T.R.N.C

TURKISH REPUBLIC OF NORTH CYPRUS

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

SURVIVAL PREDICTION IN NEWLY DIAGNOSED MULTIPLE MYELOMA PATIENTS USING GENE EXPRESSION SIGNATURES CHARACTERISED BY NEX-GENERATION SEQUENCING

MUSTAPHA TOURAY

Master of Science in Biostatistics

Supervisor:

Assist. Prof. Dr. Özgür Tosun

NICOSIA, 2021

T.R.N.C

TURKISH REPUBLIC OF NORTHERN CYPRUS

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

SURVIVAL PREDICTION IN NEWLY DIAGNOSED MULTIPLE MYELOMA PATIENTS USING GENE EXPRESSION SIGNATURES CHARACTERISED BY NEX-GENERATION SEQUENCING

MUSTAPHA TOURAY

Master of Science in Biostatistics

Supervisor:

Assist. Prof. Dr. Özgür Tosun

NICOSIA, 2021

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

THESIS APPROVAL CERTIFICATE

Thesis submitted to the Institute of Graduate Studies of Near East University in partial fulfillment of the requirement for the degree of Master of Science in Biostatistics.

Thesis Committee;

Chair of the committee:	Prof. Dr. Zeki Akkuş Dicle Üniversitesi Sig: .....
Member:	Prof. Dr. İlker Etikan Near East University Sig: .....
Advisor:	Assist. Prof. Dr. Özgür Tosun Near East University Sig: .....
Approved by:	Prof. Dr. K. Hüsnü Can Başer Director of Institute of Graduate Studies Near East University Sig: .....

STATEMENT OF DECLARATION

I declare that this thesis entitled “Survival Prediction in Newly Diagnosed Multiple Myeloma Patients Using Gene Expression Signatures Characterized by Next-Generation Sequencing” submitted to Health Sciences Institute, Near East University, in fulfillment of the requirement for the award of the degree of Masters of Science in Biostatistics is the result of my own work. From the preparation stages of the thesis to its completion, I did not engage in any unethical action. I gathered all of the data for this thesis in accordance with academic and ethical guidelines. In this study, all information taken from external sources was properly referenced. The reference list contains all of the references utilized in the writing of this study. All efforts were made to avoid infringing on patent or copyright rights during the writing of this thesis.

Mustapha Touray

Sign: ……...

ACKNOWLEDGEMENTS

First and foremost, I thank Allah the Almighty for providing me with the health, abilities, stamina, direction, and devotion necessary to complete my Master of Biostatistics degree without any hinderance.

Secondly, I would want to express my gratitude to the Near East University Admissions Committee for providing me with the scholarship that enabled me to reach this milestone.

To my family, I say thank you for your continued support and understanding throughout this journey, especially to my lovely wife, Jankeh B.S. Dibba, my wonderful children, Muhammad and Aisha, and my brother, Lamin E. Manneh. Without your encouragement and financial support, I would not have been able to reach this milestone.

I wish to express my gratitude to Asst. Prof. Dr. Özgür Tosun, my course adviser and thesis supervisor, as well as Prof. Dr. İlker Etikan, the Head of the Department of Biostatistics, for their guidance, support, and graciousness throughout my studies. I am grateful for all of your efforts in making this achievement a reality.

To my classmates, I am glad for the opportunity to share my academic adventures with you. It was a delight to learn and complete projects with all of you.

Finally, I expressed my gratitude to Ousman Bah, my roommate, and everyone else in the apartment for being so tolerant, compassionate, gracious, and accommodating. You have all helped to make my stay and study more enjoyable and memorable.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

Abbreviation	Definition
ASCT	Autologous Stem Cell Transplantation
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
FISH	Fluorescence In Situ Hybridization
GEP	Gene Expression Profile
GES	Gene Expression Signatures
HR	Hazard Ratio
IFM	Intergroupe Francophone du Myelome
IGH	Immunoglobulin Heavy Chain
IMWG	International Myeloma Working Group
ISS	International Staging System
LDH	Serum Lactate Dehydrogenase
MDE	Myeloma Defining Events
MGUS	Monoclonal Gammopathy of Undetermined Significance
MM	Multiple Myeloma
MMRF	Multiple Myeloma Research Foundation
MRI	Magnetic Resonance Imaging
NA	Not Attained
NDMM	Newly Diagnosed Multiple Myeloma
NGS	Next-Generation Sequencing
OS	Overall Survival
PFS	Progression Free Survival
RFS	Relapse Free Survival
RISS	Revise International Staging System
RNA	Ribonucleic Acid
SMM	Smoldering Multiple Myeloma
SNP	Single Nucleotide Polymorphisms
SNV	Single Nucleotide Variations
TS	Targeted Sequencing
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

ÖZET

Amaç: Bu çalışmanın birincil amacı, Yen Tanı Almış Multiple Myeloma (NDMM) hastalarının risk gruplarını sınıflandırmak ve Progresyonsuz Sağkalım (PS) ve Genel Sağkalım (GS) oranlarını tahmin edebilmek için Gen Ekspresyonu Profili (GEP) tabanlı mevcut prognostik imzaların performansını karşılaştırmaktır.

Gereç ve Yöntem: Bu çalışmada Multiple Myeloma Research Foundation'ın (MMRF) Compass araştırmasının (NCT01454297) global kayıt sürümü IA15'ten 774 adet NDMM hastasının verilerini incelenmektedir. Yeni Nesil Dizileme (YND) ile elde edilen verilerin klinik bulgularla birleştirilmesi ve gen ekspresyonu düzeylerinin PS ve GS'ı değerlendirmedeki performansını belirlemek için sağkalım analizi temelinde çözümlemeler yapılmıştır.

Bulgular: Çalışma kriterlerine uygun 767 adet NDMM hastasının ortanca PS ve GS süresi, Gen Ekspresyonu İmzaları'ndan (GES) bağımsız olarak sırasıyla 634 (259–1233) ve 1118 (409.5–1590) gün olarak bulunmuştur. Hem PS hem de GS için, değerlendirilen tüm GES için, düşük risk grubu ile yüksek risk grubu arasında istatistiksel olarak önemli farklılık tespit edilmiştir. EMC-92, IFM15, HM19 ve MRCIX6 gen imzalarının tümü, 767 NDMM hastasının 116'sında (15.12%) yüksek riskli bir popülasyon tanımlanmıştır. Benzer şekilde, UAMS70 ve UAMS17 gen imzalarının her ikisi de 767 hastadan 114 NDMM hastasından oluşan yüksek riskli bir popülasyon tanımlanmıştır (14.86%). Tanımlanabilen iki risk grubu için tüm GES'inde PS, GS'dan istatistiksel olarak önemli düzeyde daha kısadır ($p<0.05$).

Sonuçlar: Araştırmada incelenmiş olan tüm GES'ler, risk gruplarını ayırt etmede istatistiksel olarak önemli düzeyde başarılı olmuşlardır. Tüm GES'ler için sağkalım ve yüksek riskli vakaları diğer vakalardan ayırt etmeye yönelik analizlerde istatistiksel olarak önemli sonuçlar elde edilebilmiştir.

Anahtar Kelimeler: Progresyonsuz Sağkalım; Genel Sağkalım; Gen Ekspresyonu İmzaları; Yeni Nesil Dizileme

Name of the student: Mustapha Touray

Mentor: Assist. Prof. Dr. Özgür Tosun

Department: Biostatistics

ABSTRACT

Aim: The primary purpose of this study is to compare the performance of existing GEP-based prognostic signatures in stratifying risk groups and predicting NDMM patients' PFS and OS rates.

Material and Method: This study included 774 NDMM patients' data from the Multiple Myeloma Research Foundation's (MMRF) CoMMpass trial (NCT01454297), global registry version IA15, that were characterized by NGS and analyzed to determine the performance of GES in evaluating PFS and OS rate.

Findings: The median PFS and OS time in days of the 767 NDMM patients available for study were 634 (259–1233) and 1118 (409.5–1590) days, respectively, regardless of the GES. For both PFS and OS, all of the GES evaluated discriminated a high-risk group that was significantly different from the low-risk group (Figure 3A to D, Figure 4E to H, and Figure 5I to L). The EMC-92, IFM15, HM19, and MRCIX6 gene signatures all identified a high-risk population of 116 out of 767 (15.12%) NDMM patients (Table 7). Similarly, UAMS70 and UAMS17 gene signatures both identified a high-risk population of 114 NDMM patients out of 767 (14.86%) (Table 6). For each of the two categorized risk groups, the PFS was significantly shorter than the OS in all of the GES (Figure 5).

Results: All of the GES performed significantly well in distinguishing risk groups as distinctively as possible, and the proportion of classified predicted risk groups varies less among the GES, with nearly all signatures being equally sensitive in predicting survival outcomes and identifying high-risk cases in NDMM patients.

Key Words: Progression free survival; overall survival; gene expression signatures; performance; next-generation sequencing.

CHAPTER 1

INTRODUCTION

1.1 Disease Overview

Multiple myeloma (MM) is a biologically heterogeneous clonal plasma cell disease with three stages: premalignant (monoclonal gammopathy of undermining significance (MGUS)), non-symptomatic (smoldering multiple myeloma (SMM)), and lastly symptomatic disease with remissions and relapses (Prideaux et al., 2014). Several disease subtypes have been uncovered at the genetic and molecular level, despite their outward similarities. These genetic groups are linked to various clinicopathological features and outcomes (Fonseca et al., 2009). MM accounts for 10% of all hematological malignancies. It has a 160,000-case global incidence and a 106,000-case death rate. Myeloma incidence, patient empowerment, access to cancer medications, and health-care spending are all linked to the mortality-to-incidence ratio, which ranges from 9% to 64% (Ludwig et al., 2020). The elderly are disproportionately affected by this malignancy, with a median age of around 70 years at the time of diagnosis (Piazzi et al., 2016).

Patients with MM have a wide range of outcomes, with some dying weeks after diagnosis and others living for more than ten years. The causes of this heterogeneity are complex, involving interactions between host variables and disease biology characteristics. The underlying genetic characteristics of tumor cells are gradually being recognized as playing a key influence in the clinical heterogeneity of MM (Fonseca et al., 2009). Anemia, hypercalcemia, renal disease, and an increased risk of infection are some of the other common clinical symptoms (Short et al., 2011).

The disease is split into two groups based on karyotype investigations: hyperdiploid and nonhyperdiploid myeloma, with approximately 10 molecular subgroups based on gene expression studies (Hassan & Szalat, 2021, Fonseca et al., 2009 and Prideaux et al., 2014). Patients with IgH translocations, which are connected to more severe clinical features and shorter survival, make up the majority of the latter

group. t (11;14) (q13; q32), t (4;14) (p16; q32), and t (14;16) (q32; q23) are the three most common IgH translocations in myeloma. Trisomies describe hyperdiploid myeloma is a less severe form of the disease. Chromosomal 13 and 17 deletions, as well as chromosome 1 abnormalities (1p deletion and 1qamplification), have all been discovered as genetic progression factors (Fonseca et al., 2009).

The incidence of MM is projected to rise in Africa, Latin America, lower-income Asian Pacific countries, Europe, and North America. Lower-income Asia Pacific is predicted to increase in prevalence at the fastest rate: 71% by 2027 (Sharma & Kumar, 2018).

Thanks to recent methods such as gene expression profiling and array-based genome wide sequencing, researchers have gained a deeper understanding of the disease's molecular subgroups. Most myeloma cases can be categorized into one of many genetic subcategories based on the integration of data from these investigations and data gathered in other ways (Fonseca et al., 2009). Next-generation sequencing (NGS) (also known as high-throughput sequencing) methods have recently uncovered the complicated genetic landscape of MM, drastically altering our understanding of myeloma genesis (Kuiper et al., 2015 and Piazzi et al., 2016). As a result, the goal of this study is to evaluate the performance of various gene expression signatures (GES) on progression free survival (PFS) and overall survival (OS) using data characterized by NGS.

1.2 Diagnosis

The detection of 10% or more clonal plasma cells in bone marrow and M-protein in the blood has traditionally been used to diagnose MM. In recent years, the presence of at least one myeloma defining events (MDEs), as well as confirmation of at least 10% clonal plasma cells on bone marrow examination or a biopsy-proven plasmacytoma, has been necessary for a diagnosis of multiple myeloma (Table 1). The presence of CRAB features such as hypercalcemia, renal disease, anemia, or lytic bone lesions, as well as the 3 key biomarkers: clonal bone marrow plasma cells greater than 60%, serum free light chain (FLC) ratio greater than 100 mg/L, and more than one

focal lesion on MRI, define MDE (Rajkumar, 2018). In patients who do not have a large M- component, laboratory tests, imaging, and professional expertise are used to make a diagnosis (Ludwig et al., 2020).

Table 1: International myeloma working group diagnostic criteria for multiple myeloma and related plasma cell disorders

Disorder	Disease definition
Non-IgM monoclonal gammopathy of undetermined significance (MGUS)	All 3 criteria must be met: • Serum monoclonal protein (non-IgM type) <3 gm/dL • Clonal bone marrow plasma cells <10%[a] • Absence of end-organ damage such as hypercalcemia, renal insufficiency, anemia, • and bone lesions (CRAB) that can be attributed to the plasma cell proliferative disorder
Smoldering multiple myeloma	Both criteria must be met: • Serum monoclonal protein (IgG or IgA) ≥3 g/dL, or urinary monoclonal protein ≥500 mg per 24 hours and/or clonal bone marrow plasma cells 10%-60% • Absence of myeloma defining events or amyloidosis

Multiple myeloma	Both criteria must be met:
	• Clonal bone marrow plasma cells 10% or biopsy-proven bony or extramedullary Plasmacytoma
	• Any one or more of the following myeloma defining events:
	  • Evidence of end organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically:
	  • Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than the upper limit of normal or >2.75 mmol/L (>11 mg/dL)
	  • Renal insufficiency: creatinine clearance <40 mL per minute or serum creatinine >177 µmol/L (>2 mg/dL)
	  • Anemia: hemoglobin value of hemoglobin value of >2 g/dL below the lower limit of normal, or a hemoglobin value <10 g/dL
	  • Clonal bone marrow plasma cell percentage 60%
	  • Involved: uninvolved serum free light chain (FLC) ratio 100 (involved free light chain level must be 100 mg/L)
	  • >1 focal lesions on magnetic resonance imaging (MRI) studies (at least 5 mm in size)

IgM Monoclonal gammopathy of undetermined significance (IgM MGUS)	All 3 criteria must be met: • Serum IgM monoclonal protein <3 g/dL • Bone marrow lymphoplasmacytic infiltration <10% • No evidence of anemia, constitutional symptoms, hyperviscosity, lymphadenopathy, or hepatosplenomegaly that can be attributed to the underlying lymphoproliferative disorder.

Modified from (Rajkumar, 2018).

[a] Patients with low-risk MGUS (IgG type, M protein 15 gm/L, normal free light chain ratio) who have no clinical symptoms suggestive of myeloma can avoid having their bone marrow removed.

1.3 Molecular Classification

MM is actually a group of cytogenetically distinct plasma cell malignancies, despite the fact that it is still viewed as a single disease (S. K. Kumar & Vincent Rajkumar, 2018 and S. Kumar et al., 2012, Table 2). Trisomies or IgH translocations affect just a tiny fraction of the people and are the most prevalent chromosomal abnormalities seen as MGUS progresses. Additional chromosomal abnormalities such as gain(1q), del(1p), del (17p), del (13), RAS mutations, and secondary translocations involving MYC emerge over the course of multiple myeloma. Additionally, primary and secondary chromosomal anomalies can have an impact on the disease's progression, treatment response, and prognosis. Importantly, chromosomal abnormalities in multiple myeloma have different meanings and repercussions depending on the phase of the infection at which they are discovered (Rajan & Rajkumar, 2015).

Table 2: Primary molecular cytogenetic classification of multiple myeloma

Subtype	Gene(s)/chromosomes affecteda	Percentage of myeloma patients
Trisomic multiple myeloma	Recurrent trisomies involving odd-numbered chromosomes with the exception of chromosomes 1, 13, and 21	42
IgH translocated multiple myeloma		30
t(11;14) (q13;q32)	CCND1 (cyclin D1)	15
t(4;14) (p16;q32)	FGFR-3 and MMSET	6
t(14;16) (q32;q23)	C-MAF	4
t(14;20) (q32;q11)	MAFB	<1
Other IgH translocationsa	CCND3 (cyclin D3) in t(6;14) multiple myeloma	5
Combined IgH translocated/trisomic multiple myeloma	Presence of trisomies and any one of the recurrent IgH translocations in the same patient	15
Isolated Monosomy 14	Few cases may represent 14q32 translocations involving unknown partner chromosomes	4.5
Other cytogenetic abnormalities in absence of IgH translocations or trisomy or monosomy 14		5.5
Normal		3

Reproduced from (Rajkumar, 2018) who modified it from (S. Kumar et al., 2012).
[a] Includes the t(6;14)(p21;q32) translocation, and rarely, other IgH translocations involving uncommon partner chromosome.

1.4 Survival Prediction and Risk Classification

Fluorescence in-situ hybridization (FISH), G-banded karyotype, and gene expression profiling were used to determine prognosis and risk classification. These techniques have proven to be extremely useful in detecting the bulk of cytogenetic abnormalities and structural rearrangements that cause MM (*An NGS Primer for Multiple Myeloma*, 2017). Today, ECOG performance status, cancer load, and cancer biology, as defined by well-established genetic changes revealed through FISH or karyotype testing, are now used as prognostic indicators. Furthermore, prognostic indications include age, LDH levels, macroglobulin levels, and aggregate rankings such as the International Staging System (ISS). IgH translocations or genomic abnormalities such as hyperdiploidy, chromosomal gains, or deletions are considered prognostic genetic abnormalities (Piazzi et al., 2016).

Although it is estimated that 15% of MM patients have an average life span below two years, and 20% have an average life span of greater than ten years (Piazzi et al., 2016). Nonetheless, over the last decade, risk stratification methods have improved, effectively segmenting populations that respond to different therapies and extending life expectancy by a decade or more (Anderson, 2014). The Revised International Staging System (RISS) is the latest recent guideline. It takes a holistic approach to prediction, including the disease biology prognostic markers indicated in (*Table 3*). In addition, the Mayo Clinic produced another risk stratification guideline (mSMART, *Table 3*) to provide therapeutic guidance for people in various risk groups (JR et al., 2013).

While well-defined cytogenetic prognostic markers have long been included in MM guidelines, they have yet to be linked to medicines or widely employed in treatment decisions. Although other translocations, such as the t(4:14), have been associated to other changes, the Del17 alteration is regarded to be a stand-alone biological marker for poor outcomes. A few of the more rare mutations, such as the t(14;16) translocation, are gradually becoming identified as disease risk factors (Fonseca et al., 2009).

Despite the fact that a large number of Patients with MM relapse, and the prognosis for a minority of cases continues to worsen, with early relapse and poor survival, risk classification is critical to identifying these patients and developing and changing therapy regimens for them. The ISS criteria (*Table 3*) are used to assess risk upon diagnosis, which include beta-2 microglobulin and albumin levels, as well as three chromosomal abnormalities: 17p13 deletion, t(4;14), and t(14;16). These criteria can also be used to identify individuals with high-risk MM who have a poor prognosis, early relapse, or initial refractory disease, as well as a shorter survival period. This patient subgroup must be identified in order to determine the optimum treatment plan using currently existing medicines and to develop novel treatments (Szalat & Munshi, 2016). GES have recently become accessible as substitute to these risk classifiers. Despite the fact that they have yet to be widely adopted, numerous groups have independently identified them in the last five years, and at least two study groups have described them. When compared to cytogenetic tests, gene expression profiles may provide a predictive solution that is both technically superior and more accurately reflects the patient's MM biology (Piazzi et al., 2016).

Table 3: Risk Stratification Guideline

RISS (2015) Guideline	
Stage I	Serum albumin 3.5 g/dL
	Serum β-2 microglobulin <3.5 mg/L
	No high risk cytogenetics
	Normal serum lactate dehydrogenase level
Stage II	Not Stage I or III
Stage III	Serum β-2 microglobulin >5.5 mg/L
	High risk cytogenetics [t(4;14), t(14;16), or del(17p)] or Elevated serum lactate dehydrogenase level
Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSmart)	
High Risk	del(17p), t(14;16), t(14;20), High risk gene expression profiling signature
Intermediate Risk	t(4;14), del(13), hypodiploidy, plasma cell labeling index (PCLI) ≥ 3%
Standard Risk	All others including: t(11;14), t(6;14)

Reproduced from (An NGS Primer for Multiple Myeloma, 2017)

1.5 Disease Progression

Landgren et al. (2009) found that MM develops from an asymptomatic pre-malignant stage known as MGUS (*Figure 1*) and that more than half of patients identified with MGUS have had the disease for more than ten years before being diagnosed (TM et al., 2012). Over the course of five to ten years, smoldering multiple myeloma (SMM), an intermediate and more advanced asymptomatic stage, progresses to the multiple myeloma stage (Rajkumar, 2018). Plasma cell leukemia (PCL) is the final stage of the disease, in which the diseased plasma cell becomes proliferative, departing the bone marrow with rapid proliferation and eventually dying (Prideaux et al., 2014).

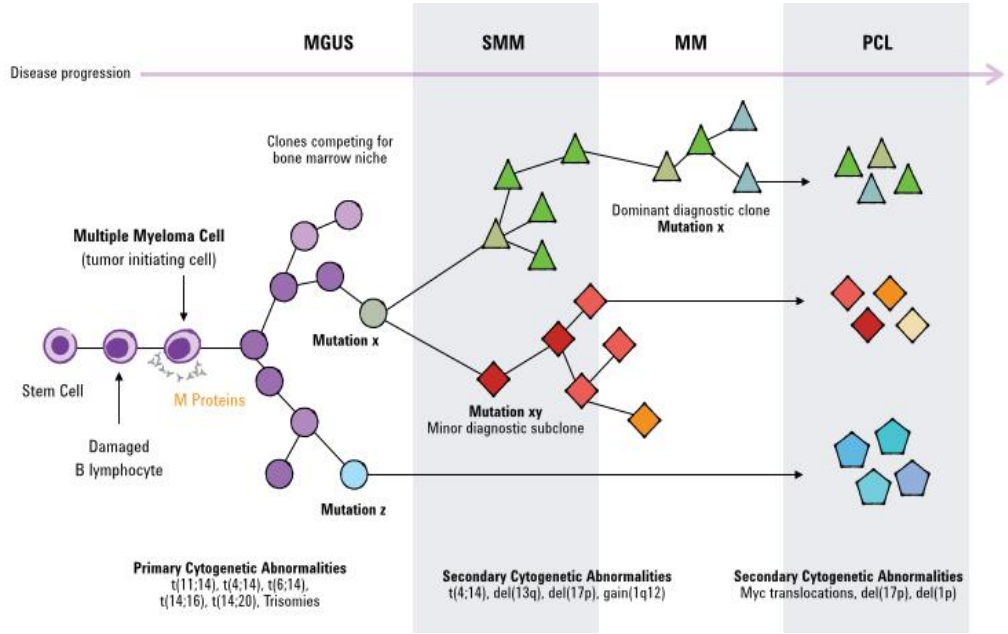

Figure 1 Clonal composition of multiple myeloma during disease progression and treatment. Derived from (Prideaux et al., 2014) *who adopted it from* (Morgan et al., 2012)

1.6 Treatment of Newly Diagnosed MM

Individual suitability with chronological age remains a crucial component in therapy selection in MM, despite the fact that adequate treatment choices are based on the reliable identification of high-risk patients at diagnosis (Manasanch et al., 2021). Over the last few years, overall survival (OS), progression-free survival (PFS), and

time to progression (TTP) have all improved significantly. In recent years, new therapeutics for MM have been developed, including innovative drugs given as combination regimens with well-established chemotherapy schemes, as well as autologous hematopoietic stem cell transplantation (Piazzi et al., 2016). However, determining which treatment will have the greatest impact is difficult due to the wide range of disease development (Martinez-Lopez et al., 2014).

As a first-line therapy, patients are given bortezomib, lenalidomide, and dexamethasone (VRd). Carfilzomib, lenalidomide, dexamethasone (KRd) is a substitute to VRd in high-risk patients. In suitable patients, first therapy is administered for 3-4 cycles, followed by autologous stem cell transplantation (ASCT). Low-risk patients can choose to have their ASCT postponed after the first relapse. VRd is administered to patients who are not transplant candidates for about 8 to 12 cycles before taking lenalidomide or lenalidomide plus dexamethasone. After ASCT, patients with standard-risk disease should stay on lenalidomide, while those with intermediate or high-risk disease should stay on a bortezomib-based regimen (Rajkumar, 2018).

Attempts to use gene expression profiling to find effective therapy have failed miserably. According to the findings, identifying effective treatments capable of attaining complete remission requires more than a basic gene expression level (Amin et al., 2014). NGS data, on the other hand, identifies novel indicators that accurately reflect prognosis, patient outcome, and treatment response, enabling for the development of tailored therapy (Szalat & Munshi, 2016).

1.7 GEP in MM

Multiple myeloma prognostication at the RNA level has been extensively examined and documented using gene expression profiling of cancerous plasma cells. Importantly, RNA sequencing has supplanted microarrays as a method of studying gene expression. This method can be used to analyze both coding and noncoding RNA, revealing not only the levels of expression of the genes but also differential splicing and isoform expression, gene mutations profiles, and fusions (Szalat & Munshi, 2016).

In 2006, GEP of tumor cells was utilized to explain the molecular classification of myeloma (Zhan et al., 2006). A study of 4750 individuals in clinical trials using multiple GEP platforms confirmed the effectiveness of GEP as a prognosis tool (Kuiper et al., 2015). According to various research, GEP is also more precise than FISH in risk categorization (Decaux et al., 2016 & Manasanch et al., 2021).

Various GES have been developed using gene expression profiling to identify patients with high- and low-risk disease. The Arkansas group described a 70-gene signature, the Intergroupe Francophone du Myélome (IFM) a 15-gene signature, and the HOVON group a 92-gene signature (Decaux et al., 2008). Approximately 20% to 25% of patients with all three signatures are at high risk of dying within the next two years. However, because each of these signatures contains diverse genes with limited intersection, it is challenging to apply them universally. In a recent study, a combined unique signature was created, which could be employed in therapeutic practice in the future (Chng et al., 2016).

Several high-throughput technologies have become available over the last 15 years (Table 4). Whole-genome and whole-exome sequencing, array comparative genomic hybridization, and high-density single nucleotide polymorphism arrays are all examples of DNA-based research. Reoccurring mutations and affected pathways, mutational signatures, clonality and clonal evolution traits, and copy number alterations can all be identified using these methods (Decaux et al., 2008).

Table 4: Genomic Methodologies and Their Practical Clinical Uses

	Possible Results	Clinical Application	Comments
RNA-Based Studies			
Microarray-Based Studies	Gene expression profile; micro-RNA expression profile	GEP-based signatures allow risk stratification	Widely and commercially available; ease of utilization of the data; being phased-out by RNA sequencing
RNA Sequencing	Gene expression profile; micro-RNA; long noncoding RNA; alternate splicing isoforms; mutations	Identify high-risk disease; may inform selection of agents in future	This method provides broader information and is more reliable for clinical trials; requires expertise to interpret data

Modified from (Szalat & Munshi, 2016)

1.8 Next-Generation Sequencing Technology Approaches

NGS profiling has emerged in MM studies, and it has the potential to move the field into a new era of precision medicine. NGS research has revealed that the genome of MM cancer is exceedingly convoluted. The discovery and application of NGS is rapidly changing cancer patient care in some indications and cancer care facilities (Piazzi et al., 2016). Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), and Targeted Sequencing (TS) are examples of NGS technologies that have been utilized extensively to analyze the genomes of multiple myeloma patients. This corpus of study has shed insight on the DNA abnormalities identified in MM cells, as well as the symptoms and course of the disease (*An NGS Primer for Multiple Myeloma*, 2017).

NGS data explains myeloma biology by providing information on disease dynamics that can be used to inform prognosis and, in particular, to define low- and high-risk disease. At this moment, a combination of validated markers, ISS, and cytogenetic characteristics can be utilized to appropriately classify low- and high-risk multiple myeloma. Three main outcomes with a dismal prognosis are amp(1q23.3), amp(5q31.3), and del(12p13.31). New prognostic markers are now based on new gene and microRNA expression profiles, as well as specific mutations and clonal alterations (Szalat & Munshi, 2016). Translational researchers have recently used next-generation sequencing (NGS) to characterize genomic changes in MM, yielding new information on indels, gene fusions, chromosomal abnormalities, single nucleotide variations (SNVs), single nucleotide polymorphisms (SNPs), and larger shifts in DNA and RNA. When it comes to diagnosis and relapse, the ability to use NGS to determine a patient's prognosis can be used to rationally construct a personalized treatment strategy (*An NGS Primer for Multiple Myeloma*, 2017).

1.9 Aim

The primary goal of this study is to analyze and contrast the performance of existing GEP-based prognostic signatures in stratifying risk groups and predicting PFS and OS in newly diagnosed multiple myeloma (NDMM) patients. The data on which

GES were evaluated on in this study was described using NGS technology. In this case, the signatures EMC-92, UAMS-70, UAMS-17, IFM-15, HM19, and MRC-IX-6 were evaluated. This research also aims to determine if there is a difference in survival between low- and high-risk patients in each of the GEP prognostic signatures stated above. Furthermore, the study tried to assess the performance of the EMC-92 gene signature in predicting survival and stratifying risk groups in comparison to other gene signatures.

1.10 Statement of the Problem

Multiple myeloma is diagnosed in over 100,000 patients worldwide each year, has an increasing incidence and prevalence in many locations, and has a relapsing course, making it a major and growing healthcare burden (Ludwig et al., 2020). Treating MM is a unique clinical challenge due to the disease's variability. Furthermore, despite major improvements in patients' outcomes, the vast majority of people with MM remain incurable, necessitating an ongoing quest for novel therapeutic options (Piazzi et al., 2016).

Several new MM treatments have improved results, and improvements in NGS technologies have allowed researchers to gain a deeper understanding of the disease's molecular underpinnings and how it progresses. Because of the disease's complexity, however, better methodologies for predicting and tracking therapy response are required (*An NGS Primer for Multiple Myeloma*, 2017). The tumor progression status could also be captured by the survival related gene signatures, which could lead to more patient-tailored therapy. Furthermore, no study has yet been published in which gene expression profiling characterized by NGS technology was used to predict survival of and stratify NDMM patients, to our knowledge. As a result, a study like this, as well as the results that come with it, could assist improve risk stratification, improve clinical decision making, and help NDMM patients find new therapy alternatives.

1.11 Significance of the Study

GEP characterized by NGS is starting to have an impact on clinical judgements and therapeutic decisions. It helps to identify high-risk patients who relapse early and have a shorter life expectancy, as well as advance our knowledge of the disease. Furthermore, GEP characterized by NGS can be used to identify those who need early identification and treatment, and NGS has recently been shown to be effective in detecting one cell out of a million in bone marrow, where the lack of myeloma indicates a better prognosis (Szalat & Munshi, 2016). Understanding gene expression signatures that are better predictors of PFS and OS in MM characterized by such a technology may thus aid clinicians in clearly separating risk groups and predicting stable groups of appropriate size. As a result, the patient and clinician may be in a better position to make an informed treatment decision, which could result in a better outcome. Furthermore, this novel technology in understanding GES performance in predicting survival and stratifying risk groups in NDMM can be used to supplement traditional prognostic markers like ISS and R-ISS stage and adverse cytogenetics, resulting in increased accuracy in outcome prediction and accurate prognosis, as well as the development of treatment schedules tailored to the individual.

CHAPTER 2

LITERATURE REVIEW

ISS was described by the IMWG to ensure a more objective classification of patients based on ß2-microglobulin and albumin levels (Table 3). These clinical features, which were chosen because they are widely available and easy to detect in blood samples, classify MM patients into three stages: stage 1 has a median OS of 62 months, stage 2 has a median OS of 44 months, and stage 3 has a median OS of 29 months (Paszekova et al., 2014).

Patients with high-risk MM have been described as having a grim prognosis, experiencing early relapse or primary refractory disease, and having a shorter life expectancy in various studies. High serum LDH and β2 microglobulin levels, low albumin levels, and any of three chromosomal abnormalities (17p13 deletion, t (4; 14), and t (14; 16) are used to diagnose high-risk myeloma in the R-ISS, which is one of the most up-to-date criteria for defining high-risk myeloma (Palumbo et al., 2015). This classification, on the other hand, does not precisely identify all high-risk patients or those with a favorable prognosis. As a result, NGS technologies have lately been used to untangle the intricate genetic landscape of myelomagenesis, substantially altering our understanding of the disease (Hassan & Szalat, 2021). According to Shaughnessy et al. (2007), established clinical prognostic markers such as serum ß2-microglobulin, creatinine, C-reactive protein, and serum lactate dehydrogenase (LDH), as well as chromosome 13 deletion and other cytogenetic abnormalities, demonstrated a significant link with the high-risk group.

Previously, microarray technology was utilized to discover gene expression indices connected to survival outcomes in order to evaluate whether an NDMM patient has a severe form of the disease. The UAMS 70 gene signature, which is made up of 51 over- and 19 under-expressed genes, is one such MM-specific indicator. Despite continuously high performance across a number of MM datasets, using this signature with RNASeq data causes problems. In their work, Penaherrera et al. (2018)

reported that GEP has predictive efficacy in a large patient group treated with contemporary medicines. They further reveal that despite triplet initial therapy and triplet maintenance therapy, high-risk patients remain to have poor medical outcomes. GEP is also a useful tool for predicting which patients would have inferior outcomes with existing treatments, according to the researchers. As a result, it was more accurate than high-risk FISH in identifying patients who died of progressive myeloma in their patient cohort.

Traditional prognostic markers like ISS and unfavorable cytogenetics, according to Kuiper et al. (2012), have been shown to be enhanced by GES in MM research to improve the accuracy of outcome prediction. Furthermore, more precise prognosis has been found to lead to the development of therapeutic strategies targeted to increase the survival of high-risk MM patients. Precisely, a signature must be able to clearly differentiate risk groups and forecast stable groupings of relevant size in order to be clinically useful. Of note, both criteria are met by the EMC-92 signature according to Kuiper et al. (2012).

According to survival projections for each patient considering the first line of treatment, Mosquera Orgueira et al. (2021) revealed that patients in their research cohort treated with the best-predicted drug combination were significantly less likely to die than those treated with alternative schemes. This was especially true for individuals who received a triplet treatment that included bortezomib, an immunomodulatory drug (ImiD), and dexamethasone. According to a study conducted by Paszekova et al. (2014), preliminary results reveal good response rates with innovative medications such as bortezomib, lenalidomide, and thalidomide in high-risk patients, implying that when using this sort of therapy, the effect of adverse prognostic factors can be overcome.

Because GEP reflects the biology of MM in individual patients, there has been an urgent need in recent years to improve the prognosis of NDMM patients. As a result, Kuiper et al. (2012), Chng et al. (2016), and EH et al. (2017) discovered that patients categorized as high-risk by the EMC-92-gene signature have a significantly shorter OS than those classified by the other signatures. Furthermore, Kuiper et al. (2012)

discovered that this signature is unrelated to presently used prognostic markers and is superior to or similar to previously described signatures in multivariate investigations.

GES has been found to be very effective in estimating event-free survival and overall survival (OS) in MM in several trials using various techniques. The GES have been established as a predictive biomarker on their own in the majority of these trials. Researchers recently published a combination of already existing prognostic signatures, characterizing it as a single reliable signature that can be used to predict outcome in MM at diagnosis and relapse, in an effort to simplify GEP application in clinical settings and produce a unique tool (Szalat et al., 2016). Kuiper et al. (2012) evaluated factors such as heterogeneity in data sets, demographic variations, and changes in methodology to all have a part in identifying which gene is most closely linked to survival in a given set. They further highlighted other explanations could be identified in the therapeutic procedures adopted, where other genes could be to blame for the poor prognosis.

Many approaches, including the use of GEP and NGS, have been developed to increase baseline classification (Bolli et al., 2018). According to Avet-Loiseau (2010) and Lonial et al. (2015), Within two years of being diagnosed, around 20% of individuals with MM will either relapse or die. As per the R-ISS, only 10% of patients are at high risk of disease progression and/or death, while 6.1% of patients are at high risk of progression and/or death according to the NGS-based "double-hit" categorization (Avet-Loiseau, 2010). It has also been reported that MM affects men slightly more than women, and African-Americans are twice as likely as Caucasians to have the disease (Landgren et al., 2009). When patients are diagnosed, they are on median 66 years old, with a median survival time of 33 months (Kyle et al., 2003).

In a study, Shaughnessy et al., 2007 employed GEPs to produce risk scores and proliferation indexes for MM disease prediction based on the expression levels of 70 and 11 genes, respectively. They discovered that higher risk ratings and greater proliferation indexes were linked to a shorter survival time for MM patients. Furthermore, using log-rank analyses of expression quartiles, they discovered that 30% of the 70 genes linked to chromosome 1 were associated with a higher risk

of disease-related death (P<.001). Essentially, most of the up-regulated genes were connected to chromosome 1q, while the majority of down-regulated genes were associated to chromosome 1p, resulting in a high-risk score in a small group of 13% of cases with a short survival period. After a median follow-up of 33 months (1.51–55.75 months) in another trial, 82 individuals were still alive. High-risk GEP patients had a significantly lower three-year RFS (41%) than low-risk GEP patients (60%) and a significantly lower three-year OS (71%) than low-risk GEP patients (83%) (P = 0.034). Patients classified as high-risk by FISH had a three-year RFS/OS that was also shorter. Furthermore, deletion 17p and t(14;16) were linked to a shorter RFS/OS (Manasanch et al., 2021).

Walker et al. (2019) revealed genetic variables that strongly relate to PFS and OS in a whole-genome and exome data exploratory study of 1273 NDMM patients. Using recursive partitioning, they discovered that a high-risk subset of the population (6.1%) has a median PFS of 15.4 months and an OS of 20.7 months. Furthermore, IGH translocations containing t(4; 14) were related with a shorter PFS but not an OS, whereas hyperdiploidy had no effect on outcome. Smadja et al. (2001) reflect similar findings, stating that chromosome ploidy number is important and that hyperdiploid and nonhyperdiploid patients had a significant difference in survival. They reported that nonhyperdiploid patients have a lower OS rate and are more likely to have structural abnormalities, such as translocations involving the IGH locus at 14q32. Furthermore, in a multivariate analysis of multiple prognostic factors, non-hyperdiploidy was revealed to be the single most significant predictor of OS.

Decaux et al. (2008) generated a 15-gene model that predicts survival in newly diagnosed MM patients. The Intergroupe Francophone du Myélome (IFM) asserts that their 15-gene signatures are more sensitive than FISH and improves ISS prognostication when it comes to stratifying MM patients based on survival. According to their findings, high-risk patients' myeloma cells overexpress genes involved in cycle advancement and monitoring, whereas low-risk patients' myeloma cells are more diverse and contain the hyperdiploid gene signature. Furthermore, the results revealed a survival-predictor score that was statistically significantly related to survival in both the training and test sets, as well as the external validation cohorts. In this investigation,

Kaplan-Meier estimates of 3-year survival rates for cases with a low- or a high-risk were 90.5% and 47.4%, respectively, independent of established prognostic markers (Decaux et al., 2016). Similarly, Chen et al. (2015) employed hierarchical clustering to divide all of the samples in one of their datasets into two groups: high-risk and low-risk groups, to see if gene expression profiles could reliably predict overall survival. In predicting OS, the found high- and low-risk categories were significantly different ($P<.01$) according to their Kaplan-Meier estimation.

The EMC92-ISS classification is a new biological and clinical prognostic tool that outperforms existing biomarkers and gives a stable, clinically useful 4-group classification. The highest risk group had a median survival time of 24 months, the intermediate risk groups 47 and 61 months, and the lowest risk group had a median survival time of 96 months. GEP classifiers showed improvement in OS than for PFS, with PFS hazard ratios (HRs) ranging from 1.8 (95% CI, 1.5-2.1; IFM15) to 2.3 (95% CI, 1.9-2.7; EMC92). EMC-92, UAMS17, and UAMS70 had 18 percent, 12 percent, and 9 percent of patients at high risk, respectively, whereas UAMS80 and HM19 had 8 percent (Rowan Kuiper et al., 2015). In their study, Dickens et al. (2010) compared three signatures: UAMS70, IFM15, and their own 97 gene cell death signature. Except for one gene, BIRC5, each signature has its own set of genes. 37 patients were identified as having a poor prognosis using all three signatures. In total, 89 cases were recognized as having a poor prognosis by their own signature, 64 by the IFM15 gene signature, and 90 by the UAMS70 gene signature.

CHAPTER 3

MATERIAL AND METHOD

1.12 Study Design

The Multiple Myeloma Research Foundation's (MMRF) CoMMpass trial (NCT01454297) is a 10-year prospective longitudinal observation study of over 1000 NDMM patients receiving different standard accepted therapies with the aim of acquiring data on patients' biopsy, genetic information, wellbeing, and various ailments and clinical outcomes throughout the trial (dbGaP Study (nih.gov)).

1.13 Patients and Treatment

Data from participants in this prospective observational trial were used in this analysis. Ethics committees or review boards at the study areas approved the study, and it was conducted in accordance with the Helsinki declaration. A formal informed written consent was signed by all of the patients (D'Agostino et al., 2020).

The data was gathered from the MMRF CoMMpass global registry version IA15, which contains 1143 NDMM patients who had clinical lab data available at diagnosis. Patients were followed up on every six months for eight years at 76 different locations throughout the world (https://themmrf.org/finding-a-cure/our-work/the-mmrf-commpass-study/). From this trial, the present study included 774 NDMM patients who were characterized using next-generation sequencing. To examine PFS and OS, patients' survival information containing their survival profiles were analyzed.

1.14 Next-Generation Sequencing

Before starting systemic therapy, baseline bone marrow CD138+ cells were collected (within 30 days before first-line treatment). The Translational Genomics Institute (TGen) performed long-insert WGS and WES. Somatic tumor alterations

were discovered by comparing cancer cells to patient-specific matched healthy cells (D'Agostino et al., 2020).

1.15 Gene Expression Signatures

In predicting survival in NDMM patients, the performance of the following gene expression signatures was evaluated in this study: EMC-92, UAMS70, UAMS17, IFM15, HM19 and MRCIX6. Normalization was done using log scaling and cutoffs were calculated using 85% quantile.

1.16 Statistical Analysis

Flowchart of the analyses is given in (Figure 2). The Bioconductor geneClassifier package was used to get the Affymetrix probe IDs for the various gene expression signatures, as well as their weighting coefficients. The following platforms were used to convert the obtained probe IDs of the respective signatures to ensembl IDs:

http://www.ensembl.org/biomart;

http://biogps.org/#goto=welcome;

https://genecards.weizmann.ac.il/geneannot/index.shtml;

https://www.affymetrix.com/site/mainPage.affx;

http://xavierlab2.mgh.harvard.edu/cgi-bin/DiscoveryDB.py.

The CoMMpass survival dataset was used to cross-validate the converted ensembl IDs of each gene expression signature. For PFS and OS, only ensembl IDs that were common to both datasets were analyzed. PFS was defined as the period from diagnosis to progression or death, while OS was defined as the period from diagnosis to death from all causes. The survival signature scores, which are the total of normalized gene expression values multiplied by the probe set-specific weighing coefficients of each gene expression signature, were also stratified using an 85 percent cut-off

criterion. Signature scores above the prescribed level were classified as high-risk, while those below it was classified as low-risk.

The Kaplan-Meier approach was used to analyze PFS and OS as time-to-event data. To examine if there is a difference in PFS and OS between low- and high-risk patients, the log-rank test was used, with a p-value of less than 0.05 indicating statistical significance. The Cox proportional hazards model was used to calculate the hazard ratios (HRs) and 95% confidence intervals (CIs). R version 4.1.0 and the Survival Analysis package version 0.2.0 were used to conduct all analyses.

1.17 Data Deposition

The Data Access Use Committee has authorized access to the Interim Analysis 15 (IA15) version of CoMMpass, which is available for download upon approval of request at https://research.themmrf.org/rp/download. The data is stored in the database of Genotypes and Phenotypes (dbGaP; Study Accession phs000748.v7.p4 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study id=phs000748.v7.p4).

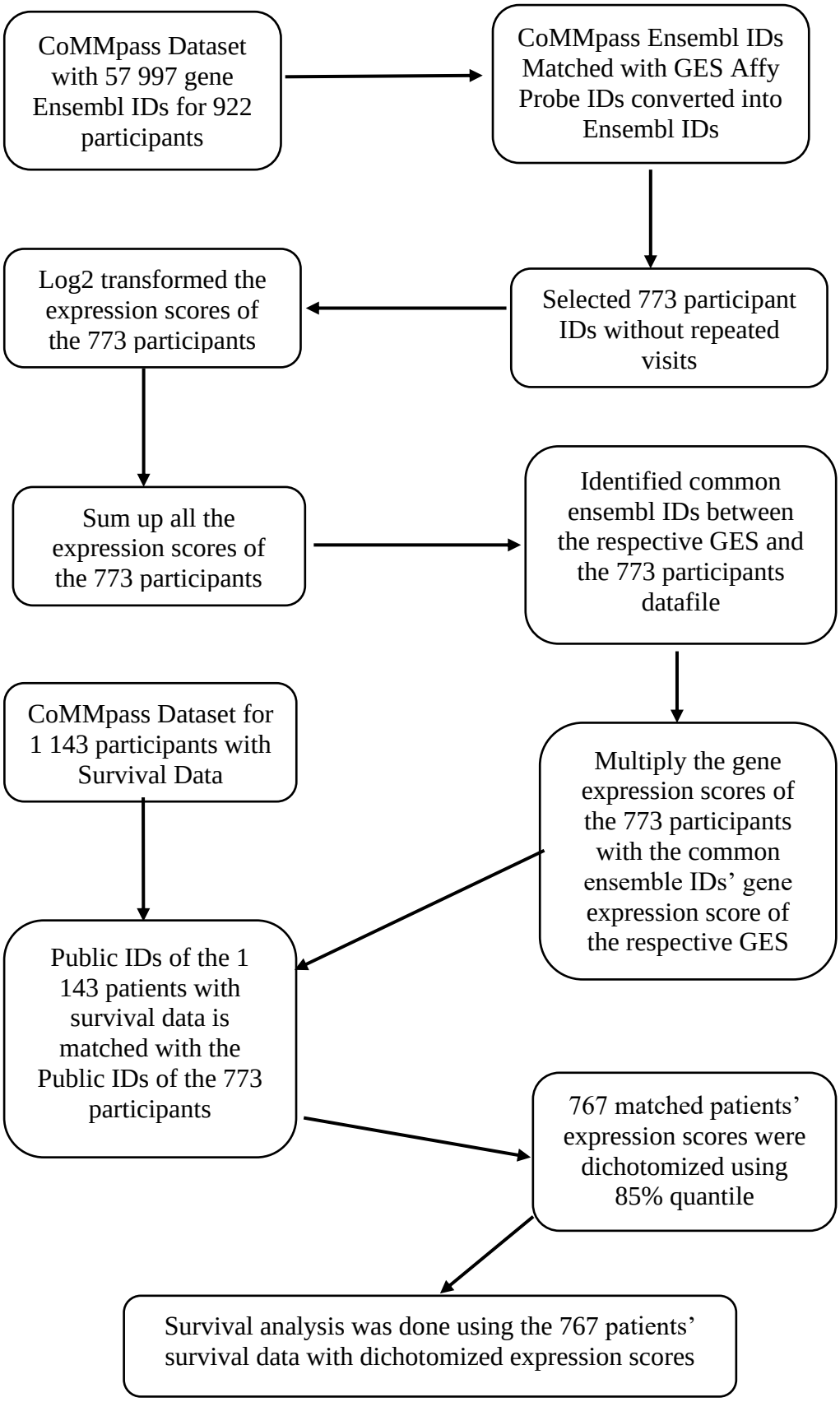

Figure 2: Flowchart of the data analysis

CHAPTER 4

FINDINGS

The performance of the six gene expression signatures (GES) was investigated using a CoMMpass dataset that included 1143 patients with survival data and 773 patients with genomic data, with only 767 NDMM patients matched. Based on the 85 percent quantile of the sum of expressed gene weighing coefficients, gene expression levels on CD138-selected plasma cells described using NGS from the 767 NDMM patients were dichotomized into low- and high-risk groups, and log-rank tests were used to determine the statistical significance difference between the classified risk groups in the various GES considered in this study (Table 5 and Table 6).

The median PFS and OS time in days of the 767 NDMM patients available for study were 634 (259–1233) and 1118 (409,5–1590) days, respectively, regardless of the GES. Only 2 (<1%) patients were uniquely commonly identified as high-risk patients by either of the signatures, while 407 (53,1%) patients were uniquely commonly identified as low-risk patients by either of the signatures. The median PFS and OS were 708 (297–163) and 1 221 (419–1 608) days, respectively, among the low-risk patients commonly identified by either of the GES. The median PFS and OS were (278 and 81 days) and (409 and 389 days), respectively, for the two patients who were commonly identified as high-risk by either of the GES.

The performance of all the GES in a survival analysis of 767 NDMM patients revealed great results, with patients classified as low-risk for each of the signatures having significantly good PFS and OS times, while those classified as high-risk had significantly worse PFS and OS times (Table 5, Table 6, Figure 3, Figure 4 and Figure 5). For both PFS and OS, all of the GES evaluated discriminated a high-risk group that was significantly different from the low-risk group (Figure 3A to D, Figure 4E to H, and Figure 5I to L).

The EMC-92, IFM15, HM19, and MRCIX6 gene signatures all identified a high-risk population of 116 out of 767 (15,12%) NDMM patients in this study (Table 7),

with PFS significantly shorter than OS for each of the two defined risk groups in each of these signatures. The median PFS of the 651 low-risk patients, 1 052, 1 069, 1 090, and 1 090 days for each of the above-mentioned signatures, respectively, was significantly greater than that of the 116 high-risk patients, 715, 769, 560, and 610 days for each of the above-mentioned signatures, respectively, with $p<0,05$ for each of them (Table 5, Figure 3A and C, and Figure 4E and G). In OS, the median significant differences between the identified risk groups of these signatures were maintained (Table 6, Figure 3B and D, and Figure 4F and H). However, there is a greater significant difference in OS than in PFS between the risk groups in all these signatures, $p<0,01$ (Figure 3 and Figure 4).

The UAMS70 and UAMS17 gene signatures both identified a high-risk population of 114 NDMM patients out of 767 (14,86%) (Table 6), with PFS significantly shorter than OS in each signature's two categorized groups (Figure 5). The median PFS of the 653 low-risk patients (1 065 and 1 082 days, respectively) was significantly greater than that of the 114 high-risk patients (776 and 553 days, respectively) for each of these signatures, with $p<0,05$ (Table 5, Figure 3I and K). Similarly, the median significant differences observed between the classified risk groups of these two signatures were maintained in OS too (Table 6, Figure 3J and L). Nonetheless, there was a higher significant difference between the risk groups in OS than in PFS, $p<0,05$, in both signatures (Table 6 and Figure 3). Furthermore, median survival was not reached after 2 190 days in Kaplan-Meier estimations of OS for the lowest risk categories in all signatures (Figure 3B and D, Figure 4F and H, and Figure 5J and L).

Table 5: Summary Statistics of Kaplan-Meier Estimates of Progression-free Survival

Signature	Dichotomous (Cutoff value)	Progression-free Survival				
		n(Events) Low-risk	n(Events) High-risk	Median PFS (95% CI) Low-risk	Median PFS (95% CI) High-risk	p-Value
EMC92	0,5380177	335 (51,5%)	67 (57,8%)	1 052 (908-1 186)	715 (612-1 090)	0,0483
UAMS70	0,3922542	339 (51,9%)	63 (55,3%)	1 065 (917-1 176)	776 (677-1 052)	0,023
UAMS17	-1,242702	326 (49,9%)	76 (66,7%)	1 082 (964-1 190)	553 (445 - 774)	4e-04
HM19	30,55749	323 (49,6%)	79 (68,1%)	1 090 (964-1 222)	560 (454 - 791)	4e-06
IFM15	6,349037	329 (50,5%)	73 (62,9%)	1 069 (938 -1 176)	769 (572 - 978)	0,008
MRCIX6	29,41472	326 (50,1%)	76 (65,5%)	1 090 (964-215)	610 (449 - 833)	2e-06

Table 6:Summary Statistics of Kaplan-Meier Estimates of Overall Survival

Signature	Dichotomous (Cutoff value)	Overall Survival				
		n(Event) Low-risk	n(Event) High-risk	Median OS (95% CI) Low-risk	Median OS (95% CI) High-risk	p-Value
EMC92	0,5380177	152 (23,3%)	43 (37,1%)	NA (2207 –NA)	1 704 (1 360 –NA)	8e-04
UAMS70	0,3922542	157 (24,0%)	38 (33,3%)	NA	2 207 (345 - NA)	0,014
UAMS17	-1,242702	147 (22,5%)	48 (42,1%)	NA (2207 - NA)	1 670 (1 033 –NA)	1e-05
HM19	30,55749	146 (22,4%)	49 (42,2%)	NA	2 207 (1 094 –NA)	2e-06
IFM15	6,349037	152 (23,3%)	42 (36,2%)	NA	2 207 (1 500 –NA)	0,004
MRCIX6	29,41472	148 (22,7%)	47 (40,5%)	NA	1 590 (1 094 - NA)	3e-06

The performance of the classified risk groups in the various signatures were significantly better for OS than PFS (Table 7, Table 8, Figure 3, Figure 4, and Figure 5) with HRs relative to the low-risk patients of the PFS ranging from 1,303 (95% CI, 1,002-1,694; EMC-92) to 1,8326 (95% CI, 1,426-2,355; MRCIX-6). In terms of OS, HRs relative to the low-risk patients range from 1,553 (95% CI, 1,089-2,213; UAMS70) to 2,153 (95% CI, 1,549-2,99; MRCIX6). The proportion of high-risk

patients identified among some signatures are the same: 15,12% (EMC-92, IFM15, HM19, and MRCIX6), and 14,86% (UAMS70 and UAMS17).

Among all the signatures, UAMS17 classified high-risk relative to low-risk patients had significantly the poorest PFS (median 553 vs. 1 082 days; HR 1,5658, 95% CI: 1,219–2,011; P= $4{,}45\times 10^{-4}$, estimated two-year PFS 44% vs. 63%; Figure 5 I) while MRC-IX6 high-risk relative to low-risk patients had significantly the poorest OS (median 1 590 days vs. not reached; HR 2,1525, 95% CI: 1,549–2,99; P=$4{,}87\times 10^{-6}$, estimated two-years OS 44% vs. 63%; Figure 4 H) (Table 5, Table 6, Table 7, and Table 8). UAMS70 high-risk patients had considerably better PFS (median 776 vs. 1 065 days; HR 1,3667, 95% CI: 1,044–1,79; P= 0,0232; estimated 2 years PFS 51% vs. 63%; Figure 5 K) and OS (median 2 207 days vs. not reached; HR 1,5526, 95% CI: 1,089–2,213; P= 0,015; estimated 2 years OS 78% vs. 81%; Figure 5 L) when it came to the best-performing signatures. Furthermore, UAMS70 high-risk patients have the same median OS time as HM19 (HR = 2,1398, 95% CI: 1,548-2,958; P= $4{,}14\times 10^{-6}$; estimated 2 years OS 65% vs. 85%; Figure 3 D), and IFM15 (HR= 1,6515, 95% CI: 1,173-2,324; P= $4{,}01\times 10^{-3}$; estimated 2 years OS 75% vs. 81%; Figure 4F) high-risk patients. However, their HRs and 95% CI differ (Table 5, Table 6, and Table 7, and Table 8).

In analyzing the performance of the signatures in predicting the patients' two-year survival rate, we discovered that the two-year probability of PFS in this study was, on average, around 63% for low-risk patients, which is lower than the two-year OS probability, which was, on average, roughly 84% for the same group of patients. Similarly, we discovered that the two-year probability of PFS for high-risk patients was around 47% on average, which is much lower than the two-year probability of OS for the same group of patients, which was around 71%. Thus, regardless of the defined risk groups, the two-year probability of survival in OS was significantly higher than in PFS (Figure 3, Figure 4, and Figure 5).

Table 7:Summary Statistics of Cox Regression Analysis of Progression-free Survival

High-risk signatures	Proportion of High-risk	Progression-free Survival		
		Hazard ratio	95% CI	Wald p-value
EMC92	15,12%	1,3030	(1,002 – 1,694)	0,0483
UAMS70	14,86%	1,3667	(1,044 – 1,79)	0,0232
UAMS17	14,86%	1,5658	(1,219 – 2,011)	4,45e-04
HM19	15,12%	1,7739	(1,386 – 2,27)	5,13e-06
IFM15	15,12%	1,4039	(1,089 – 1,81)	0,00881
MRCIX6	15,12%	1,8326	(1,426 – 2,355)	2,33e-06

Table 8:Summary Statistics of Cox Regression Analysis of Overall Survival

High-risk signatures	Proportion of High-risk	Overall Survival		
		Hazard ratio	95% CI	Wald p-value
EMC92	15,12%	1,7749	(1,264 – 2,491)	0,00092
UAMS70	14,86%	1,5526	(1,089 – 2,213)	0,015
UAMS17	14,86%	2,0320	(1,467 – 2,815)	2,03e-05
HM19	15,12%	2,1398	(1,548 – 2,958)	4,14e-06
IFM15	15,12%	1,6515	(1,173 – 2,324)	0,00401
MRCIX6	15,12%	2,1525	(1,549 – 2,99)	4,87e-06

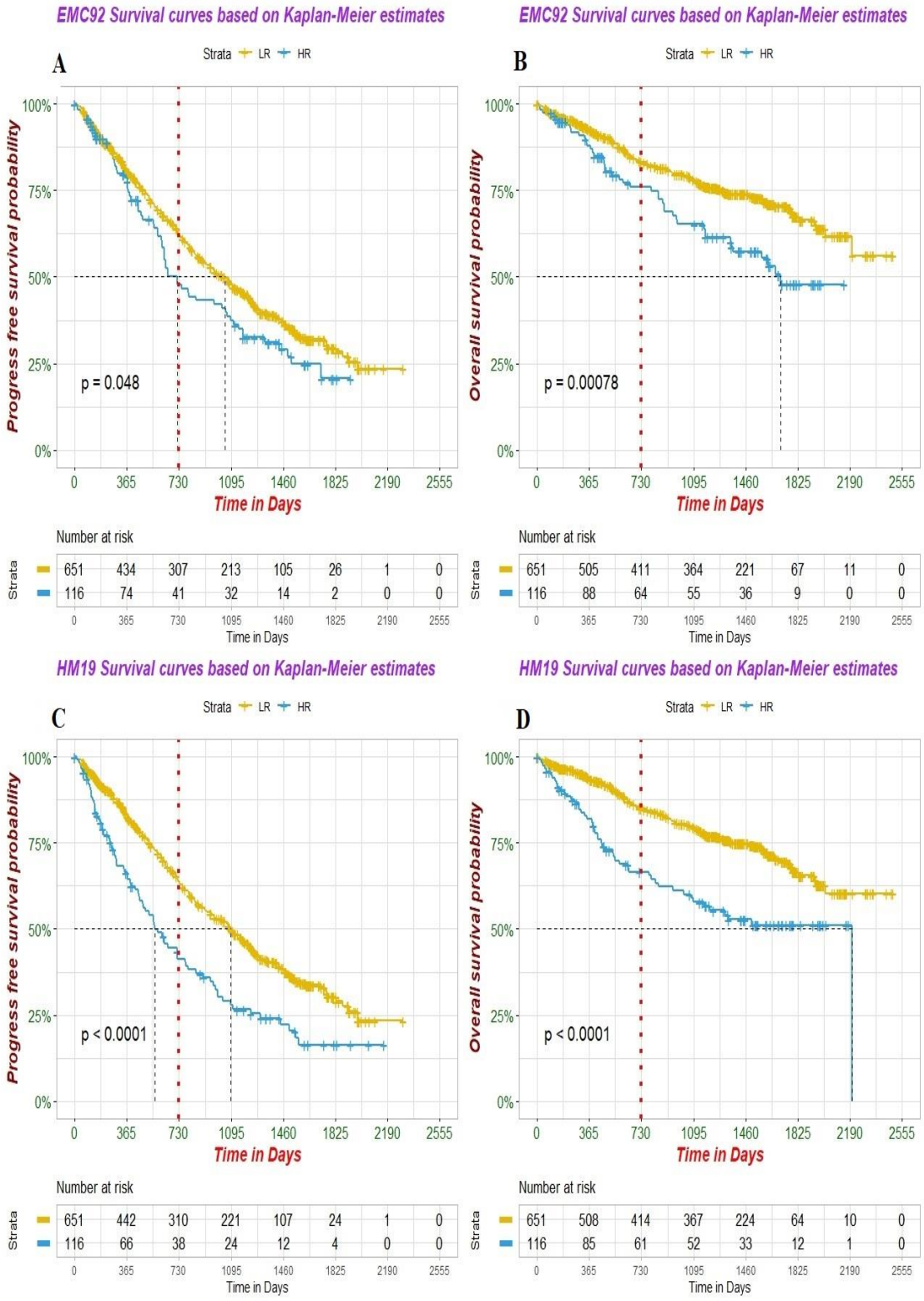

Figure 3: Performance of the gene expression signatures in determining progression-free and overall survival of the defined two risk groups.

LR means low-risk group; HR means high-risk group; p means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

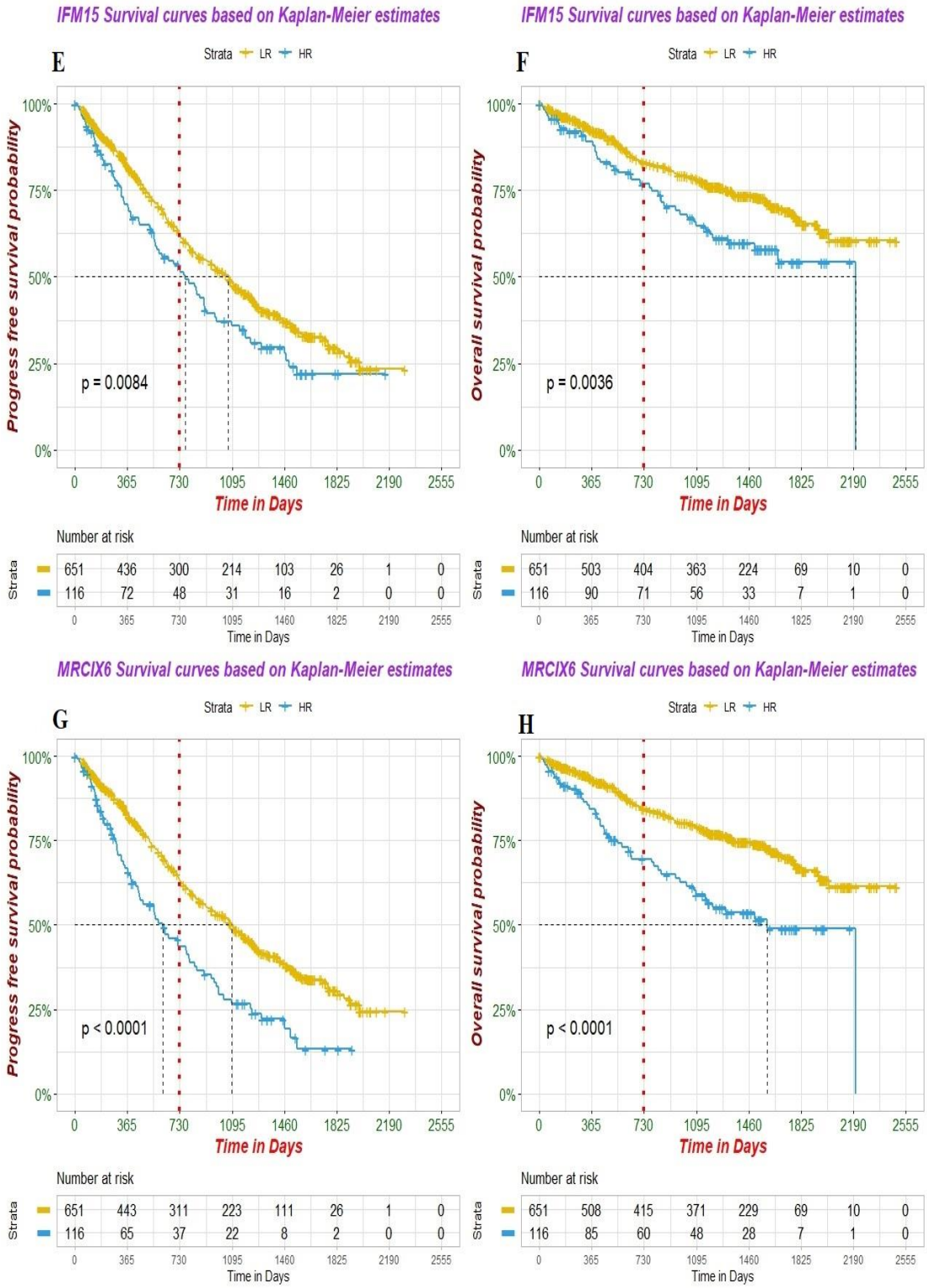

Figure 4: Performance of the gene expression signatures in determining progression-free and overall survival of the defined two risk groups.

LR means low-risk group; HR means high-risk group; p means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

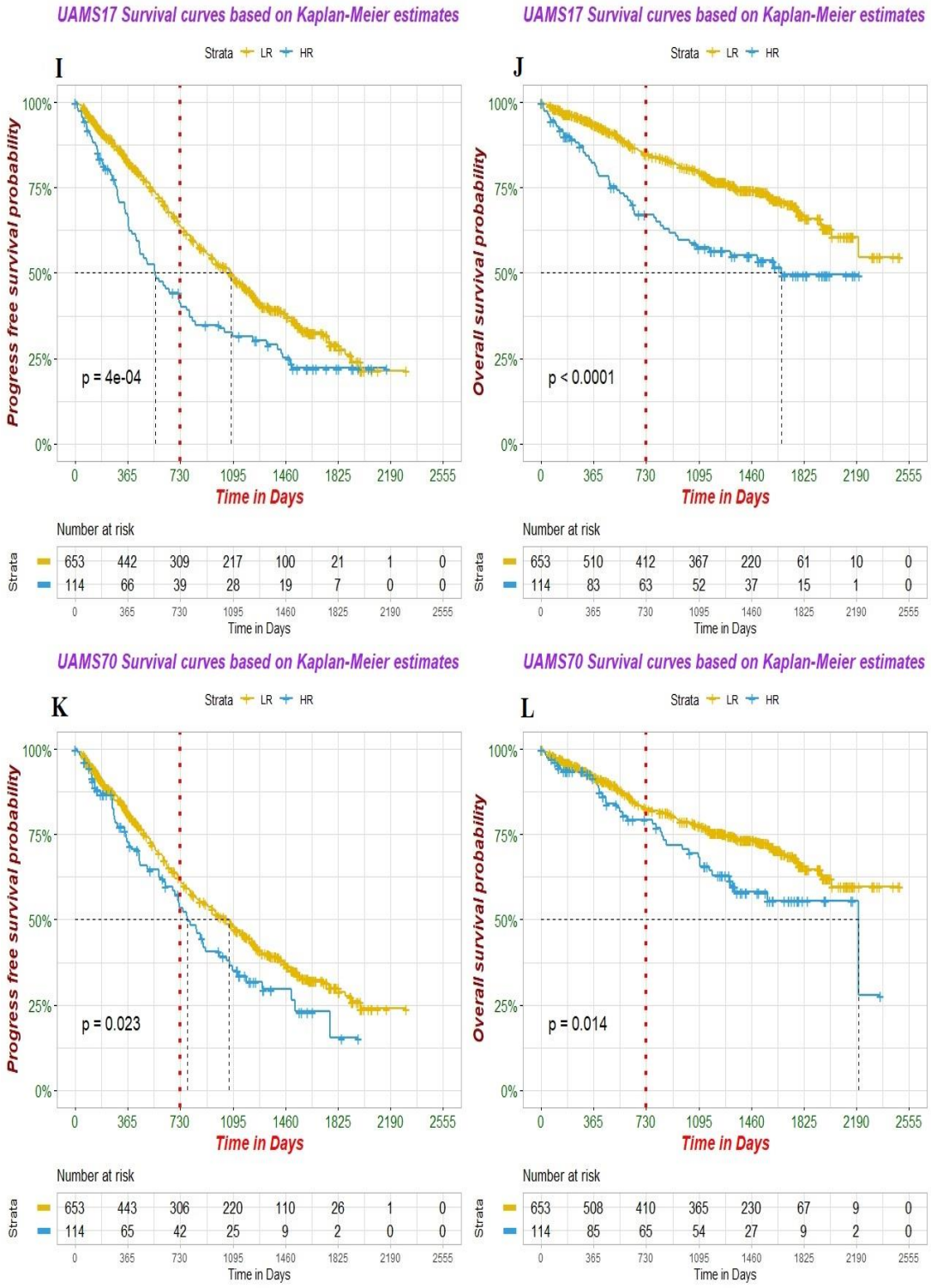

Figure 5: Performance of the gene expression signatures in determining progression-free and overall survival of the defined two risk groups.

LR means low-risk group; HR means high-risk group; p means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

The EMC-92 gene signature was compared to the other gene signatures using pair-wise comparison to evaluate which one better explained survival prediction and risk classification. The intersection of high-risk patients between the EMC-92 gene and UAMS-70, UAMS-17, IFM15, HM19, and MRCIX-6 gene signatures was 24 (3.1%), 37 (4,8%), 16 (2,1%), 32 (4,2%), and 34 (4,4%), respectively, of the overall population, according to the comparison. Furthermore, the EMC-92 signature classified 92(12%), 79(10,3%), 100(13%), 84(11%), and 82(10,7%) patients as high-risk patients, while the UAMS-70, UAMS-17, IFM15, HM19, and MRCIX-6 signatures identified them as low-risk patients, respectively (Figure 7). Similarly, in the stated comparable signatures, 90(11,7%), 77(10%), 100(13%), 84(11%), and 82(10,7%) of the patients were exclusively categorized as high-risk, but in EMC-92 they were regarded low-risk (Figure 7).

When EMC-92 was compared to each of the other GES in this study, patients exclusively classified as high-risk by EMC-92 had a significantly better prognosis than those identified as high-risk by EMC-92 and any of the other signatures (EMC92 vs. UAMS70: median 800 vs 581 days; HR 0,536, 95% CI: 0,3–0,918; P= 0,035; estimated 2 years PFS 52% vs. 32%, EMC-92 vs. UAMS17: median 1066 vs. 393 days; HR 0,425, 95% CI: 0,261–0,695; P= $6{,}4\times 10^{-4}$, estimated 2 years PFS 61% vs. 23%, EMC-92 vs. IFM15: median 750 vs. 552 days; HR 0,6051, 95% CI: 0,33–1,11; P= 0,105 , estimated 2 years PFS 50% vs. 32%, EMC-92 vs. HM19: median 1 066 vs. 393 days; HR 0,446, 95% CI: 0,269–0,740; P= 0,00178 , estimated 2 years PFS 60% vs. 19%, EMC-92 vs. MRCIX6: median 1 057 vs. 475 days; HR 0,504, 95% CI: 0,302–0,842; P= 0,0088 , estimated 2 years PFS 58% vs. 20%; Table 9, Table 10, Figure 8, Figure 9, and Figure 6). Similarly, patients who were exclusively classified as high-risk by the EMC-92 gene signature had a significantly better prognosis than those who were classified as high-risk by EMC-92 intersection with any of the following signatures (EMC-92 vs. UAMS17: median Not reached after 1 825 vs. 796 days; HR 0,295, 95% CI: 0,162–0,540; P= $7{,}5\times 10^{-5}$, estimated 2 years OS 88% vs. 52%, EMC-92 vs. HM19: median 1 704 vs. 847 days; HR 0,469, 95% CI: 0,254–0,867; P= 0,0157 , estimated 2 years OS 83% vs. 56%; Table 9, Table 10, Figure 8, and Figure 9). On the contrary, the OS rate of patients exclusively identified as high-risk by EMC-92 gene expression signature had insignificantly better prognosis than high-risk patients commonly

identified between EMC-92 and anyone of the following signatures (EMC-92 vs. UAMS70: median Not reached after 1 825 vs. 1 170 days; HR 0,6116; 95% CI: 0,307–1,217; P= 0,161, estimated 2 years OS 77% vs. 70%, EMC-92 vs. IFM15: median 1 704 vs. 983 days; HR 0,771; 95% CI: 0,357–1,664; P= 0,508, estimated 2 years OS 76% vs. 74%; EMC-92 vs. MRCIX6: median 1 704 vs. 983 days; HR 0,566; 95% CI: 0,304–1,054; P= 0,0727 , estimated 2 years OS 81% vs. 62%; Table 9, Table 10, Figure 8, Figure 9, and Figure 6).

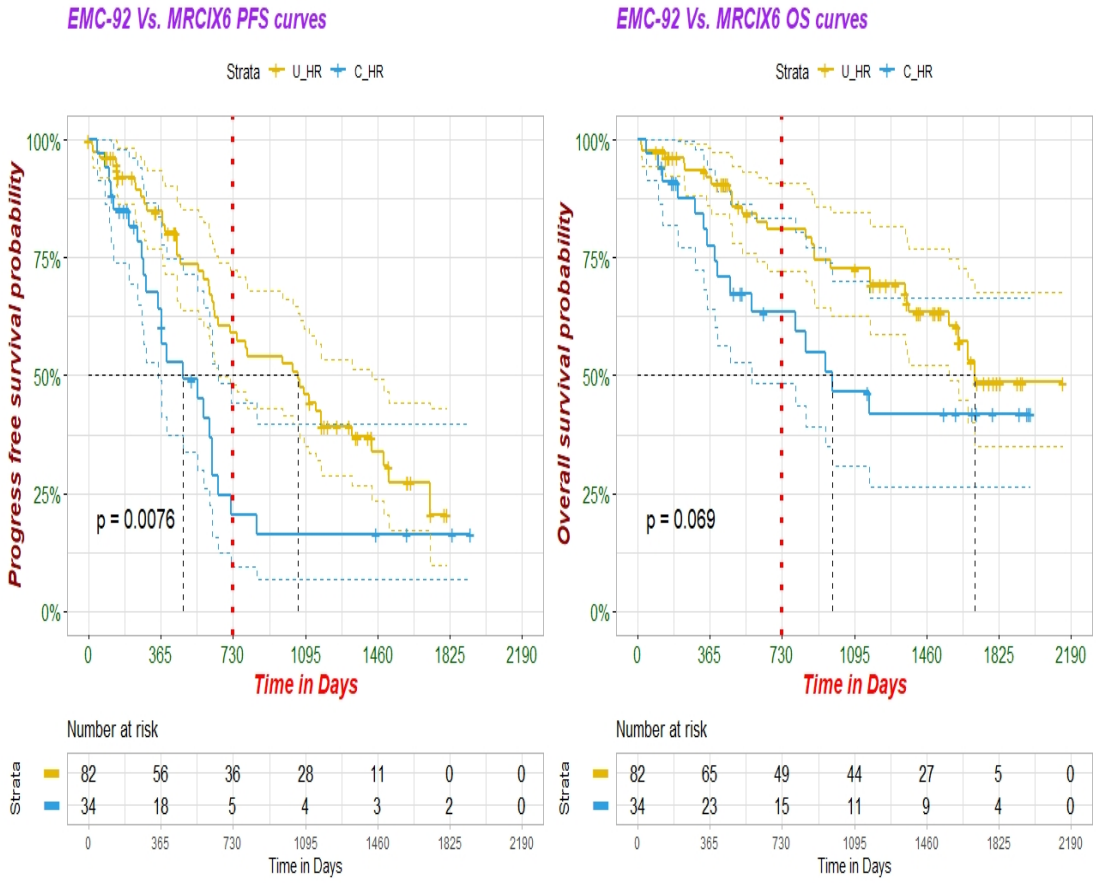

Figure 6: Kaplan-Meier curves of the comparison between uniquely identified HR patients in EMC-92 and its intersection with the other GES.

U-HR means uniquely identified high-risk patients by EMC-92; C-HR means the intersecting HR patients between EMC-92 and the other gene expression signature; HR means high-risk patients; P means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

UAMS17

EMC-92	Total	Low-risk	High-risk
Low-risk	651 (84,8%)	574 (74,8%)	77 (10%)
High-risk	116 (15,1%)	79 (10,3%)	37 (4,8%)

IFM15

EMC-92	Total	Low-risk	High-risk
Low-risk	651 (84,8%)	551 (71,8%)	100 (13%)
High-risk	116 (15,1%)	100 (13%)	16 (2,1%)

UAMS70

EMC-92	Total	Low-risk	High-risk
Low-risk	651 (84,8%)	561 (73,1%%)	90 (11,7%)
High-risk	116 (15,1%)	92 (12%)	24 (3,1%)

HM19

EMC-92	Total	Low-risk	High-risk
Low-risk	651 (84,8%)	567 (73,9%)	84 (11%)
High-risk	116 (15,1%)	84 (11%)	32 (4,1%)

MRCIX6

EMC92	Total	Low-risk	High-risk
Low-risk	651 (84,8%)	569 (74,2%)	82 (10,7%)
High-risk	116 (15,1%)	82 (10,7%)	34 (4,4%)

Figure 7: Confusion matrixes between EMC92 and the other five signatures

Table 9: Comparison of median survival between uniquely identified HR patients in EMC-92 and its intersection with the other GES

Compared Signatures	Survival	n(Event) C_HR	n(Event) U_HR	Median PFS (95% CI) C_HR	Median PFS (95% CI) U_HR	p-Value
EMC-92 vs. UAMS70	PFS	15	32	581 (268 - NA)	800 (634 – 1 176)	0,03
	OS	11	32	1 170 (796 - NA)	Not reached after 1 825 days	0,2
EMC-92 vs. UAMS17	PFS	28	39	393 (349 - 624)	1066 (715 – 1 518)	4e-4
	OS	23	20	796 (469 – 1 670)	Not reached after 1 825 days	3e-5
EMC-92 vs. IFM15	PFS	13	54	552 (278 – 1 149)	750 (625 – 1 176)	0,1
	OS	8	35	983 (849 - NA)	1 704 (1 369 - NA)	0,5
EMC-92 vs. HM19	PFS	24	43	393 (291 - 625)	1 066 (715 – 1 493)	0,001
	OS	17	26	849 (401 - NA)	1 704 (1 574 - NA)	0,01
EMC-92 vs. MRCIX6	PFS	23	44	475 (349 - 652)	1 057 (654 – 1 432)	0,008
	OS	16	27	983 (574 - NA)	1 704 (1 574 - NA)	0,07

Table 10: Comparison of hazard ratios between uniquely identified HR patients in EMC-92 and its intersection with the other GES

Signatures	Survival	2-years survival rate C_HR vs. UE_HR (%)	Hazard ratio	95% CI	Wald p-value
EMC92 vs. UAMS70	PFS	32/52	0,536	(0,3 – 0,958)	0,0352
	OS	70/77	0,612	(0,307 – 1,217)	0,161
EMC92 vs. UAMS17	PFS	23/61	0,425	(0,261 – 0,695)	6,4e-4
	OS	52/88	0,295	(0,162 – 0,540)	7,5e-5
EMC92 vs. IFM15	PFS	32/50	0,605	(0,33 – 1,11)	0,105
	OS	74/76	0,771	(0,357 – 1,664)	0,508
EMC92 vs. HM19	PFS	19/60	0,446	(0,269 – 0,740)	0,00178
	OS	56/83	0,469	(0,254 – 0,867)	0,0157
EMC92 vs. MRCIX6	PFS	20/58	0,504	(0,302 – 0,842)	0,00882
	OS	62/82	0,566	(0,304 – 1,054)	0,0727

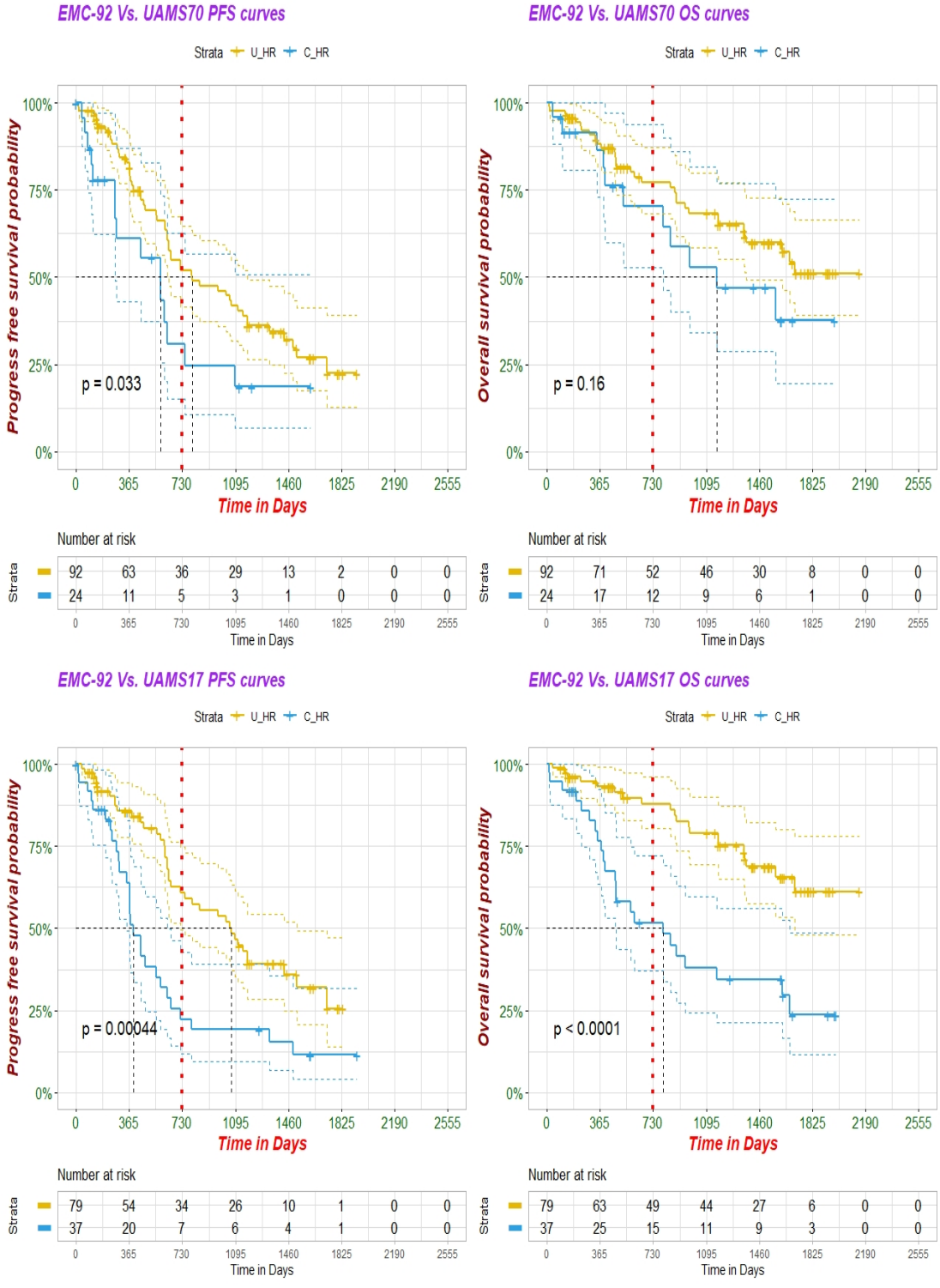

Figure 8: Kaplan-Meier curves of the comparison between uniquely identified HR patients in EMC-92 and its intersection with the other GES.

U-HR means uniquely identified high-risk patients by EMC-92; C-HR means the intersecting HR patients between EMC-92 and the other gene expression signature; HR means high-risk patients; P means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

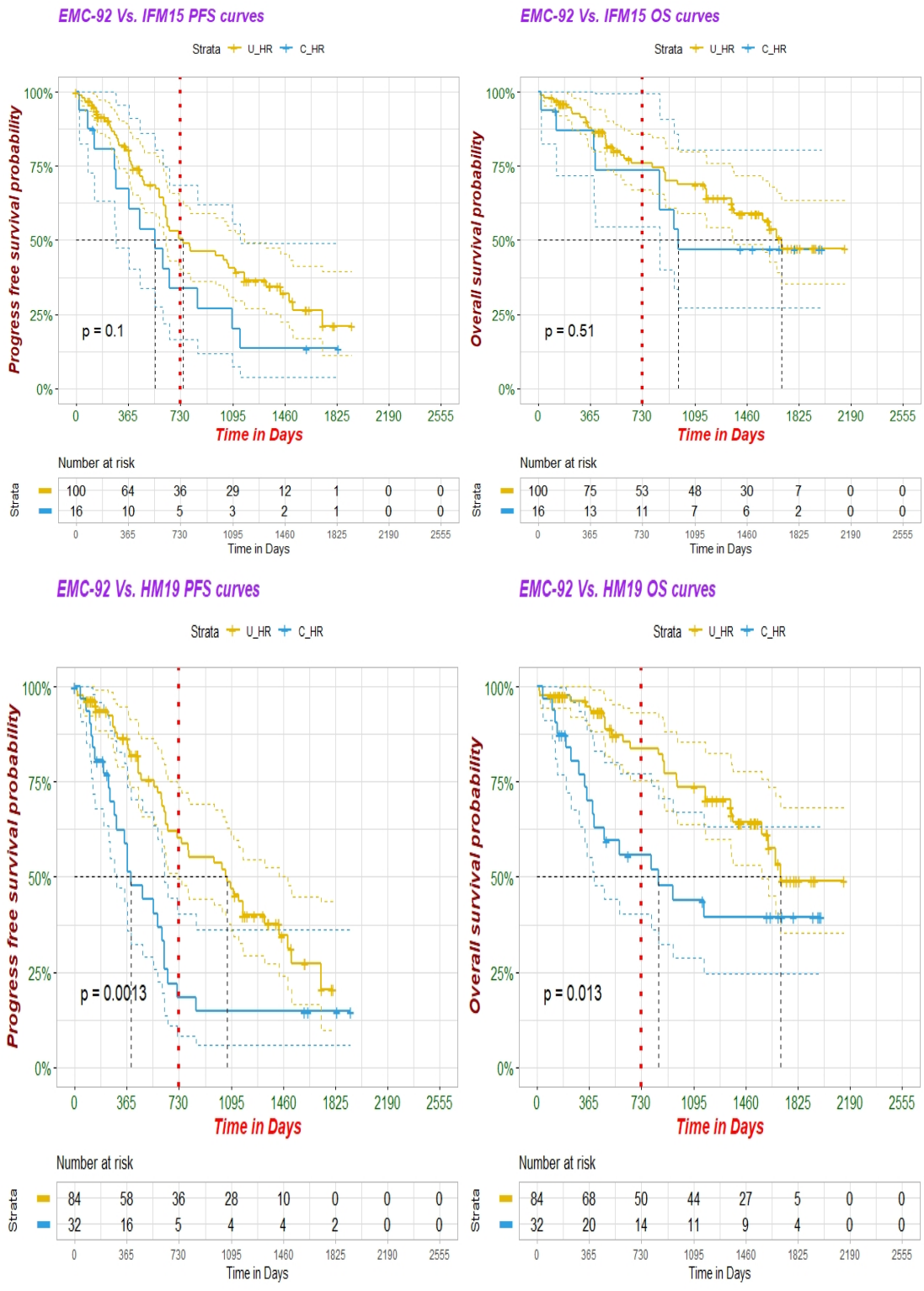

Figure 9: Kaplan-Meier curves of the comparison between uniquely identified HR patients in EMC-92 and its intersection with the other GES.

U-HR means uniquely identified high-risk patients by EMC-92; C-HR means the intersecting HR patients between EMC-92 and the other gene expression signature; HR means high-risk patients; P means the log-rank p-value; alpha (α) = 0.05; the vertical red dotted lines indicate survival at two years; and the horizontal black dotted lines indicates median survival.

CHAPTER 5

DISCUSSION OF RESULTS

The research discussed in this paper adds to our understanding of how well GES performs in risk classification and survival prognostication in multiple myeloma. Clinical and cytogenetic characteristics are being used to predict poor prognosis. These indications, however, do not provide all predictive information. Gene expression analysis can be used to identify patients with a poor prognosis. Thus, based on patients' GEP characterized using NGS, the performance of GES in predicting OS and PFS of NDMM patients was evaluated in this study.

The main findings of particular interest were, first and foremost, that survival performance in OS was significantly better than in PFS, regardless of the GES. This finding is consistent with what Rowan Kuiper et al. (2015) have reported. Furthermore, all GES performed significantly well in categorizing patients into their respective risk groups, with the categorized low-risk patients having significantly longer survival than the high-risk patients, which is highly significantly discriminated by all GES in both PFS and OS (Figure 3, Figure 4, and Figure 5). Chen et al. (2015) confirmed this conclusion, reporting a significant difference in OS between the low-risk and high-risk groups ($p<0,01$) in their study. Despite the fact that Shah et al. (2020) reported that EMC-92 and UAMS70 high-risk patients had considerably shorter OS, only EMC92 had significantly shorter PFS in their study, which contradicts our findings.

Dickens et al. (2010) and R. Kuiper et al. (2012) found that the signatures in their research intersected with a substantially larger proportion of cases with a poor prognosis. Their findings, however, are in contrast to ours, as only two individuals were commonly recognized as high-risk patients across all signatures in our investigation. Furthermore, all of the signatures in our study commonly identified 407 patients as low-risk. In general, the patients in this study had a median PFS and OS of 634 (259 – 1 233) and 1 184 (409 – 1 590) days, respectively, which is considerably less than what is reported by Walker et al. (2019) and Palumbo et al. (2015).

When comparing the number of high-risk patients identified by the various signatures, EMC-92, IFM15, HM19, and MRCIX6, our findings and those of Van Rhee et al. (2014) reveal that GES scores can identify about 15% of high-risk patients. Rowan Kuiper et al. (2015) and R. Kuiper et al. (2012), on the other hand, found that the proportion of high-risk patients varied by signature in their study, with roughly 13% of patients identified as high-risk by either signature. This contradicts our findings because UAMS 70 and UAMS17 both identified the same proportion of high-risk patients (Table 7), which Shaughnessy et al. (2007) also confirmed. Similarly, in our study, the proportion of patients classified as having the worst prognosis by IFM15 and UAMS70 was considerably higher than that found by Dickens et al. (2010). Furthermore, these discrepancies could be due to the prognostic stratification methods used and the platforms used in describing the signatures.

When it comes to establishing how well the GES predicted PFS rate, low-risk patients identified by HM19 and MRCIX6 had a significantly better prognosis, while high-risk patients identified by UAMS70 and IFM15 had a significantly better prognosis (Table 5). Similarly, in OS, the low-risk patients identified by all GES had significantly much better survival prediction, with median survival not reached after 2190 days in each. Similar findings were previously reported by Shaughnessy et al. (2007) and Zhan et al. (2014). Furthermore, in OS, the high-risk patients identified by HM19 and IFM15 performed significantly better in terms of median survival time (Table 6). Additionally, patients identified as high-risk by MRCIX6 had significantly highest hazard ratio in both PFS and OS, whereas those classified as high-risk by EMC-92 and UAMS70 had the significantly lowest hazard ratio in both PFS and OS (Table 7). This finding contradicts the findings of R. Kuiper et al. (2012), but it is consistent with the findings of Rowan Kuiper et al., 2015.

When the predicted 2-year survival rates of the different gene expression signatures were compared, the high-risk patients classified by UAMS17 and MRCIX6 had significantly lower 2-year PFS and OS rates, with 44% in each (Figure 4 H, Figure 5 I). UAMS70 categorized high-risk patients performed significantly better in both survival events, with 2-year survival rates of 52% and 78%, respectively, than any other GES in predicting 2-year survival probability (Figure 5 K and L). Furthermore,

the MRCIX6 signature-classified low-risk patients had a significantly higher predicted 2-year survival probability (Figure 5 G and H). In summary, in all of the GES in this study, the 2-year survival probability was significantly higher in OS than in PFS for all risk groups, notably in the low-risk patients (Figure 3, Figure 4, and Figure 5). Moreau et al. (2014), reported a similar finding too.

In a pair-wise comparison to determine the GES that best explained the observed survival between the EMC-92 gene signature and each of the other GES, the result of our findings regarding the proportion of intersecting high-risk patients (3,1%) between EMC-92 and UAMS70 gene signatures, which is slightly higher than that between EMC-92 and the others (Table 9 and Table 10), is in discordance with the results of (R. Kuiper et al., 2012). Furthermore, the proportion of patients identified exclusively as high-risk by EMC-92 was either greater than or equal to that categorized exclusively by the other GES (Figure 7). Similarly, patients exclusively categorized as high-risk by EMC-92 had a significantly better prognosis for both PFS and OS events than the intersection of patients classified as high-risk between EMC-92 and any of the other GES (Table 9, Table 10, Figure 8, Figure 9, and Figure 6). R. Kuiper et al. (2012) also reported similar findings.

Many GES have been studied and their association with outcomes has been reported in several publications (R. Kuiper et al., 2012; Moreaux et al., 2011). However, only EMC-92 and UAMS70 have been validated and approved into clinical practice so far. That said, in this study, our focus was centered on evaluating NDMM patients' survival using prognostication and stratification performance of the GES based on 85% quantile dichotomization of the express genes weighting coefficients. This, of itself, is a limitation of this study because we do not evaluate the biological aspects of the GES, the therapies patients get, or the clinical and demographic factors of patients that may be associated with survival. Furthermore, GES probe-ids in this study were converted from microarray to RNASeq, resulting in the exclusion of some GES probe-ids (EMC-92 less by 4 and UAMS less by 1) because they could not be matched. This could have an impact on the proportion of patients classified in the risk groups of the said GES. For further research, comparison between the GES and other known methods of survival prediction and risk stratification in MM patients should be

considered together with the signatures' biological features associated with survival to provide a wholistic understanding and best method applicable.

In conclusion, according to the findings of this study, all of the GES performed significantly well in distinguishing risk groups as distinctively as possible, and the proportion of classified predicted risk groups varies less among the GES, with nearly all signatures being equally sensitive in predicting survival outcomes and identifying high-risk cases in NDMM patients. Research using cutting-edge GEP technology is thought to have generated an important new insight into the molecular biology of myeloma. However, these breakthroughs have not resulted in better patient outcomes or care. Nonetheless, in this and previous studies, GEP has been shown to be a suitable candidate for predicting NDMM patients' survival rate and stratifying them for treatment options in clinical trials, so we expect it will be considered useful in clinical settings someday soon. Moreover, using GES as an accurate stratification technique would also be a huge step forward in the clinical care of MM patients, with significant implications for improving their progression-free and overall survival rates, as well as quality of life.

REFERENCES

Amin, S. B., Yip, W. K., Minvielle, S., Broyl, A., Li, Y., Hanlon, B., Swanson, D., Shah, P. K., Moreau, P., Van Der Holt, B., Van Duin, M., Magrangeas, F., Pieter Sonneveld, P., Anderson, K. C., Li, C., Avet-Loiseau, H., & Munshi, N. C. (2014). Gene expression profile alone is inadequate in predicting complete response in multiple myeloma. *Leukemia*, *28*(11), 2229–2234. https://doi.org/10.1038/leu.2014.140

An NGS primer for Multiple Myeloma. (2017).

Anderson, K. (2014). Highlights of the NCCN 19th Annual Conference Anderson chairs the NCCN Guidelines Panel for Multiple Myeloma, and he described the latest treatment advances and what's on the horizon at the NCCN 19th Annual Conference. Multiple Myeloma: Advances Reported. In *Journal of the National Comprehensive Cancer Network* (Vol. 12).

Avet-Loiseau, H. (2010). Ultra high-risk myeloma. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, *2010*, 489–493. https://doi.org/10.1182/asheducation-2010.1.489

Bolli, N., Biancon, G., Moarii, M., Gimondi, S., Li, Y., de Philippis, C., Maura, F., Sathiaseelan, V., Tai, Y. T., Mudie, L., O'Meara, S., Raine, K., Teague, J. W., Butler, A. P., Carniti, C., Gerstung, M., Bagratuni, T., Kastritis, E., Dimopoulos, M., … Munshi, N. C. (2018). Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia*, *32*(12), 2604–2616. https://doi.org/10.1038/s41375-018-0037-9

Chen, F., Xiang, C. X., Zhou, Y., Ao, X. S., Zhou, D. Q., Peng, P., Zhang, H. Q., Liu, H. D., & Huang, X. (2015). Gene expression profile for predicting survival of patients with meningioma. *International Journal of Oncology*, *46*(2), 791–797. https://doi.org/10.3892/ijo.2014.2779

Chng, W. J., Chung, T.-H., Kumar, S., Usmani, S., Munshi, N., Avet-Loiseau, H., Goldschmidt, H., Durie, B., & Sonneveld, P. (2016). Gene signature combinations improve prognostic stratification of multiple myeloma patients. *Leukemia*, *30*, 1071–1078. https://doi.org/10.1038/leu.2015.341

D'Agostino, M., Zaccaria, G. M., Ziccheddu, B., Rustad, E. H., Genuardi, E., Capra, A., Oliva, S., Auclair, D., Yesil, J., Colucci, P., Keats, J. J., Gambella, M., Bringhen, S., Larocca, A., Boccadoro, M., Bolli, N., Maura, F., & Gay, F. (2020). Early Relapse Risk in Patients with Newly Diagnosed Multiple Myeloma Characterized by Next-generation Sequencing. *Clinical Cancer Research*, *26*(18), 4832–4841. https://doi.org/10.1158/1078-0432.CCR-20-0951

Decaux, O., Lodé, L., Magrangeas, F., Charbonnel, C., Gouraud, W., Jézéquel, P.,

Attal, M., Harousseau, J.-L. L., Moreau, P., Bataille, R., Campion, L., Avet-Loiseau, H., & Minvielle, S. (2008). Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: A study of the Intergroupe Francophone du Myélom. *Journal of Clinical Oncology*, *26*(29), 4798–4805. https://doi.org/10.1200/JCO.2007.13.8545

Decaux, O., Lodé, L., Magrangeas, F., Charbonnel, C., Gouraud, W., Jézéquel, P., Attal, M., Harousseau, J.-L., Moreau, P., Bataille, R., Campion, L., Avet-Loiseau, H., & Minvielle, S. (2016). Prediction of Survival in Multiple Myeloma Based on Gene Expression Profiles Reveals Cell Cycle and Chromosomal Instability Signatures in High-Risk Patients and Hyperdiploid Signatures in Low-Risk Patients: A Study of the Intergroupe Francophone du Myélom. *Https://Doi.Org/10.1200/JCO.2007.13.8545*, *26*(29), 4798–4805. https://doi.org/10.1200/JCO.2007.13.8545

Dickens, N. J., Walker, B. A., Leone, P. E., Johnson, D. C., Brito, J. L., Zeisig, A., Jenner, M. W., Boyd, K. D., Gonzalez, D., Gregory, W. M., Ross, F. M., Davies, F. E., & Morgan, G. J. (2010). Homozygous deletion mapping in myeloma samples identifies genes and an expression signature relevant to pathogenesis and outcome. *Clinical Cancer Research*, *16*(6), 1856–1864. https://doi.org/10.1158/1078-0432.CCR-09-2831

EH, van B., MH, van V., R, K., L, de B., KC, A., A, C., S, J., A, J., SK, K., JB, L., D, A., S, L., D, R., P, R., DS, S., AK, S., S, T., R, V., TM, Z., & R, F. (2017). Prognostic Validation of SKY92 and Its Combination With ISS in an Independent Cohort of Patients With Multiple Myeloma. *Clinical Lymphoma, Myeloma & Leukemia*, *17*(9), 555–562. https://doi.org/10.1016/J.CLML.2017.06.020

Fonseca, R., Bergsagel, P. L., Drach, J., Shaughnessy, J., Gutierrez, N., Stewart, A. K., Morgan, G., Ness, V., Chesi, M., Minvielle, S., Neri, A., Barlogie, B., Kuehl, W. M., Liebisch, P., Davies, F., Chen-Kiang, S., Durie, B., Carrasco, R., Sezer, O., … Avet-Loiseau, H. (2009). International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. *Leukemia*, *23*, 2210–2221. https://doi.org/10.1038/leu.2009.174

Hassan, H., & Szalat, R. (2021). Genetic predictors of mortality in patients with multiple myeloma. *Application of Clinical Genetics*, *14*, 241–254. https://doi.org/10.2147/TACG.S262866

JR, M., D, D., V, R., CB, R., FK, B., SR, H., A, D., R, F., T, S., RA, K., Y, L., SJ, R., S, K., PL, B., SR, Z., N, L., MT, D., P, K., SM, A., … MQ, L. (2013). Management of newly diagnosed symptomatic multiple myeloma: updated Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) consensus guidelines 2013. *Mayo Clinic Proceedings*, *88*(4), 360–376. https://doi.org/10.1016/J.MAYOCP.2013.01.019

Kuiper, R., Broyl, A., De Knegt, Y., Van Vliet, M. H., Van Beers, E. H., Van Der Holt, B., El Jarari, L., Mulligan, G., Gregory, W., Morgan, G., Goldschmidt, H., Lokhorst, H. M., Van Duin, M., & Sonneveld, P. (2012). A gene expression signature for high-risk multiple myeloma. *Leukemia*, *26*(11), 2406–2413. https://doi.org/10.1038/leu.2012.127

Kuiper, Rowan, Van Duin, M., Van Vliet, M. H., Broijl, A., Van Der Holt, B., Jarari, L. El, Van Beers, E. H., Mulligan, G., Avet-Loiseau, H., Gregory, W. M., Morgan, G., Goldschmidt, H., Lokhorst, H. M., & Sonneveld, P. (2015). Prediction of high- and low-risk multiple myeloma based on gene expression and the International Staging System. *Blood*, *126*(17), 1996–2004. https://doi.org/10.1182/blood-2015-05-644039

Kumar, S., Fonseca, R., Ketterling, R. P., Dispenzieri, A., Lacy, M. Q., Gertz, M. A., Hayman, S. R., Buadi, F. K., Dingli, D., Knudson, R. A., Greenberg, A., Russell, S. J., Zeldenrust, S. R., Lust, J. A., Kyle, R. A., Bergsagel, L., & Rajkumar, S. V. (2012). Trisomies in multiple myeloma: Impact on survival in patients with high-risk cytogenetics. *Blood*, *119*(9), 2100–2105. https://doi.org/10.1182/blood-2011-11-390658

Kumar, S. K., & Vincent Rajkumar, S. (2018). The multiple myelomas — current concepts in cytogenetic classification and therapy. *Nature Reviews Clinical Oncology*. https://doi.org/10.1038/s41571-018-0018-y

Kyle, R. A., Gertz, M. A., Witzig, T. E., Lust, J. A., Lacy, M. Q., Dispenzieri, A., Fonseca, R., Rajkumar, S. V., Offord, J. R., Larson, D. R., Plevak, M. E., Therneau, T. M., & Greipp, P. R. (2003). Review of 1027 patients with newly diagnosed multiple myeloma. *Mayo Clinic Proceedings*, *78*(1), 21–33. https://doi.org/10.4065/78.1.21

Landgren, O., Kyle, R. A., Pfeiffer, R. M., Katzmann, J. A., Caporaso, N. E., Hayes, R. B., Dispenzieri, A., Kumar, S., Clark, R. J., Baris, D., Hoover, R., & Rajkumar, S. V. (2009). Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: A prospective study. *Blood*, *113*(22), 5412–5417. https://doi.org/10.1182/blood-2008-12-194241

Lonial, S., Boise, L. H., & Kaufman, J. (2015). How I treat high-risk myeloma. *Blood*, *126*(13), 1536–1543. https://doi.org/10.1182/BLOOD-2015-06-653261

Ludwig, H., Durie, S. N., Meckl, A., Hinke, A., & Durie, B. (2020). *Multiple Myeloma Incidence and Mortality Around the Globe; Interrelations Between Health Access and Quality, Economic Resources, and Patient Empowerment*. https://doi.org/10.1634/theoncologist.2020-0141

Manasanch, E. E., Berrios, D., Fountain, E., Claussen, C. M., Chuang, T., Kaufman, G., Amini, B., Bashir, Q., Nieto, Y., Qazilbash, M., Patel, K., Thomas, S. K., Weber, D. M., Berkova, Z., Toruner, G., Lin, P., Feng, L., Lee, H. C., Orlowski, R. Z., … C, K. (2021). Gene expression profiling predicts relapse-free and

overall survival in newly diagnosed myeloma patients treated with novel therapies. *British Journal of Haematology*, *192*(4), e115–e120. https://doi.org/10.1111/bjh.17344

Martinez-Lopez, J., Lahuerta, J. J., Pepin, F., GonzálezGonz, M., Barrio, S., Ayala, R., Puig, N., Montalban, M. A., Paiva, B., Weng, L., JiménezJim, C., Sopena, M., Moorhead, M., Cedena, T., Rapado, I., Victoria Mateos, M., Rosiñ ol, L., Oriol, A., Blanchard, M. J., … García-Sanz, R. (2014). *Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma*. https://doi.org/10.1182/blood-2014-01

Moreau, P., Cavo, M., Sonneveld, P., Rosinol, L., Attal, M., Pezzi, A., Goldschmidt, H., Lahuerta, J. J., Marit, G., Palumbo, A., Van Der Holt, B., Bladé, J., Petrucci, M. T., Neben, K., San Miguel, J., Patriarca, F., Lokhorst, H., Zamagni, E., Hulin, C., … Mary, J.-Y. (2014). Combination of International Scoring System 3, High Lactate Dehydrogenase, and t(4;14) and/or del(17p) Identifies Patients With Multiple Myeloma (MM) Treated With Front-Line Autologous Stem-Cell Transplantation at High Risk of Early MM Progression-Related Death. *J Clin Oncol*, *32*, 2173–2180. https://doi.org/10.1200/JCO.2013.53.0329

Moreaux, J., Klein, B., Bataille, R., Descamps, G., Maïga, S., Hose, D., Goldschmidt, H., Jauch, A., Rème, T., Jourdan, M., Amiot, M., & Pellat-Deceunynck, C. (2011). A high-risk signature for patients with multiple myeloma established from the molecular classification of human myeloma cell lines. *Haematologica*, *96*(4), 574–582. https://doi.org/10.3324/haematol.2010.033456

Morgan, G. J., Walker, B. A., & Davies, F. E. (2012). *The genetic architecture of multiple myeloma*. https://doi.org/10.1038/nrc3257

Mosquera Orgueira, A., González Pérez, M. S., Díaz Arias, J. Á., Antelo Rodríguez, B., Alonso Vence, N., Bendaña López, Á., Abuín Blanco, A., Bao Pérez, L., Peleteiro Raíndo, A., Cid López, M., Pérez Encinas, M. M., Bello López, J. L., & Mateos Manteca, M. V. (2021). Survival prediction and treatment optimization of multiple myeloma patients using machine-learning models based on clinical and gene expression data. *Leukemia*. https://doi.org/10.1038/s41375-021-01286-2

Palumbo, A., Avet-Loiseau, H., Oliva, S., Lokhorst, H. M., Goldschmidt, H., Rosinol, L., Richardson, P., Caltagirone, S., Lahuerta, J. J., Facon, T., Bringhen, S., Gay, F., Attal, M., Passera, R., Spencer, A., Offidani, M., Kumar, S., Musto, P., Lonial, S., … Moreau, P. (2015). Revised international staging system for multiple myeloma: A report from international myeloma working group. *Journal of Clinical Oncology*, *33*(26), 2863–2869. https://doi.org/10.1200/JCO.2015.61.2267

Paszekova, H., Kryukov, F., Kubiczkova, L., Hajek, R., & Sevcikova, S. (2014). High-risk multiple myeloma: Different definitions, different outcomes? *Clinical Lymphoma, Myeloma and Leukemia*, *14*(1), 24–30.

https://doi.org/10.1016/j.clml.2013.09.004

Penaherrera, D., Skerget, S., Christofferson, A., Aldrich, J., Nasser, S., Legendre, C., Boateng, M., Cuyugan, L., Adkins, J., Tassone, E., Yesil, J., Auclair, D., Liang, W., Lonial, S., & Keats, J. J. (2018). Development and Validation of a High Risk Multiple Myeloma Gene Expression Index from RNA Sequencing: An Mmrf Commpass Analysis. *Blood*, *132*(Supplement 1), 1895. https://doi.org/10.1182/BLOOD-2018-99-119610

Piazzi, G., Acosta, J., Smith, B., & Faye, N. (2016). *Multiple Myeloma Overview: Novel therapies, the role of biomarkers and imaging, operational aspects*. www.quintilesims.com

Prideaux, S. M., Conway O'brien, E., & Chevassut, T. J. (2014). *The Genetic Architecture of Multiple Myeloma*. https://doi.org/10.1155/2014/864058

Rajan, A. M., & Rajkumar, S. V. (2015). Interpretation of cytogenetic results in multiple myeloma for clinical practice. *Blood Cancer Journal*, *5*, 365. https://doi.org/10.1038/bcj.2015.92

Rajkumar, S. V. (2018). Multiple myeloma: 2018 update on diagnosis, risk-stratification, and management. *Am J Hematol*, *93*, 1091–1110. https://doi.org/10.1002/ajh.25117

Shah, V., Sherborne, A. L., Johnson, D. C., Ellis, S., Price, A., Chowdhury, F., Kendall, J., Jenner, M. W., Drayson, M. T., Owen, R. G., Gregory, W. M., Morgan, G. J., Davies, F. E., Cook, G., Cairns, D. A., Houlston, R. S., Jackson, G., & Kaiser, M. F. (2020). Predicting ultrahigh risk multiple myeloma by molecular profiling: an analysis of newly diagnosed transplant eligible myeloma XI trial patients. *Leukemia 2020 34:11*, *34*(11), 3091–3096. https://doi.org/10.1038/s41375-020-0750-z

Sharma, A., & Kumar, N. (2018). *The global incidence and prevalence of multiple myeloma over the next 10 years (2017-2027)*. *08*. https://doi.org/10.4172/2161-0681-C1-045

Shaughnessy, J. D., Zhan, F., Burington, B. E., Huang, Y., Colla, S., Hanamura, I., Stewart, J. P., Kordsmeier, B., Randolph, C., Williams, D. R., Xiao, Y., Xu, H., Epstein, J., Anaissie, E., Krishna, S. G., Cottler-Fox, M., Hollmig, K., Mohiuddin, A., Pineda-Roman, M., … Barlogie, B. (2007). A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*, *109*(6), 2276–2284. https://doi.org/10.1182/BLOOD-2006-07-038430

Short, K. D., Rajkumar, S. V, Larson, D., Buadi, F., Hayman, S., Dispenzieri, A., Gertz, M., Kumar, S., Mikhael, J., Roy, V., Kyle, R. A., & Lacy, M. Q. (2011). Incidence of extramedullary disease in patients with multiple myeloma in the era of novel therapy, and the activity of pomalidomide on extramedullary

myeloma. *Leukemia*, *25*, 906–908. https://doi.org/10.1038/leu.2011.29

Smadja, N. V., Bastard, C., Brigaudeau, C., Leroux, D., & Fruchart, C. (2001). Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood*, *98*(7), 2229–2238. https://doi.org/10.1182/blood.V98.7.2229

Szalat, R., Avet-Loiseau, H., & Munshi, N. C. (2016). Gene expression profiles in myeloma: Ready for the real world? *Clinical Cancer Research*, *22*(22), 5434–5442. https://doi.org/10.1158/1078-0432.CCR-16-0867

Szalat, R., & Munshi, N. C. (2016). Next-Generation Sequencing Informing Therapeutic Decisions and Personalized Approaches. *American Society of Clinical Oncology Educational Book*, *36*(36), e442–e448. https://doi.org/10.14694/edbk_159017

TM, T., RA, K., LJ, M., DR, L., JT, B., CL, C., A, D., S, K., JA, K., JR, C., & SV, R. (2012). Incidence of monoclonal gammopathy of undetermined significance and estimation of duration before first clinical recognition. *Mayo Clinic Proceedings*, *87*(11), 1071–1079. https://doi.org/10.1016/J.MAYOCP.2012.06.014

Van Rhee, F., Giralt, S., & Barlogie, B. (2014). *The future of autologous stem cell transplantation in myeloma*. https://doi.org/10.1182/blood-2014-03-561985

Walker, B. A., Mavrommatis, K., Wardell, C. P., Ashby, T. C., Bauer, M., Davies, F., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., … Morgan, G. (2019). A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*, *33*(1), 159–170. https://doi.org/10.1038/s41375-018-0196-8

Zhan, F., Barlogie, B., Mulligan, G., Jr, J. D. S., & Bryant, B. (2014). *High-risk myeloma : a gene expression – based risk-stratification model for newly diagnosed diagnosed multiple myeloma treated with high-dose therapy is predictive of outcome in relapsed disease treated with single-agent bortezomib or high-dose dexamethas*. *111*(2), 968–970.

Zhan, F., Huang, Y., Colla, S., Stewart, J. P., Hanamura, I., Gupta, S., Epstein, J., Yaccoby, S., Sawyer, J., Burington, B., Anaissie, E., Hollmig, K., Pineda-Roman, M., Tricot, G., Van Rhee, F., Walker, R., Zangari, M., Crowley, J., Barlogie, B., & Shaughnessy, J. D. (2006). *The molecular classification of multiple myeloma*. https://doi.org/10.1182/blood-2005-11-013458