand the complex immunological dynamics in these immune-related diseases. On the other hand, they can provide guidance in making clinical surveillance and treatment decisions for KD and UVM, and yield novel insights for the exploitation of immune microenvironment for clinical application. Looking ahead, it would be interesting to obtain multi-center data with a larger sample size and sufficient clinical information specific to the diagnosis of KD and the prognosis of UVM, as well as to carry out laboratory experiments, in order to significantly optimize and comprehensively validate our models. Moreover, studies on the association of our models with response to immunotherapy are needed to explore their potential to predict immunotherapeutic efficacy and potentially, to pave a way for novel immunotherapy development. In addition, given the immune-related characteristics of KD and UVM, we believe that extending these models to other autoimmune and cancerous conditions with similar features is a promising avenue worth exploring.

References

1. Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994 Sep 30;265(5181):2037-48.

2. Muller B, Wilcke A, Boulesteix AL, Brauer J, Passarge E, Boltze J, Kirsten H. Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives. Hum Genet. 2016 Mar;135(3):259-72.

3. Oh SM, Stefani KM, Kim HC. Development and application of chronic disease risk prediction models. Yonsei Med J. 2014 Jul;55(4):853-60.

4. Albert-Vega C, Tawfik DM, Trouillet-Assant S, Vachot L, Mallet F, Textoris J. Immune Functional Assays, From Custom to Standardized Tests for Precision Medicine. Front Immunol. 2018;9:2367.

5. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek S, Carpenter AE, Shrikumar A, Xu J, Cofer EM, Lavender CA, Turaga SC, Alexandari AM, Lu Z, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Boca SM, Swamidass SJ, Huang A, Gitter A, Greene CS. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018 Apr;15(141).

6. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, Morris Q, Barash Y, Krainer AR, Jojic N, Scherer SW, Blencowe BJ, Frey BJ. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science. 2015 Jan 9;347(6218):1254806.

7. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 May;12(5):453-7.

8. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.

9. Kawasaki T. [Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children]. Arerugi. 1967 Mar;16(3):178-222.

10. Burns JC, Glodé MP. Kawasaki syndrome. The Lancet. 2004 2004/08;364(9433):533-44.

11. Makino N, Nakamura Y, Yashiro M, Ae R, Tsuboi S, Aoyama Y, Kojo T, Uehara R, Kotani K, Yanagawa H. Descriptive epidemiology of Kawasaki disease in Japan, 2011-2012: from the results of the 22nd nationwide survey. J Epidemiol. 2015;25(3):239-45.

12. Rowley AH, Shulman ST. The Epidemiology and Pathogenesis of Kawasaki Disease. Front Pediatr. 2018;6:374.

13. Singh S, Vignesh P, Burgner D. The epidemiology of Kawasaki disease: a global update. Arch Dis Child. 2015 Nov;100(11):1084-8.

14. Juan CC, Hwang B, Lee PC, Lin YJ, Chien JC, Lee HY, Meng CC. The clinical manifestations and risk factors of a delayed diagnosis of Kawasaki disease. J Chin Med Assoc. 2007 Sep;70(9):374-9.

15. McCrindle BW, Rowley AH, Newburger JW, Burns JC, Bolger AF, Gewitz M, Baker AL, Jackson MA, Takahashi M, Shah PB, Kobayashi T, Wu M-H, Saji TT, Pahl E. Diagnosis,

Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association. Circulation. 2017 2017/04/25;135(17).

16. Singh S, Sharma A, Jiao F. Kawasaki Disease: Issues in Diagnosis and Treatment--A Developing Country Perspective. Indian J Pediatr. 2016 Feb;83(2):140-5.

17. Weiss PF. Pediatric vasculitis. Pediatr Clin North Am. 2012 Apr;59(2):407-23.

18. Marginean CO, Melit LE, Marginean MO. The peculiarities of Kawasaki disease at the extremes of age: Two case reports. Medicine (Baltimore). 2019 Oct;98(42):e17595.

19. Bayers S, Shulman ST, Paller AS. Kawasaki disease: part II. Complications and treatment. J Am Acad Dermatol. 2013 Oct;69(4):513 e1-8; quiz 21-2.

20. Hara T, Yamamura K, Sakai Y. The up-to-date pathophysiology of Kawasaki disease. Clin Transl Immunology. 2021;10(5):e1284.

21. Group JCSJW. Guidelines for diagnosis and management of cardiovascular sequelae in Kawasaki disease (JCS 2013). Digest version. Circ J. 2014;78(10):2521-62.

22. Takahashi K, Oharaseki T, Yokouchi Y. Pathogenesis of Kawasaki disease. Clin Exp Immunol. 2011 May;164 Suppl 1(Suppl 1):20-2.

23. Noval Rivas M, Arditi M. Kawasaki disease: pathophysiology and insights from mouse models. Nat Rev Rheumatol. 2020 Jul;16(7):391-405.

24. Hara T, Nakashima Y, Sakai Y, Nishio H, Motomura Y, Yamasaki S. Kawasaki disease: a matter of innate immunity. Clin Exp Immunol. 2016 Nov;186(2):134-43.

25. Ikeda K, Yamaguchi K, Tanaka T, Mizuno Y, Hijikata A, Ohara O, Takada H, Kusuhara K, Hara T. Unique activation status of peripheral blood mononuclear cells at acute phase of Kawasaki disease. Clin Exp Immunol. 2010 May;160(2):246-55.

26. Biezeveld MH, van Mierlo G, Lutter R, Kuipers IM, Dekker T, Hack CE, Newburger JW, Kuijpers TW. Sustained activation of neutrophils in the course of Kawasaki disease: an association with matrix metalloproteinases. Clin Exp Immunol. 2005 Jul;141(1):183-8.

27. Hoang LT, Shimizu C, Ling L, Naim AN, Khor CC, Tremoulet AH, Wright V, Levin M, Hibberd ML, Burns JC. Global gene expression profiling identifies new therapeutic targets in acute Kawasaki disease. Genome Med. 2014;6(11):541.

28. Alphonse MP, Duong TT, Shumitzu C, Hoang TL, McCrindle BW, Franco A, Schurmans S, Philpott DJ, Hibberd ML, Burns J, Kuijpers TW, Yeung RS. Inositol-Triphosphate 3-Kinase C Mediates Inflammasome Activation and Treatment Response in Kawasaki Disease. J Immunol. 2016 Nov 1;197(9):3481-9.

29. Kusuda T, Nakashima Y, Murata K, Kanno S, Nishio H, Saito M, Tanaka T, Yamamura K, Sakai Y, Takada H, Miyamoto T, Mizuno Y, Ouchi K, Waki K, Hara T. Kawasaki disease-specific molecules in the sera are linked to microbe-associated molecular patterns in the biofilms. PLoS One. 2014;9(11):e113054.

30. Saguil A, Fargo M, Grogan S. Diagnosis and management of kawasaki disease. Am Fam Physician. 2015 Mar 15;91(6):365-71.

31. Seo YM, Kang HM, Lee SC, Yu JW, Kil HR, Rhim JW, Han JW, Lee KY. Clinical implications in laboratory parameter values in acute Kawasaki disease for early diagnosis and proper treatment. Korean J Pediatr. 2018 May;61(5):160-6.

32. Ruan Y, Ye B, Zhao X. Clinical characteristics of Kawasaki syndrome and the risk factors for coronary artery lesions in China. Pediatr Infect Dis J. 2013 Oct;32(10):e397-402.

33. Ha KS, Jang G, Lee J, Lee K, Hong Y, Son C, Lee J. Incomplete clinical manifestation as a risk factor for coronary artery abnormalities in Kawasaki disease: a meta-analysis. Eur J Pediatr. 2013 Mar;172(3):343-9.

34. Jaggi P, Kajon AE, Mejias A, Ramilo O, Leber A. Human adenovirus infection in Kawasaki disease: a confounding bystander? Clin Infect Dis. 2013 Jan;56(1):58-64.

35. Wright VJ, Herberg JA, Kaforou M, Shimizu C, Eleftherohorinou H, Shailes H, Barendregt AM, Menikou S, Gormley S, Berk M, Hoang LT, Tremoulet AH, Kanegaye JT, Coin LJM, Glode MP, Hibberd M, Kuijpers TW, Hoggart CJ, Burns JC, Levin M, Immunopathology of Respiratory I, Infectious Disease Study C, the Pediatric Emergency Medicine Kawasaki Disease Research G. Diagnosis of Kawasaki Disease Using a Minimal Whole-Blood Gene Expression Signature. JAMA Pediatr. 2018 Oct 1;172(10):e182293.

36. Wang H, Huang Z, Zhang D, Arief J, Lyu T, Tian J. Integrating Co-Clustering and Interpretable Machine Learning for the Prediction of Intravenous Immunoglobulin Resistance in Kawasaki Disease. IEEE Access. 2020;8:97064-71.

37. Jaggi P, Mejias A, Xu Z, Yin H, Moore-Clingenpeel M, Smith B, Burns JC, Tremoulet AH, Jordan-Villegas A, Chaussabel D, Texter K, Pascual V, Ramilo O. Whole blood transcriptional profiles as a prognostic tool in complete and incomplete Kawasaki Disease. PLoS One. 2018;13(5):e0197858.

38. Jager MJ, Shields CL, Cebulla CM, Abdel-Rahman MH, Grossniklaus HE, Stern MH, Carvajal RD, Belfort RN, Jia R, Shields JA, Damato BE. Uveal melanoma. Nat Rev Dis Primers. 2020 Apr 9;6(1):24.

39. Rossi E, Croce M, Reggiani F, Schinzari G, Ambrosio M, Gangemi R, Tortora G, Pfeffer U, Amaro A. Uveal Melanoma Metastasis. Cancers (Basel). 2021 Nov 13;13(22).

40. Kaliki S, Shields CL. Uveal melanoma: relatively rare but deadly cancer. Eye (Lond). 2017 Feb;31(2):241-57.

41. Xu Y, Lou L, Wang Y, Miao Q, Jin K, Chen M, Ye J. Epidemiological Study of Uveal Melanoma from US Surveillance, Epidemiology, and End Results Program (2010-2015). J Ophthalmol. 2020;2020:3614039.

42. Houtzagers LE, Wierenga APA, Ruys AAM, Luyten GPM, Jager MJ. Iris Colour and the Risk of Developing Uveal Melanoma. Int J Mol Sci. 2020 Sep 28;21(19).

43. Krantz BA, Dave N, Komatsubara KM, Marr BP, Carvajal RD. Uveal melanoma: epidemiology, etiology, and treatment of primary disease. Clin Ophthalmol. 2017;11:279-89.

44. Branisteanu DC, Bogdanici CM, Branisteanu DE, Maranduca MA, Zemba M, Balta F, Branisteanu CI, Moraru AD. Uveal melanoma diagnosis and current treatment options (Review). Exp Ther Med. 2021 Dec;22(6):1428.

45. Dono M, Angelini G, Cecconi M, Amaro A, Esposito AI, Mirisola V, Maric I, Lanza F, Nasciuti F, Viaggi S, Gualco M, Bandelloni R, Truini M, Coviello DA, Zupo S, Mosci C, Pfeffer U. Mutation frequencies of GNAQ, GNA11, BAP1, SF3B1, EIF1AX and TERT in uveal melanoma: detection of an activating mutation in the TERT gene promoter in a single case of uveal melanoma. Br J Cancer. 2014 Feb 18;110(4):1058-65.

46. Ewens KG, Kanetsky PA, Richards-Yutz J, Purrazzella J, Shields CL, Ganguly T, Ganguly A. Chromosome 3 status combined with BAP1 and EIF1AX mutation profiles are associated with metastasis in uveal melanoma. Invest Ophthalmol Vis Sci. 2014 Jun 26;55(8):5160-7.

47. Staby KM, Gravdal K, Mork SJ, Heegaard S, Vintermyr OK, Krohn J. Prognostic impact of chromosomal aberrations and GNAQ, GNA11 and BAP1 mutations in uveal melanoma. Acta Ophthalmol. 2018 Feb;96(1):31-8.

48. Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA, Council ML, Matatall KA, Helms C, Bowcock AM. Frequent mutation of BAP1 in metastasizing uveal melanomas. Science. 2010 Dec 3;330(6009):1410-3.

49. Angi M, Kalirai H, Prendergast S, Simpson D, Hammond DE, Madigan MC, Beynon RJ, Coupland SE. In-depth proteomic profiling of the uveal melanoma secretome. Oncotarget. 2016 Aug 2;7(31):49623-35.

50. Populo H, Soares P, Rocha AS, Silva P, Lopes JM. Evaluation of the mTOR pathway in ocular (uvea and conjunctiva) melanoma. Melanoma Res. 2010 Apr;20(2):107-17.

51. Vader MJC, Madigan MC, Versluis M, Suleiman HM, Gezgin G, Gruis NA, Out-Luiting JJ, Bergman W, Verdijk RM, Jager MJ, van der Velden PA. GNAQ and GNA11 mutations and downstream YAP activation in choroidal nevi. Br J Cancer. 2017 Sep 5;117(6):884-7.

52. Field MG, Decatur CL, Kurtenbach S, Gezgin G, van der Velden PA, Jager MJ, Kozak KN, Harbour JW. PRAME as an Independent Biomarker for Metastasis in Uveal Melanoma. Clin Cancer Res. 2016 Mar 1;22(5):1234-42.

53. Ding Y, Li G, Xiong LJ, Yin W, Liu J, Liu F, Wang RG, Xia K, Zhang SL, Zhao L. Profiles of responses of immunological factors to different subtypes of Kawasaki disease. BMC Musculoskelet Disord. 2015 Oct 23;16:315.

54. Geng Z, Tao Y, Zheng F, Wu L, Wang Y, Wang Y, Sun Y, Fu S, Wang W, Xie C, Zhang Y, Gong F. Altered Monocyte Subsets in Kawasaki Disease Revealed by Single-cell RNA-Sequencing. J Inflamm Res. 2021;14:885-96.

55. Basile MS, Mazzon E, Fagone P, Longo A, Russo A, Fallico M, Bonfiglio V, Nicoletti F, Avitabile T, Reibaldi M. Immunobiology of Uveal Melanoma: State of the Art and Therapeutic Targets. Front Oncol. 2019;9:1145.

56. Bronkhorst IH, Jager MJ. Uveal melanoma: the inflammatory microenvironment. J Innate Immun. 2012;4(5-6):454-62.

57. de Waard-Siebinga I, Hilders CG, Hansen BE, van Delft JL, Jager MJ. HLA expression and tumor-infiltrating immune cells in uveal melanoma. Graefes Arch Clin Exp Ophthalmol. 1996 Jan;234(1):34-42.

58. Lorenzo D, Piulats JM, Ochoa M, Arias L, Gutierrez C, Catala J, Cobos E, Garcia-Bru P, Dias B, Padron-Perez N, Caminal JM. Clinical predictors of survival in metastatic uveal melanoma. Jpn J Ophthalmol. 2019 Mar;63(2):197-209.

59. Singh AD, Topham A. Survival rates with uveal melanoma in the United States: 1973-1997. Ophthalmology. 2003 May;110(5):962-5.

60. Shields CL, Furuta M, Thangappan A, Nagori S, Mashayekhi A, Lally DR, Kelly CC, Rudich DS, Nagori AV, Wakade OA, Mehta S, Forte L, Long A, Dellacava EF, Kaplan B, Shields JA. Metastasis of uveal melanoma millimeter-by-millimeter in 8033 consecutive eyes. Arch Ophthalmol. 2009 Aug;127(8):989-98.

61. Rietschel P, Panageas KS, Hanlon C, Patel A, Abramson DH, Chapman PB. Variates of survival in metastatic uveal melanoma. J Clin Oncol. 2005 Nov 1;23(31):8076-80.

62. Zloto O, Pe'er J, Frenkel S. Gender differences in clinical presentation and prognosis of uveal melanoma. Invest Ophthalmol Vis Sci. 2013 Jan 23;54(1):652-6.

63. Kodjikian L, Grange JD, Baldo S, Baillif S, Garweg JG, Rivoire M. Prognostic factors of liver metastases from uveal melanoma. Graefes Arch Clin Exp Ophthalmol. 2005 Oct;243(10):985-93.

64. Valpione S, Moser JC, Parrozzani R, Bazzi M, Mansfield AS, Mocellin S, Pigozzo J, Midena E, Markovic SN, Aliberti C, Campana LG, Chiarion-Sileni V. Development and external validation of a prognostic nomogram for metastatic uveal melanoma. PLoS One. 2015;10(3):e0120181.

65. Shields CL, Kaliki S, Furuta M, Fulco E, Alarcon C, Shields JA. American Joint Committee on Cancer classification of posterior uveal melanoma (tumor size category) predicts prognosis in 7731 patients. Ophthalmology. 2013 Oct;120(10):2066-71.

66. Prescher G, Bornfeld N, Hirche H, Horsthemke B, Jockel KH, Becher R. Prognostic implications of monosomy 3 in uveal melanoma. Lancet. 1996 May 4;347(9010):1222-5.

67. Kaliki S, Shields CL, Shields JA. Uveal melanoma: estimating prognosis. Indian J Ophthalmol. 2015 Feb;63(2):93-102.

68. Xu Y, Han W, Xu WH, Wang Y, Yang XL, Nie HL, Yao J, Shen GL, Zhang XF. Identification of differentially expressed genes and functional annotations associated with metastases of the uveal melanoma. J Cell Biochem. 2019 Nov;120(11):19202-14.

69. Detorakis ET, Engstrom RE, Jr., Wallace R, Straatsma BR. Iris and anterior chamber angle neovascularization after iodine 125 brachytherapy for uveal melanoma. Ophthalmology. 2005 Mar;112(3):505-10.

70. Singh M, Durairaj P, Yeung J. Uveal Melanoma: A Review of the Literature. Oncol Ther. 2018 Jun;6(1):87-104.

71. Carvajal RD, Piperno-Neumann S, Kapiteijn E, Chapman PB, Frank S, Joshua AM, Piulats JM, Wolter P, Cocquyt V, Chmielowski B, Evans TRJ, Gastaud L, Linette G, Berking C, Schachter J, Rodrigues MJ, Shoushtari AN, Clemett D, Ghiorghiu D, Mariani G, Spratt S, Lovick S, Barker P, Kilgour E, Lai Z, Schwartz GK, Nathan P. Selumetinib in Combination With Dacarbazine in Patients With Metastatic Uveal Melanoma: A Phase III, Multicenter, Randomized Trial (SUMIT). J Clin Oncol. 2018 Apr 20;36(12):1232-9.

72. Carvajal RD, Schwartz GK, Tezel T, Marr B, Francis JH, Nathan PD. Metastatic disease from uveal melanoma: treatment options and future prospects. Br J Ophthalmol. 2017 Jan;101(1):38-44.

73. Augsburger JJ, Correa ZM, Shaikh AH. Effectiveness of treatments for metastatic uveal melanoma. Am J Ophthalmol. 2009 Jul;148(1):119-27.

74. Rossi E, Pagliara MM, Orteschi D, Dosa T, Sammarco MG, Caputo CG, Petrone G, Rindi G, Zollino M, Blasi MA, Cassano A, Bria E, Tortora G, Schinzari G. Pembrolizumab as first-line treatment for metastatic uveal melanoma. Cancer Immunol Immunother. 2019 Jul;68(7):1179-85.

75. Carvajal RD, Sosman JA, Quevedo JF, Milhem MM, Joshua AM, Kudchadkar RR, Linette GP, Gajewski TF, Lutzky J, Lawson DH, Lao CD, Flynn PJ, Albertini MR, Sato T, Lewis K, Doyle A, Ancell K, Panageas KS, Bluth M, Hedvat C, Erinjeri J, Ambrosini G, Marr B, Abramson DH, Dickson MA, Wolchok JD, Chapman PB, Schwartz GK. Effect of selumetinib vs chemotherapy on progression-free survival in uveal melanoma: a randomized clinical trial. JAMA. 2014 Jun 18;311(23):2397-405.

76. Graziani G, Tentori L, Navarra P. Ipilimumab: a novel immunostimulatory monoclonal antibody for the treatment of cancer. Pharmacol Res. 2012 Jan;65(1):9-22.

77. Luke JJ, Callahan MK, Postow MA, Romano E, Ramaiya N, Bluth M, Giobbie-Hurder A, Lawrence DP, Ibrahim N, Ott PA, Flaherty KT, Sullivan RJ, Harding JJ, D'Angelo S,

Dickson M, Schwartz GK, Chapman PB, Wolchok JD, Hodi FS, Carvajal RD. Clinical activity of ipilimumab for metastatic uveal melanoma: a retrospective review of the Dana-Farber Cancer Institute, Massachusetts General Hospital, Memorial Sloan-Kettering Cancer Center, and University Hospital of Lausanne experience. Cancer. 2013 Oct 15;119(20):3687-95.

78. Tsai KK, Zarzoso I, Daud AI. PD-1 and PD-L1 antibodies for melanoma. Hum Vaccin Immunother. 2014;10(11):3111-6.

79. Kim DW, Anderson J, Patel SP. Immunotherapy for uveal melanoma. Melanoma Manag. 2016 Jun;3(2):125-35.

80. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. Nat Rev Immunol. 2015 Nov;15(11):669-82.

81. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017 Feb 28;387:61-8.

82. Hao NB, Lu MH, Fan YH, Cao YL, Zhang ZR, Yang SM. Macrophages in tumor microenvironments and the progression of tumors. Clin Dev Immunol. 2012;2012:948098.

83. Facciabene A, Motz GT, Coukos G. T-regulatory cells: key players in tumor immune escape and angiogenesis. Cancer Res. 2012 May 1;72(9):2162-71.

84. Nishikawa H, Sakaguchi S. Regulatory T cells in cancer immunotherapy. Curr Opin Immunol. 2014 Apr;27:1-7.

85. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov. 2019 Mar;18(3):197-218.

86. Lei S, Zhang Y. Identification of survival-related genes and a novel gene-based prognostic signature involving the tumor microenvironment of uveal melanoma. Int Immunopharmacol. 2021 Jul;96:107816.

87. Peltanova B, Raudenska M, Masarik M. Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: a systematic review. Mol Cancer. 2019 Mar 30;18(1):63.

88. Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, Bin J, Liao Y, Liao W. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. Cancer Immunol Immunother. 2019 Mar;68(3):433-42.

89. Predina J, Eruslanov E, Judy B, Kapoor V, Cheng G, Wang LC, Sun J, Moon EK, Fridlender ZG, Albelda S, Singhal S. Changes in the local tumor microenvironment in recurrent cancers may explain the failure of vaccines after surgery. Proc Natl Acad Sci U S A. 2013 Jan 29;110(5):E415-24.

90. Fay KT, Ford ML, Coopersmith CM. The intestinal microenvironment in sepsis. Biochim Biophys Acta Mol Basis Dis. 2017 Oct;1863(10 Pt B):2574-83.

91. Herrada AA, Escobedo N, Iruretagoyena M, Valenzuela RA, Burgos PI, Cuitino L, Llanos C. Innate Immune Cells' Contribution to Systemic Lupus Erythematosus. Front Immunol. 2019;10:772.

92. Zhong Y, Zhang W, Hong X, Zeng Z, Chen Y, Liao S, Cai W, Xu Y, Wang G, Liu D, Tang D, Dai Y. Screening Biomarkers for Systemic Lupus Erythematosus Based on Machine Learning and Exploring Their Expression Correlations With the Ratios of Various Immune Cells. Front Immunol. 2022;13:873787.

93. Edovitsky E, Elkin M, Zcharia E, Peretz T, Vlodavsky I. Heparanase gene silencing, tumor invasiveness, angiogenesis, and metastasis. J Natl Cancer Inst. 2004 Aug 18;96(16):1219-30.

94. Xu L, Nirwane A, Yao Y. Basement membrane and blood-brain barrier. Stroke Vasc Neurol. 2019 Jul;4(2):78-82.

95. Reuten R, Zendehroud S, Nicolau M, Fleischhauer L, Laitala A, Kiderlen S, Nikodemus D, Wullkopf L, Nielsen SR, McNeilly S, Prein C, Rafaeva M, Schoof EM, Furtwangler B, Porse BT, Kim H, Won KJ, Sudhop S, Zornhagen KW, Suhr F, Maniati E, Pearce OMT, Koch M, Oddershede LB, Van Agtmael T, Madsen CD, Mayorca-Guiliani AE, Bloch W, Netz RR, Clausen-Schaumann H, Erler JT. Basement membrane stiffness determines metastases formation. Nat Mater. 2021 Jun;20(6):892-903.

96. Sathyanarayana UG, Padar A, Suzuki M, Maruyama R, Shigematsu H, Hsieh JT, Frenkel EP, Gazdar AF. Aberrant promoter methylation of laminin-5-encoding genes in prostate cancers and its relationship to clinicopathological features. Clin Cancer Res. 2003 Dec 15;9(17):6395-400.

97. Jayadev R, Morais M, Ellingford JM, Srinivasan S, Naylor RW, Lawless C, Li AS, Ingham JF, Hastie E, Chi Q, Fresquet M, Koudis NM, Thomas HB, O'Keefe RT, Williams E, Adamson A, Stuart HM, Banka S, Smedley D, Genomics England Research C, Sherwood DR, Lennon R. A basement membrane discovery pipeline uncovers network complexity, regulators, and human disease associations. Sci Adv. 2022 May 20;8(20):eabn2265.

98. Green DR, Ferguson T, Zitvogel L, Kroemer G. Immunogenic and tolerogenic cell death. Nat Rev Immunol. 2009 May;9(5):353-63.

99. Galluzzi L, Buque A, Kepp O, Zitvogel L, Kroemer G. Immunogenic cell death in cancer and infectious disease. Nat Rev Immunol. 2017 Feb;17(2):97-111.

100. Wang Q, Ju X, Wang J, Fan Y, Ren M, Zhang H. Immunogenic cell death in anticancer chemotherapy and its impact on clinical studies. Cancer Lett. 2018 Dec 1;438:17-23.

101. Sun Y, Feng X, Wan C, Lovell JF, Jin H, Ding J. Role of nanoparticle-mediated immunogenic cell death in cancer immunotherapy. Asian J Pharm Sci. 2021 Mar;16(2):129-32.

102. Garg AD, Martin S, Golab J, Agostinis P. Danger signalling during cancer cell death: origins, plasticity and regulation. Cell Death Differ. 2014 Jan;21(1):26-38.

103. Ahmed A, Tait SWG. Targeting immunogenic cell death in cancer. Mol Oncol. 2020 Dec;14(12):2994-3006.

104. Fucikova J, Kepp O, Kasikova L, Petroni G, Yamazaki T, Liu P, Zhao L, Spisek R, Kroemer G, Galluzzi L. Detection of immunogenic cell death and its relevance for cancer therapy. Cell Death Dis. 2020 Nov 26;11(11):1013.

105. Kroemer G, Galluzzi L, Kepp O, Zitvogel L. Immunogenic cell death in cancer therapy. Annu Rev Immunol. 2013;31:51-72.

106. Choi JU, Maharjan R, Pangeni R, Jha SK, Lee NK, Kweon S, Lee HK, Chang KY, Choi YK, Park JW, Byun Y. Modulating tumor immunity by metronomic dosing of oxaliplatin incorporated in multiple oral nanoemulsion. J Control Release. 2020 Jun 10;322:13-30.

107. Fahrmeir L, Kneib T, Lang S, Marx BD, editors. Regression: Models, Methods and Applications2013.

108. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1-22.

109. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. 1972;34(2):187-220.

110. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society. 1996;58(1):267-88.

111. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, Lingjaerde OC. Predicting survival from microarray data--a comparative study. Bioinformatics. 2007 Aug 15;23(16):2080-7.

112. Boulesteix AL, De Bin R, Jiang X, Fuchs M. IPF-LASSO: Integrative L(1)-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. Comput Math Methods Med. 2017;2017:7691937.

113. Zhao Z, Zucknick M. Structured penalized regression for drug sensitivity prediction. Journal of the Royal Statistical Society. 2020;69(3):525-45.

114. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982 Apr;143(1):29-36.

115. Heagerty PJ, Lumley T, Pepe MS. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. Biometrics. 2000;56(2):337-44.

116. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12:77.

117. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. Machine Learning. 2003;52:99-118.

118. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572-3.

119. McDermaid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. Brief Bioinform. 2019 Nov 27;20(6):2044-54.

120. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

121. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol. 2012;8(2):e1002375.

122. Maleki F, Ovens K, Hogan DJ, Kusalik AJ. Gene Set Analysis: Challenges, Opportunities, and Future Research. Front Genet. 2020;11:654.

123. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

124. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284-7.

125. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-25.

126. Du S, Mansmann U, Geisler BP, Li Y, Hornung R. A Diagnostic Model for Kawasaki Disease Based on Immune Cell Characterization From Blood Samples. Front Pediatr. 2022;9:769937.

127. Dominguez SR, Anderson MS, El-Adawy M, Glode MP. Preventing coronary artery abnormalities: a need for earlier diagnosis and treatment of Kawasaki disease. Pediatr Infect Dis J. 2012 Dec;31(12):1217-20.

128. Anderson MS, Todd JK, Glode MP. Delayed diagnosis of Kawasaki syndrome: an analysis of the problem. Pediatrics. 2005 Apr;115(4):e428-33.

129. Minich LL, Sleeper LA, Atz AM, McCrindle BW, Lu M, Colan SD, Printz BF, Klein GL, Sundel RP, Takahashi M, Li JS, Vetter VL, Newburger JW, Pediatric Heart Network I. Delayed diagnosis of Kawasaki disease: what are the risk factors? Pediatrics. 2007 Dec;120(6):e1434-40.

130. Chaudhary H, Nameirakpam J, Kumrah R, Pandiarajan V, Suri D, Rawat A, Singh S. Biomarkers for Kawasaki Disease: Clinical Utility and the Challenges Ahead. Front Pediatr. 2019;7:242.

131. Herberg JA, Kaforou M, Wright VJ, Shailes H, Eleftherohorinou H, Hoggart CJ, Cebey-López M, Carter MJ, Janes VA, Gormley S, Shimizu C, Tremoulet AH, Barendregt AM, Salas A, Kanegaye J, Pollard AJ, Faust SN, Patel S, Kuijpers T, Martinón-Torres F, Burns JC, Coin LJM, Levin M, Consortium I. Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children. JAMA. 2016;316(8):835-45.

132. Ling XB, Lau K, Kanegaye JT, Pan Z, Peng S, Ji J, Liu G, Sato Y, Yu TT, Whitin JC, Schilling J, Burns JC, Cohen HJ. A diagnostic algorithm combining clinical and molecular data distinguishes Kawasaki disease from other febrile illnesses. BMC Med. 2011 Dec 6;9:130.

133. Popper SJ, Watson VE, Shimizu C, Kanegaye JT, Burns JC, Relman DA. Gene transcript abundance profiles distinguish Kawasaki disease from adenovirus infection. J Infect Dis. 2009 Aug 15;200(4):657-66.

134. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008 Jul 1;24(13):1547-8.

135. Youden WJ. Index for rating diagnostic tests. Cancer. 1950 Jan;3(1):32-5.

136. Zhao Y, Wong L, Goh WWB. How to do quantile normalization correctly for gene expression data analyses. Scientific Reports. 2020 2020/09/23;10(1):15534.

137. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012 Mar 15;28(6):882-3.

138. Furukawa S, Matsubara T, Tsuji K, Motohashi T, Okumura K, Yabuta K. Serum soluble CD4 and CD8 levels in Kawasaki disease. Clin Exp Immunol. 1991 Oct;86(1):134-9.

139. Bengio Y, Grandvalet Y. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. J Mach Learn Res. 2004;5:1089–105.

140. Hicar MD. Antibodies and Immunity During Kawasaki Disease. Front Cardiovasc Med. 2020;7:94.

141. Menikou S, Langford PR, Levin M. Kawasaki Disease: The Role of Immune Complexes Revisited. Front Immunol. 2019;10:1156.

142. Xu M, Jiang Y, Wang J, Liu J, Liu C, Liu D, Yang S. Distinct variations of antibody secreting cells and memory B cells during the course of Kawasaki disease. BMC Immunol. 2019 Jun 3;20(1):16.

143. Wilder MS, Palinkas LA, Kao AS, Bastian JF, Turner CL, Burns JC. Delayed Diagnosis by Physicians Contributes to the Development of Coronary Artery Aneurysms in Children With Kawasaki Syndrome. Pediatric Infectious Disease Journal. 2007 2007/03;26(3):256-60.

144. Fu Y, Xiao W, Mao Y. Recent Advances and Challenges in Uveal Melanoma Immunotherapy. Cancers (Basel). 2022 Jun 23;14(13).

145. Heppt MV, Steeb T, Schlager JG, Rosumeck S, Dressler C, Ruzicka T, Nast A, Berking C. Immune checkpoint blockade for unresectable or metastatic uveal melanoma: A systematic review. Cancer Treat Rev. 2017 Nov;60:44-52.

146. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell. 2012 Mar 20;21(3):309-22.

147. Durante MA, Rodriguez DA, Kurtenbach S, Kuznetsov JN, Sanchez MI, Decatur CL, Snyder H, Feun LG, Livingstone AS, Harbour JW. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. Nat Commun. 2020 Jan 24;11(1):496.

148. Van den Eynden GG, Majeed AW, Illemann M, Vermeulen PB, Bird NC, Hoyer-Hansen G, Eefsen RL, Reynolds AR, Brodt P. The multifaceted role of the microenvironment in liver metastasis: biology and clinical implications. Cancer Res. 2013 Apr 1;73(7):2031-43.

149. Singh AD, Shields CL, Shields JA. Prognostic factors in uveal melanoma. Melanoma Res. 2001 Jun;11(3):255-63.

150. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, Dawson N, O'Donnell PH, Balmanoukian A, Loriot Y, Srinivas S, Retz MM, Grivas P, Joseph RW, Galsky MD, Fleming MT, Petrylak DP, Perez-Gracia JL, Burris HA, Castellano D, Canil C, Bellmunt J, Bajorin D, Nickles D, Bourgon R, Frampton GM, Cui N, Mariathasan S, Abidoye O, Fine GD, Dreicer R. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016 May 7;387(10031):1909-20.

151. Liu X, Qiao Y, Chen J, Ge G. Basement membrane promotes tumor development by attenuating T cell activation. J Mol Cell Biol. 2022 May 20;14(2).

152. Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. Matrix Biol. 2017 Jan;57-58:1-11.

153. Galluzzi L, Vitale I, Warren S, Adjemian S, Agostinis P, Martinez AB, Chan TA, Coukos G, Demaria S, Deutsch E, Draganov D, Edelson RL, Formenti SC, Fucikova J, Gabriele L, Gaipl US, Gameiro SR, Garg AD, Golden E, Han J, Harrington KJ, Hemminki A, Hodge JW, Hossain DMS, Illidge T, Karin M, Kaufman HL, Kepp O, Kroemer G, Lasarte JJ, Loi S, Lotze MT, Manic G, Merghoub T, Melcher AA, Mossman KL, Prosper F, Rekdal O, Rescigno M, Riganti C, Sistigu A, Smyth MJ, Spisek R, Stagg J, Strauss BE, Tang D, Tatsuno K, van Gool SW, Vandenabeele P, Yamazaki T, Zamarin D, Zitvogel L, Cesano A, Marincola FM. Consensus guidelines for the definition, detection and interpretation of immunogenic cell death. J Immunother Cancer. 2020 Mar;8(1).

154. Fucikova J, Moserova I, Urbanova L, Bezu L, Kepp O, Cremer I, Salek C, Strnad P, Kroemer G, Galluzzi L, Spisek R. Prognostic and Predictive Value of DAMPs and DAMP-Associated Processes in Cancer. Front Immunol. 2015;6:402.

155. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012 Dec;131(4):281-5.

156. Nidheesh N, Abdul Nazeer KA, Ameer PM. An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. Comput Biol Med. 2017 Dec 1;91:213-21.

157. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008 Mar 10;26(8):1364-70.

158. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. Journal of Statistical Software. 2010 09/16;36(11):1 - 13.

159. Garg AD, De Ruysscher D, Agostinis P. Immunological metagene signatures derived from immunogenic cancer cell death associate with improved survival of patients with lung, breast or ovarian malignancies: A large-scale meta-analysis. Oncoimmunology. 2016 Feb;5(2):e1069938.

160. Liu Y, Kimura K, Yanai R, Chikama T, Nishida T. Cytokine, chemokine, and adhesion molecule expression mediated by MAPKs in human corneal fibroblasts exposed to poly(I:C). Invest Ophthalmol Vis Sci. 2008 Aug;49(8):3336-44.

161. Harlin H, Meng Y, Peterson AC, Zha Y, Tretiakova M, Slingluff C, McKee M, Gajewski TF. Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment. Cancer Res. 2009 Apr 1;69(7):3077-85.

162. Bagheri H, Pourhanifeh MH, Derakhshan M, Mahjoubin-Tehran M, Ghasemi F, Mousavi S, Rafiei R, Abbaszadeh-Goudarzi K, Mirzaei HR, Mirzaei H. CXCL-10: a new candidate for melanoma therapy? Cell Oncol (Dordr). 2020 Jun;43(3):353-65.

163. Wang Y, Xu Y, Dai X, Lin X, Shan Y, Ye J. The prognostic landscape of adaptive immune resistance signatures and infiltrating immune cells in the tumor microenvironment of uveal melanoma. Exp Eye Res. 2020 Jul;196:108069.

164. Triozzi PL, Schoenfield L, Plesec T, Saunthararajah Y, Tubbs RR, Singh AD. Molecular profiling of primary uveal melanomas with tumor-infiltrating lymphocytes. Oncoimmunology. 2019;8(10):e947169.

165. Maat W, Ly LV, Jordanova ES, de Wolff-Rouendaal D, Schalij-Delfos NE, Jager MJ. Monosomy of chromosome 3 and an inflammatory phenotype occur together in uveal melanoma. Invest Ophthalmol Vis Sci. 2008 Feb;49(2):505-10.

166. Riley RS, June CH, Langer R, Mitchell MJ. Delivery technologies for cancer immunotherapy. Nat Rev Drug Discov. 2019 Mar;18(3):175-96.

167. Murciano-Goroff YR, Warner AB, Wolchok JD. The future of cancer immunotherapy: microenvironment-targeting combinations. Cell Res. 2020 Jun;30(6):507-19.

168. Oliva M, Rullan AJ, Piulats JM. Uveal melanoma as a target for immune-therapy. Ann Transl Med. 2016 May;4(9):172.

169. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautes-Fridman C, Fridman WH, de Reynies A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016 Oct 20;17(1):218.

170. Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, Guo AY. ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. Adv Sci (Weinh). 2020 Apr;7(7):1902880.

171. Tosi A, Cappellesso R, Dei Tos AP, Rossi V, Aliberti C, Pigozzo J, Fabozzi A, Sbaraglia M, Blandamura S, Del Bianco P, Chiarion-Sileni V, Rosato A. The immune cell landscape of metastatic uveal melanoma correlates with overall survival. J Exp Clin Cancer Res. 2021 May 4;40(1):154.

Acknowledgements

I have experienced and gained a lot through the 4+ years of conducting research for this project. It was not an easy but precious time. I would like to thank many people who have helped me to get through or complete it. I would like to express my deepest gratitude to my first supervisor, **Prof. Ulrich Mansmann**. He has not only given me the opportunity to start such interesting research in Germany, but also provided me with great help and advice along the way, both in my research and in my daily life. Then, I would like to thank **Prof. Jutta Engel**, a very easy-going female professor, who has offered me many valuable advice for my dissertation as my second TAC member. Unfortunately, she left us before the completion of this dissertation. We all mourned her passing and missed her very much. As a new second TAC member, **Prof. Michael Ingrisch** also deserves my thanks. His presence kept me from worrying about my Ph.D. process. In addition, I am very grateful to **Dr. Roman Hornung**, who is both my third TAC member and the mentor of my first published article, for his meticulous guidance throughout my Ph.D. dissertation. In short, words cannot express all the gratitude to my all TAC members.

In addition, the completion of this dissertation is also thanks to the following people. Firstly, **Dr. Benjamin P. Geisler**, a researcher at both Harvard Medical School and IBE, who provided rich advice on clinical and epidemiological research in my first paper, and who improved my English writing skills. **Dr. Aarif Mohamed Nazeer Batcha and Yingxia Li**, my colleagues at IBE, have been extremely helpful in my machine learning and R programming and I would like to thank them for that. On the administrative side, I would also like to thank the two program coordinators of the Ph.D.—EPH office, **Dr. Annette Hartmann and Monika Darchinger**, who have provided each Ph.D. student with all kinds of help and interesting activities, and have taken the trouble to answer my various questions about the program

Last but not least, I would like to thank **all my family and friends** for their unconditional encouragement and support, they have been my harbor during the storms, especially **Pfr. Dr. Xi Kang**, who guided me to adapt to life in Germany. I would also like to give special thanks to **the China Scholarship Council (CSC)**, for offering me with a 4-year funding and German language training, making my study and life in Germany much smoother. Once again, I really appreciate everyone mentioned above.

List of Publications

Published papers

1 **Du S.**; Mansmann U.; Geisler BP.; Li Y.; Hornung R. A Diagnostic Model for Kawasaki Disease Based on Immune Cell Characterization From Blood Samples. Frontiers in Pediatrics, 2022, 9:769937. doi: 10.3389/fped.2021.769937. **Impact factor = 3.569**

2 Li, Y.; Mansmann, U.; **Du, S.**; Hornung, R. Benchmark study of feature selection strategies for multi-omics data. *BMC Bioinformatics*, 23, 412 (2022). doi: 10.1186/s12859-022-04962-x. **Impact factor = 3.307**

3 Li, Y.; Mansmann, U.; **Du, S.**; Hornung, R. Synergistic Effects of Different Levels of Genomic Data for the Staging of Lung Adenocarcinoma: An Illustrative Study. *Genes*, 2021, 12, 1872. doi: 10.3390/genes12121872. **Impact factor = 4.141**

4 Mei Q.; Wang AY.; Bryant A.; Yang Y.; Li M.; Wang F.; **Du S.**; Kurts C.; Wu P.; Ma K.; Wu L.; Chen H.; Luo J.; Li Y.; Hu G.; Yuan X.; Li J. Survival Factors and Metabolic Pathogenesis in Elderly Patients (≥65) With COVID-19: A Multi-Center Study. *Frontiers in Medicine (Lausanne)*, 2021 Jan 7;7:595503. doi: 10.3389/fmed.2020.595503. **Impact factor = 5.058**

5 Mei Q.; Wang AY.; Bryant A.; Yang Y.; Li M.; Wang F.; Zhao JW.; Ma K.; Wu L.; Chen H.; Luo J.; **Du S.**; Halfter K.; Li Y.; Kurts C.; Hu G.; Yuan X.; Li J. Development and validation of prognostic model for predicting mortality of COVID-19 patients in Wuhan, China. Scientific Reports, 2020 Dec 31;10(1):22451. doi: 10.1038/s41598-020-78870-6. **Impact factor = 4.379**

Manuscript under review

1 Mei Q.; **Du S.**; Halfter K.; Li X.; Zhang G.; Li Y.; Saad C.; Huang H.; Qi S.; Cheng W.; Hu G.; Mansmann U.; Li J. Genome-scale Pathway Flux Analysis Predicts Efficacy of anti-PD-1 Therapy in Melanoma. *Genomics, Proteomics & Bioinformatics (Under review)*

Manuscript in preparation

1 Integration of basement membrane and immunogenic cell death patterns predicts the prognosis and immunotherapy response in uveal melanoma.