國立臺灣大學醫學院臨床醫學研究所

碩士論文

Graduate Institute of Clinical Medicine
College of Medicine
National Taiwan University

Master Thesis

等位基因突變頻率於骨髓化生不良症候群患者之預後意義

Effect of mutation allele frequency on the risk stratification of myelodysplastic syndrome patients

李婉瑄

Wan-Hsuan Lee

指導教授：侯信安 博士

Advisor: Hsin-Ann Hou, Ph.D.

中華民國 112 年 6 月

June 2023

doi:10.6342/NTU202301887

誠摯感謝

我大學就讀國立陽明大學醫學系，畢業後於台大醫院接受完整的畢業後一般醫學訓練(post-graduate year training, PGY)以及內科住院醫師訓練，在踏入臨床之後，除了學習照護病人，也看到更多人情冷暖，同時堅定了我對於內科醫學的熱忱。於內科專科醫師考試通過後，我選擇了台大血液科做為內科次專科訓練，科內的老師們作為我未來生涯的榜樣，此時經由學長姐及師長們的帶領，開始初步接觸血液學相關研究，參與了由田蕙芬教授、周文堅教授及侯信安教授帶領建構的骨髓疾病資料庫整理，從做研究最基礎的資料收集與整理開始學習，因此開始產生做研究的興趣，並對研究與資料分析有了基礎的概念。

在過去兩年，就讀台大臨床醫學研究所醫學研究組碩士班，系統性的課程設計，包括生物統計學、分子生物學、臨床試驗概論及臨床研究方法等，讓我可以完整並逐步掌握做研究需要的方法與切入點，除了學術知識以外，老師們皆是各領域研究中的佼佼者，透過他們在課堂上的經驗分享與傳承，讓我了解做研究所需要的堅持、努力、機遇與靈感。針對自己研究過程中遭遇的困難，侯教授及眾多同事都提供許多解決的方法與經驗的傳承，特別感謝臨醫所邵文逸老師及陳祈玲老師於統計領域的指導與建議，同時在染色體分析，次世代基因定序分析的過程中，得到諸多同事與師長的協助，包括蔡承宏醫師、林建嶔醫師、郭遠燁博士等等，因為有上述同仁前輩的協助，才能在這兩年中在侯信安教授的指導下，完成了三篇論文，並將我們的研究成果透過血液病年會，跟全台血液學專家分享，此外亦參與了很多國際上重要的會議，透過跟國外專家的交流，可以更清楚國內研究的優勢與弱勢，強化自己做研究的想法與能力，並更深刻的了解醫學研究之於照護病人的重要性。最後感謝我的父母從小對我的栽培與教導，還有我先生祥瑋這段時間對我的支持及包容。

doi:10.6342/NTU202301887

中文摘要

關鍵詞：骨髓化生不良症候群、風險分類模型、等位基因突變頻率、預後、次世代基因定序

一、背景

骨髓化生不良症候群 (myelodysplastic neoplasms, MDS) 是一種造血幹細胞疾病，臨床表徵與預後差異性大，以造血功能不良、血球低下、細胞化生不良與特定染色體變化為主要表徵，部分患者會快速轉變為急性骨髓性白血病，而基因突變模式與預後息息相關，它是導致疾病發生的重要機轉之一，也是使患者疾病惡化的關鍵因素。如何正確區分高風險患者並給予積極治療是目前臨床上急待解決之問題，目前已有數個可以將患者做風險分級的模型包括：International Prognostic Scoring System (IPSS)、revised IPSS (IPSS-R)、World Health Organization Classification-based Prognostic Scoring System 以及 MD Anderson Prognostic Scoring System，然而這些模型並不包括基因突變模式，包括等位基因突變頻率 (variant allele frequency, VAF) 之表現，因此本研究分析等位基因突變頻率在骨髓化生不良症候群患者之預後意義，並探討去甲基化藥物治療與異體骨髓幹細胞移植的角色。

二、方法與程序

根據 2016 年世界衛生組織針對骨髓化生不良症候群的疾病分類與診斷標準，排除先前接受過化學治療或是惡性血液腫瘤病史之患者，總共收集 698 位原發性骨髓化生不良症候群患者的骨髓檢體，使用 TruSight myeloid sequencing panel (Illumina)及 HiSeq 平台，以次世代基因定序 (next generation sequencing) 的方式完成 54 個基因突變及等位基因突變頻率之分析，針對 *CEBPA* 基因另外使用 Sanger 定序方式做確認。而 *FLT3*-ITD 則使用 PCR (polymerase chain reaction)，及毛細管電泳 (fluorescence capillary electrophoresis)方式分析。此研究經臺大醫院研究倫理委員會核可，所有患者皆簽屬臨床試驗同意書 (核可號 201709072RINC)。

三、統計分析

連續變相使用曼惠特尼檢定，類別變相使用費雪或卡方檢定，皮爾森相關係數檢定用以判定等位基因突變頻率與臨床指標的相關性，相關係數大於/小於 0.4/-0.4 被認定為有意義的正/負相關；無白血病存活率 (leukemia-free survival) 定義為診斷至轉變為急性骨髓性白血病，或診斷至死亡的時間；整體存活率 (overall survival) 定義為診斷至死亡的時間。最大選擇統計量 (maximally selected rank statistics) 用以尋找並檢定等位基因突變頻率與預後有統計相關的閾值，並進一步使用自助法 (bootstrapping) 做內部驗證，所有檢定以 p 值小於 0.05 視為有統計意義。

四、結果

甲、病人特性

診斷年齡中位數為 66.5 歲，男性居多 (63.3%)，追蹤中位數為 54.7 個月，根據

doi:10.6342/NTU202301887

2016 年世界衛生組織診斷標準，52%的患者屬於高風險族群 (MDS-excess blasts)，若根據 IPSS-R 風險模型分類，22.5%及 22.2%患者分別屬於高或極高風險類別；24.4%患者有接受去甲基化藥物治療，16.1%患者接受異體骨髓幹細胞移植；23%患者於追蹤期間轉變為急性骨髓性白血病，49.3%患者於追蹤期間內死亡。

乙、基因突變模式

71.5%患者至少帶有一個基因突變，最常見的基因突變為 *ASXL1* (20.3%)、*TET2* (14.3%)、 *SF3B1* (13.8%)、*RUNX1* (12.6%)、*STAG2* (12.5%) 及 *TP53* (12.3%)。

丙、預後分析

單變相預後分析發現年紀較大，IPSS-R 高風險類別，曾接受去甲基化藥物治療，*TET2*、 *IDH2*、*ASXL1*、*EZH2*、*CBL*、*RUNX1*、*U2AF1*、*SRSF2*、*ZRSR2*、*STAG2* 及 *TP53* 突變與較短的無白血病存活率和整體存活率有關，而 *DNMT3A*、*BCORL1* 及 *NRAS* 突變則與較短的無白血病存活率相關；接受異體幹細胞移植、女性及 *SF3B1* 突變與較長的無白血病存活率和整體存活率相關。而在多變項分析中，IPSS-R 高風險類別、*DNMT3A*、*TET2*、*IDH2*、*CBL* 及 *TP53* 為較差的獨立預後因子，而女性及接受異體幹細胞移植為較好的獨立預後因子。

丁、等位基因突變頻率與臨床指標及預後的關聯

高 *IDH2* 等位基因突變頻率與較低的血色素 (r = -0.496, p = .009) 與骨髓芽細胞比例 (r = -0.432, p = .024) 相關，帶有高等位基因突變頻率之 *DNMT3A* (閾值 40%, hazard ratio [HR] 2.87, p < 0.001)、*TET2* (45%, HR 2.55, p < 0.001)、*ASXL1* (20%, HR 2.24, p < 0.001)、*EZH2* (40%, HR 2.12, p = 0.036)、*SETBP1* (15%, HR 1.94, p = 0.024)、*BCOR* (80%, HR 2.49, p = 0.043)、*SRSF2* (50%, HR 3.65, p = 0.002)、*ZRSR2* (60%, HR 2.91, p < 0.001) 及 *TP53* (25%, HR 7.84, p < 0.001)，相較於無突變患者其無白血病存活率較短，除了 *EZH2*、*SETBP1* 及 *BCOR* 之外，其他基因之高等位基因突變頻率亦與較短的整體存活率相關。在多變項分析中，女性及有接受異體幹細胞移植患者預後較好，年紀大、高風險 IPSS-R 類別、帶有 *IDH2* 及 *CBL* 突變的患者無白血病存活率和整體存活率較短，而帶有 *U2AF1* 突變、*DNMT3A* 或 *ZRSR2* 高等位基因突變頻率患者整體存活率較短。若患者帶有等位基因突變頻率> 25%的 *TP53* 突變，其預後相較等位基因突變頻率≤ 25%或不帶有 *TP53* 突變的患者差。

若將上述帶有獨立預後預測意義的基因 (與無白血病存活率相關之基因：*IDH2*、*CBL* 及 *TP53* 突變；與整體存活率相關之基因：高等位基因突變頻率之 *DNMT3A*、*ZRSR2* 突變，與 *IDH2*、*CBL*、*U2AF1* 及 *TP53* 突變) 納入現有的風險分類模型 (IPSS-R) 中，可以將患者做更好的分類，舉例來說，原本同樣被分類為低或極低風險的 IPSS-R 患者，若帶有不好的基因突變，相較於其他同樣類別的患者，預後顯著較差。針對整體存活率，若將這些不好的預後因子納入 IPSS-R 風險分類模型，分別將有 8.9%、17.9%及 34.4%自原屬於低或極低風險、中等風險或高風險 IPSS-R 分類中被重新歸類。

戊、去甲基化藥物及異體幹細胞移植之影響

進一步分析治療對於預後的影響，使用去甲基化藥物及異體幹細胞移植無法改

doi:10.6342/NTU202301887

善帶有不好基因的患者預後，然而若針對個別特定基因作分析，若帶有 *U2AF1* 突變的患者，接受去甲基化藥物治療可顯著改善其預後。

五、討論及結論

針對等位基因突變頻率與預後之關聯，此研究為目前已知最完整的分析之一，在本研究中發現若帶有高等位基因突變頻率之 *DNMT3A*、*TET2*、*ASXL1*、*SRSF2*、*ZRSR2* 及 *TP53* 突變患者預後較差，患者若帶有上述基因之低等位基因突變頻率，其預後與未帶有上述突變之患者相當，而 *TP53* 的等位基因突變頻率閾值 (25%) 可以將患者分為高、低及未帶有突變三群，三群患者預後不同，以高者為最差，未帶有突變者最好。在多變項分析中，高等位基因突變頻率之 *DNMT3A*、*ZRSR2* 突變，與 *IDH2*、*CBL*、*U2AF1* 及 *TP53* 突變為獨立預後因子，若將這些高風險分子預後因子納入現有風險分類系統 (IPSS-R)，可以將病患做更好的風險分級。除此之外，我們首次證實帶有 *U2AF1* 突變的患者，接受去甲基化藥物治療可以顯著改善其預後。

骨髓化生不良症候群的病生理機轉複雜，不同分類其預後差異大，隨著基因定序技術的進展，可以更加完整分析基因突變模式對於疾病進展與預後的影響，除了基因突變的有無外，目前已有眾多證據指出等位基因突變頻率跟預後亦有相關，然而目前大多研究都著重分析 *TP53* 突變，先前有研究發現分別以 20%及 50%為閾值，可將帶有 *TP53* 突變的患者分成預後不同的三個族群，亦有統合分析 (meta-analysis) 認為以 20%為分界，區分高或低突變負荷量 (mutation burden)。除此之外，因骨髓化生不良症候群患者僅有少數人帶有 *ZRSR2* 突變，因此先前的研究無法分析 *ZRSR2* 之等位基因突變頻率與預後的關係，而本研究首次得以證實 *ZRSR2* 高等位基因突變頻率與較差的預後有關。

本篇研究的限制包括其為一回顧性研究，且針對等位基因突變頻率之閾值與風險分類模型的預測力缺少外部效度 (external validation) 的驗證，然這項研究成果仍可提供未來發展新的風險分類模型的雛型與方向，也促進骨髓化生不良症候群患者個人化且精準醫療發展之可能性。

doi:10.6342/NTU202301887

英文摘要

Keywords: myelodysplastic syndrome, risk stratification, variant allele frequency, prognosis, next generation sequencing

A. Introduction

Myelodysplastic syndrome (MDS) is a diverse group of clonal myeloid neoplasms, characterized by clinical and genetic heterogeneity, and increased risk of acute myeloid leukemia (AML) transformation. The accumulation of mutations is involved in MDS pathogenesis, which gives rise to clonal architecture and leads to disease progression. Several prognostic models, including the International Prognostic Scoring System (IPSS), revised IPSS (IPSS-R), World Health Organization Classification-based Prognostic Scoring System and MD Anderson Prognostic Scoring System have been developed to risk-stratify MDS patients. Mounting evidences demonstrate that the addition of mutation data improves the prognostic stratification. In addition to mutational profiles, the variant allele frequency (VAF) of individual mutations also influence the prognosis in MDS patients. In the present study, we performed comprehensive VAF analyses, focusing on the correlation between VAF and survival. We further analyzed the impacts of allogenic hematopoietic stem cell transplantation (HSCT) and hypomethylating agents (HMA) on outcomes considering various VAF in different genes.

B. Material and Method

A total of 698 primary MDS patients with adequate cryopreserved bone marrow samples for deep-targeted sequencing and IPSS-R data were recruited. The diagnoses were based on the 2016 World Health Organization (WHO) classification. Patients with antecedent chemotherapy/radiotherapy or hematologic malignancies were excluded. This study was approved by the Research Ethics Committee of the National Taiwan University Hospital; and written informed consent was obtained from all participants (approval number: 201709072RINC).

TruSight myeloid sequencing panel (Illumina) and the HiSeq platform were adopted to analyze the gene alterations and mutant allele burdens of 54 myeloid-neoplasm relevant genes. Because of the sequencing sensitivity issue, we verified *CEBPA* mutations via Sanger sequencing. Analysis of *FLT3*-ITD was performed via polymerase chain reaction (PCR), followed by fluorescence capillary electrophoresis

C. Statistical analysis

Pairwise comparison between continuous variables was performed using the Mann–Whitney U test, and the Fisher's exact test or the χ^2 test was performed for discrete variables. Pearson's correlation coefficient was used to assess the strength and direction of the linear relationships between VAF and clinical parameters. The correlation coefficient (r) greater or lower than 0.4/-0.4 was thought to be positive/negative correlated. Leukemia-free survival (LFS) was defined as the duration from the date of diagnosis to the last follow-up, documented leukemia transformation, or death from any cause, whichever occurred first. Overall survival (OS) was the duration from the date of diagnosis to the last follow-up or death from any cause, whichever occurred first. Maximally selected rank statistics were applied for VAF exploration. All *P* values were two-sided and considered statistically significant if

doi:10.6342/NTU202301887

<0.05.

D. Results

a. Demographic features

The median age was 66.5 years, with male predominance (63.3%). The median follow-up time was 54.7 months. When categorized by the 2016 WHO classification, over half (52.0%) of the patients had myelodysplastic syndrome with excess blasts, including EB1 (23.8%) and EB2 (28.2%). A total of 71.1% patients had IPSS-R intermediate (26.4%), high (22.5%), or very high-risk disease (22.2%). Regarding treatments, 24.4% of patients received HMA and 16.1% underwent allogeneic HSCT. Twenty three percent of patients experienced leukemic transformation and 49.3% died at the end of follow-up. Overall, 71.5% had at least one gene mutation. The most common mutation in the cohort was *ASXL1* mutation (20.3%), followed by *TET2* (14.3%), *SF3B1* (13.8%), *RUNX1* (12.6%), *STAG2* (12.5%), and *TP53* mutations (12.3%).

b. Survival analyses

In univariable analysis, older age, higher-risk IPSS-R, HMA treatment, and presence of mutations in *TET2*, *IDH2*, *ASXL1*, *EZH2*, *CBL*, *RUNX1*, *U2AF1*, *SRSF2*, *ZRSR2*, *STAG2*, and *TP53* were significantly associated with both shorter LFS and OS, while *DNMT3A*, *BCORL1*, and *NRAS* mutations conferred shorter LFS. Receiving HSCT, female sex, and mutated *SF3B1* were favorable factors for LFS and OS. Multivariable analysis showed that older age, higher-risk IPSS-R, and *DNMT3A*, *TET2*, *IDH2*, *CBL*, and *TP53* mutations were independent poor risk factors, while female sex and receiving HSCT were good risk factors for both LFS and OS.

c. VAF of mutations and the correlation with clinical parameters and outcomes

VAF of *IDH2* mutation had impact on clinical features; higher VAF of *IDH2* was associated with lower hemoglobin level (r=-0.496, p=0.009) and lower bone marrow blasts percentage (r=-0.432, p=0.024). Compared with wild-type genes, high VAF of mutations in 9 genes, including *DNMT3A* (cutoff value 40%, HR 2.87, p<0.001), *TET2* (45%, HR 2.55, p<0.001), *ASXL1* (20%, HR 2.24, p<0.001), *EZH2* (40%, HR 2.12, p=0.036), *SETBP1* (15%, HR 1.94, p=0.024), *BCOR* (80%, HR 2.49, p=0.043), *SRSF2* (50%, HR 3.65, p=0.002), *ZRSR2* (60%, HR 2.91, p<0.001) and *TP53* (25%, HR 7.84, p<0.001) were significantly associated with shorter LFS. With the exception of *EZH2*, *SETBP1*, and *BCOR* mutations, high VAF of all other six mutations were also associated with shorter OS. In multivariable analysis, female sex and receiving HSCT were independent favorable factors for both LFS and OS, while older age and higher-risk IPSS-R predicted shorter LFS and OS. Mutant *IDH2* and *CBL* predicted both shorter LFS and OS, while *U2AF1* mutation and *DNMT3A* and *ZRSR2* mutations with high VAF were associated poorer OS. Regarding *TP53* mutations, patients with high VAF had the worst outcomes compared to those with wild type *TP53* or low VAF.

The presence of poor-risk mutations (*DNMT3A* and *ZRSR2* mutations with high VAF, mutant *TP53*, *IDH2*, *CBL* and *U2AF1* for OS; mutant *TP53*, *IDH2* and *CBL* for LFS) could re-stratify the IPSS-R risk groups. For instance, in the IPSS-R low and very low-risk group, the patients with poor-risk mutations had an OS significantly shorter than those without poor-risk mutations. Considering OS, the incorporation of the molecular data in the IPSS-R could reclassify 8.9% (18/202) of IPSS-R very low/low-risk patients to intermediate-risk subgroup, 17.9% (33/184) of IPSS-R intermediate to high-risk

doi:10.6342/NTU202301887

subgroup, and 34.4% (54/157) of IPSS-R high to very high-risk subgroup.

d. Impact of hypomethylating agents and HSCT on survival in patients with poor-risk mutations

The use of HMA or HSCT could not significantly improve LFS or OS in patients with at least one of the poor-risk mutations. However, focusing on specific mutations, patients harboring *U2AF1* mutation had similar LFS and OS compared with those with wild-type *U2AF1* if they received HMA treatment.

E. Discussion and Conclusion

To the best of our knowledge, the present study was one of the most comprehensive researches that investigated the clinical significance of VAF in a large number of myeloid-malignancies related gene mutations in MDS patients. We demonstrated that high VAF of *DNMT3A, TET2, ASXL1, SRSF2, ZRSR2* and *TP53* mutations were significantly associated with shorter LFS and OS. Patients with low VAF of *DNMT3A, TET2, ASXL1, SRSF2,* and *ZRSR2* mutation, respectively had outcome comparable to those with the wild-type gene. For *TP53* mutation, the VAF level could separate patients into three risk groups with distinct outcomes. Further, in multivariable analysis, high VAF of *DNMT3A* and *ZRSR2* mutations independently predicted poorer OS. Moreover, the presence of mutations in *TP53, IDH2, CBL,* and *U2AF1* also conferred a worse prognosis.

With the advances of next-generation sequencing technologies, which are more powerful for comprehensive mutation analysis and more sensitive to identify rare variants and mutations with low-frequency. In addition to mutations, mounting evidences have shown VAF of mutations also have clinical significances. One of the previous studies demonstrated that *TP53* mutated patients could be segregated into three groups with distinct outcomes using 20% and 50% as cutoff values for VAF. A meta-analysis suggested a threshold of 20% as a rough line between high and low clone burden of *TP53* mutation. Due to the relatively lower rate of *ZRSR2* mutation in MDS, data for its clinical impacts are limited. Here, we demonstrated the association between *ZRSR2* mutation and poor outcomes and found that the prognostic impact of *ZRSR2* mutation depended on its VAF.

The limitations of our study include its retrospective nature and the lack of external validation to confirm the prognostic significance of the VAF cutoff levels we set. However, our data fostered our understanding of the mutation burden of the diseases and provided future patient-tailored therapeutic avenues.

doi:10.6342/NTU202301887

目錄

doi:10.6342/NTU202301887

表目錄

圖目錄

doi:10.6342/NTU202301887

Introduction

Myelodysplastic syndrome (MDS) is a diverse group of clonal myeloid neoplasms, characterized by clinical and genetic heterogeneity, and increased risk of acute myeloid leukemia (AML) transformation. With the cardinal features of ineffective hematopoiesis, it is characterized by cytopenias, dysplastic hematopoietic cells, and recurrent chromosomal abnormalities (1).

The accumulation of mutations in stem cell compartments is involved in MDS pathogenesis, which gives rise to clonal architecture and leads to disease progression (2). In the genomic era, mutation landscapes have illuminated many recurrent mutations in MDS (3-7), in which some are important predictors for clinical outcomes. Several prognostic models, including the International Prognostic Scoring System (IPSS) (8), revised IPSS (IPSS-R) (9), World Health Organization Classification-based Prognostic Scoring System (10) and MD Anderson Prognostic Scoring System (11) have been developed to risk-stratify MDS patients and guide treatment. These scoring systems are mainly based on the severity of cytopenia, transfusion requirement, percentage of bone marrow blasts, and chromosomal abnormalities. Mounting evidences demonstrate that the addition of mutation data improves the prognostic stratification in MDS patients (5, 6, 12-14). Recently, Bernard E et al. proposed a clinical-molecular prognostic model (IPSS-Molecular [IPSS-M]) combining clinical parameters, cytogenetic abnormalities, and somatic mutations of 31 genes (15). They established a six-risk category schema, resulting in the re-stratification of 46% of patients from their original IPSS-R classifications. Notably, 6% of patients in the IPSS-R very low/low groups were reclassified into the IPSS-M very high/high-risk groups. In addition to mutational profiles, the variant allele frequency (VAF) of individual mutations also influence the prognosis in MDS patients (16-20). However, most studies focused on VAF in *TP53* mutation and the data concerning the clinical implication of VAF in other recurrent mutations are limited.

Additionally, in 2022, two distinct classification systems for MDS have been proposed, 2022 International Consensus Classification (ICC) (21) and 2022 WHO (22). The main innovative changes of MDS in the ICC include the reclassification of MDS with blasts of 10-19 % as MDS/AML, MDS with mutated *SF3B1* without excess blasts as MDS-*SF3B1* irrespective of the number of ring sideroblasts, and the introduction of novel molecular-defining categories including myeloid neoplasms with mutated *TP53*, and MDS/AML with MDS-related gene mutations. At the same time, under the aegis of the International Agency for Research on Cancer, the 5th edition of WHO (2022 WHO)

doi:10.6342/NTU202301887

was released. By emphasizing the molecular features and incorporating tissue architecture and histologic appearance, 2022 WHO proposed new categories including MDS with biallelic *TP53* inactivation (MDS-bi*TP53*), MDS, hypoplastic (MDS-h) and MDS with fibrosis (MDS-f)

In the present study, we performed comprehensive VAF analyses of 54 myeloid malignancies-related gene mutations in MDS patients, focusing on the correlation between VAF and survival. We further analyzed the impacts of allogenic hematopoietic stem cell transplantation (HSCT) and hypomethylating agents (HMA) on outcomes considering various VAF in different genes.

Material and Methods

(A) Patients and samples

A total of 698 primary MDS patients diagnosed between January 1986 and May 2021 at the National Taiwan University Hospital who had adequate cryopreserved bone marrow samples for deep-targeted sequencing and IPSS-R data were recruited. The diagnoses were based on the 2016 WHO classification (23) and patients were further categorized by the 2022 WHO classification (22) and 2022 International Consensus Classification (ICC) (21). Patients with antecedent chemotherapy/radiotherapy or hematologic malignancies were excluded to homogenize the cohort since the mutational landscapes differed between primary and secondary MDS (24, 25). This study was approved by the Research Ethics Committee of the National Taiwan University Hospital; and written informed consent was obtained from all participants in accordance with the Declaration of Helsinki (approval number: 201709072RINC).

(B) Cytogenetic study

Bone marrow cells are harvested directly or after 1-3 days of non-stimulated culture. Metaphase chromosomes are banded by the conventional trypsin-Giemsa banding technique and karyotyped according to the International System for Human Cytogenetic Nomenclature. Clonal cytogenetic abnormalities were defined as (i) chromosomal loss in ≥3 metaphases; (ii) chromosomal gain in ≥2 metaphases; or (iii) chromosomal structural abnormality (including deletion, translocation, and inversion, etc.) in ≥2 metaphases (26, 27). The results were interpreted according to the International System for Human Cytogenetic Nomenclature (26, 27).

(C) Gene mutation analysis

doi:10.6342/NTU202301887

TruSight myeloid sequencing panel (Illumina, San Diego, CS, USA) and the HiSeq platform (Illumina, San Diego, CA, USA) were adopted to analyze the gene alterations and mutant allele burdens of 54 myeloid-neoplasm relevant genes (28) (Supplemental Table 1). The library preparation and sequencing were performed in accordance with the manufacturer's instructions. The median reading depth was 10550x. We used COSMIC database version 86, dbSNP version 151, ClinVar, PolyPhen-2, and SIFT to evaluate the consequence of every variant. The minimum VAF for diagnostic samples was 5% (29). Because of the sequencing sensitivity issue, we verified *CEBPA* mutations via Sanger sequencing. Analysis of *FLT3*-ITD was performed via polymerase chain reaction, followed by fluorescence capillary electrophoresis.

(D) Statistical analysis

Pairwise comparison between continuous variables was performed using the Mann–Whitney U test, and the Fisher's exact test or the χ^2 test was performed for discrete variables. Pearson's correlation coefficient was used to assess the strength and direction of the linear relationships between VAF and clinical parameters. The correlation coefficient (r) greater or lower than 0.4/-0.4 was thought to be positive/negative correlated (30, 31).

Leukemia-free survival (LFS) was defined as the duration from the date of diagnosis to the last follow-up, documented leukemia transformation, or death from any cause, whichever occurred first. Overall survival (OS) was the duration from the date of diagnosis to the last follow-up or death from any cause, whichever occurred first.

Survival curves were plotted using the Kaplan-Meier analysis, and the statistical significance was calculated using the log-rank test. The Cox proportional hazards models were used in univariable and multivariable analyses. Maximally selected rank statistics were applied for VAF exploration (32, 33). This method was an appropriate standardized two-sample linear rank statistic to identify the maximum of the standardized statistics of all possible cutoffs, which can provide the best separation of the results into two groups (34-39) when we analyzed the correlation between the VAF of mutated genes and survival.

Bootstrapping repeated the process of sample generation from an underlying population by drawing samples with replacements from the original dataset, of the same size as that of the original data set. The developed results were tested in the original sample or those subjects not included in the bootstrap sample (40). All *P* values were

doi:10.6342/NTU202301887

two-sided and considered statistically significant if <0.05. All analyses were performed with IBM SPSS Statistics v23 for Windows and *jamovi.* 2.3.12.

Results

(A) Demographic features

The patient characteristics are shown in Supplemental Table 2. The median age was 66.5 years, with male predominance (63.3%). The median follow-up time was 54.7 months (0.1-329.9 months). When categorized by the 2016 WHO classification, over half (52.0%) of the patients had myelodysplastic syndrome with excess blasts, including EB1 (23.8%) and EB2 (28.2%). Of note, there were lower proportions of MDS patients with isolated del(5q) (MDS-5q; 0.7%), MDS with ring sideroblasts and single lineage dysplasia (MDS-RS-SLD, 6.6%) and MDS-RS and multilineage dysplasia (MDS-RS-MLD, 3.9%) in this cohort, compared to those in Western population (41). When categorized by the 2022 WHO classification, the disease of 33 (4.7%) patients were classified as AML, whereas 14 (2.0%) were so according to the 2022 ICC (Supplemental Table 3). In the 2022 ICC, the name of previous MDS-EB2 is changed to MDS/AML, defined as cytopenic myeloid neoplasm and 10-19% of blasts in the BM or blood, which distinguished from MDS-EB. The comparison of clinical characteristics between patients with MDS and MDS/AML was shown in Supplemental Table 4. Overall, patients with MDS/AML had significant lower white blood cells counts, absolute neutrophil counts and higher blasts percentage in the bone marrow and peripheral blood. A total of 71.1% of 698 patients had IPSS-R intermediate (26.4%), high (22.5%), or very high-risk disease (22.2%) and patients with MDS/AML more frequently had higher risk IPSS-R compared with those of MDS (Supplemental Table 4).

Regarding treatments, 24.4% of patients received HMA and 16.1% underwent allogeneic HSCT. Treatment modalities were different among subgroups based on the 2016 WHO classification and IPSS-R risk group (Supplemental Figure 1). Patients with intermediate, high, or very high-risk IPSS-R received HMA, intensive chemotherapy, or HSCT more frequently than those with very low or low-risk IPSS-R. Twenty three percent of patients experienced leukemic transformation and 49.3% died at the end of follow-up.

(B) Genetic profiles

Overall, 545 patients (78.1%) had at least one gene mutation (71.5%) in the 54

doi:10.6342/NTU202301887

genes analyzed or abnormal cytogenetic change (42.0%). The average number of mutations was 1.7 (range 0-8). As shown in Supplemental Table 5 and Supplemental Figure 2a, the most common mutation in the cohort was *ASXL1* mutation (20.3%), followed by *TET2* (14.3%), *SF3B1* (13.8%), *RUNX1* (12.6%), *STAG2* (12.5%), and *TP53* mutations (12.3%). When stratified based on the biological function of the affected genes, mutations in genes involved in epigenetic modifications (45.6%), including DNA methylation-related genes (26.5%) and chromatin modifying genes (28.8%), were the most common, followed by mutations in spliceosome-complex genes (33.8%) (Supplemental Figure 2b). Comparison of genetic alternations between patients with MDS and MDS/AML defined by the 2022 ICC were shown in Supplemental Table 4. Patients with MDS/AML had significantly higher number of mutations at diagnosis and more frequently had mutations in *IDH1, IDH2, ASXL1, BCOR, BCORL1*, *PHF6, NRAS, RUNX1, GATA2, IKZF1, SRSF2, STAG2* and *TP5*3, while less commonly *SF3B1* mutation than patients with MDS.

(C) Correlation between clinical outcomes and mutational status

In univariable Cox regression analyses for LFS and OS, we tested the variables including age, sex, IPSS-R, mutation status, and receiving HSCT/ HMA or not. Older age, higher-risk IPSS-R, HMA treatment, and presence of mutations in *TET2, IDH2, ASXL1, EZH2, CBL, RUNX1, U2AF1, SRSF2, ZRSR2, STAG2,* and *TP53* were significantly associated with both shorter LFS and OS, while *DNMT3A*, *BCORL1*, and *NRAS* mutations conferred shorter LFS and a trend of worse OS. Receiving HSCT, female sex, and mutated *SF3B1* were favorable factors for LFS and OS (Supplemental Table 6). Multivariable analysis including clinical parameters and mutations that had a *p* value<0.1 in univariable analysis showed that older age, higher-risk IPSS-R, and *DNMT3A, TET2, IDH2, CBL,* and *TP53* mutations were independent poor risk factors, while female sex and receiving HSCT were good risk factors for both LFS and OS (Supplemental Table 7).

(D) VAF of mutations and the correlation with clinical parameters and outcomes

The median VAF and distribution of each gene mutation is presented in Supplemental Table 5 and Supplemental Figure 3 respectively. *ZRSR2* mutation had the highest median VAF (66.5%, range 5.2-93.7%), followed by *IKZF1* mutation (52.7%, range 8.1-56.1%), *KIT* mutation (50.7%, range 49.1-99.8%) and *TP53* mutation (46.2%, range 5.2-93.9%). Spliceosome gene of *ZRSR2* is known to locate on the X

doi:10.6342/NTU202301887

chromosome and somatic alterations are observed predominantly in males (42). Sex-biased *ZRSR2* mutations were noted in our study since 83.3% patients with mutant *ZRSR2* were males (male *vs.* female, 83.7% *vs.* 16.7%, p=0.020). Furthermore, all of the patients with high VAF of *ZRSR2* mutations (n=14) were males (male vs. female, 100% *vs.* 0%, p=0.026).

Among the gene mutations, VAF of *IDH2* mutation had impact on clinical features; higher VAF of *IDH2* was associated with lower hemoglobin level (r=-0.496, p=0.009) and lower bone marrow blasts percentage (r=-0.432, p=0.024). There was no correlation of VAF of other mutations with clinical characteristics. Further exploration of the correlation between prognosis and VAF of individual genes showed that compared with wild-type genes, high VAF of mutations in 9 genes, including *DNMT3A* (cutoff value 40%, HR 2.87, p<0.001), *TET2* (45%, HR 2.55, p<0.001), *ASXL1* (20%, HR 2.24, p<0.001), *EZH2* (40%, HR 2.12, p=0.036), *SETBP1* (15%, HR 1.94, p=0.024), *BCOR* (80%, HR 2.49, p=0.043), *SRSF2* (50%, HR 3.65, p=0.002), *ZRSR2* (60%, HR 2.91, p<0.001) and *TP53* (25%, HR 7.84, p<0.001) were significantly associated with shorter LFS (Table 1). With the exception of *EZH2*, *SETBP1*, and *BCOR* mutations, high VAF of all other six mutations were also associated with shorter OS (Table 1). In addition, the patients with *SRSF2* and *TP53* mutations, even at lower VAF, had poorer prognosis for both LFS and OS compared with those with wild-type genes (Table 1). For other 7 above-mentioned mutations, only higher VAF, but not lower one, conferred poorer outcomes in univariable analysis; patients with these mutations at lower VAF had similar survival to those without mutation. The proportion of patients with high VAF of these nine gene mutations in which VAF levels had impacts on prognosis are shown in Supplemental Figures 4.

In multivariable analysis (Table 2), all the clinical parameters and genetic alterations with a p value<0.1 in univariable Cox regression analysis were used as covariates. For the 9 mutations in which VAF had impact on survival (Table 2), high VAF were included as covariates and for those in which VAF had no prognostic implication, the mutations themselves were used as covariates. Female sex and receiving HSCT were independent favorable factors for both LFS and OS, while older age and higher-risk IPSS-R predicted shorter LFS and OS (Table 2). Mutant *IDH2* and *CBL* predicted both shorter LFS and OS, while *U2AF1* mutation and *DNMT3A* and *ZRSR2* mutations with high VAF were associated poorer OS. *NRAS* mutation and mutant *BCOR* and *ZRSR2* with high VAF were associated with a trend of shorter LFS. Regarding *TP53* mutations, patients with high VAF had the worst outcomes compared

doi:10.6342/NTU202301887

to those with wild type *TP53* or low VAF. Even patients with low VAF (HR 2.179 for LFS, HR 2.607 for OS, $p<0.05$) had shorter LFS and OS compared to those with wild type *TP53* (Table 2). In other words, the mutation status and VAF levels of *TP53* mutations could distinguish patients into three hierarchical groups with different prognoses (Supplemental Figure 5a). Among the subgroup of patients with *TP53* VAF ≤ 25%, the presence of complex karyotype or chromosome 17 deletion conferred negative impact on survivals. Patients with *TP53* VAF ≤ 25% and complex karyotype/chromosome 17 deletion had similar dismal outcomes to those having VAF > 25% (Supplemental Figure 5b).

(E) Incorporating poor-risk mutations into IPSS-R re-stratified patients.

The presence of poor-risk mutations (*DNMT3A* and *ZRSR2* mutations with high VAF, mutant *TP53*, *IDH2*, *CBL* and *U2AF1* for OS; mutant *TP53*, *IDH2* and *CBL* for LFS) as identified in the multivariable Cox analysis shown above could re-stratify the IPSS-R risk groups (Figure 1). For instance, in the IPSS-R low and very low-risk group, the patients with poor-risk mutations had an OS significantly shorter than those without poor-risk mutations (median OS, 69.9 *vs.* 156.0 months, $p=0.001$; Figure 1a), but similar to the IPSS-R intermediate-risk patients (median OS, 69.9 *vs.* 53.8 months, $p=0.562$; Figure 1a), and similarly, the LFS was shorter in the former group than the latter (median LFS, 16.6 *vs.* 155.7 months, $p<0.001$; Figure 1e). The same were also true for patients with IPSS-R intermediate, high, or very high-risk MDS (Figures 1b–1d and 1f-1h). Notably, the LFS of IPSS-R very low/low-risk patients harboring poor-risk mutations were even shorter than the IPSS-R intermediate-risk patients (median LFS, 16.6 months *vs.* 50.6 months, $p=0.038$). In other words, patients with these unfavorable mutations in each IPSS-R risk subgroup had survivals worse than other patients of the same risk but similar to or even worse than those in the next higher-risk subgroup. Intriguingly, considering OS, the incorporation of the molecular data in the IPSS-R could reclassify 8.9% (18/202) of IPSS-R very low/low-risk patients to intermediate-risk subgroup, 17.9% (33/184) of IPSS-R intermediate to high-risk subgroup, and 34.4% (54/157) of IPSS-R high to very high-risk subgroup.

(F) Impact of treatment with hypomethylating agents and HSCT on survival in patients with poor-risk mutations

The impact of HMA and HSCT on survival in patients with poor-risk mutations was analyzed. The use of HMA or HSCT could not significantly improve LFS or OS in

doi:10.6342/NTU202301887

patients with at least one of the poor-risk mutations (Supplemental Figures 6a-d, 7a-d). However, focusing on specific mutations, patients harboring *U2AF1* mutation had poorer LFS and OS if they were not treated with HMA, but had similar LFS (wild-type *vs.* mutant-type, 14.3 *vs.* 15.5 months, p=0.537) and OS (wild-type *vs*. mutant-type, 22.2 *vs.* 40.1 months, p=0.450) compared with those with wild-type *U2AF1* if they received HMA treatment (Figure 2).

Discussion

To the best of our knowledge, the present study was one of the most comprehensive researches that investigated the clinical significance of VAF in a large number of myeloid-malignancies related gene mutations in MDS patients. We demonstrated that high VAF of *DNMT3A, TET2, ASXL1, SRSF2, ZRSR2* and *TP53* mutations were significantly associated with shorter LFS and OS (Table 1). Patients with low VAF of *DNMT3A, TET2, ASXL1, SRSF2,* and *ZRSR2* mutation, respectively had outcome comparable to those with the wild-type gene, similar to what happened to *FLT3*-ITD in AML (41). For *TP53* mutation, the VAF level could separate patients into three risk groups with distinct outcomes. Further, in multivariable analysis, high VAF of *DNMT3A* and *ZRSR2* mutations independently predicted poorer OS (Table 2). Moreover, the presence of mutations in *TP53, IDH2, CBL*, and *U2AF1* also conferred a worse prognosis. While the 2022 WHO and 2022 ICC have been recently proposed, we showed the pattern of case allocation, differences of genetics profiles and clinical manifestations according to these novel classification systems (Supplemental Tables 3 and 4).

Somatic mutations in leukemia driver genes result in clonal hematopoiesis. It is known to have a premalignant potential which is derived from an expansion of a single hematopoietic stem cell and is associated with increased risks of hematologic cancer and death (43, 44). However, the incidence of hematopoietic malignancies for most individuals with clonal hematopoiesis is low (44, 45). Risk stratification focused on distinguishing high-risk patients who may require early intervention while avoiding overdiagnosing, subjecting low-risk patients to unnecessary monitoring, or treatments. Thus, Lachelle D. et al. had proposed a clonal hematopoiesis risk score which was a simple prognostic model for individuals with clonal hematopoiesis of indeterminate potential or clonal cytopenia of unknown significance (46).

Based on the heterogeneity of MDS and varied outcomes, even in the same IPSS-R risk group, it is reasonable to explore prognostic molecular markers to enhance the

doi:10.6342/NTU202301887

prognostication of novel scoring systems. We and other researchers previously proposed models integrating the mutation status of certain genes into IPSS or IPSS-R which improved risk stratification of MDS and could identify poor-risk patients in the same IPSS or IPSS-R risk groups (6, 12, 13, 47). With the advances of next-generation sequencing technologies, which are more powerful for comprehensive mutation analysis and more sensitive to identify rare variants and mutations with low-frequency compared to Sanger sequencing (48), the relevance of mutational burdens and their clinical implications can be deeply evaluated. Recently, in studying 2,957 patients under the aegis of the International Working Group for Prognosis in MDS, Bernard et al. proposed a clinical-molecular prognostic model, IPSS-Molecular (IPSS-M) that combines clinical parameters, cytogenetic abnormalities, and somatic mutations of 31 genes (15). Six risk category schema was established that resulted in the reclassification of 46% of the patients from their original IPSS-R classifications. This model was validated in an external cohort of 754 Japanese patients with MDS.

In addition to mutations, mounting evidences have shown VAF of mutations also have clinical significances (15, 18-20, 49, 50). Montalban-Bravo et al demonstrated that high VAF of *TP53* mutation was associated with a worse prognosis and *TP53* mutated patients could be segregated into three groups with distinct outcomes using 20% and 50% as cutoff values for VAF (18). Moreover, they showed lower VAF of *TP53* mutation was correlated with a higher overall response rate to HMA compared to higher VAF. *TP53* VAF could stratify distinct prognostic groups independent of clinical prognostic scoring systems (16, 18, 19, 50). However, the cutoff values of mutant *TP53* VAF that could discriminate outcomes varied among studies. A meta-analysis including 11 studies suggested a threshold of 20% as a rough line between high and low clone burden and 40% as a cutoff point to guide treatment (51). In our study, we found the threshold of 25% could separate MDS patients into three risk groups with distinct survival. Though patients with mutant *TP53* VAF ≤ 25% had better survival than those with VAF > 25%, their prognosis remained dismal compared with *TP53*-wild patients (Supplemental Figure 5a). Furthermore, subgroup analysis disclosed that patients with mutant *TP53* VAF ≤ 25% accompanying with complex karyotype or chromosome 17 deletion had similar dismal prognosis to those with VAF > 25% (Supplemental Figure 5b). Thus, based on our results and the existing research, we believed that 20-25% was a proper cutoff for mutant *TP53* VAF.

Spliceosome mutations, including those of the *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* genes, were founding genetic lesions and the most common acquired genetic

doi:10.6342/NTU202301887

alterations in MDS patients (13, 52). Due to the relatively lower rate of *ZRSR2* mutation in MDS, data for its clinical impacts are limited (52). Here, we demonstrated the association between *ZRSR2* mutation and poor outcomes in a relatively large cohort, and found that the prognostic impact of *ZRSR2* mutation depended on its VAF. It was reported that *U2AF1* mutation with VAF>40% was associated with shorter OS (20) and a higher risk of leukemic evolution. In this study, we found that *U2AF1* mutation was associated with poorer OS, but the use of HMA could improve the clinical outcome in patients with this mutation. However, we could not demonstrate mutant *U2AF1* VAF had impact on the prognosis in our cohort. More investigations are required to clarify the clinical significance of VAF of mutations and the choice of treatment in patients with poor-risk mutations.

Regarding the clinical relevance of our findings, we showed that mutational screening of *IDH2, CBL, U2AF1,* and *TP53* and VAF of two mutations, including *DNMT3A* and *ZRSR2* mutations, could refine risk stratification of IPSS-R. Patients with poor-risk mutations had an OS worse than others in the same IPSS-R risk subgroup, but similar to those in the next higher-risk subgroup. These findings explain the clinical heterogeneity in the same IPSS-R risk groups. A substantial portion of patients in each IPSS-R risk group could be adjusted to different prognostic groups based on the integrated prognostic system; 8.9% of the IPSS-R very low and low-risk patients could be redistributed into the intermediate-risk group, 17.9% of IPSS-R intermediate-risk patients to the high-risk group, and 34.4% of IPSS-R high-risk to the very high-risk group. The incorporation of the mutation status and VAF of these poor-risk mutations into the survival analysis would be especially helpful to identify patients with poorer prognoses for more aggressive treatment in the lower-risk MDS patients defined by conventional scoring systems.

The limitations of our study include the lack of external validation to confirm the prognostic significance of the VAF cutoff levels we set. However, the bootstrap method we used has been shown good for internal validation (40), which can best separate the two groups. The pathobiology and causal relationship between VAF and poorer prognosis still require further confirmation since the dynamic nature of VAF during disease progression or during treatments. In the past decades, there have been rapid advancements in the treatment of MDS, with numerous novel therapies emerging. The characteristics of the disease and prognoses may be modified, so validation of the clinical significance of the mutation VAF we found in this study needs further prospective studies. Moreover, due to limited cases, whether HMA or HSCT can

doi:10.6342/NTU202301887

remedy the dismal outcomes of patients with poor-risk mutations also needs further clarification.

In conclusion, VAF is critical for risk stratification in MDS patients and should be considered in novel scoring systems. Our data fostered our understanding of the mutation burden of the diseases and provided future patient-tailored therapeutic avenues.

This study has been published in *American Journal of Hematology*, in December 2022.

doi:10.6342/NTU202301887

References

1. Cazzola M. Myelodysplastic Syndromes. N Engl J Med. 2020;383(14):1358-74.
2. Chen J, Kao YR, Sun D, Todorova TI, Reynolds D, Narayanagari SR, et al. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. Nat Med. 2019;25(1):103-10.
3. Malcovati L, Stevenson K, Papaemmanuil E, Neuberg D, Bejar R, Boultwood J, et al. SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS. Blood. 2020;136(2):157-70.
4. Bejar R, Stevenson KE, Caughey B, Lindsley RC, Mar BG, Stojanov P, et al. Somatic mutations predict poor outcome in patients with myelodysplastic syndrome after hematopoietic stem-cell transplantation. J Clin Oncol. 2014;32(25):2691-8.
5. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-7.
6. Nazha A, Narkhede M, Radivoyevitch T, Seastone DJ, Patel BJ, Gerds AT, et al. Incorporation of molecular data into the Revised International Prognostic Scoring System in treated patients with myelodysplastic syndromes. Leukemia. 2016;30(11):2214-20.
7. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616-27; quiz 99.
8. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997;89(6):2079-88.
9. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120(12):2454-65.
10. Malcovati L, Germing U, Kuendgen A, Della Porta MG, Pascutto C, Invernizzi R, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. J Clin Oncol. 2007;25(23):3503-10.
11. Kantarjian H, O'Brien S, Ravandi F, Cortes J, Shan J, Bennett JM, et al. Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. Cancer. 2008;113(6):1351-61.

doi:10.6342/NTU202301887

12. Nazha A, Komrokji R, Meggendorfer M, Jia X, Radakovich N, Shreve J, et al. Personalized Prediction Model to Risk Stratify Patients With Myelodysplastic Syndromes. J Clin Oncol. 2021;39(33):3737-46.
13. Hou HA, Tsai CH, Lin CC, Chou WC, Kuo YY, Liu CY, et al. Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. Blood Cancer J. 2018;8(4):39.
14. Nazha A, Al-Issa K, Hamilton BK, Radivoyevitch T, Gerds AT, Mukherjee S, et al. Adding molecular data to prognostic models can improve predictive power in treated patients with myelodysplastic syndromes. Leukemia. 2017;31(12):2848-50.
15. Bernard E, Tuechler H, Greenberg PL, Hasserjian RP, Ossa JEA, Nannya Y, et al. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. NEJM Evidence. 2022;1(7):EVIDoa2200008.
16. Belickova M, Vesela J, Jonasova A, Pejsova B, Votavova H, Merkerova MD, et al. TP53 mutation variant allele frequency is a potential predictor for clinical outcome of patients with lower-risk myelodysplastic syndromes. Oncotarget. 2016;7(24):36266-79.
17. Jiang L, Wang L, Shen C, Zhu S, Lang W, Luo Y, et al. Impact of mutational variant allele frequency on prognosis in myelodysplastic syndromes. Am J Cancer Res. 2020;10(12):4476-87.
18. Montalban-Bravo G, Kanagal-Shamanna R, Benton CB, Class CA, Chien KS, Sasaki K, et al. Genomic context and TP53 allele frequency define clinical outcomes in TP53-mutated myelodysplastic syndromes. Blood Adv. 2020;4(3):482-95.
19. Sallman DA, Komrokji R, Vaupel C, Cluzeau T, Geyer SM, McGraw KL, et al. Impact of TP53 mutation variant allele frequency on phenotype and outcomes in myelodysplastic syndromes. Leukemia. 2016;30(3):666-73.
20. Wang H, Guo Y, Dong Z, Li T, Xie X, Wan D, et al. Differential U2AF1 mutation sites, burden and co-mutation genes can predict prognosis in patients with myelodysplastic syndrome. Sci Rep. 2020;10(1):18622.
21. Arber DA, Orazi A, Hasserjian RP, Borowitz MJ, Calvo KR, Kvasnicka HM, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood. 2022;140(11):1200-28.
22. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703-19.

doi:10.6342/NTU202301887

23. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-405.
24. Ok CY, Patel KP, Garcia-Manero G, Routbort MJ, Fu B, Tang G, et al. Mutational profiling of therapy-related myelodysplastic syndromes and acute myeloid leukemia by next generation sequencing, a comparison with de novo diseases. Leuk Res. 2015;39(3):348-54.
25. Singhal D, Wee LYA, Kutyna MM, Chhetri R, Geoghegan J, Schreiber AW, et al. The mutational burden of therapy-related myeloid neoplasms is similar to primary myelodysplastic syndrome but has a distinctive distribution. Leukemia. 2019;33(12):2842-53.
26. Hou HA, Kuo YY, Liu CY, Chou WC, Lee MC, Chen CY, et al. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. Blood. 2012;119(2):559-68.
27. International Standing Committee on Human Cytogenomic Nomenclature M-JJHRJMSSKG. ISCN 2020 an International System for Human Cytogenomic Nomenclature (2020) : recommendations of the International Standing Committee on Human Cytogenomic Nomenclature including revised sequence-based cytogenomic nomenclature developed in collaboration with the Human Genome Variation Society (HGVS) Sequence Variant Description Working Group2020.
28. Grimwade D, Ivey A, Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood. 2016;127(1):29-41.
29. Tsai CH, Tang JL, Tien FM, Kuo YY, Wu DC, Lin CC, et al. Clinical implications of sequential MRD monitoring by NGS at 2 time points after chemotherapy in patients with AML. Blood Adv. 2021;5(10):2456-66.
30. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71.
31. Overholser BR, Sowinski KM. Biostatistics primer: part 2. Nutr Clin Pract. 2008;23(1):76-84.
32. Laska E, Meisner M, Wanderling J. A maximally selected test of symmetry about zero. Stat Med. 2012;31(26):3178-91.
33. Zheng BW, Huang W, Liu FS, Zhang TL, Wang XB, Li J, et al. Clinicopathological and Prognostic Characteristics in Spinal Chondroblastomas: A Pooled Analysis of Individual Patient Data From a Single Institute and 27 Studies. Global Spine J. 2021:21925682211005732.

doi:10.6342/NTU202301887

34. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Computational Statistics & Data Analysis. 2003;43(2):121-37.
35. Lausen B, Schumacher M. Maximally Selected Rank Statistics. Biometrics. 1992;48(1):73-85.
36. Lausen B, Hothorn T, Bretz F, Schumacher M. Assessment of Optimal Selected Prognostic Factors. Biometrical Journal. 2004;46(3):364-74.
37. Terry M, Thomas L. survival: Survival Analysis. [R package]. https:// CRAN.R-project.org/package=survival. Accessed December 19, 2021.
38. Jonathon L, Damian D, Ravi S, et al. The jamovi project (2021). jamovi. (Version 1.6) [Computer Software]. https://www.jamovi.org. Accessed December 19, 2021.
39. Institute for Statistics and Mathematics WWW. R Core Team (2020). R: A Language and environment for statistical computing. (Version 4.0) [Computer software]. https://cran.r-project.org. Accessed December 19, 2021.
40. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
41. Jiang Y, Eveillard JR, Couturier MA, Soubise B, Chen JM, Gao S, et al. Asian Population Is More Prone to Develop High-Risk Myelodysplastic Syndrome, Concordantly with Their Propensity to Exhibit High-Risk Cytogenetic Aberrations. Cancers (Basel). 2021;13(3).
42. Madan V, Kanojia D, Li J, Okamoto R, Sato-Otsubo A, Kohlmann A, et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat Commun. 2015;6:6042.
43. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. New England Journal of Medicine. 2014;371(26):2477-87.
44. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. New England Journal of Medicine. 2014;371(26):2488-98.
45. Fabre MA, de Almeida JG, Fiorillo E, Mitchell E, Damaskou A, Rak J, et al. The longitudinal dynamics and natural history of clonal haematopoiesis. Nature. 2022;606(7913):335-42.
46. Weeks LD, Niroula A, Neuberg D, Wong W, Lindsley RC, Luskin MR, et al. Prediction of Risk for Myeloid Malignancy in Clonal Hematopoiesis. NEJM Evidence. 2023;2(5):EVIDoa2200310.

doi:10.6342/NTU202301887

47. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011;364(26):2496-506.
48. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
49. Jiang L, Luo Y, Zhu S, Wang L, Ma L, Zhang H, et al. Mutation status and burden can improve prognostic prediction of patients with lower-risk myelodysplastic syndromes. Cancer Sci. 2020;111(2):580-91.
50. Jiang L, Ye L, Ma L, Ren Y, Zhou X, Mei C, et al. Predictive values of mutational variant allele frequency in overall survival and leukemic progression of myelodysplastic syndromes. J Cancer Res Clin Oncol. 2022;148(4):845-56.
51. Deng J, Wu X, Ling Y, Liu X, Zheng X, Ye W, et al. The prognostic impact of variant allele frequency (VAF) in TP53 mutant patients with MDS: A systematic review and meta-analysis. Eur J Haematol. 2020;105(5):524-39.
52. Thol F, Kade S, Schlarmann C, Löffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood. 2012;119(15):3578-84.

doi:10.6342/NTU202301887

Table 1. Univariable Cox regression analysis incorporating variant allele frequencies on the leukemia-free survival and overall survival of myelodysplastic syndrome patients

Variable	Cutoff VAF (%)	LFS		OS	
		HR (95% CI)	*P* value	HR	*P* value
DNMT3A	40				
High *vs.* wild		2.87 (1.64-4.81)	**<0.001**	2.75 (1.61-4.72)	**<0.001**
Low *vs.* wild		1.34 (0.90-2.00)	0.152	1.14 (0.74-1.76)	0.560
High *vs.* low		2.21 (1.14-4.29)	**0.019**	2.41 (1.22-4.74)	**0.011**
TET2	45				
High *vs.* wild		2.55 (1.58-4.12)	**<0.001**	2.32 (1.40-3.86)	**0.001**
Low *vs.* wild		1.39 (0.99-1.94)	0.055	1.32 (0.93-1.88)	0.122
High *vs.* low		1.87 (1.06-3.31)	**0.032**	1.94 (1.06-3.56)	**0.033**
ASXL1	20				
High *vs.* wild		2.24 (1.71-2.93)	**<0.001**	2.01 (1.52-2.64)	**<0.001**
Low *vs.* wild		0.99 (0.58-1.69)	0.965	1.10 (0.64-1.89)	0.721
High *vs.* low		2.55 (1.43-4.55)	**0.001**	1.99 (1.11-3.54)	**0.020**
EZH2	40				
High *vs.* wild		2.12 (1.05-4.29)	**0.036**	1.84 (0.87-3.91)	0.110
Low *vs.* wild		1.49 (0.77-2.90)	0.237	1.45 (0.72-2.93)	0.299
High *vs.* low		1.45 (0.53-3.94)	0.469	1.23 (0.42-3.64)	0.706
SETBP1	15				
High *vs.* wild		1.94 (1.09-3.45)	**0.024**	1.55 (0.85-2.84)	0.151
Low *vs.* wild		1.45 (0.47-4.53)	0.520	1.43 (0.46-4.47)	0.535
High *vs.* low		1.88 (0.50-7.03)	0.348	1.25 (0.33-4.77)	0.740
BCOR	80				
High *vs.* wild		2.49 (1.03-6.05)	**0.043**	1.91 (0.79-4.63)	0.151
Low *vs.* wild		0.80 (0.47-1.37)	0.420	0.88 (0.51-1.50)	0.636
High *vs.* low		3.44 (1.16-10.20)	**0.026**	2.16 (0.75-6.26)	0.154
SRSF2	50				
High *vs.* wild		3.65 (1.62-8.23)	**0.002**	2.76 (1.22-6.22)	**0.015**
Low *vs.* wild		1.83 (1.19-2.84)	**0.006**	1.64 (1.04-2.59)	**0.035**
High *vs.* low		2.63 (1.00-6.88)	**0.049**	2.40 (0.90-6.46)	0.082
ZRSR2	60				
High *vs.* wild		2.91 (1.63-5.18)	**<0.001**	2.54 (1.39-4.65)	**0.002**
Low *vs.* wild		1.35 (0.67-2.72)	0.406	1.47 (0.73-2.97)	0.281
High *vs.* low		2.51 (0.96-6.55)	0.060	2.15 (0.80-5.78)	0.129
TP53	25				
High *vs.* wild		7.84 (5.73-10.72)	**<0.001**	10.02 (7.25-13.86)	**<0.001**
Low *vs.* wild		3.48 (1.99-6.01)	**<0.001**	4.62 (2.63-8.12)	**<0.001**
High *vs.* low		2.30 (1.22-4.34)	**0.010**	2.06 (1.09-3.89)	**0.026**

Those mutations in which VAF had no impact on prognosis were not shown in this table.

Abbreviations: HR, Hazard ratios; CI, confidence interval; VAF, variant allele frequency; LFS, leukemia-free survival; OS, overall survival

P values of <0.05 are statistically significant.

doi:10.6342/NTU202301887

Table 2. Multivariable analysis for leukemia-free survival and overall survival in myelodysplastic syndrome patients

	LFS				OS			
Variable	**HR**	**95% CI**		***P***	**HR**	**95% CI**		***P***
Age*	1.021	1.011	1.030	**<0.001**	1.023	1.013	1.033	**<0.001**
IPSS-R†	2.611	1.949	3.499	**<0.001**	2.772	2.050	3.750	**<0.001**
Female	0.779	0.613	0.990	**0.041**	0.737	0.575	0.945	**0.016**
HSCT	0.650	0.447	0.945	**0.024**	0.624	0.421	0.923	**0.018**
HMA	1.26	0.922	1.577	0.171	1.08	0.760	1.337	0.954
Mutational VAF**								
DNMT3A								
High VAF *vs.* low and wild	1.649	0.904	3.006	0.103	1.864	1.015	3.422	**0.044**
TET2								
High VAF *vs.* low and wild	1.563	0.890	2.746	0.120	1.451	0.794	2.649	0.226
ASXL1								
High VAF *vs.* low and wild	0.966	0.639	1.459	0.868	0.960	0.630	1.463	0.849
EZH2								
High VAF *vs.* low and wild	0.940	0.350	2.530	0.903	0.835	0.297	2.351	0.733
SETBP1								
High VAF *vs.* low and wild	1.464	0.707	3.033	0.305	-	-	-	-
BCOR								
High VAF *vs.* low and wild	2.369	0.902	6.218	0.080	-	-	-	-
SRSF2				0.276				0.602
Low *vs.* wild	1.092	0.644	1.853	0.743	1.183	0.691	2.027	0.540
High *vs.* low	1.927	0.689	5.391	0.112	1.259	0.454	3.487	0.658
ZRSR2								
High VAF *vs.* low and wild	2.602	0.994	6.811	0.051	2.905	1.119	7.539	**0.028**
TP53				**<0.001**				**<0.001**
Low *vs.* wild	2.179	1.191	3.990	**0.012**	2.607	1.420	4.787	**0.002**
High *vs.* low	2.609	1.356	5.020	**<0.001**	2.918	1.510	5.639	**0.001**
Mutation status#								
IDH2	1.953	1.173	3.252	**0.010**	1.873	1.079	3.251	**0.026**
BCORL1	1.277	0.573	2.847	0.550	1.073	0.478	2.409	0.865
NRAS	1.804	0.994	3.275	0.052	1.498	0.804	2.792	0.203
CBL	3.025	1.503	6.088	**0.002**	2.909	1.443	5.864	**0.003**

doi:10.6342/NTU202301887

RUNX1	1.064	0.741	1.528	0.738	1.054	0.727	1.529	0.780
U2AF1	1.409	0.935	2.124	0.101	1.772	1.177	2.669	**0.006**
SF3B1	0.851	0.592	1.224	0.384	0.854	0.588	1.241	0.407
STAG2	1.260	0.865	1.835	0.229	1.197	0.820	1.748	0.352

Abbreviations: HR, Hazard ratios; CI, confidence interval; LFS, leukemia free survival; OS, overall survival

P values of <0.05 are statistically significant.

Genes with limited mutated events were not included.

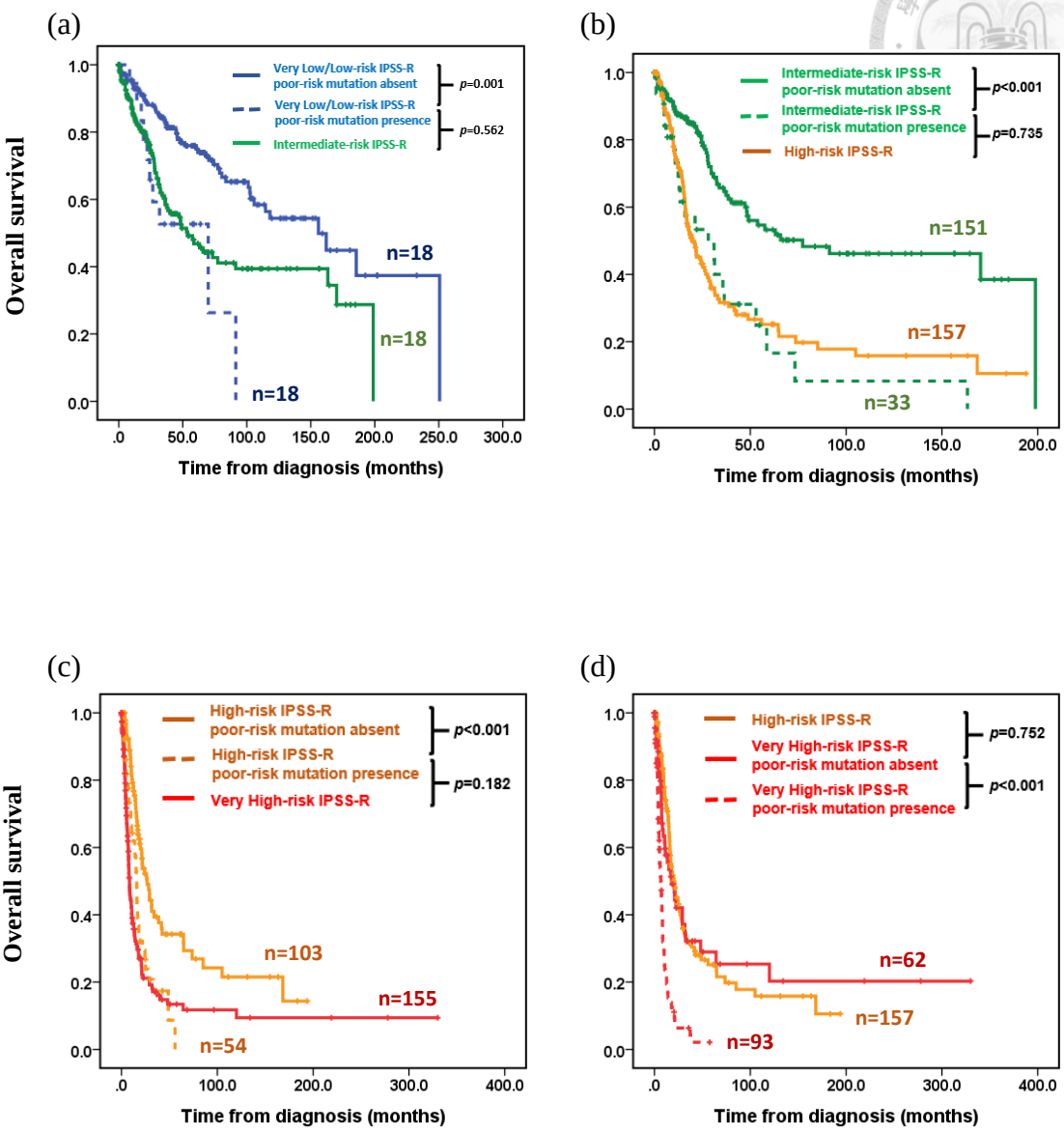
*Age, as a continuous variable analysis.

†IPSS-R higher-risk (high, very high) group vs. lower-risk (very low, low, intermediate) group.

**Mutations of *DNMT3A, TET2, ASXL1, EZH2, SETBP1, BCOR* and *ZRSR2* with higher VAF, but not lower VAF, predicted poorer prognosis (Table 2), so mutations of these genes with high VAF were chosen as covariables, while mutations in *SRSF2* and *TP53* at either high or low VAF conferred poorer survival compared with wild-type gene, so *SRSF2* and *TP53* mutations with both low and high VAF were included as covariables.

#VAF of mutations in the 8 genes below had no impact on clinical outcomes, so mutations in these genes, no matter the VAF levels, were covariables in the analysis.

doi:10.6342/NTU202301887



doi:10.6342/NTU202301887

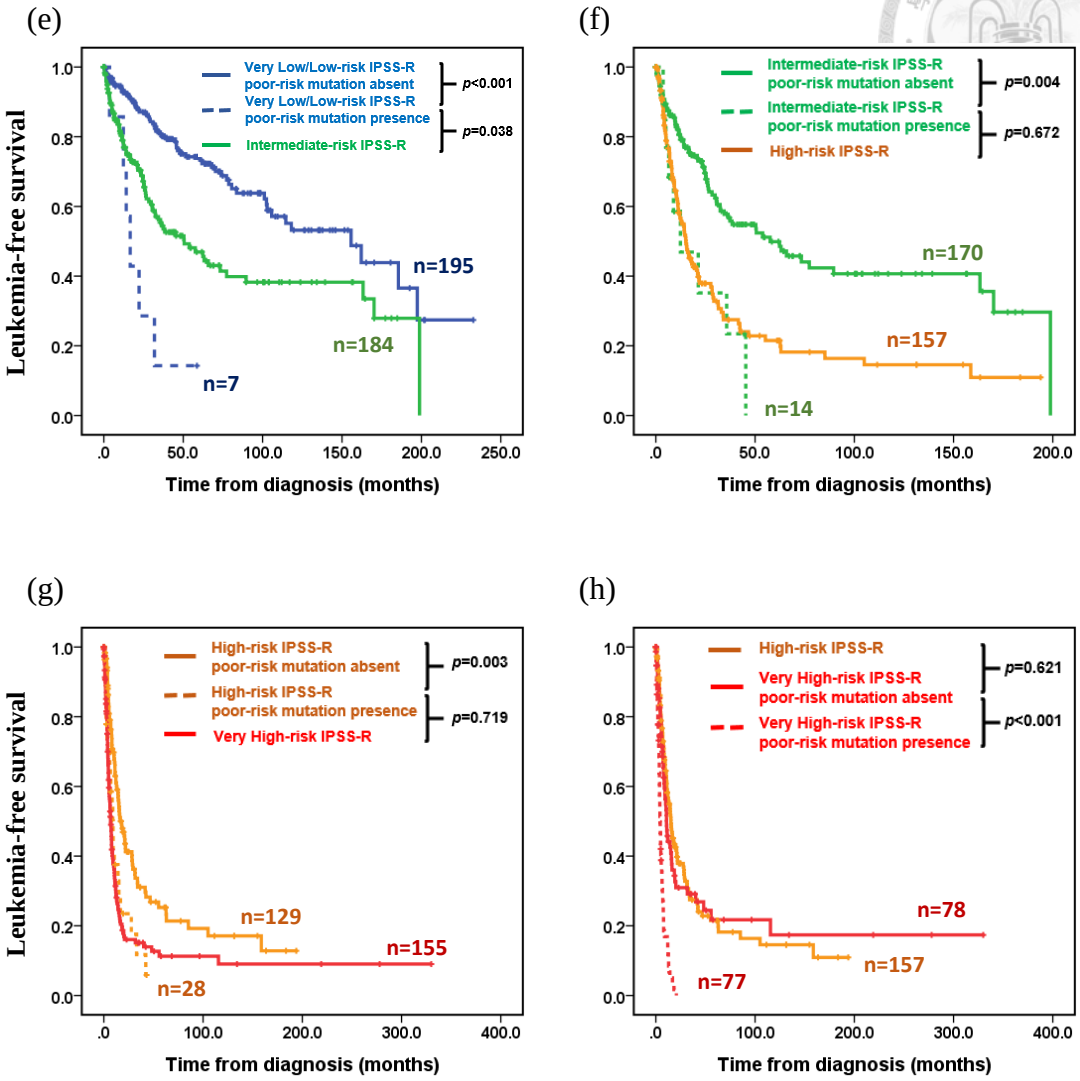


Figure 1. Kaplan-Meier curves for overall survival and leukemia-free survival stratified by the revised International Prognostic Scoring System (IPSS-R) risk categories and mutational status.

(a–d) Overall survival stratified by the presence or absence of poor-risk mutations (*DNMT3A* and *ZRSR2* mutations with high VAF, and mutant *TP53*, *IDH2*, *CBL* and *U2AF1* for OS) and IPSS-R.

(e–h) Leukemia-free survival stratified by the presence or absence of poor-risk mutations (mutant *TP53*, *IDH2* and *CBL* for LFS) and IPSS-R

doi:10.6342/NTU202301887

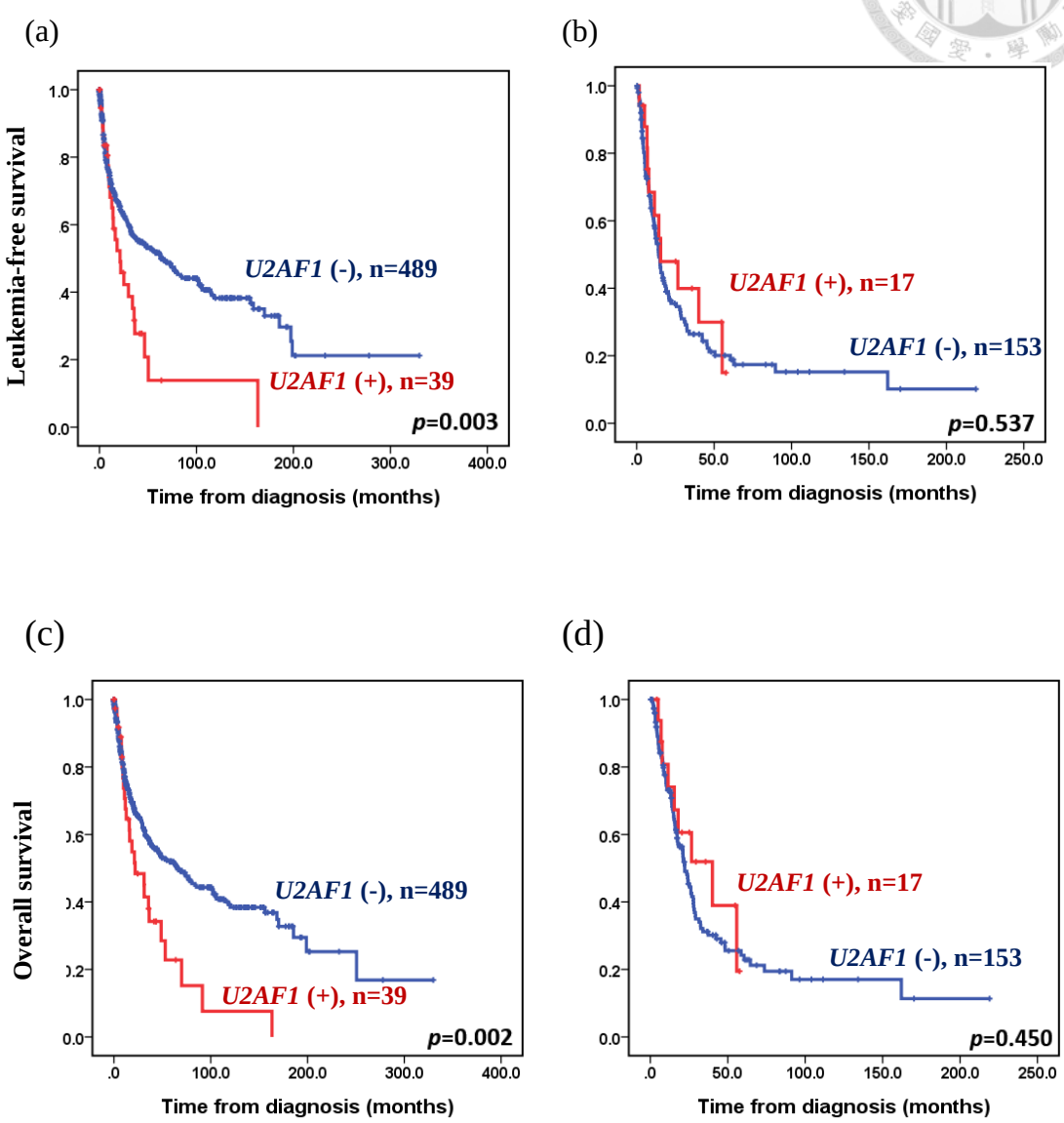


Figure 2. Kaplan–Meier curves of patients with or without *U2AF1* mutation: impact of HMA treatment

(a) Leukemia-free survival for patients not receiving HMA treatment, stratified by *U2AF1* mutation status.

(b) Leukemia-free survival for patients receiving HMA treatment, stratified by *U2AF1* mutation status.

(c) Overall survival for patients not receiving HMA treatment, stratified by *U2AF1* mutation status.

(d) Overall survival for patients receiving HMA treatment, stratified by *U2AF1* mutation status.

doi:10.6342/NTU202301887

Supplemental Material

Supplemental Table 1. List of 54 myeloid neoplasm-relevant genes studied in targeted NGS sequencing

Gene name	Target region (exon)	Gene name	Target region (exon)
ABL	4-6	***JAK3***	13
ASXL1	12	***KDM6A***	full
ATRX	8-10 and 17-31	***KIT***	2, 8-11, 13+17
BCOR	full	***KRAS***	2+3
BCORL1	full	***MLL***	5-8
BRAF	15	***MPL***	10
CALR	9	***MYD88***	3-5
CBL	8+9	***NOTCH1***	26-28, 34
CBLB	9, 10	***NPM1***	12
CBLC	9, 10	***NRAS***	2+3
CDKN2A	full	***PDGFRA***	12, 14, 18
CEBPA	full	***PHF6***	full
CSF3R	14-17	***PTEN***	5+7
CUX1	full	***PTPN11***	3+13
DNMT3A	full	***RAD21***	full
ETV6	full	***RUNX1***	full
EZH2	full	***SETBP1***	4 (partial)
FBXW7	9+10+11	***SF3B1***	13-16
FLT3	14+15+20	***SMC1A***	2, 11, 16+17
GATA1	2	***SMC3***	10, 13, 19, 23, 25+28
GATA2	2-6	***SRSF2***	1
GNAS	8+9	***STAG2***	full
HRAS	2+3	***TET2***	3-11
IDH1	4	***TP53***	2-11
IDH2	4	***U2AF1***	2+6
IKZF1	full	***WT1***	7+9
JAK2	12+14	***ZRSR2***	full

doi:10.6342/NTU202301887

Supplemental Table 2. Clinical characteristics of patients with myelodysplastic syndrome

Clinical characters	Total (n=698)	%/range	Clinical characters	Total (n=698)	%/range
Sex			**IPSS-R**‡		
Female	256	36.7	Very low	22	3.1
Male	442	63.3	Low	180	25.8
Age (years)*	66.5	18.4-94.5	Int	184	26.4
Laboratory data*			High	157	22.5
WBC, X 10^9 /L	3.31	0.56-26.33	Very high	155	22.2
ANC, X 10^9 /L	1.50	0-16.59	**Treatment**†		
Hb, g/dL	8.2	2.6-17.1	HMA	170	24.4
PLT, X 10^9 /L	78	1-607	Intensive chemotherapy	23	3.3
BM blast (%)	4.6	0.0-19.5	Clinical trial	25	3.6
PB blast (%)	0.0	0.0-19.0	HSCT	111	16.1
2016 WHO			Supportive care	309	44.8
MDS-5q	5	0.7	Other treatment§	154	22.4
MDS-SLD	100	14.3	**AML transformation**	165	23.6
MDS-MLD	149	21.3	**Death**	344	49.3
MDS-RS-SLD	46	6.6	**Early mortality**¶	88	12.6
MDS-RS-MLD	27	3.9			
MDS-U	8	1.1			
MDS-EB	363	52.0			
MDS-EB1	166	23.8			
MDS-EB2	197	28.2			

*Median (range).

‡IPSS-R: Very low, ≦1.5; Low, >1.5-3; intermediate (INT),>3-4.5; High, >4.5-6; and Very high, >6.

†Patients may receive more than one treatment.

§Other treatment: include low-dose cytarabine, rabbit-derived anti-thymocyte globulin (rATG), cyclosporine, danazol, eltrombopag, erythropoietin-stimulating agents (ESA), thalidomide, steroid, venetoclax-based therapy and oral chemotherapy.

¶Death within 3 months of diagnosis.

Abbreviations: ANC, absolute neutrophil count; Hb, hemoglobin; MDS-5q, MDS with isolated del(5q); MDS-RS, MDS with ring sideroblasts; MDS-EB, MDS with excess blasts; MDS-SLD, MDS with single lineage dysplasia; MDS-MLD, MDS with multilineage dysplasia; MDS-RS-SLD, MDS with ring

doi:10.6342/NTU202301887

sideroblasts and single lineage dysplasia; MDS-RS-MLD, MDS with ring sideroblasts and multilineage dysplasia; MDS-U, MDS, unclassifiable; PLT: platelet; IPSS-R, revised international prognosis scoring system; HMA, hypomethylation agent; HSCT, allogeneic hematopoietic stem cell transplantation; AML, acute myeloid leukemia

doi:10.6342/NTU202301887

Supplemental Table 3. Case allocation from 4th World Health Organization (WHO) to 2022 WHO and 2022 International Consensus Classification (ICC)

2022 ICC	Number (%)	2016 WHO	Number	2022 WHO	Number (%)
MDS with del(5q)	5 (100%)	**MDS-del(5q)**	5	**MDS-5q**	5 (100%)
MDS, NOS, with SLD	96 (96.0%)	**MDS-SLD**	100	**MDS-*SF3B1***	4 (4.0%)
MDS with mutated *SF3B1*	4 (4.0%)			**MDS-LB**	53 (53.0%)
				MDS-h	42 (42.0%)
				AML with *NUP98* rearrangement	1 (1.0%)
MDS, NOS, with MLD	146 (98.0%)	**MDS-MLD**	149	**MDS-*SF3B1***	4 (2.7%)
MDS with mutated *SF3B1*	3 (2.0%)			**MDS-LB**	88 (59.1%)
				MDS-h	55 (36.9%)
				AML with *NPM1*	1 (0.7%)
				AML with *NUP98* rearrangement	1 (0.7%)
MDS with mutated *SF3B1*	37 (80.4%)	**MDS-RS-SLD**	4	**MDS-*SF3B1***	37 (80.4%)
MDS, NOS, with SLD	9 (19.6%)			**MDS-LB and RS**	9 (19.6%)
MDS with mutated *SF3B1*	17 (63.0%)	**MDS-RS-MLD**	27	**MDS-*SF3B1***	19 (70.4%)
MDS, NOS, with MLD	9 (33.3%)			**MDS-LB and RS**	7 (25.9%)
MDS with mutated *TP53*	1 (3.7%)			**MDS-bi*TP53***	1 (3.7%)
MDS with EB	138 (83.1%)	**MDS-EB1**	166	**MDS-bi*TP53***	17 (10.2%)
MDS with mutated *TP53*	28 (16.9%)			**MDS-IB1**	128 (77.1%)

doi:10.6342/NTU202301887

				MDS-f	6 (3.6%)
				AML with *NPM1*	9 (5.4%)
				AML with *NUP98* rearrangement	6 (3.6%)
MDS with EB	11 (5.6%)	**MDS-EB2**	19	**MDS-bi*TP53***	30 (15.2%)
MDS with mutated *TP53*	3 (1.5%)			**MDS-IB2**	146 (74.1%)
MDS/AML with mutated *TP53*	41 (20.8%)			**MDS-f**	6 (3.0%)
MDS/AML with MDS-G*	95 (48.2%)			**AML with *NPM1***	10 (5.1%)
MDS/AML with MDS-C[†]	12 (6.1%)			**AML with *MECOM* rearrangement**	1 (0.5%)
MDS/AML, NOS	21 (10.7%)			**AML with *NUP98* rearrangement**	4 (2.0%)
AML with mutated *NPM1*	10 (5.1%)				
AML with in-frame bZIP *CEBPA* mutation	3 (1.5%)				
AML with inv(3)(q21.3q26.2)	1 (0.5%)				
MDS, NOS, with SLD	7 (87.5%)	**MDS-U**	8	**MDS-LB**	4 (50.0%)
MDS with EB	1 (12.5%)			**MDS-h**	4 (50.0%)

Abbreviations: MDS, myelodysplastic syndrome; SLD, single lineage dysplasia; MLD, multilineage dysplasia; EB, excess of blasts; RS, ring sideroblast; U, unclassifiable; AML, acute myeloid leukemia; LB, low blasts; h, hypoplastic; f, fibrosis; bi*TP53*, biallelic *TP53* inactivation; IB, increased blasts; G, MDS-related gene mutations; C, MDS-related cytogenetic abnormalities

*MDS-related gene mutations: *ASXL1*, *BCOR*, *EZH2*, *RUNX1*, *SF3B1*, *SRSF2*, *STAG2*, *U2AF1*, or *ZRSR2*

[†] MDS-related cytogenetic abnormalities: complex (≥3 clones) karyotype (in the absence of a *TP53* mutation), del(5q)/t(5q)/add(5q), -7/del(7q), +8, del(12p)/t(12p)/add(12p), i(17q), -17/add(17p) or del(17p), del(20q), and/or idic(X)(q13) clonal abnormalities

doi:10.6342/NTU202301887

Supplemental Table 4. Comparison of clinical characteristics and genetic alterations between patients with myelodysplastic syndrome (MDS) and myelodysplastic syndrome/acute myeloid leukemia (MDS/AML), based on the 2022 International Consensus Classification

Clinical characters	Total (n=684)	MDS (n=515)	MDS/AML (n=169)	*P* value
Sex				0.516
Female	249 (36.4%)	191 (37.1%)	58 (34.3%)	
Male	435 (63.6%)	324 (62.9%)	111 (65.7%)	
Age*	66.5 (18.4-94.5)	66.6 (18.4-94.5)	66.0 (23.6-91.3)	0.887
Laboratory data*				
WBC, $\times 10^9$ /L	3.31 (0.59-26.33)	3.47 (0.66-26.33)	3.00 (0.59-20.44)	**0.022**
ANC, $\times 10^9$ /L	1.50 (0.0-16.59)	1.60 (0.02-16.59)	1.05 (0.0-9.91)	**<0.001**
Hb, g/dL	8.2 (2.6-17.1)	8.2 (3.2-17.1)	8.1 (2.6-14.6)	0.963
Platelet, $\times 10^9$ /L	78 (1-607)	76 (1-607)	78 (3-460)	0.137
BM blast (%)	4.6 (0.0-19.5)	2.5 (0.0-9.8)	13.4 (2.0-19.5)	**<0.001**
PB blast (%)	0.0 (0.0-19.0)	0.0 (0.0-9.1)	1.0 (0.0-19.0)	**<0.001**
IPSS-R[†]				
Very low	22 (3.2%)	22 (4.3%)	0 (0.0%)	**0.004**
Low	180 (26.3%)	180 (34.9%)	0 (0.0%)	**<0.001**
Int	179 (26.2%)	168 (32.6%)	11 (6.5%)	**<0.001**
High	152 (22.2%)	85 (16.5%)	67 (39.6%)	**<0.001**
Very high	151 (22.1%)	60 (11.7%)	91 (53.9%)	**<0.001**
Epigenetics modifiers	310 (45.3%)	205 (39.8%)	105 (62.1%)	**<0.001**
DNA methylation	179 (26.2%)	119 (23.1%)	60 (35.5%)	**0.001**
DNMT3A	62 (9.1%)	42 (8.2%)	20 (11.8%)	0.148
TET2	99 (14.5%)	72 (14.0%)	27 (16.0%)	0.522
IDH1	5 (0.7%)	0 (0.0%)	5 (3.0%)	**0.001**
IDH2	27 (3.9%)	13 (2.5%)	14 (8.3%)	**0.001**
WT1	11 (1.6%)	8 (1.6%)	3 (1.8%)	0.738
Chromatin modifiers	199 (29.1%)	126 (24.5%)	73 (43.2%)	**<0.001**
ASXL1	142 (20.8%)	89 (17.3%)	53 (31.4%)	**<0.001**
EZH2	28 (4.1%)	18 (3.5%)	10 (5.9%)	0.168

doi:10.6342/NTU202301887

MLL	9 (1.3%)	6 (1.2%)	3 (1.8%)	0.697
SETBP1	18 (2.6%)	13 (2.5%)	5 (3.0%)	0.783
BCOR	37 (5.4%)	21 (4.1%)	16 (9.5%)	**0.007**
BCORL1	12 (1.8%)	5 (1.0%)	7 (4.1%)	**0.013**
PHF6	11 (1.6%)	4 (0.8%)	7 (4.1%)	**0.007**
Activated signaling	62 (9.1%)	36 (7.0%)	26 (15.4%)	**0.001**
***FLT3*-ITD**	2 (0.3%)	1 (0.2%)	1 (0.6%)	0.433
***FLT3*-TKD**	3 (0.4%)	1 (0.2%)	2 (1.2%)	0.153
KIT	4 (0.6%)	2 (0.4%)	2 (1.2%)	0.256
KRAS	4 (0.6%)	3 (0.6%)	1 (0.6%)	>0.999
NRAS	23 (3.4%)	11 (2.1%)	12 (7.1%)	**0.002**
PTPN11	8 (1.2%)	4 (0.8%)	4 (2.4%)	0.109
JAK2	10 (1.5%)	9 (1.7%)	1 (0.6%)	0.465
CBL	14 (2.0%)	8 (1.6%)	6 (3.6%)	0.112
GNAS	2 (0.3%)	2 (0.4%)	0 (0.0%)	>0.999
Transcription factor	132 (19.3%)	75 (14.6%)	57 (33.7%)	**<0.001**
CEBPA	27 (3.9%)	17 (3.3%)	10 (5.9%)	0.130
RUNX1	87 (12.7%)	49 (9.5%)	38 (22.5%)	**<0.001**
GATA2	10 (1.5%)	4 (0.8%)	6 (3.6%)	**0.018**
ETV6	19 (2.8%)	11 (2.1%)	8 (4.7%)	0.075
IKZF1	5 (0.7%)	1 (0.2%)	4 (2.4%)	**0.015**
Spliceosome-complex	234 (34.2%)	171 (33.2%)	63 (37.3%)	0.333
U2AF1	55 (8.0%)	38 (7.4%)	17 (10.1%)	0.266
SRSF2	65 (9.5%)	41 (8.0%)	24 (14.2%)	**0.016**
ZRSR2	30 (4.4%)	19 (3.7%)	11 (6.5%)	0.120
SF3B1	95 (13.9%)	80 (15.5%)	15 (8.9%)	**0.030**
Cohesin complex	92 (13.5%)	53 (10.3%)	39 (23.1%)	**<0.001**
RAD21	3 (0.4%)	3 (0.6%)	0 (0.0%)	>0.999
SMC1A	2 (0.3%)	1 (0.2%)	1 (0.6%)	0.433
SMC3	1 (0.1%)	1 (0.2%)	0 (0.0%)	>0.999
STAG2	87 (12.7%)	49 (9.5%)	38 (22.5%)	**<0.001**
Tumor suppressor	93 (13.6%)	50 (9.7%)	43 (25.4%)	**<0.001**
TP53	86 (12.6%)	44 (8.5%)	42 (24.9%)	**<0.001**
CUX1	7 (1.0%)	6 (1.2%)	1 (0.6%)	>0.999
NPM1	10 (1.5%)	10 (1.9%)	0 (0.0%)	0.131

doi:10.6342/NTU202301887

Median number of mutations	1 (0-8)	1 (0-8)	2 (0-7)	**<0.001**

P values of <0.05 are statistically significant.

*Median (range).

†IPSS-R: very low, ≤1.5; low, >1.5-3; intermediate (INT), >3-4.5; high, >4.5-6; and very high, >6.

Data are presented as n (%)

Note: Large insertion in *FLT3*-ITD and high GC content in *CEBPA* limited the detection and quantification by NGS.

Abbreviations: ANC, absolute neutrophil count; Hb, hemoglobin; IPSS-R, revised international prognosis scoring system

doi:10.6342/NTU202301887

Supplemental Table 5. Mutations and VAF in genes categorized by functional groups in myelodysplastic syndrome patients

Genes	Total (n=698)	Median VAF (%)	Range of VAF (%)
Epigenetics modifiers	318 (45.6%)		
DNA methylation-related	185 (26.5%)		
DNMT3A	67 (9.6%)	34.6	5.0-91.3
TET2	100 (14.3%)	38.8	5.1-78.4
IDH1	5 (0.7%)	13.5	6.4-21.1
IDH2	27 (3.9%)	30.6	6.8-47.3
WT1	11 (1.6%)	10.8	6.4-48.4
Chromatin modifiers	201 (28.8%)		
ASXL1	142 (20.3%)	30.7	5.1-63.8
EZH2	28 (4.0%)	38.2	5.0-90.1
MLL	9 (1.3%)	5.7	5.1-56.0
SETBP1	18 (2.6%)	25.3	7.4-48.5
BCOR	38 (5.4%)	31.4	5.4-94.8
BCORL1	12 (1.7%)	16.7	5.3-99.5
PHF6	12 (1.7%)	12.9	6.4-89.0
Activated signaling	66 (9.5%)		
***FLT3*-ITD**	3 (0.4%)	-	-
***FLT3*-TKD**	3 (0.4%)	30.7	9.1-88.7
KIT	4 (0.6%)	50.7	49.1-99.8
KRAS	4 (0.6%)	29.4	12.4-38.1
NRAS	24 (3.4%)	10.7	5.0-45.4
PTPN11	9 (1.3%)	20.8	5.8-29.2
JAK2	11 (1.6%)	31.6	6.8-96.6
CBL	14 (2.0%)	27.3	7.3-85.5
GNAS	2 (0.3%)	-	-
Transcription factor	138 (19.8%)		
CEBPA	32 (4.6%)	11.7	5.1-96.0
RUNX1	88 (12.6%)	30.5	5.0-61.9
GATA2	10 (1.4%)	31.4	7.9-47.9
ETV6	19 (2.7%)	32.1	6.5-48.0
IKZF1	6 (0.9%)	52.7	8.1-56.1
Spliceosome-complex	236 (33.8%)		

doi:10.6342/NTU202301887

U2AF1	56 (8.0%)	34.0	8.1-58.3
SRSF2	65 (9.3%)	45.9	17.9-66.2
ZRSR2	30 (4.3%)	66.5	5.2-93.7
SF3B1	96 (13.8%)	33.9	5.6-47.8
Cohesin complex	93 (13.3%)		
RAD21	4 (0.6%)	13.7	7.3-29.2
SMC1A	2 (0.3%)	63.2	63.2
SMC3	2 (0.3%)	6.7	6.6-6.8
STAG2	87 (12.5%)	29.8	5.0-97.6
Tumor suppressor	93 (13.3%)		
TP53	86 (12.3%)	46.2	5.2-93.9
CUX1	7 (1.0%)	34.8	19.5-93.1
NPM1	20 (2.9%)	26.2	5.6-39.7

Note: Large insertion in *FLT3*-ITD and high GC content in *CEBPA* limited the detection and quantification by NGS.

Abbreviations: VAF, variant allele frequency

doi:10.6342/NTU202301887

Supplemental Table 6. Univariable Cox regression analysis of the impact of different variables on the leukemia-free survival and overall survival of myelodysplastic syndrome patients

	LFS		OS	
Variable	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**
Age*	1.023 (1.016-1.029)	**<0.001**	1.025 (1.018-1.032)	**<0.001**
IPSS-R†	3.609 (2.911-4.473)	**<0.001**	3.537 (2.838-4.409)	**<0.001**
Female	0.702 (0.566-0.872)	**0.001**	0.668 (0.531-0.836)	**<0.001**
HSCT	0.738 (0.531-0.970)	**0.030**	0.634 (0.474-0.848)	**0.002**
HMA	1.971 (1.587-2.450)	**<0.001**	1.671 (1.331-2.099)	**<0.001**
Genetic alteration				
DNMT3A	1.516 (1.102-1.227)	**0.010**	1.388 (0.994-1.938)	0.054
TET2	1.614 (1.481-1.549)	**0.001**	1.546 (1.160-2.060)	**0.003**
IDH2	2.266 (1.481-3.467)	**<0.001**	1.970 (1.250-3.105)	**0.003**
ASXL1	1.951 (1.549-2.458)	**<0.001**	1.830 (1.442-2.321)	**<0.001**
EZH2	1.688 (1.075-2.653)	**0.023**	1.637 (1.018-2.634)	**0.042**
SETBP1	1.831 (1.092-3.072)	**0.022**	1.545 (0.905-2.637)	0.111
BCOR	0.985 (0.621-1.564)	0.951	1.036 (0.653-1.646)	0.880
BCORL1	1.946 (1.003-3.773)	**0.049**	1.898 (0.978-3.683)	0.058
NRAS	1.944 (1.223-3.089)	**0.005**	1.590 (0.945-2.592)	0.063
JAK2	0.796 (0.329-1.924)	0.612	0.875 (0.362-2.116)	0.766
WT1	0.857 (0.382-1.922)	0.708	0.581 (0.217-1.558)	0.281
CBL	2.535 (1.305-4.926)	**0.006**	2.327 (1.199-4.520)	**0.013**
CEBPA	1.423 (0.925-2.190)	0.109	1.352 (0.861-2.123)	0.190
RUNX1	1.845 (1.392-2.446)	**<0.001**	1.741 (1.306-2.320)	**<0.001**
GATA2	1.079 (0.445-2.618)	0.866	0.659 (0.211-2.057)	0.473
ETV6	1.588 (0.892-2.829)	0.116	1.133 (0.603-2.127)	0.698
U2AF1	1.483 (1.067-2.061)	**0.019**	1.513 (1.080-2.122)	**0.016**
SRSF2	1.895 (1.382-2.597)	**<0.001**	1.730 (1.248-2.397)	**0.001**
ZRSR2	1.715 (1.114-2.639)	**0.014**	1.704 (1.096-2.650)	**0.018**
SF3B1	0.671 (0.483-0.931)	**0.017**	0.698 (0.500-0.974)	**0.035**
STAG2	2.199 (1.693-2.855)	**<0.001**	2.030 (1.552-2.656)	**<0.001**
PHF6	1.442 (0.716-2.907)	0.306	1.570 (0.779-3.166)	0.207
TP53	6.443 (4.891-8.488)	**<0.001**	8.177 (6.152-10.869)	**<0.001**
NPM1	2.428 (1.444-4.081)	**0.001**	1.589 (0.870-2.901)	0.132

Abbreviations: HR, Hazard ratios; CI, confidence interval; LFS, leukemia-free survival; OS, overall survival

P values of <0.05 are statistically significant.

13 genes (*IDH1*, *FLT3*-ITD/TKD, *KIT*, *KRAS*, *PTPN11*, *GNAS*, *IKZF1*, *RAD21*, *SMC1A*, *SMC3*, *CUX1*, *MLL*) with less than 10 mutated events were not included.

*Age, as a continuous variable analysis.

†IPSS-R higher-risk (high, very high) group vs. lower-risk (very low, low, intermediate) group.

doi:10.6342/NTU202301887

Supplemental Table 7. Multivariable analysis for leukemia-free survival and overall survival in myelodysplastic syndrome patients

	LFS		OS	
Variable	**HR (95% CI)**	**P value**	**HR (95% CI)**	**P value**
Age*	1.015 (1.006-1.024)	**0.001**	1.017 (1.008-1.027)	**<0.001**
IPSS-R†	2.645 (2.027-3.450)	**<0.001**	2.826 (2.150-3.715)	**<0.001**
Female	0.795 (0.634-0.996)	**0.047**	0.755 (0.598-0.954)	**0.018**
HSCT	0.665 (0.465-0.951)	**0.025**	0.592 (0.405-0.866)	**0.007**
HMA	1.208 (0.940-1.551)	0.140	1.032 (0.795-1.341)	0.812
DNMT3A	1.572 (1.102-2.244)	**0.013**	1.498 (1.029-2.180)	**0.035**
TET2	1.556 (1.155-2.097)	**0.004**	1.441 (1.055-1.968)	**0.022**
IDH2	1.988 (1.233-3.205)	**0.005**	1.830 (1.108-3.023)	**0.018**
ASXL1	1.040 (0.756-1.430)	0.811	1.014 (0.732-1.404)	0.935
EZH2	1.043 (0.617-1.763)	0.875	1.119 (0.644-1.947)	0.690
SETBP1	1.532 (0.888-2.642)	0.126	-	-
BCORL1	1.243 (0.585-2.638)	0.572	1.118 (0.521-2.402)	0.774
NRAS	1.695 (1.009-2.849)	**0.046**	1.491 (0.863-2.576)	0.152
CBL	3.389 (1.696-6.774)	**0.001**	3.127 (1.556-6.282)	**0.001**
RUNX1	1.124 (0.818-1.545)	0.470	1.126 (0.813-1.559)	0.477
U2AF1	1.236 (0.838-1.822)	0.285	1.392 (0.940-2.060)	0.099
SRSF2	1.208 (0.827-1.765)	0.329	1.174 (0.790-1.745)	0.426
ZRSR2	1.510 (0.925-2.463)	0.099	1.593 (0.959-2.645)	0.072
SF3B1	0.793 (0.554-1.135)	0.205	0.797 (0.551-1.152)	0.227
STAG2	1.289 (0.929-1.787)	0.129	1.229 (0.881-1.714)	0.226
TP53	4.765 (3.445-6.590)	**<0.001**	5.878 (4.203-8.221)	**<0.001**

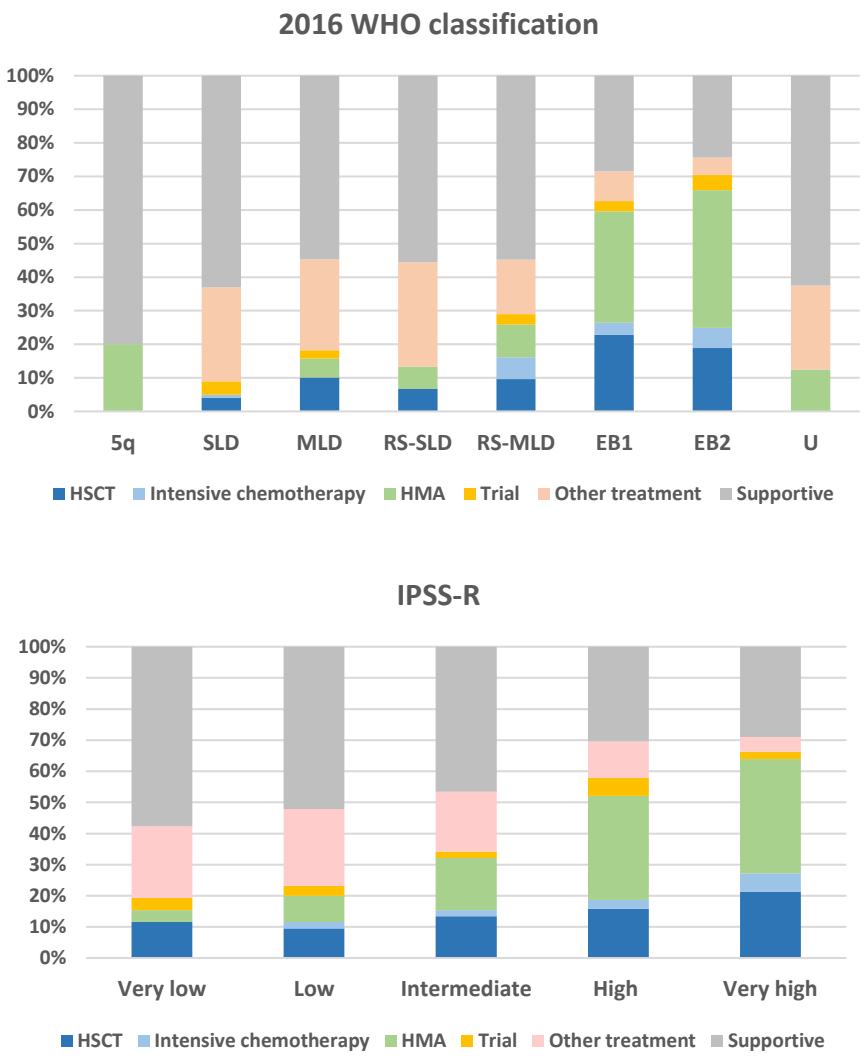
Abbreviations: HR, Hazard ratios; CI, confidence interval; LFS, leukemia-free survival; OS, overall survival

P values of <0.05 are statistically significant. Mutations with limited mutated events were not included.

*Age, as a continuous variable analysis.

†IPSS-R higher-risk (high, very high) group vs. lower-risk (very low, low, intermediate) group.

doi:10.6342/NTU202301887



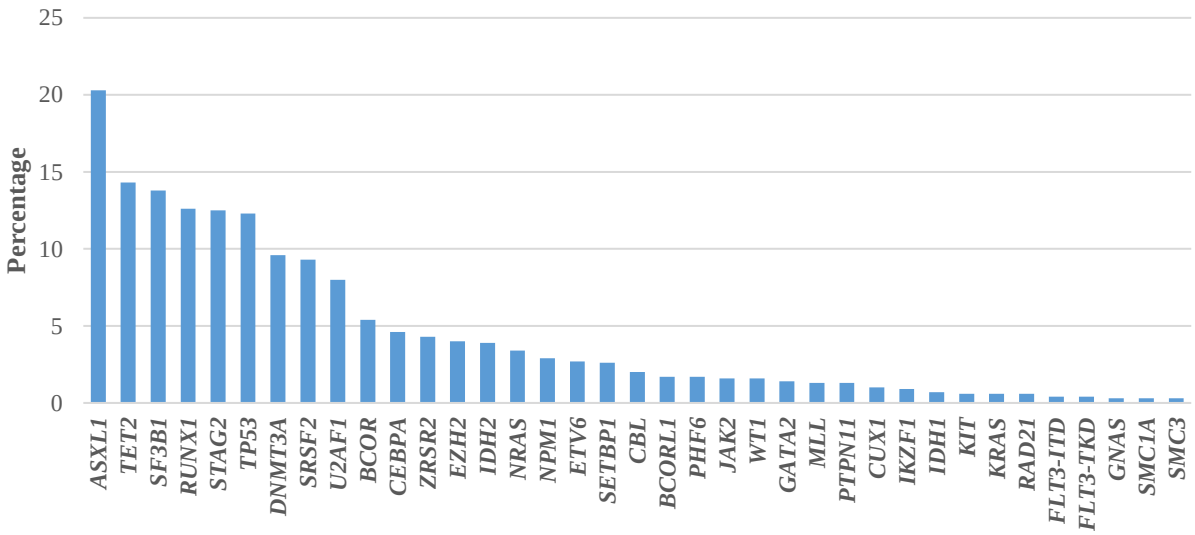
Supplemental Figure 1. Differences in treatment among subgroups according to the 2016 WHO or IPSS-R classification

(a) Differences in treatments among subgroups according to the 2016 WHO classification

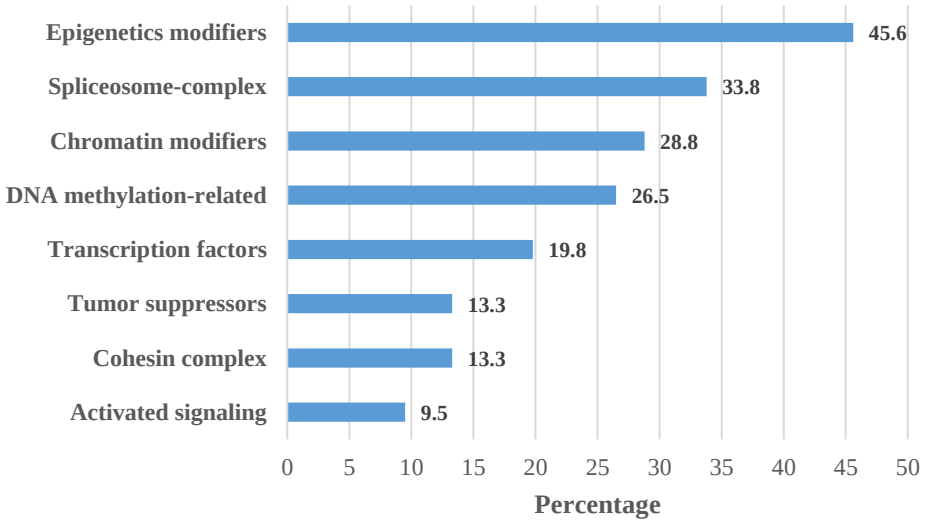
(b) Differences in treatments among subgroups according to IPSS-R risk group

doi:10.6342/NTU202301887

(a)

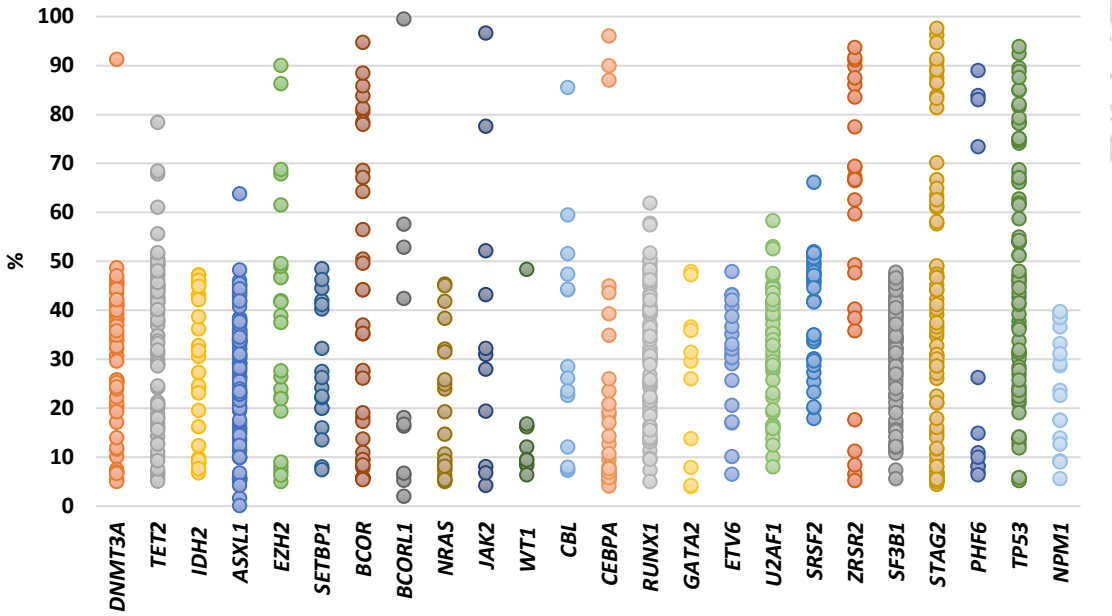


(b)



Supplemental Figure 2. Frequencies of the 37 commonly occurred mutations (a) and frequencies of mutations categorized by the functional groups (b) in MDS patients

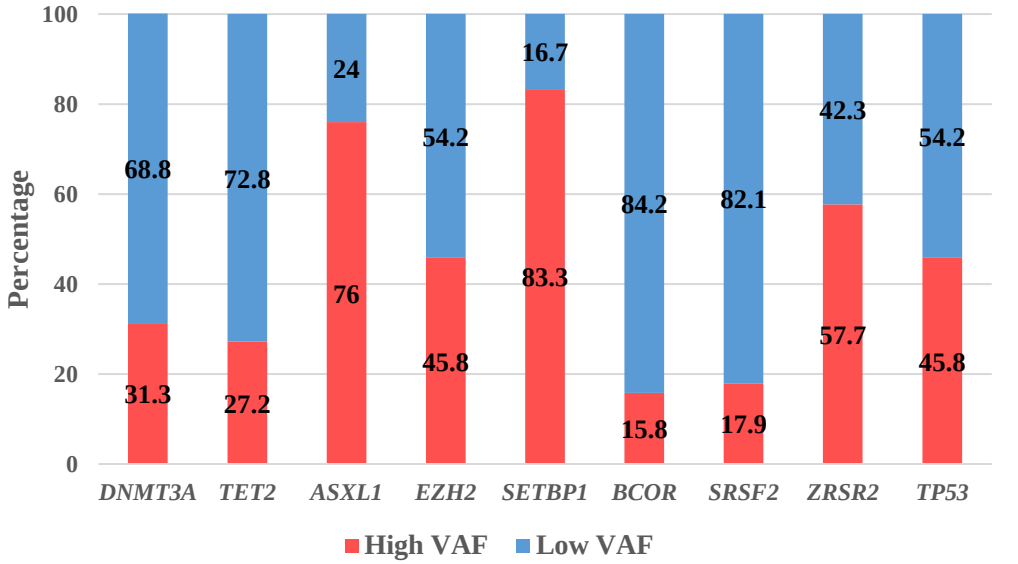
doi:10.6342/NTU202301887



Note: 13 genes (*IDH1*, *FLT3*-ITD/TKD, *KIT*, *KRAS*, *PTPN11*, *GNAS*, *IKZF1*, *RAD21*, *SMC1A*, *SMC3*, *CUX1*, *MLL*) with less than 10 mutated events were not included.

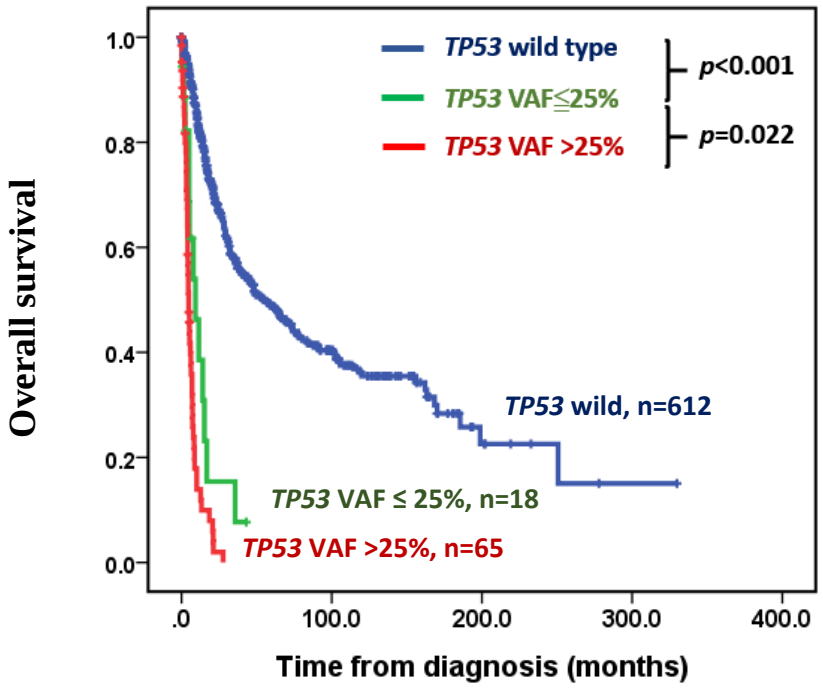
Supplemental Figure 3. Scatter plots of mutation variant allele frequencies (VAF) for the commonly occurred gene mutations

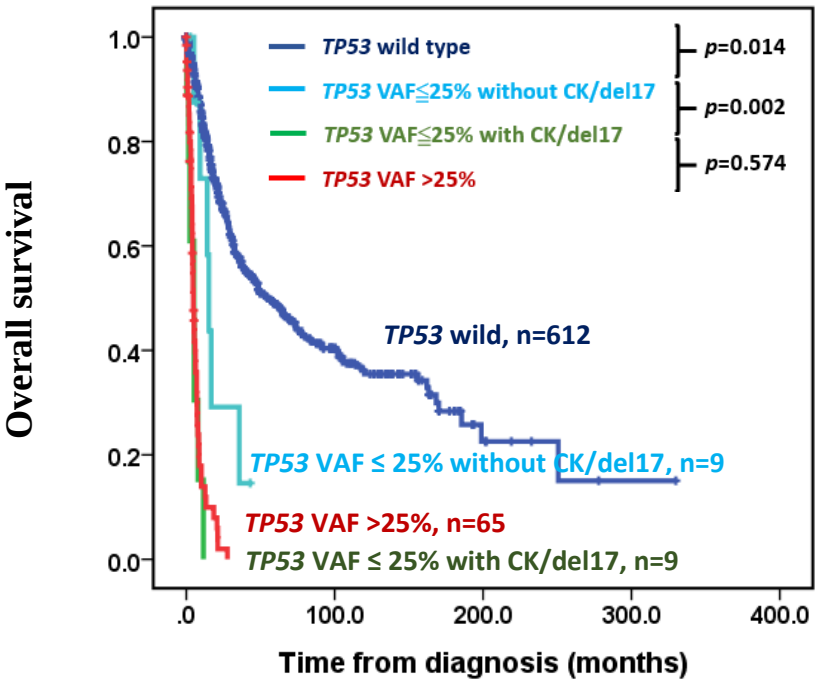
doi:10.6342/NTU202301887



Supplemental Figure 4. The distribution of high/low variant allelic frequency (VAF) of the 9 gene mutations in which VAF had impact on the survival

doi:10.6342/NTU202301887



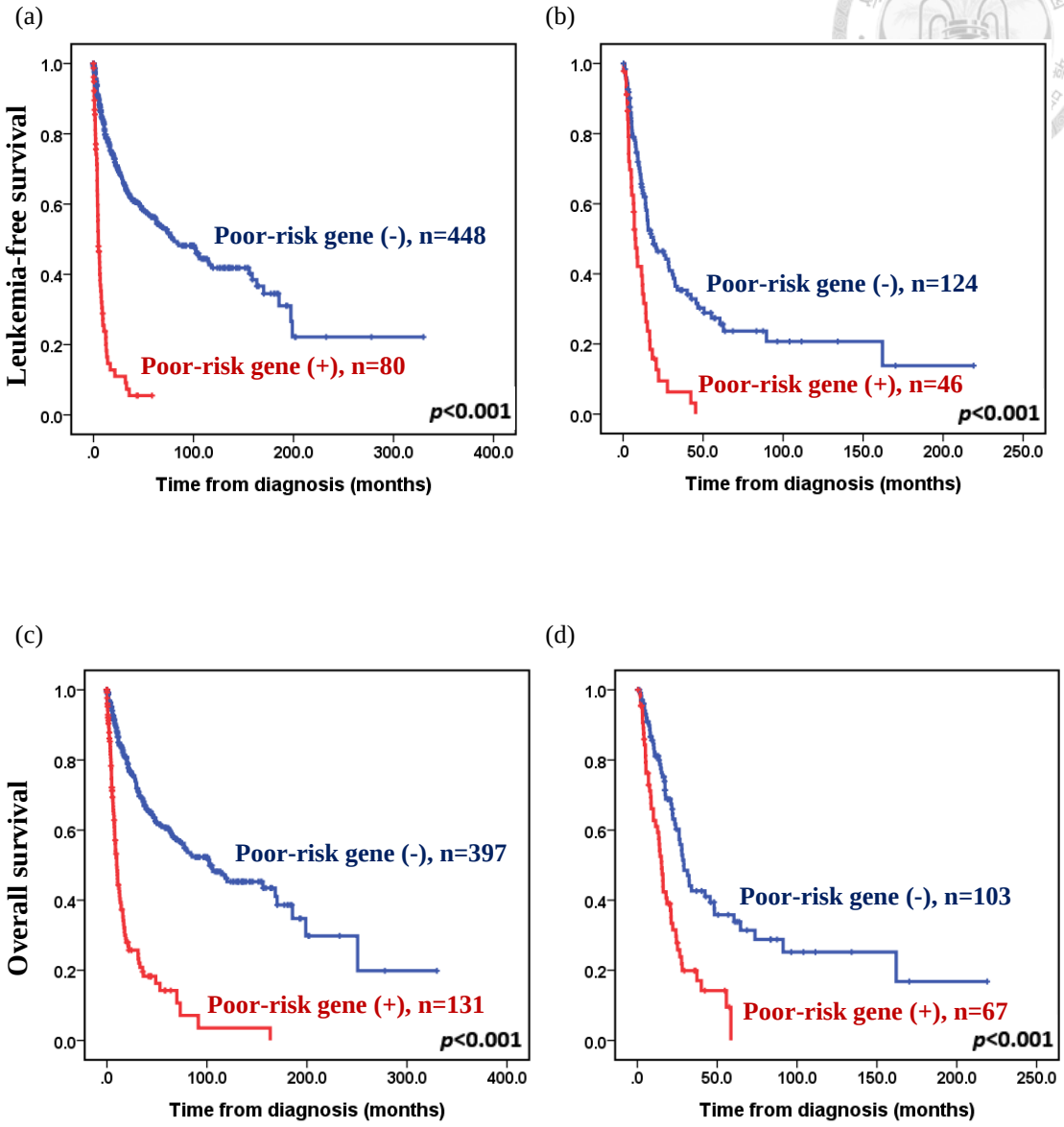


Supplemental Figure 5. Kaplan–Meier curves of myelodysplastic syndrome patients with different status of TP53 mutations and complex karyotype (CK) or chromosome 17 deletion (del17)

(a) Patients were stratified by mutational status and variant allelic frequency (VAF) of *TP53*.

(b) Patients were stratified by mutational status, VAF of *TP53* and the presence/absence of CK/del17.

doi:10.6342/NTU202301887



*Note: poor-risk gene mutations for LFS: mutant *TP53*, *IDH2*, and *CBL*; for OS: *DNMT3A* and *ZRSR2* mutations with high VAF, and mutant *TP53*, *IDH2*, *CBL* and *U2AF1*

Supplemental Figure 6. Kaplan–Meier curves of patients with or without at least one of the poor-risk mutations: impact of HMA treatment

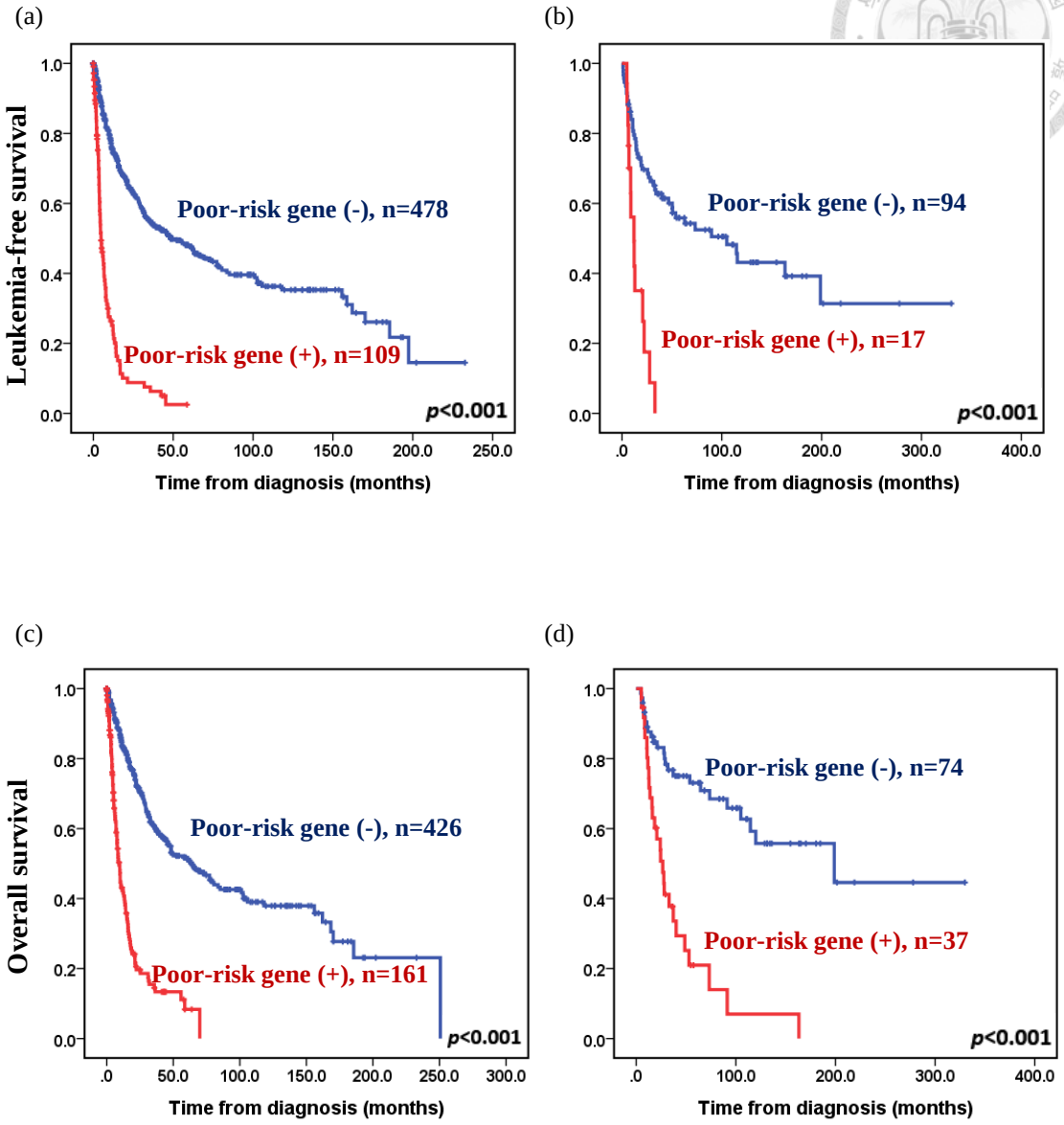
(a) Leukemia-free survival for patients not receiving HMA treatment.

(b) Leukemia-free survival for patients receiving HMA treatment.

(c) Overall survival for patients not receiving HMA treatment.

(d) Overall survival for patients receiving HMA treatment.

doi:10.6342/NTU202301887



*Note: poor-risk genes for LFS: mutant *TP53*, *IDH2*, and *CBL*; for OS: *DNMT3A* and *ZRSR2* mutations with high VAF, and mutant *TP53*, *IDH2*, *CBL* and *U2AF1*

Supplemental Figure 7. Kaplan–Meier curves of patients with or without one of the poor-risk mutations: impact of HSCT

(a) Leukemia-free survival for patients not receiving HSCT treatment.

(b) Leukemia-free survival for patients receiving HSCT treatment.

(c) Overall survival for patients not receiving HSCT treatment.

(d) Overall survival for patients receiving HSCT treatment.

doi:10.6342/NTU202301887