Genetically inspired prognostic scoring system (GIPSS) outperforms dynamic international prognostic scoring system (DIPSS) in myelofibrosis patients

Andrew T. Kuykendall | Chetasi Talati | Eric Padron | Kendra Sweet | David Sallman | Alan F. List | Jeffrey E. Lancet | Rami S. Komrokji
H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida
Correspondence
Andrew T. Kuykendall, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612.
Email: andrew.kuykendall@moffitt.org

Abstract
A genetically inspired prognostic scoring system (GIPSS) that stratifies primary myelofibrosis (PMF) patients by genetic variants alone was recently proposed. While non-inferior to the dynamic international prognostic scoring system (DIPSS), the lack of overlapping prognostic variables between the models leads to increased risk for disagreement between two valid prognostic models and presents a challenging clinical situation. In an external cohort of 266 molecularly annotated myelofibrosis (MF) patients, we demonstrated that the GIPSS model significantly differentiated between four risk groups (low, int-1, int-2, high) with median OS that was not reached, not reached, 60.5 and 28.9 months, respectively. High-risk patients had significantly inferior leukemia-free survival (LFS) \( (P < 0.0001) \). We identified a cohort of prognostically ambiguous patients \( (n = 39) \) in which GIPSS and DIPSS models differed by \( \geq 2 \) risk groups. Among these patients, a similar proportion were upstaged by DIPSS \( (n = 19) \) and GIPSS \( (n = 20) \). Patients upstaged by GIPSS (genetically high-risk) had a trend toward inferior OS compared with patients upstaged by DIPSS (clinically high-risk) \( (P = .08) \) and significantly worse LFS \( (P = .04) \). Patients deemed intermediate-2 and high-risk by GIPSS who underwent allogeneic transplant had improved OS compared with those that did not \( (P = .04) \). GIPSS is a valid disease-specific prognostic system and outperforms DIPSS in patients where the two models disagree. Additionally, while GIPSS was developed for PMF; the current study shows, however, that the contemporary genetic model performs equally well for both primary and secondary myelofibrosis.

1 | INTRODUCTION

Primary myelofibrosis (PMF) is a chronic myeloproliferative neoplasm marked by constitutional symptoms, cytopenias, extramedullary hematopoiesis, and a propensity to transform into acute myeloid leukemia (AML). While the median survival after diagnosis has been reported as 6-10 years, the disease course is heterogeneous with some patients quickly transforming to AML while others survive decades without treatment.\(^1\)\(^-\)\(^3\) To date, the only treatment with curative potential is allogeneic stem cell transplant (AH SCT) which comes with a significant risk for morbidity and mortality. Numerous prognostic scoring systems have been developed and validated in an effort to better identify patients with high-risk disease who would most benefit from AH SCT.\(^1\)\(^,\)\(^4\)\(^-\)\(^9\)

Historically, prognostic systems in myelofibrosis (MF) have relied solely upon clinical variables with arbitrary cutpoints often utilized to separate “high” from “low.” The Lille score, developed in 1996, stratified patients into three risk groups based solely on hemoglobin and white blood cell count.\(^10\) Constitutional symptoms, anemia, leukocytosis, peripheral blasts, and age were used in the international prognostic scoring system (IPSS) which has served as the backbone to numerous subsequent models.\(^1\)\(^,\)\(^4\)\(^-\)\(^6\)\(^,\)\(^8\) While consistently shown to be prognostic, clinical variables are not disease-specific, are often treated as dichotomous variables despite their continuous nature, and can fluctuate due to disease-related and disease-independent factors.

Approximately 10 different prognostic models have been developed in abstract or published form over the past decade—each incorporating new variables that have been shown to be independently prognostic in MF (Supporting Information Figure S1A). Increasingly, these models have incorporated molecular information that has been shown to have disease-specific prognostic value. The progressive utilization of molecular information culminated with the publication of the genetically inspired prognostic scoring system for primary myelofibrosis (GIPSS) which relies solely on genomic inputs to stratify patients into low,
intermediate-1 (int-1), intermediate-2 (int-2), and high risk categories. While the development of an ostensibly disease-specific prognostic model is certainly a substantive achievement, the lack of shared prognostic inputs with traditional models (eg, IPSS, DIPSS, etc.) raises the potential for significant disagreement between models. Resultantly, a subpopulation of prognostically ambiguous patients emerges that are deemed higher risk by one model and lower risk by another. This presents a clinical—and ethical—challenge in terms of counseling patients on treatment options and ultimately deciding whether or not to recommend a morbid, but potentially curative, therapy.

In this study, we aim to validate the recently published GIPSS model and assess its performance compared with the clinically based dynamic international prognostic model (DIPSS) in a group of prognostically ambiguous patients.

## METHODS

Using a comprehensive database of MF patients treated at our institution, we identified patients with a diagnosis of PMF, post-polycythemia (post-PV), or post-essential thrombocytopenia (post-ET) MF by WHO 2016 or IWG-MRT criteria, respectively, who were fully annotated for DIPSS and GIPSS prognostic modeling. Clinical variables were assessed at time of first presentation to our center. Cytogenetic assessment was performed at time of diagnostic bone marrow biopsy.
regarding gene sequencing for driver mutations (JAK2, MPL, and CALR) and cooperating high risk mutations was performed as a part routine clinical care with the former often performed at time of diagnosis and first presentation to our center, while the latter was performed at time of first presentation. High-risk mutations were defined as ASXL1, SRSF2, and U2AF1 Q157 given their previously reported prognostic value and inclusion in the GIPSS prognostic model.11–16 Cooperating high risk mutations were tested via targeted amplicon next generation sequencing, while driver mutations were tested via both targeted amplicon next generation sequencing and direct sequencing.

Differences in the distribution of continuous variables were analyzed by t-test. Differences in nominal variables were compared by chi-squared test. Survival curves were created by Kaplan–Meier method and compared by log-rank test. Overall survival (OS) and leukemia-free survival (LFS) were calculated from date of diagnosis until date of death with patients being censored at last follow-up or date of AHSCT. In the analysis of the benefit of AHSCT in high-molecular risk patients, censoring at time of AHSCT was not performed. P-values of <.05 were considered significant. GraphPad Prism 7.04 was used for all calculations and production of survival curves.

3 | RESULTS

We analyzed 266 MF (PMF = 177, post-PV = 36, and post-ET MF = 51) patients who were fully annotated for GIPSS and DIPSS modeling. Median OS for the entire cohort was 98 months. With a median follow-up of 30.5 months, 67 (25%) patients had died and 19 (7%) had undergone AHSCT. Applying the GIPSS model, four risk groups were categorized with low, int-1, int-2, and high-risk groups comprised of 29, 104, 87, and 46 patients, respectively. Median OS by risk group was not reached (low), not reached (int-1), 60.5 months (int-2), and 28.9 months (high). Each risk group was statistically different than the risk group(s) that were directly adjacent (low vs int-1, \( P = .03 \); int-1 vs int-2, \( P < .0001 \); int-2 vs high, \( P = .03 \)) (Figure 1A). Median LFS by risk group was not reached in the low, int-1 or int-2 risk groups and was 35.4 months in the high-risk group (\( P < .0001 \)) with a 3-year LFS of 46% (Figure 1B).

Given that GIPSS was developed for PMF, we performed a separate analysis of post-PV and post-ET MF patients. Due to small numbers, int-2 and high-risk groups were combined resulting in three risk groups categorized as low \( (n = 7) \), intermediate \( (n = 46) \), and high \( (n = 34) \). High-risk patients had mOS of 32.7 months compared with a median survival that was not reached in the intermediate and low-risk groups (\( P = .009 \)) (Supporting Information Figure S1B).

We then assessed the concordance of GIPSS and DIPSS modeling in our cohort. A *significantly discordant* patient was defined as a patient in whom GIPSS and DIPSS models differed by ≥2 risk groups (ie, high by DIPSS and low/int-1 by GIPSS, int-2 by GIPSS, and low by DIPSS). Among 266 patients, 39 (14.6%) were found to be significantly discordant, with 20 (51%) up-staged by GIPSS (genomically high-risk) and 19 (49%) up-staged by DIPSS (clinically high-risk) (Figure 2). In the genomically high-risk cohort, 17 (85%) were high-risk
by GIPSS and 3 (15%) were int-2 risk. In the clinically high-risk group, 10 (53%) were high-risk by DIPSS and 9 (47%) were int-2 risk.

Patient demographics, clinical, and genomic characteristics are summarized in Table 1. In general, the genomically high-risk cohort was enriched for high-risk molecular features while the clinically high-risk cohort was older with a higher incidence of constitutional symptoms and hematologic abnormalities. A similar proportion of secondary MF patients were present in each cohort.

Direct comparison between the two cohorts of significantly discordant patients revealed a trend toward inferior OS in genomically high-risk patients compared with clinically high-risk patients (mOS 37.5 vs 109 months, P = .08) (Supporting Information Figure S1C).

Genomically high-risk patients had a significantly inferior LFS compared with clinically high-risk patients (5-year LFS 54% vs 100%, respectively) (OR: 6.3 [1.3-32], P = .04) (Supporting Information Figure S1D). OS curves were then compared with curves created by applying DIPSS and GIPSS modeling to the larger cohort of fully annotated patients. Genomically high-risk patients trended closely with higher-risk groups as defined by both DIPSS and GIPSS, while the clinically high-risk patients trended more closely with intermediate risk groups as defined by both DIPSS and GIPSS (Figure 3).

Lastly, we evaluated patients deemed int-2 and high-risk by GIPSS who underwent AH SCT. Among 133 patients, 11 underwent AH SCT. Those undergoing AH SCT had improved mOS compared with those who did not undergo AH SCT (OR 0.40 [0.16-0.96], P = .04) suggesting, in general, AH SCT can overcome high-risk genomic features (Supporting Information Figure S1E). Notably, none of the higher risk GIPSS patients who underwent AH SCT were part of the significantly discordant cohort.

### 4 | DISCUSSION

With the benefit of large patient databases, and the widespread of next generation sequencing technology for somatic mutations, the discovery of variables that independently impact outcomes in MF has flourished over the past decade. This, in turn, has led to the proliferation of prognostic models that build upon their predecessors in an effort to better risk-stratify patients who share a diagnosis that is both phenotypically and prognostically heterogeneous. Efforts to make models age-independent, disease-specific, and dynamic result in models that vary slightly from their source. Nevertheless, the evolutionary result of continued small manipulations is the production of models that markedly differ from their distant, but still oft utilized, predecessors. In a subset of patients, this results in a clinically challenging situation whereby a patient is categorized to be both high and low risk, with survival estimates by two models that can differ by more than a decade.

In this study, we externally validated the genomically-driven GIPSS model in our cohort of primary, post-PV, and post-ET MF patients and demonstrated it is able to identify a high-risk subgroup of patients that frequently transforms to AML. The value of a disease-specific model cannot be understated as it eliminates many of the confounding factors built on clinical variables. Clinical variables have always served as nonspecific surrogates to estimate disease-risk. Often, clinical variables perform better in identifying a high-risk patient rather than high-risk disease. This is especially important as the primary goal of risk-stratification, as it pertains to MF, is to identify which patients should be referred for AH SCT—a potentially curative therapy that provides net benefit to higher-risk disease, but relies on otherwise healthy patients to withstand a highly morbid, and potentially mortal, treatment. This is not to say that the GIPSS model offers variable frequencies, mutation combinations, rare genomic insults, and determine what exactly constitutes a "true mutation" when it does not occur at a well-described DNA hotspot.

We also showed that the potential for discordant finding between GIPSS and DIPSS is significant. Prior analyses of the MIPSS70, MIPSS70-plus, and GIPSS have included comparisons to traditional clinical models to assess concordance. The MIPSS70 was compared with IPSS, MIPSS70-plus to DIPSS-plus and GIPSS to MIPSS70-plus.6,7 A high rate of concordance was seen in these
analyses, though comparison of MIPSS70 to IPSS compares a 3-tiered model to a 4-tiered model making concordance assessments more difficult. Moreover, concordance should be expected in these cases as they share many prognostic variables. In contrast, the lack of common variables between GIPSS and DIPSS accounts for the increased rate of discordance, highlighting that MF can be validly viewed in a clinically-specific and genomically-specific manner, and often these two views do not agree. The concordance is observed when the clinical phenotype reflects the biological phenotype. Decision trees have been proposed to guide usage of newer prognostic models in clinical practice, but these do not fully address the minority of patients who are prognostically ambiguous.6,7

Importantly, we also showed that int-2 and high-risk patients by GIPSS had improved survival when receiving AHSCT compared with
those who did not receive AHSCT. While this is an incomplete analysis with the potential for selection bias, it is nevertheless important as it addresses the basis of why we risk-stratify patients. If AHCT is unable to overcome high-risk disease, our prognostic systems no longer have interventional predictive power. The use of clinically-driven prognostic systems has been challenged in MF and is one of the driving forces behind the proliferation of new models.2 Genomically-driven models are likely to outperform clinical models in identifying transplant candidates given the disease-specific nature of the model, though this only holds true if the prognostic risk factors are sensitive to immune-based therapies, such as AHSCT. Recent studies have suggested that AHSCT can overcome high-risk disease in terms of cytogenetics17 and high-risk somatic mutations,18 however, larger studies are needed to confirm and refine these findings since all high-risk genomic findings are unlikely to be equally sensitive to transplant.

Finally, we acknowledge our study has several limitations, including its retrospective nature, limited follow-up time and inclusion of secondary MF patients who were not included in the development of the DIPSS or GIPSS. We chose to include these patients since the contribution of abnormal cytogenetics and driver mutations has been demonstrated in these populations,9,19 though we understand the impact of cooperating high-risk mutations has not been shown to the same degree.20 Nevertheless, while GIPSS was developed for PMF, the current study suggests that the contemporary genetic model performs reasonably well for both primary and secondary MF.

To conclude, we have validated the GIPSS model in an independent cohort of primary, post-ET, and post-PV MF patients. Additionally, we have shown that a small, but significant, population of patients have highly variable risk-classifications when subjected to both GIPSS and DIPSS modeling. In these patients, GIPSS appears to outperform DIPSS, reinforcing the importance of disease-specific variables when evaluating disease risk.

ORCID
Andrew T. Kuykendall https://orcid.org/0000-0002-9040-7415
Kendra Sweet https://orcid.org/0000-0001-6248-3317
Jeffrey E. Lancet https://orcid.org/0000-0002-7751-7310
Rami S. Komrokji https://orcid.org/0000-0002-1876-5269

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.