Published in final edited form as:
Expert Rev Hematol. 2014 April ; 7(2): 191–194. doi:10.1586/17474086.2014.891437.

Risk stratification in myelodysplastic syndromes: is there a role for gene expression profiling?

Amer M Zeidan[1,*], **Thomas Prebet**[2], **Ehab Saad Aldin**[3], and **Steven David Gore**[4]

[1]Department of Oncology, the Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21287, USA

[2]Département d'hématologie, Aix-Marseille University, Marseille, France

[3]Department of Internal Medicine, Medstar Good Samaritan Hospital, Baltimore, MD, USA

[4]Department of Medicine, Yale University, New Haven, CT, USA

Abstract

Patients with myelodysplastic syndromes (MDS) exhibit wide heterogeneity in clinical outcomes making accurate risk-stratification an integral part of the risk-adaptive management paradigm. Current prognostic schemes for MDS rely on clinicopathological parameters. Despite the increasing knowledge of the genetic landscape of MDS and the prognostic impact of many newly discovered molecular aberrations, none to date has been incorporated formally into the major risk models. Efforts are ongoing to use data generated from genome-wide high-throughput techniques to improve the 'individualized' outcome prediction for patients. We here discuss an important paper in which gene expression profiling (GEP) technology was applied to marrow $CD34^+$ cells from 125 MDS patients to generate and validate a standardized GEP-based prognostic signature.

Keywords

gene expression profiling; international prognostic scoring system; myelodysplastic syndromes; prognostication; risk stratification

Myelodysplastic syndromes (MDSs) comprise a group of clonal hematopoietic stem malignancies characterized by dysplastic differentiation and ineffective hematopoiesis, leading to variable degrees of peripheral blood (PB) cytopenias and risk of progression to acute myeloid leukemia [1]. The clinical course of patients with MDS is highly variable with outcomes ranging from survival limited to few months to a near-normal life expectancy [2]. The clinical and pathological heterogeneity of MDS are a reflection of an underlying network of highly complex and diverse disrupted critical biologic pathways [1,3]. The

*Author for correspondence: Tel.: +1 410 614 4459, Fax: +1 410 955 0185, azeidan1@jhmi.edu.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

clinical heterogeneity of MDS resulted in the development of a risk-adaptive management paradigm with therapeutic options ranging from observation all the way to upfront allogeneic stem cell transplantation [2,4]. Given the wide range of aggressiveness of available therapeutic options, a core component of optimal MDS management is the ability to accurately predict the prognostic outlook for individual patients [5]. Such evidence-based risk stratification approach would help put in context the benefit ratio of any proposed intervention especially considering the lack of any validated clinical or laboratory predictors of benefit from aggressive therapies [2,6].

Over the last three decades, multiple prognostic scoring systems have been developed and validated; the most widely used of which is the International Prognostic Scoring System (IPSS) [7]. These models depend on clinical (e.g., transfusion dependency) and laboratory criteria (e.g., peripheral blood [PB] cell counts, hemoglobin levels, percentage of blasts in PB and bone marrow [BM], cytogenetics and others) to separate patients into diverse prognostic groups with significantly different projected survivals and probabilities of leukemic progression [6,8]. Although these prognostic models are widely used clinically for therapeutic decisions and for trial eligibility determination, some of the parameters used to calculate the scores are evaluated subjectively (e.g., blast percentage). Approximately 10% of cases had their diagnosis IPSS score reclassified upon review in a tertiary center [9]. Moreover, while these scores tend to perform well for groups of patients, the accuracy of prognostication for any individual patient is more limited.

The last few years have witnessed an exponential increase in our knowledge of the underlying genetic lesions of MDS and their impact on the prognosis of the disease [2,10]. Although attempts have been made to incorporate some of the better-characterized prognostic molecular alterations into the commonly used clinical risk models to improve their prognostic discrimination, none of these modified risk models has been widely validated or adopted for routine clinical use [10]. In contrast to MDS, standardized molecular testing to identify patients with projected worse prognosis who should be considered for more aggressive therapy is already in routine clinical use for some cancers including acute myeloid leukemia. For example, gene expression profiling (GEP) signatures and in colon cancer are used for predictive and prognostic purposes in some patients with breast (e.g., Oncotype DX® and mammaPrint®) and colon cancers (ColoPrint®) [11,12]. We here discuss an important report by Pellagatti *et al*. [13] of the use of GEP-based prognostic approach to refine risk stratification in MDS.

Summary of methods & results

In this study, the transcriptome of $CD34^+$ cells from 125 patients with MDS and a minimum of 1 year of follow-up from the date of BM sampling was evaluated [13]. The median age of the cohort was 71 years, and the median follow-up was 47 months. Of the patients 36, 43, 17 and 4% had IPSS low, intermediate-1, intermediate-2 and high risk, respectively; 31, 31, 17 and 21% had refractory anemia (RA), RA with ring sideroblasts (RARS), RA with excess blasts-1 (RAEB-1) and RAEB-2, respectively.

$CD34^+$ cells were isolated from the mononuclear cell fraction of the BM using magnetic cell separation, and their purity was confirmed to be ≥90% using fluorescence-activated cell sorting. GEP was performed using amplified RNA and the U133 plus 2.0 arrays (affymetrix®). This microchip covers >47,000 transcripts representing 39,000 human genes. Overall survival (OS), the primary endpoint, was calculated from the date of BM sample collection. Supervised principal components (SuperPC) and lasso penalized Cox proportional hazards regression models applying the Coxnet algorithm were used to predict OS. Clinical covariates evaluated in the Cox models included age, gender and the IPSS score at the time of BM sampling.

For the SuperPC method, patients were split randomly into a training (n = 84) and validation (n = 41) sets. Coefficients of the univariate Cox regression models (Cox scores) were calculated for every gene in the training set, and then principal components were calculated for genes with Cox scores exceeding an absolute cross-validated threshold. The scores were then validated and an importance score was calculated in correlation with the SuperPC predictor. Using this approach, the authors identified 139 genes of highest importance in prediction of survival. Among these genes were *LEF1* and *CDH1* whose high expression was associated with increased OS, and *WT1*, *MN1* and *PTH2R* whose high expression was associated with reduced OS. A lasso penalized Cox proportional hazard was then generated to identify a smaller set of prognostic genes. A predictor of 20 probe sets, corresponding to 20 genes, was identified of which eight probe sets matched those identified by the SuperPC method (*ADHFE1*, *BTBD6*, *CPT1B*, *LEF1*, *FRMD6*, *GPR114*, *C7orf58* and *LOC100505956*).

The prognostic accuracy of these models was then compared using prediction error curves. The prediction error was lowest using the Coxnet signature, which performed better than all other predictors including the one that additionally used the clinical data. There was a significant association between the Coxnet signature and the IPSS ($p < 0.001$) and the IPSS cytogenetic groups ($p = 0.045$). Interestingly, when the Coxnet 20-gene signature was applied to an independent GEP data set obtained from unsorted BM mononuclear cells using a univariate Cox regression model, there was a statistically significant separation of patients with good or bad prognosis ($p = 0.007$).

Expert commentary & five-year view

Cytogenetic abnormalities have long been known to be key prognostic determinants in MDS and as such have been incorporated in all major prognostic models. The technologic advancements in genome-wide high-throughput approaches in the last decade have revolutionized the discovery of novel molecular lesions and allowed the dissection of the genetic landscape of the disease at unprecedented depths [3,14]. High hopes have been placed on this explosion of knowledge translating into a better understanding of the complex pathophysiology of MDS, discovery of novel therapeutic targets, refinement of the current prognostic tools and the introduction of therapy-specific predictive models [6]. Indeed, approximately 90% of MDS patients now have detectable clonal genetic aberrations [14]. The driver oncogenic mutations in MDS tend to occur in genes involved in epigenetic

modulation, signal transduction, pre-mRNA splicing and transcription regulation, DNA repair and the cohesin complex [1,15].

Efforts are ongoing to use genome-wide approaches to refine the risk stratification schemes in MDS. Clonal aberrations detected using array-based comparative genomic hybridization and single-nucleotide polymorphism arrays were found to have additional prognostic information [16,17]. Using a combination of genomic approaches on samples from 439 MDS patients, Bejar *et al*. [18] identified recurrent somatic mutations in 18 genes. After adjustment for age, IPSS score and other identified mutations in a multivariate regression model, five mutations maintained an independent negative impact on OS: *TP53*, *EZH2*, *ETV6*, *RUNX1* and *ASXL1*. Incorporating these five mutations in the IPSS upstaged patients with any mutation to the next-highest IPSS risk group [18]. In another report by the same group, mutations in four genes (*EZH2*, *RUNX1*, *TP53* and *ASXL1*) were independently associated with a worsened OS in patients with low and intermediate-1 IPSS [19]. The researchers found that the prognostic discrimination of the MD Anderson low-risk prognostic scoring system could be improved by accounting for the *EZH2* mutational status [19].

In another recent large study, samples from 944 MDS patients were screened for aberrations in 104 genes using targeted massive parallel sequencing [14]. Genetic aberrations were detected in 91.5% of patients and 89.5% harbored at least one mutation (median three per patient). Forty-seven genes were significantly mutated with *TET2*, *SF3B1*, *ASXL1*, *SRSF2*, *DNMT3A* and *RUNX1* being the most commonly mutated (>10% of cases each). Twenty-five genes affected survival on univariate analysis. The authors constructed two prognostic models using this data. In the first model, the status of 14 genes was combined with conventional factors to separate patients into four risk groups with 3-year OS of 95.2, 69.3, 32.8 and 5.3% ($p < 0.001$). In the second ‘gene-only model,’ 14 genes also significantly separated patients into 4 prognostic groups ($p < 0.001$). The first model was found to outperform the revised IPSS (IPSS-R) and the ‘gene-only’ model in a validation set of 175 patients [14].

Another recent genome-wide study of 603 MDS patients, 78% of the patients had at least one molecular aberration [15]. Consistent with other studies, eight genes were associated with significantly inferior leukemia-free survival. The authors found that clonal and subclonal mutations affected prognosis equally and that outcomes correlated inversely with the number of oncogenic driver mutations. The incorporation of point mutations resulted in a marginal nonsignificant increase in prognostic information compared with the IPSS [15].

Pellagatti *et al*. [13] took a different approach and evaluated GEP signatures in $CD34^+$ cells from MDS patients. GEP is a robust technology that allows for the detection of the expression products of thousands of genes simultaneously. This technique has already been used to generate validated prognostic expression signatures in other malignancies such as breast and colon cancers. GEP allows the evaluation of downstream effects of many other genetic and epigenetic aberrations that affect expression of relevant prognostic genes. Sridhar *et al.* [20] previously used this methodology to study BM $CD34^+$ cells from 35 MDS patients (24 with low/Int-1 and 11 with int-2 and high IPSS). Based on the GEP data, the

authors defined a poor-risk six-gene signature that was associated with leukemic progression and provided additive prognostic information for the IPSS Int-1 patients. Ebert *et al*. used this approach to derive an erythroid differentiation GEP-based signature that predicts response to lenalidomide in MDS without 5q deletion [21]. Pellagatti *et al*. [13] confirmed the utility of GEP approach in a significantly larger sample of patients and demonstrated its robustness by applying it to unsorted BM mononuclear cells therefore increasing its potential applicability to routine clinical use.

The MDS patient cohort in the Pellagatti *et al*. study was pathologically heterogonous, but the prognostic inference of GEP was applied to the group as a whole. Since clinical heterogeneity in MDS appears to result from the specific underlining of genetic alterations (e.g., *SF3B1* mutation association with RARS), it might have been more informative to study MDS subsets (e.g., RARS, RAEB, etc.) separately. Additionally, OS was measured from the time of BM sampling, and similarly the IPSS was calculated at the time of BM sampling. It is not clear if the BM sampling was done at the time of diagnosis, whether all patients started follow-up at a uniform point in the natural history of their disease and what type of disease-modifying therapies they received. Third, the authors used only the IPSS as a clinical risk model. It would have been interesting to determine whether the GEP-based signature provided superior prognostic information above that of the newer risk models (e.g., the revised IPSS).

Compared with next-generation DNA sequencing, a common problem to the use of GEP is the lack of reproducibility of the same relevant prognostic signatures; therefore, making external validation by other groups and by different GEP platforms very important. Similarly, the genotype/phenotype associations (e.g., *SF3B1*/RARS) appear more reliable and the potential for therapeutic targeting might be more straightforward using DNA-sequencing platforms. Pellagatti and colleagues noted that mutations in relevant prognostic genes that do not alter expression levels will not be detected by GEP; therefore, explaining some of the discrepancy in the determination of relevant prognostic genes as determined by GEP versus other high-throughput genome-wide techniques [13]. Nonetheless, one would expect that mutations in some genes such as transcription factors (e.g., *RUNX1*) would be associated with specific GEP signatures. Indeed, two of the prognostic genes in the Coxnet signature (*WT1* and *LEF1*) have been already associated with worse outcomes in patients with MDS [22,23].

How to best incorporate genomic and expression aberrations in the risk stratification of MDS in routine clinical practice remains an open question. Novel approaches are needed for the integration of data obtained from GEP and various DNA and RNA-sequencing techniques [24]. Laboratory assays still require standardization, external wide scale validation and resolution of several logistic issues [6]. The issues of mutational burden, driver versus secondary aberrations, interactions between different mutations, intratumoral heterogeneity and the importance of genetic subclones in MDS need further investigation [1,15,25]. Given that many genes are rarely mutated and are associated with a complex pattern of comutations, much larger data sets are likely needed to be studied to fully elucidate these complex interactions [15].

Once the technical issues are solved, reaching a consensus on how to use this data clinically will be of paramount importance. Overall, it appears that adding the weighted independent prognostic effects of well-characterized molecular lesions to validated clinicopathological prognostic schemes outperforms 'molecular aberrations-only' models. The molecular predictors whose independent prognostic effect is not reflected in other known prognostic clinicopatholgoical features will probably be the most valuable to refine the prognostic precision of prognostic schemes [18]. While many challenges still lie ahead, we have already taken the first few steps toward the era of true 'individualized' MDS care.

References

1. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood. 2013; 122(25):4021–34. [PubMed: 24136165]
2. Zeidan AM, Linhares Y, Gore SD. Current therapy of myelodysplastic syndromes. Blood Rev. 2013; 27(5):243–59. [PubMed: 23954262]
3. Raza A, Galili N. The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. Nat Rev Cancer. 2012; 12(12):849–59. [PubMed: 23175121]
4. Faltas B, Zeidan A, Gergis U. Myelodysplastic syndromes: toward a risk-adapted treatment approach. Expert Rev Hematol. 2013; 6(5):611–24. [PubMed: 24094045]
5. Zeidan AM, Smith BD, Komrokji RS, Gore SD. Prognostication in myelodysplastic syndromes: beyond the International Prognostic Scoring System (IPSS). Am J Med. 2013; 126(4):e25. [PubMed: 23507216]
6. Zeidan AM, Komrokji RS. There's risk, and then there's risk: the latest clinical prognostic risk stratification models in myelodysplastic syndromes. Curr Hematol Malig Rep. 2013; 8(4):351–60. [PubMed: 23979829]
7. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997; 89(6):2079–88. [PubMed: 9058730]
8. Bejar R. Prognostic models in myelodysplastic syndromes. Hematology Am Soc Hematol Educ Program. 2013; 2013:504–10. [PubMed: 24319225]
9. Naqvi K, Jabbour E, Bueso-Ramos C, et al. Implications of discrepancy in morphologic diagnosis of myelodysplastic syndrome between referral and tertiary care centers. Blood. 2011; 118(17):4690–3. [PubMed: 21868570]
10. Zeidan AM, Kharfan-Dabaja MA, Komrokji RS. Beyond hypomethylating agents failure in patients with myelodysplastic syndromes. Curr Opin Hematol. 2013 Epub ahead of print.
11. Kittaneh M, Montero AJ, Gluck S. Molecular profiling for breast cancer: a comprehensive review. Biomark Cancer. 2013; 5:61–70. [PubMed: 24250234]
12. Maak M, Simon I, Nitsche U, et al. Independent validation of a prognostic genomic signature (ColoPrint) for patients with stage II colon cancer. Ann Surg. 2013; 257(6):1053–8. [PubMed: 23295318]
13. Pellagatti A, Benner A, Mills KI, et al. Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. J Clin Oncol. 2013; 31(28):3557–64. [PubMed: 24002510]
14. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2013; 28(2):241–7. [PubMed: 24220272]
15. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013; 122(22):3616–27. [PubMed: 24030381]
16. Starczynowski DT, Vercauteren S, Telenius A, et al. High-resolution whole genome tiling path array CGH analysis of CD34+ cells from patients with low-risk myelodysplastic syndromes reveals cryptic copy number alterations and predicts overall and leukemia-free survival. Blood. 2008; 112(8):3412–24. [PubMed: 18663149]

17. Tiu RV, Gondek LP, O'Keefe CL, et al. Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. Blood. 2011; 117(17):4552–60. [PubMed: 21285439]

18. Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011; 364(26):2496–506. [PubMed: 21714648]

19. Bejar R, Stevenson KE, Caughey BA, et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J Clin Oncol. 2012; 30(27): 3376–82. [PubMed: 22869879]

20. Sridhar K, Ross DT, Tibshirani R, et al. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. Blood. 2009; 114(23):4847–58. [PubMed: 19801443]

21. Ebert BL, Galili N, Tamayo P, et al. An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. PLoS Med. 2008; 5(2):e35. [PubMed: 18271621]

22. Cilloni D, Gottardi E, Messa F, et al. Significant correlation between the degree of WT1 expression and the International Prognostic Scoring System Score in patients with myelodysplastic syndromes. J Clin Oncol. 2003; 21(10):1988–95. [PubMed: 12743153]

23. Pellagatti A, Marafioti T, Paterson JC, et al. Marked downregulation of the granulopoiesis regulator LEF1 is associated with disease progression in the myelodysplastic syndromes. Br J Haematol. 2009; 146(1):86–90. [PubMed: 19438482]

24. Chavan SS, Bauer MA, Peterson EA, et al. Towards the integration, annotation and association of historical microarray experiments with RNA-seq. BMC Bioinformatics. 2013; 14(Suppl 14):S4. [PubMed: 24268045]

25. Bejar R, Abdel-Wahab O. The importance of subclonal genetic events in MDS. Blood. 2013; 122(22):3550–1. [PubMed: 24263953]

Key issues

- Accurate risk stratification is a vital component of the risk-adaptive management paradigm for myelodysplastic syndromes (MDS).
- Current prognostic schemes for MDS rely on clinicopathological parameters.
- Despite the increasing knowledge of the prognostic impact of molecular aberrations in MDS, none has yet been incorporated into the major risk models.
- Genome-wide high-throughput techniques, including gene expression profiling, have the potential to improve the ‘individualized’ outcome prediction for MDS patients.