

Machine learning derived genomics driven prognostication for acute myeloid leukemia with *RUNX1-RUNX1T1*

Anam Fatima Shaikh, Chinmayee Kakirde, Chetan Dhamne, Prasanna Bhanshe, Swapnali Joshi, Shruti Chaudhary, Gaurav Chatterjee, Prashant Tembhare, Maya Prasad, Nirmalya Roy Moulik, Anant Gokarn, Avinash Bonda, Lingaraj Nayak, Sachin Punatkar, Hasmukh Jain, Bhausaheb Bagal, Dhanalaxmi Shetty, Manju Sengar, Gaurav Narula, Navin Khattri, Shripad Banavali, Sumeet Gujral, Subramanian P. G. & Nikhil Patkar

To cite this article: Anam Fatima Shaikh, Chinmayee Kakirde, Chetan Dhamne, Prasanna Bhanshe, Swapnali Joshi, Shruti Chaudhary, Gaurav Chatterjee, Prashant Tembhare, Maya Prasad, Nirmalya Roy Moulik, Anant Gokarn, Avinash Bonda, Lingaraj Nayak, Sachin Punatkar, Hasmukh Jain, Bhausaheb Bagal, Dhanalaxmi Shetty, Manju Sengar, Gaurav Narula, Navin Khattri, Shripad Banavali, Sumeet Gujral, Subramanian P. G. & Nikhil Patkar (2020) Machine learning derived genomics driven prognostication for acute myeloid leukemia with *RUNX1-RUNX1T1*, *Leukemia & Lymphoma*, 61:13, 3154-3160, DOI: [10.1080/10428194.2020.1798951](https://doi.org/10.1080/10428194.2020.1798951)

To link to this article: <https://doi.org/10.1080/10428194.2020.1798951>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 05 Aug 2020.



[Submit your article to this journal](#)



Article views: 3346



[View related articles](#)



[View Crossmark data](#)



Citing articles: 7 [View citing articles](#)



Machine learning derived genomics driven prognostication for acute myeloid leukemia with *RUNX1-RUNX1T1*

Anam Fatima Shaikh^{a*}, Chinmayee Kakirde^{a*}, Chetan Dhamne^{b,c}, Prasanna Bhanshe^a, Swapnali Joshi^a, Shruti Chaudhary^a, Gaurav Chatterjee^a, Prashant Tembhare^{a,b}, Maya Prasad^{b,c}, Nirmalya Roy Moulik^{b,c}, Anant Gokarn^{b,d}, Avinash Bonda^{b,d}, Lingaraj Nayak^{b,d}, Sachin Punatkar^{b,d}, Hasmukh Jain^{b,d}, Bhausaheb Bagal^{b,d}, Dhanalaxmi Shetty^e, Manju Sengar^{b,d}, Gaurav Narula^{b,c}, Navin Khattray^{b,d}, Shripad Banavali^{b,c}, Sumeet Gujral^{a,b}, Subramanian P. G.^{a,b} and Nikhil Patkar^{a,b}

^aHaematopathology Laboratory, ACTREC, Tata Memorial Centre, Navi Mumbai, India; ^bHomi Bhabha National Institute (HBNI), Mumbai, India; ^cPediatric Haematolymphoid Disease Management Group, Tata Memorial Centre, Mumbai, India;

^dAdult Haematolymphoid Disease Management Group, Tata Memorial Centre, Mumbai, India; ^eDepartment of Cytogenetics, ACTREC, Tata Memorial Centre, Navi Mumbai, India

ABSTRACT

Panel based next generation sequencing was performed on a discovery cohort of AML with *RUNX1-RUNX1T1*. Supervised machine learning identified *NRAS* mutation and absence of mutations in *ASXL2*, *RAD21*, *KIT* and *FLT3* genes as well as a low mutation to be associated with favorable outcome. Based on this data patients were classified into favorable and poor genetic risk classes. Patients classified as poor genetic risk had a significantly lower overall survival (OS) and relapse free survival (RFS). We could validate these findings independently on a validation cohort ($n=61$). Patients in the poor genetic risk group were more likely to harbor measurable residual disease. Poor genetic risk emerged as an independent risk factor predictive of inferior outcome. Using an unbiased computational approach based we provide evidence for gene panel-based testing in AML with *RUNX1-RUNX1T1* and a framework for integration of genomic markers toward clinical decision making in this heterogeneous disease entity.

ARTICLE HISTORY

Received 15 June 2020

Revised 14 July 2020

Accepted 15 July 2020

KEYWORDS

Acute myeloid leukemia (AML) with *RUNX1-RUNX1T1*; machine learning; genomic risk stratification; gene mutations in AML with *RUNX1-RUNX1T1*; gene mutations in AML with *t(8;21)*

Introduction

Acute myeloid leukemia (AML) with *t(8;21)(q22;q22)* that results in the *RUNX1-RUNX1T1* chimeric gene fusion is one of the commonest subtypes of AML. It is characterized by a distinct morphology and a unique immunophenotype and is thus recognized as a specific entity amongst 'AML with recurrent genetic abnormalities' [1]. Traditionally, this AML has been recognized as having a favorable outcome as evident by superior survival rates when compared to intermediate and poor cytogenetic risk AMLs [2]. Unfortunately, the treatment outcome in these cases is not homogeneous as evident by relapse in a significant number of patients despite achievement of morphological complete remission (CR) [3,4]. In fact, studies claim that only half the patients of AML with *RUNX1-RUNX1T1*

get cured [5,6]. This heterogeneous outcome has been explained, in part, by cooperating somatic mutations in genes involved in signaling pathways such as *FLT3* and *KIT* [7].

In the last few years, largely due to next generation sequencing (NGS) technologies, we have identified somatic mutations affecting diverse cellular pathways in AML [8]. Some of these mutations are clinically relevant affecting prognosis or are amenable to targeted therapy. Somatic mutations have been identified in nearly 90% of AML with *t(8;21)* and commonly include genes encoding for chromatin modifiers (e.g. *ASXL1*, *ASXL2*, *EZH2*, *KDM6A*), cohesin complex (e.g. *RAD21*, *SMC3*, *SMC1A*) and signaling pathways (e.g. *KIT*, *FLT3*, *NRAS*) [4,9]. Whereas a general consensus exists amongst researchers that the above sets of genes are recurrently mutated in AML with *RUNX1-RUNX1T1*, an accurate prognostication

CONTACT Nikhil Patkar nvpatkar@gmail.com Wellcome Trust – DBT Intermediate Fellow, Haematopathology Laboratory, CCE Building, Advanced Centre for Treatment Research and Education in Cancer, Tata Memorial Centre, Kharharg, 410210, Maharashtra, India

*These authors contributed equally to this work.

Supplemental data for this article can be accessed [here](#).

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

scheme that guides a treating physician is largely lacking. It is imperative that better approaches be developed so that we can identify patients who are at a high risk of relapse in this rather common subtype of AML.

Machine learning (ML) is a subset of artificial intelligence that holds promise in deciphering complex genomic datasets. ML has been used to develop algorithms for diverse applications such as identification of regulatory regions in the genome to prediction of cancer susceptibility, recurrence and survival [10,11]. ML has also been recently used for prediction of drug response in AML based on gene expression profiles as well as discovery of novel antibiotics [12,13]. We have recently developed a supervised ML based algorithm for prognostication of AML with mutated *NPM1* based on the underlying genomic data [14]. To decipher the clinical significance of the large numbers of genetic variables we used an *unbiased* computational approach and identified that *NPM1* mutation type and corrected *NPM1* variant allele fraction (VAF), presence of *DNMT3A* R882 mutation, *FLT3* internal tandem duplication VAF and *IDH2* mutations were clinically relevant. Based on these ML derived variables we developed a scoring system. The genetic score could classify AML with mutated *NPM1* into three classes with vastly different outcomes.

Given the heterogeneity in treatment outcomes that is observed in AML with *RUNX1-RUNX1T1* we questioned if such ML based approaches can be applied to AML with *RUNX1-RUNX1T1*. In this manuscript, we develop a ML based genomics driven prognostication model for AML with *RUNX1-RUNX1T1* and demonstrate that this model correlates with measurable residual disease (MRD) and clinical outcome.

Methods

A. PATIENT DETAILS:

- a. Patient Accrual: The study was cleared by the institutional ethics board (IEC III Project 163 and IEC III Project 900613). We accrued a total of 131 patients of AML with *RUNX1-RUNX1T1*. These patients were accrued over a 7-year period from March 2012 to April 2019. Diagnosis, immunophenotyping and cytogenetic analysis were performed as previously described [3].
- b. Patient Treatment and Evaluation of Outcome: Patients were divided into a discovery cohort ($n=70$) and an independent validation cohort ($n=61$). All patients were treated

with conventional induction '3 + 7' chemotherapy consisting of daunorubicin (60 mg/m^2 D1-D3) and cytarabine ($100 \text{ mg/m}^2/\text{day}$ D1-D7). For the validation cohort, 13 out of 61 patients had baseline fungal pneumonia or multidrug resistant bacterial colonization and were treated with oral metronomic chemotherapy to stabilize the patient prior to intensive '3 + 7' chemotherapy. Complete remission (CR), overall survival (OS) and relapse free survival (RFS) were calculated as previously described [3,14,15]. One out of 13 patients not achieving morphological CR was treated with palliation and the rest were treated with conventional therapy, at discretion of the treating physician.

B. GENETIC TESTING ON DIAGNOSTIC SAMPLE:

- a. Cytogenetics: Only patients who were confirmed to have *RUNX1-RUNX1T1* by conventional karyotyping and/or fluorescence in-situ hybridization (FISH) were included.
- b. Panel Based NGS and Data Analysis: Details of the single molecule molecular inversion probe (smMIP) based myeloid sequencing panel and bioinformatics approaches used to analyze this dataset are as previously described in detail (please see [supplementary methods](#)) [14].
- c. Machine learning based genetic score: We developed a supervised ML based approach for identification of prognostic variables most likely to influence outcome in AML with *RUNX1-RUNX1T1* as described previously [14]. Additional details pertaining to ML can be seen in the [Supplementary Methods](#) that accompanies this manuscript. Based on the results of the ML model we scored each variable as '-1' if the results were predictive of an unfavorable outcome and '+1' if otherwise. A sum of these scores was finally derived to generate a final score. Based on this final score patients were classified into favorable and poor genetic risk (GR).
- C. MEASURABLE RESIDUAL DISEASE ASSESSMENT USING MULTIPARAMETRIC FCM (FCM-MRD): FCM-MRD was detected using a two tube 10 color assay as described previously by our group [3]. Patients were called as FCM-MRD negative if they were negative on two consecutive MRD time points (post induction and post consolidation). Everyone else was MRD positive.
- D. CORRELATION OF ML DERIVED GR WITH FCM-MRD AND TREATMENT OUTCOME: Chi squared test was used to correlate FCM-MRD

with ML derived GR classes. The impact of GR defined classes was also evaluated against OS and RFS using Kaplan–Meier technique and log-rank test.

Results

Table 1 is a summary of clinical and laboratory parameters of the entire cohort.

A. PATIENT DETAILS:

The median follow-up for the entire cohort (131 patients) was 27.6 months. The median overall survival (OS) was 30.7 months (95% CI: 23.0–38.4 months) and the median relapse free survival (RFS) was 32.9 months (95% CI: 27.7–38.1 months). Out of 131 patients only 4 underwent allogeneic bone marrow transplantation. Due to small numbers their outcome was not different from the rest with respect to OS ($p=0.25$) or RFS ($p=0.9$). These patients are therefore not considered separately. The clinical and laboratory parameters of the discovery and validation cohorts can be seen in **Supplementary Tables 1 & 2** respectively.

B. GENETIC TESTING ON THE DIAGNOSTIC SAMPLE:

- Cytogenetics and gene mutations:** Details of conventional karyotyping and FISH can be seen in **supplementary data** accompanying this manuscript. At least one somatic mutation was detected in 85.5% of all patients (median coverage: 983.5, range: 402–2793X). An overview of these mutations can be seen in **Figure 1**. **Supplementary Figure 1** highlights the frequencies of commonly occurring mutations seen in our cohort.
- Machine Learning Based Modeling:** Performance characteristics of ML model as well as results of feature selection can be seen in **supplementary methods** (**Supplementary Tables 3–5** and **Supplementary Figure 2**) accompanying this manuscript. Based on the ML modeling on the discovery cohort, mutations in *FLT3*, *NRAS*, *ASXL2*, *RAD21*, *KIT* genes as well as mutation burden (≥ 2 as high mutation burden) were determined as important variables likely to predict outcome.
- Machine Learning Derived Genetic Score:** High mutation burden, mutations in *FLT3*, *RAD21*, *ASXL2* and *KIT* genes were associated with an inferior prognosis and assigned a negative score whereas mutations in *NRAS*

were associated with a favorable outcome and a positive score if present. These features were used to generate a scoring system as is seen in **Figure 1**. Based on the prognostic impact of these variables a score was allotted to each feature [14] the sum of which resulted in classification of patients into favorable (score ≥ 4) and poor genetic risk (score ≤ 3). In the discovery cohort, patients classified as poor genetic risk were associated with inferior OS and RFS (**Supplementary Table 1** and **Supplementary Figure 3**). We reconfirmed the clinical relevance of these findings by an independent validation on a cohort of pediatric *RUNX1-RUNX1T1* rearranged AML (**Supplementary Table 2** and **Supplementary Figure 4**).

C. MEASURABLE RESIDUAL DISEASE ASSESSMENT USING MULTIPARAMETRIC FCM (FCM-MRD):

The presence of FCM-MRD was significantly associated with inferior OS and RFS (**Supplementary Figure 5**).

D. CORRELATION OF ML DERIVED GR WITH FCM-MRD AND TREATMENT OUTCOME:

We observed a strong correlation of ML derived genetic risk classes with FCM-MRD where cases which were classified as poor genetic risk were more likely to be MRD positive (**Supplementary Figure 6**). Lastly ML derived GR based risk classes were highly predictive of outcome as seen in **Figure 1** and **Table 1**. The clinical relevance of individual components of the ML derived scoring system can be visualized in **Supplementary Figure 7**.

Discussion

In the last decade several studies have analyzed the prognostic relevance of signaling pathway mutations in core binding factor AML as reviewed by Boissel et al. [16]. However, there is little consensus amongst investigators with respect to clinical relevance of these mutations. This could be attributed to various reasons, such as technical issues associated with low sensitivity assays like sanger sequencing. Some investigators have indicated that allelic abundances of mutations are important [9,17]. Others have indicated that even within a single gene, mutational hotspots such as *KIT* D816 [18] or *FLT3-TKD* [19] may have different prognostic connotation in core binding factor AML. Krauth and colleagues indicated mutation burden may be an additional determinant of outcome [20]. Studies

Table 1. Prognostic significance of machine learning derived genetic risk in AML with t(8;21).

Parameter	Observation (%)	
Demographics:		
Age	Range: 2–60 years; Median: 20 years Male:Female : 2.2:1	
Sex		
Clinical characteristics:		
Total number of patients accrued	131	
Cases not in morphological remission	13	
Remission characteristics:		
Complete remission (CR)	50	
CR with incomplete hematologic recovery (CRi)	81	
Bone marrow transplantation:		
Patients who underwent BMT	04	
Laboratory characteristics:		
<i>Blood counts at presentation</i>		
1. More than 50,000/mm ³	08	
2. Less than 50,000/mm ³	123	
Individual parameters of genetic risk score:		
1. Mutation burden (≥ 2)	53 (40.4%)	
2. Any <i>KIT</i> mutation	44 (33.5%)	
3. <i>NRAS</i> mutation	25 (19.1%)	
4. <i>RAD21</i> mutation	11 (8.4%)	
5. Any <i>FLT3</i> mutation	11 (8.4%)	
6. <i>ASXL2</i> mutation	17 (13%)	
Classification according to genetic risk:		
Favorable genetic risk (Fav GR)	63 (48.1%)	
Poor genetic risk (Poor GR)	68 (51.9%)	
Post induction flow MRD ($n = 131$):		
MRD positive	58 (44.2%)	
MRD negative	73 (55.7%)	
Post consolidation flow MRD ($n = 99$):		
MRD positive	13 (13.1%)	
MRD negative	86 (86.8%)	
Paired MRD analysis ($n = 87$):		
Any MRD positive	46 (52.8%)	
Dual time point MRD negative	41 (47.1%)	
Univariate Cox analysis		
Machine learning derived genetic risk	Overall survival (OS) <hr/> HR (95% CI) p	
Favorable genetic risk	1	0.0001
Poor genetic risk	3.5 (1.88–6.55)	
Machine learning derived genetic risk	Overall survival (OS) <hr/> HR (95% CI) p	
Favorable genetic risk	Mean OS: 44.6 months; 95% CI (40.6–48.6 months), Median OS: not reached	0.0001
Poor genetic risk	Mean OS: 29.6 months; 95% CI (23.7–35.6 months), Median OS: 30.5 months; 95% CI (16.0–42.1 months)	
Dual time point FCM-MRD	Overall survival (OS) <hr/> HR (95% CI) p	
MRD Negative	1	0.01
MRD Positive	2.5 (1.16–5.50)	
Dual time point FCM-MRD	Overall survival (OS) <hr/> HR (95% CI) p	
MRD Negative	Mean OS: 62.8 months; 95% CI (53.1–72.4 months), Median OS: 74.1 months; 95% CI (74.1–74.1 months)	0.01
MRD Positive	Mean OS: 52.9 months; 95% CI (40.3–65.4 months), Median OS: not reached	
Multivariate cox analysis	Overall survival (OS) <hr/> HR (95% CI) p	
Dual MRD positive	1.6 (0.77–3.47)	0.202
Poor genetic risk	3.7 (1.69–8.08)	0.001
Relapse free survival (RFS)		
Favorable genetic risk	1	0.0008
Poor genetic risk	2.5 (1.43–4.33)	
Machine learning derived genetic risk	Relapse free survival (RFS) <hr/> HR (95% CI) p	
Favorable genetic risk	Mean RFS: 37.7 months; 95% CI (32.8–42.6 months), Median RFS: 72.6 months 95% CI (35.1–72.6 months)	0.0008
Poor genetic risk	Mean RFS: 24.4 months; 95% CI (19.0–29.9 months), Median RFS: 16.7 months; 95% CI (12.7–32.6 months)	
Dual time point FCM-MRD	Relapse free survival (RFS) <hr/> HR (95% CI) p	
MRD Negative	1	0.01
MRD Positive	1.6 (0.87–3.18)	
Dual time point FCM-MRD	Relapse free survival (RFS) <hr/> HR (95% CI) p	
MRD Negative	Mean RFS: 49.5 months; 95% CI (39.4–59.7 months), Median RFS: 72.6 months 95% CI (29.2–72.6 months)	0.01
MRD Positive	Mean RFS: 43.0 months; 95% CI (30.3–55.7 months), Median RFS: 23.5 months; 95% CI (15.4–32.6 months)	
Multivariate cox analysis	Relapse free survival (RFS) <hr/> HR (95% CI) p	
Dual MRD positive	1.3 (0.67–2.58)	0.41
Poor genetic risk	2.3 (1.17–4.60)	0.01

OS: Overall Survival; RFS: Relapse Free Survival; HR: Hazards ratio; CI: confidence interval; MRD: Measurable Residual Disease. FCM-MRD was assessed in 87 patients in morphological CR (<5% blasts).

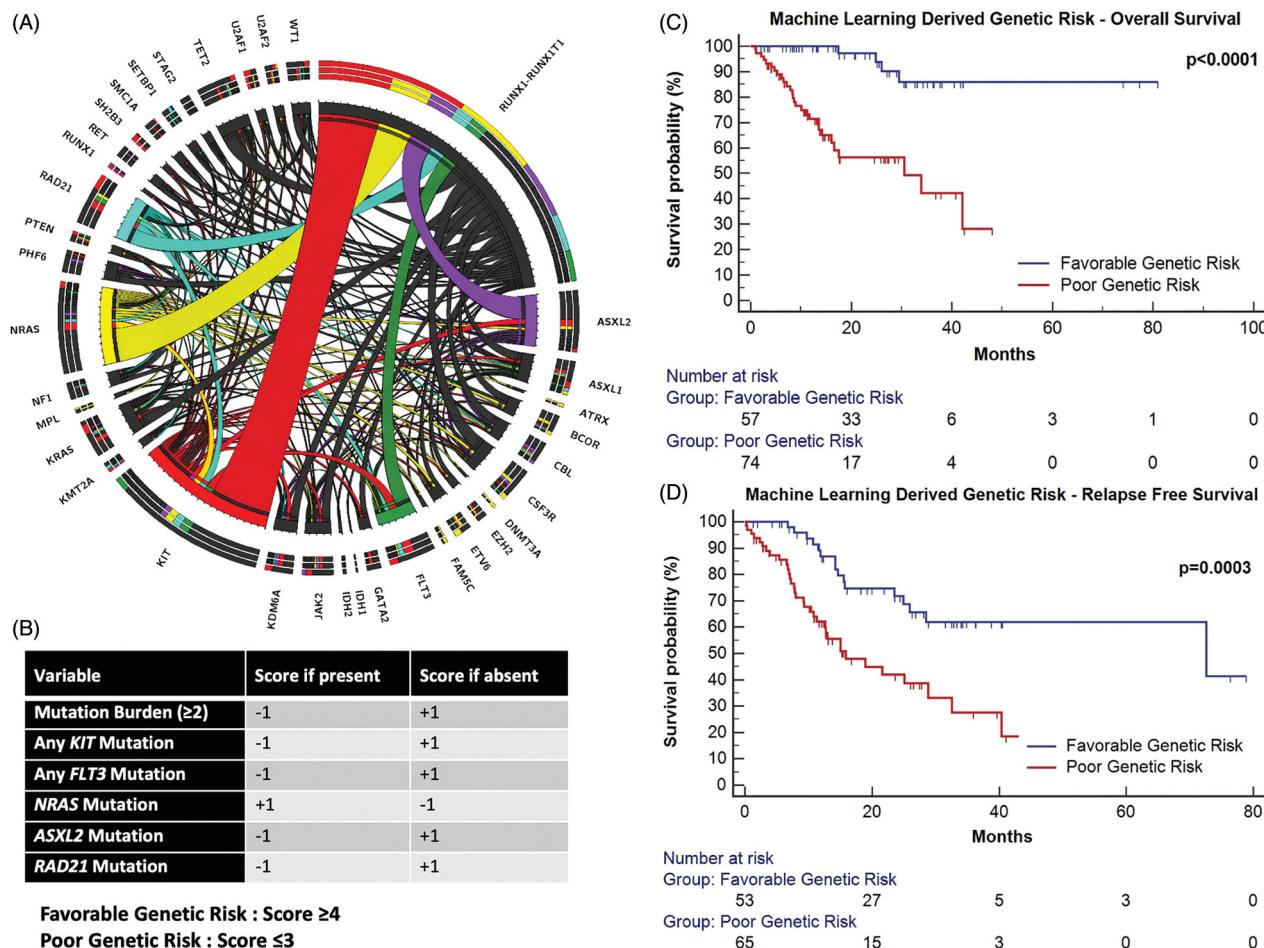


Figure 1. The above circos plot (A) highlights the spectrum of mutations and their interaction in AML with *RUNX1-RUNX1T1*. Commonly occurring gene mutations are colored. The machine learning derived scoring system is described in (B). The Kaplan–Meier plot in the top right section (C) shows the clinical impact on overall survival (OS) and for relapse free survival (RFS, D), lower right).

employing high throughput sequencing technologies have indicated that beyond signaling pathways the mutational landscape of AML with t(8;21) may be unique characterized by high frequencies of mutations in genes encoding for cohesin complex and chromatin modeling pathways [9,21]. As a result of lack of consensus, current guidelines group AML with *RUNX1-RUNX1T1* as a single disease entity [22,23].

Instead of selecting individual genes, we analyzed all commonly ($>5\%$) occurring mutations in AML with *RUNX1-RUNX1T1* in an unbiased manner using a supervised machine learning algorithm. The threshold of including mutations occurring at a frequency of $>5\%$ (from a 50 gene panel) is chosen empirically. This is keeping in mind a balance between applicability of the model to AML with *RUNX1-RUNX1T1* and inclusion of nongeneralizable data (typically seen with rare mutations). This will presumably prevent ‘overfitting’ into the dataset [24]. ML approaches allow us to identify interactions between data that are not readily

visible using legacy approaches. Our approach enabled us to develop a scoring system based which we could classify AML with *RUNX1-RUNX1T1* into two prognostic subgroups with different outcomes. In a multivariate analysis this was found to be an independently important predictor of outcome.

Based on these data, we propose a risk stratification of AML with *RUNX1-RUNX1T1* that incorporates somatic mutations in *FLT3*, *NRAS*, *ASXL2*, *RAD21*, *KIT* genes as well as mutation burden. From data that has been published previously we expected that *KIT* [18] and *FLT3* (exon 20) [19] gene mutations would be prognostically relevant. In addition, based on scant published data we also suspected that mutation burden would influence outcome [20]. However, using AI we could additionally infer the prognostic impact of *NRAS*, *RAD21*, *FLT3-ITD*, *ASXL2* mutations. The latter were not expected from legacy data. Micol et al and other studies, have previously demonstrated a high frequency of *ASXL2* mutations in this subset of AML and possible

inferior outcome [20,25]. Recently Ishikawa et al indicated that only exon 17 mutations were prognostically relevant in AML with *RUNX1-RUNX1T1*. In comparison, we determined that all *KIT* mutations may be relevant as determined by ML modeling. Cohesin gene mutations have been associated with inferior outcome in myeloid malignancies [26]. In our study, we identify *RAD21* mutations as a mutation associated with possible inferior outcome, especially in the context of ML derived scoring system. *NRAS* mutations have been described in AML with *RUNX1-RUNX1T1*, however, have failed to demonstrate a clear survival advantage [27,28]. We think, a more global approach which takes into account the complex interaction of these mutations rather than a simplistic evaluation as evident by our scoring system is warranted for prognostication of this seemingly homogeneous AML.

A disadvantage of our study could possibly be not including recently described newer gene mutations including *ZBTB7A* [29]. Nonetheless, our approach provides additional evidence for gene panel-based testing in AML with *RUNX1-RUNX1T1* and a general framework for the integration of genomic markers toward clinical decision making. The potential limitations of this study include a retrospective analysis and a limited number of patients. This machine learning derived genomics score for AML with *RUNX1-RUNX1T1* should be validated prospectively by other investigators.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Wellcome Trust/DBT India Alliance Fellowship [grant number IA/CPHI/14/1/501485] awarded to Dr Nikhil Patkar.

References

- [1] Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016; 127(20):2391–2405.
- [2] Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood*. 1998; 92(7):2322–2333.
- [3] Patkar N, Kakirde C, Bhanshe P, et al. Utility of immunophenotypic measurable residual disease in adult acute myeloid leukemia-real-world context. *Front Oncol*. 2019;9:450.
- [4] Christen F, Hoyer K, Yoshida K, et al. Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. *Blood*. 2019;133(10):1140–1151.
- [5] Marcucci G, Mrozek K, Ruppert AS, et al. Prognostic factors and outcome of core binding factor acute myeloid leukemia patients with t(8;21) differ from those of patients with inv(16): a Cancer and Leukemia Group B study. *J Clin Oncol*. 2005;23(24):5705–5717.
- [6] Schlenk RF, Benner A, Krauter J, et al. Individual patient data-based meta-analysis of patients aged 16 to 60 years with core binding factor acute myeloid leukemia: a survey of the German Acute Myeloid Leukemia Intergroup. *J Clin Oncol*. 2004;22(18):3741–3750.
- [7] Care RS, Valk PJ, Goodeve AC, et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br J Haematol*. 2003;121(5):775–777.
- [8] Ley TJ, Miller C, Ding L, et al. Genomic and epigenetic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059–2074.
- [9] Duployez N, Marceau-Renaud A, Boissel N, et al. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood*. 2016;127(20):2451–2459.
- [10] Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
- [11] Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015; 16(6):321–332.
- [12] Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun*. 2018;9(1):42.
- [13] Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4): 688–702.e13.
- [14] Patkar N, Shaikh AF, Kakirde C, et al. A novel machine-learning-derived genetic score correlates with measurable residual disease and is highly predictive of outcome in acute myeloid leukemia with mutated NPM1. *Blood Cancer J*. 2019;9(10):79.
- [15] Patkar N, Kodgule R, Kakirde C, et al. Clinical impact of measurable residual disease monitoring by ultra-deep next generation sequencing in NPM1 mutated acute myeloid leukemia. *Oncotarget*. 2018;9(93): 36613–36624.
- [16] Boissel N, Leroy H, Brethon B, et al. Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia*. 2006;20(6):965–970.
- [17] Allen C, Hills RK, Lamb K, et al. The importance of relative mutant level for evaluating impact on outcome of KIT, FLT3 and CBL mutations in core-binding factor acute myeloid leukemia. *Leukemia*. 2013;27(9): 1891–1901.
- [18] Schnittger S, Kohl TM, Haferlach T, et al. KIT-D816 mutations in AML1-ETO-positive AML are associated

- with impaired event-free and overall survival. *Blood*. 2006;107(5):1791–1799.
- [19] Cher CY, Leung GM, Au CH, et al. Next-generation sequencing with a myeloid gene panel in core-binding factor AML showed KIT activation loop and TET2 mutations predictive of outcome. *Blood Cancer J*. 2016;6(7):e442–e442.
- [20] Krauth MT, Eder C, Alpermann T, et al. High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. *Leukemia*. 2014;28(7):1449–1458.
- [21] Faber ZJ, Chen X, Gedman AL, et al. The genomic landscape of core-binding factor acute myeloid leukemias. *Nat Genet*. 2016;48(12):1551–1556.
- [22] Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447.
- [23] Tallman MS, Wang ES, Altman JK, OCN, et al. Acute myeloid leukemia, version 3.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2019;17(6):721–749.
- [24] Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol*. 2020;7(7):e541–e550.
- [25] Micol JB, Duployez N, Boissel N, et al. Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations. *Blood*. 2014;124(9):1445–1449.
- [26] Thota S, Viny AD, Makishima H, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*. 2014;124(11):1790–1798.
- [27] Bowen DT, Frew ME, Hills R, et al. RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years. *Blood*. 2005;106(6):2113–2119.
- [28] Berman JN, Gerbing RB, Alonso TA, et al. Prevalence and clinical implications of NRAS mutations in childhood AML: a report from the Children's Oncology Group. *Leukemia*. 2011;25(6):1039–1042.
- [29] Opatz S, Bamopoulos SA, Metzeler KH, et al. The clinical mutatome of core binding factor leukemia. *Leukemia*. 2020;34(6):1553–1562.