Genetic alterations of the cohesin complex genes in myeloid malignancies

Swapna Thota MD* 1, Aaron D. Viny MD MS* 3, Hideki Makishima MD PhD1, Barbara Spitzer MD3, Tomas Radivoyevitch2, Bartlomiej Przychodzen1, Mikkael A. Sekeres MD MS1, Ross L. Levine MD3, Jaroslaw P. Maciejewski MD PhD1

* These authors contributed equally to this work

1Translational Hematology and Oncology Research, Taussig Cancer Institute, and 2Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, 44195, 3Memorial Sloan-Kettering Cancer Center, Human Oncology & Pathogenesis Program and Leukemia Service, Department of Medicine, New York, NY 10065

Key words: Cohesin, Acute Myeloid Leukemia, Myelodysplastic Syndrome, Myeloproliferative Neoplasm

Corresponding author:
Jaroslaw Maciejewski, M.D., Ph.D.,
Translational Hematology & Oncology Research
Taussig Cancer Institute R/40, Cleveland Clinic
9500 Euclid Avenue, Cleveland, Ohio
Tel 216-445-5962; Fax 216-636-2495
E-mail: maciejj@ccf.org
Key Points:

- Recurrent hypomorphic cohesin defects as well as cohesin low expression were identified in a significant proportion of patients with MDS and AML
- Cohesin mutations likely represent secondary events in clonal hierarchy and contribute to clonal transformation

Abstract

Somatic cohesin mutations have been reported in myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). To account for the morphologic and cytogenetic diversity of these neoplasms, a well-annotated cohort of 1060 patients with myeloid malignancies including MDS (n=386), myeloproliferative neoplasms (MPN, n=55), MDS/MPN (n=169) and AML (n=450) were analyzed for cohesin gene mutational status, gene expression, therapeutic and survival outcomes. Somatic cohesin defects were detected in 12% of patients with myeloid malignancies, while low expression of these genes was present in an additional 15% of patients. Mutations of cohesin genes were mutually exclusive and mostly resulted in predicted loss-of-function. Patients with low cohesin gene expression showed similar expression signatures as those with somatic cohesin mutations. Cross-sectional deep sequencing analysis for clonal hierarchy demonstrated STAG2, SMC3, RAD21 mutations to be ancestral in 18%, 18% and 47% of cases, respectively, and each expanded to clonal dominance concordant with disease transformation. Cohesin mutations were significantly associated with RUNX1, Ras-family oncogenes, BCOR and ASXL1 mutations and were most prevalent in high-risk MDS and secondary AML. Cohesin defects were associated with poor overall survival (27.2 vs. 40 months; p=0.023) especially in STAG2 mutant MDS patients surviving >12 months (median survival 35 vs. 50 months; p=0.017).
Introduction

The use of high throughput next generation sequencing (NGS) for somatic mutational profiling has led to the identification of recurrent somatic mutations within the cohesin complex in patients with acute myeloid leukemia (AML)\textsuperscript{1,2}, myelodysplastic syndrome (MDS)\textsuperscript{3}, glioblastoma multiforme\textsuperscript{4}, Ewing's sarcoma\textsuperscript{5}, colorectal\textsuperscript{6} and bladder carcinomas\textsuperscript{7,8}. Cohesin is a multi-protein complex that has an established role as an effector of sister-chromatid cohesin during metaphase. The cohesin complex is comprised of two long structural maintenance proteins, SMC1A and SMC3, which heterodimerize at the hinge domain, forming a closed loop by binding at the α-klesin end to RAD21 and adapter proteins STAG1/STAG2\textsuperscript{9}. Despite the demonstrated role of cohesin in mitosis, the mechanism by which altered cohesin function leads to clonal expansion and malignant transformation has not been elucidated. Given the role of cohesin in alignment of sister chromatids during cellular replication,\textsuperscript{10} cohesin mutations could theoretically lead to chromosomal instability\textsuperscript{5,6}. However, in AML, cohesin mutations have not been associated with aneuploidy or complex cytogenetics\textsuperscript{2}, suggesting an alternate mechanism of transformation\textsuperscript{11}.

Perturbation of other functions of cohesin complex genes may contribute to hematopoietic transformation, particularly through their role in regulating gene expression\textsuperscript{12,13} and DNA-loop formation\textsuperscript{14,15}. More recently, cohesin has been found to localize at super-enhancers, which are locus-control regions with broad peaks of H3K27-acetylation. These sites regulate the expression of specific genes that govern physiologic cell identity and pluripotency/self-renewal in malignancy and in embryonic stem cells\textsuperscript{14,16}. These data suggest the possibility that loss-of-function cohesin mutations alter transcriptional orchestration of differentiation/proliferation and may have a role in clonal evolution and tumorigenesis. Notably, germline cohesin mutations in NIBPL, SMC1A, SMC3, and HDAC8, occur in patients with Cornelia deLange syndrome\textsuperscript{17} hallmarked by craniofacial abnormalities and cognitive deficits. Although no propensity to malignancy has been described, these children have a markedy abbreviated lifespan, abrogating the oncogenic propensity of these defects. Somatic loss-of-function of HDAC8 impairs SMC3 deacetylation, leading to decreased cohesin occupancy and altered transcriptional programs\textsuperscript{18}, underscoring a role for cohesin in gene regulation\textsuperscript{14,19}.

Somatic cohesin mutations have been reported in several cohorts of patients with de novo AML.\textsuperscript{2,3} Given the molecular heterogeneity of myeloid neoplasms, precise establishment of the clinical impact of these leukemia disease alleles and their relevance to clinical phenotype, somatic genotype, and clinical outcome, requires large patient cohorts with detailed clinical annotation. We therefore studied a cohort of 1060 patients with MDS, myeloproliferative neoplasms (MPN), MDS/MPN overlap syndromes and pAML/sAML (primary and secondary) to elucidate the clinical characteristics of patients
with myeloid neoplasms harboring cohesin mutations with respect to disease phenotype, genetic background, therapeutic response, and clinical outcome.

**MATERIALS & METHODS**

**Patient population.** The mutational status for cohesin complex genes (**STAG2, STAG1, RAD21, SMC3, SMC1A, PDS5B, WAPAL, MAU2, REC8, NIPBL, SMC5 and ESCO2**) were analyzed in bone marrow and blood specimens of 1060 patients with myelodysplastic syndrome (MDS) (N=386), myeloproliferative neoplasm (MPN) (N=55), MDS/MPN (N=169), secondary AML (sAML) (N=149) that evolved from these conditions, and primary AML (pAML) (N=301) including 200 patients in The Cancer Genome Atlas (TCGA) AML cohort (Supplemental Table 1). MDS subtypes of refractory anemia (RA), RA with ringed sideroblasts (RARS) and refractory cytopenia with multilineage dysplasia (RCMD) were classified as low-risk MDS. RA with excess blasts (RAEB1 & 2) cases were considered as high-risk MDS. Informed consent was obtained according to protocols approved by the institutional review boards and in accordance with the Declaration of Helsinki. Diagnosis was confirmed at each institution according to World Health Organization classification criteria. Clinical parameters studied included age, gender, overall survival (OS) and blood counts. These parameters were all available for 780 patients of this cohort after excluding 200 pAML TCGA cases and 80 patients with incomplete clinical details. Metaphase cytogenetics details were retrieved from 892 patients. The median follow up of the cohort was 20.4 months (range, 1-282 months).

**Whole exome sequencing and targeted multi-amplicon deep sequencing.** For whole exome sequencing (WES), the 50 Mb of protein coding sequences were enriched from total genomic DNA by liquid phase hybridization using SureSelect® ver. 4 (Agilent Technology, Santa Clara, CA), followed by massively parallel sequencing with HiSeq 2000 (Illumina, San Diego, CA). Data were analyzed using our in-house pipeline for somatic mutation calling (Genomon) as previously described (http://genomon.hgc.jp/exome/en/index.html)20. To minimize false positives and focus on the most prevalent/relevant somatic events, we implemented a rational bio-analytic filtering approach and applied heuristic bio-analytic pipelines. We used two independent pipelines for identifying somatic and germline alterations. For confirmation of somatic mutations, we analyzed paired germline DNA from CD3+ lymphocytes. The selected observations were validated by targeted deep sequencing using MiSeq. Our sequence library for deep sequencing was generated by TruSeqCustom Amplicon (Illumina, San Diego, CA). Moreover, cohesin gene mutations were screened using whole exome sequencing results available through TCGA (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp). To detect germline variants of
cohesin genes our strategy was to select non-synonymous variants present in both somatic and germline samples that are possibly deleterious. Further, the putative candidates were prioritized based on population genotypic frequency. For the purpose of this study, non-synonymous variants known to be pathogenic, rare or with unknown genotypic frequency variants in general population are included. The mean coverage for cohesin genes (STAG2, SMC3, RAD21) is 265x for the deep sequencing samples.

**Expression array and RNA sequencing:** RNA sequencing from TCGA patients was downloaded from the TCGA repository. They were aligned to hg19 using STAR\textsuperscript{21}. Differential expression analysis for this data set was performed using the Bioconductor DESeq2 package\textsuperscript{22}. Patients with t(15;17) were excluded from any analyses. Cut-offs for low expression of STAG2 was considered the bottom 7.5% transcript levels, and high expressors were considered the top 25%.

Candidate genes identified were validated using two external validation cohorts. Affymetrix U133plus2 cel files for 183 MDS patients and 17 healthy controls in GEO GSE19429 were pooled using Bioconductor’s gcrma tools for quantile normalization. Histograms of log2 expression values were then used to identify cut-offs that defined low expression of STAG2, RAD21, and SMC3 as compared to the cohort of healthy donors. Genes differentially expressed between these two groups were then identified using the Bioconductor package limma and Benjamini-Hochberg adjusted (i.e. to correct for multiple hypothesis testing\textsuperscript{23}) P values of ≤0.001; specifically, we used the limma pipeline of voom() ->lmFit() ->eBayes() which can be applied to both microarray and RNASeq data.\textsuperscript{24} Further analyses were performed using oncomine (https://www.oncomine.org/).

**Statistical analyses.** Comparisons of proportions were performed by the χ² test and by Fisher’s exact test, and differences in values and in ranks were assessed by t tests and Mann-Whitney U tests, respectively. Cox models were used to identify correlates with overall survival (OS). Examination of log (−log) survival plots and partial residuals confirmed that the underlying assumption of proportional hazards was met except in the case wherein survival curves are equal during the first 12 months and separate drastically thereafter (Figure 3); we therefore dichotomized follow up into before and after 12 months. Significant covariates were dichotomized at their medians and resulting two-group log rank test differences were visualized using Kaplan-Meier methods. Throughout, 2-sided tests were used with significance defined as α < .05. These analyses were performed using JMP9 (SAS, Cary, NC) and R-statistical language.
RESULTS

Cohesin gene family sequence alterations in myeloid neoplasms. After excluding all known single nucleotide polymorphisms (SNP) and filtering out alterations present in paired germline samples, somatic alterations in cohesin complex genes were identified in 11.6% of patients, including STAG2 (5.9%), RAD21 (2%), and SMC3 (2%). Somatic mutations in other cohesin genes (SMC1A, PDS5B, WAPAL, MAU2, REC8, NIPBL, SMC5, ESCO2) were present in <1% of our cohort (Figure 1, Supplemental Table 2-4). We identified 2 cases with germline known rare non synonymous alterations, specifically a germline alteration in STAG1 (p.E760D) with population allelic frequency of 0.01, and a novel alteration in WAPAL (p.G230V, Supplemental Table 4) All other sequence variants were confirmed to be somatic by sequencing paired flow sorted CD3+ cells. Cohesin complex mutations were almost always mutually exclusive (Figure 2a). All cohesin mutations were heterozygous in nature except for mutations in STAG2 and SMC1A, which are located on chromosome X and thus result in the absence of expression of a wild-type (WT) allele. In STAG2, we detected mutations in 17 females (14 heterozygous and 3 homozygous) and in 44 males (hemizygous). In addition, we identified 8 deletions involving STAG2 (Chr X) or RAD21 (Chr 8) (Supplemental Figure 3, Supplemental Table 3). Among STAG2 sequence alterations, 62/63 predicted loss-of-function, including frame shift (n=19), splice site (n=10) and/or nonsense (n=33) mutations. Similarly, 15/20 RAD21 sequence alterations were either nonsense (n=2), frame shift (n=11) or splice site (n=2) mutations causing low expression of the gene; 5 were missense mutations. Among SMC3 alterations, 8/19 were nonsense (n=1), frame shift (n=1) or splice site (n=6) while 11/19 were missense mutations.

Clinical features of cases harboring cohesin complex gene mutations. Cohesin defects were found in 11% of lower-risk MDS patients and 4% of MDS/MPD overlap patients (Supplemental Table 4). A higher frequency of cohesin defects were found in sAML and high-risk MDS patients, with somatic alterations in 20% and 17% of these disease phenotypes, respectively (Figure 1c). STAG2 was associated with higher-risk disease (p=0.0023), and RAD21 and SMC3 defects were most frequently identified in sAML (3.5% and 4%, respectively; Supplemental Figure 1). We did not identify any other clinically relevant differences in the distribution of cohesin defects. Cohesin mutant MDS patients were more likely to be associated with high International Prognostic Scoring System (IPSS) scores (Table 1), and were more prevalent in patients who transformed to sAML.

Cytogenetic and molecular abnormalities associated with cohesin defects. To examine the potential association of cohesin complex mutations with other somatic defects, cases with mutated
cohesin complex genes were compared to WT. Normal cytogenetics were present in 48% of mutant cases while 10% had complex cytogenetics, a distribution comparable to WT patients. Among cases with an abnormal karyotype, trisomy 8 was more common in cohesin mutant patients (19% vs. 9%, p=0.004, Supplemental Table 6), while -5/del(5q) and -7/del(7q) were less common (Table 1). When compared to the cohesin WT cohort, co-occurring mutations in *RUNX1* (p=0.0001), Ras-family oncogenes (p=0.0001) *BCOR* (n=15; p=0.003) and *ASXL1* (p=0.03) were significantly enriched in patients with cohesin mutations (Figure 2, Supplemental Table 7). When stratifying by normal/abnormal karyotype, no particular genetic lesion was enriched based on karyotype, although 5/9 patients (55%) with *RUNX1* mutations and abnormal karyotype had trisomy 8.

**Therapeutic response and clinical outcomes in patients with cohesin alterations.** Within our cohort, we analyzed response to therapeutic interventions of lenalidomide (n=92, Supplemental Table 7) or the hypomethylating agents (HMA), decitabine and azacitidine, (n=140, Table 2). Patients who received at least 4 cycles of treatment were assessed by modified International working group criteria. For this study we considered complete remission, partial remission and hematologic improvement (HI) in any cell lineage as response, while stable disease, death, progression or relapse during therapy as non-response. Patients with *STAG2/RAD21* mutations were more likely to respond to treatment with HMA compared to WT cases (79% vs. 47%; p=0.04) and were comparable in their age and IPSS scores to the non-responders. With respect to lenalidomide-treated patients (8 cohesin mutant and 84 WT), we did not observe a significant difference in response rate in cohesin mutant (50%) vs. cohesin WT (52%) patients. No overall survival advantage was observed despite the increased response to HMA. Patients with cohesin gene alterations had poor overall survival compared to patients without cohesin lesions (27.2 vs. 39.9; p=0.023). Further we performed cox multivariable analysis of myeloid malignancy subtypes and identified that presence of *STAG2* mutation predicted poor survival independent of age, sex and IPSS scores in high and low risk MDS patients: HR=1.5 (0.9,2.6), p=0.11, but in patients with MDS who survived beyond 12 months, HR=2.1 (1.1, 3.8), p=0.017, (Figure 3).

**Clonal acquisition of cohesin complex gene mutations.** Clonal hierarchy analysis was performed to determine if cohesin complex gene variants were ancestral or secondary genetic events. This involved first, variant allelic frequency (VAF) analyses of co-occurring mutations in samples, adjusted for copy number. The adjusted VAF of *STAG2*, *RAD21* and *SMC3* mutations were significantly lower than those for other concomitantly present *TET2* gene mutations (p=0.01). In this cross-sectional analysis of the cohesin mutant cohort *STAG2* and *SMC3* mutant clones were ancestral in only 18% of the cases. We
also performed serial analyses for 12 cases with \textit{STAG2} and \textit{SMC3} mutations and one case with a mutation of \textit{RAD21} (Figure 4a-b). This enabled us to determine clonal architecture, and we observed that cohesin mutant clones expanded or remained stable in all 13 patients with a statistically significant increase in clone size ($p=0.0032$). Findings were similar for each individual mutation, with clonal expansion in 7 of 8 patients with \textit{STAG2} mutations ($p=0.05$), 2 of 4 \textit{SMC3} mutant patients, as well as in the single patient with a \textit{RAD21} mutation. In 2 cases of sAML that evolved from RAEB, cohesin mutations were observed only after transformation to AML (Figure 4c-d), indicating these mutations can occur as a late event that is associated with AML transformation. In all, 10 of 13 patients had cohesin present in the dominant clone after transformation (Supplemental Table 10-11). Three AML cases with cohesin mutations (\textit{SMC3}, \textit{STAG2} and \textit{RAD21}) at diagnosis achieved remission with cytarabine and idarubicin treatment. Analysis for mutational burden after therapy demonstrated the \textit{STAG2} and \textit{SMC3} mutant clones in two cases were undetectable and the third case had persistence of a small \textit{RAD21} clone (44% to 7%).

\textit{Expression signatures associated with cohesin defects.} The presence of hypomorphic mutations suggests that somatic alterations of cohesin lead to loss-of-function. It is thus possible that functionally decreased expression may produce a clinical phenocopy of cohesin mutations. When dichotomized using a threshold expression of $<$mean-2SD (95\% confidence interval) as a cut off value, 10-20\% of cases showed decreased cohesin expression (Figures 5a-b). A subset of these patients had mutations in cohesin complex members and reduced expression of \textit{STAG2/RAD21/SMC3}, indeed we noted that most patients with cohesin mutations had reduced expression of the entire cohesin complex. However, we also noted reduced expression of cohesin genes in patients without mutations, including, 15\% of lower- and higher-risk MDS patients with reduced expression of \textit{STAG2} and \textit{RAD21}. We also observed a subset of AML patients with marked reduction in \textit{STAG2} and \textit{PDS5B} expression in the absence of cohesin mutation. We also investigated expression signatures of genes associated with cohesin mutations or with reduced expression of \textit{STAG2/RAD21/SMC3}. Our analysis showed a similar expression profile in patients with \textit{STAG2} mutations or patients with reduced expression of cohesin complex members (Figure 6a). We observed decreased expression of \textit{DCK} and increased expression of \textit{UCK} in patients with cohesin mutations or with non-mutational suppression of cohesin expression, indicating common alterations of downstream pathways (Supplemental Table 12). Notably, expression of the entire cohesin complex was reduced in patients with \textit{STAG2} mutations in a manner similar to that seen in patients with reduced cohesin expression in the absence of mutations (Figure 6b). When comparing patients with reduced expression or mutation of \textit{STAG2} to patients with intact expression of \textit{WT STAG2}, we identified 28 genes, which were differentially expressed in patients with \textit{STAG2}...
alterations regardless of whether this was due to somatic mutations or transcriptional downregulation (Supplemental Table 13). This includes decreased expression of signaling effectors \((NRAS, JAK1, CBL)\), and reduced expression of \(HIF1A\), suggesting a novel mechanism of transformation for patients with myeloid malignancies characterized by alterations in the cohesin complex. Previous work in a zebrafish model identified runx dependence on rad21 expression\(^{26}\) and given the high association of \(RUNX1\) mutations in cohesin mutant myeloid disease, we queried if this relationship would be evident in our patient cohort. We were unable to find any correlation between \(RAD21\) and \(RUNX1\) expression, nor between mutational status and expression, in our patient cohorts, suggesting that this interaction in humans may be subject to pathway redundancy.

**Discussion**

We have demonstrated that somatic alterations in cohesin complex genes occur in 11.6% of patients with myeloid neoplasms in our cohort, consistent with recent reports in smaller patient cohorts\(^1,2\). The observation that most mutations are nonsense and frame shift types suggests that cohesin mutations lead to decreased function of the cohesin complex. As low expression of these genes may have functionally equivalent consequences, we defined a subset of patients with altered expression of the cohesin complex in an additional 15% of patients. Such non-mutated cohesin low-expressors have been described in both AML and Ewing’s sarcoma cell lines\(^3,5\). Additionally, this study expands the current knowledge of cohesin mutations across the spectrum of myeloid diseases, identifying cohesin mutation frequencies in different disease subtypes, most prominently in sAML (20%). Clonal hierarchy analysis rarely identified cohesin mutations in founding clones; rather cohesins were present in subclones that underwent clonal expansion over time. This subclonal architecture occurs in the setting of co-occurring genetic and molecular abnormalities, including mutations in chromatin modifiers (\(ASXL1\)) and key transcriptional regulators (\(RUNX1\), Ras family genes, \(BCOR\)). Our data further suggest that cohesin has a cooperative role in epigenetic regulation and transcriptional programming by virtue of the higher response rate in cohesin mutant MDS patients treated with hypomethylating agents.

In line with other studies, cohesin mutations were not associated with an increase in the frequency of complex cytogenetic abnormalities. Despite cohesin’s role in alignment of sister chromatids during metaphase, these data provide further evidence that the leukemogenic consequences of cohesin defects do not involve disrupted mitosis and global chromosomal instability. The expanded sample size of our cohort also enabled precise analysis of clinical associations. Cohesin mutations did not alter the IPSS cytogenetic risk of MDS patients (average IPSS cytogenetics score 0.55 points for cohesin mutant cohort and 0.60 points for WT). Nonetheless, overall IPSS risk was higher in cohesin-mutated patients, driven by high blast count, suggesting a stem cell defect in the mutant clone.
The vast majority of mutations identified across cohesin proteins were heterozygous nonsense and frame shift events, which would be anticipated to cause expression of a truncated protein. In addition to clearly hypomorphic mutations, splice site mutations were also identified (e.g., STAG2 5’ of exon 4 and 8). Splice site mutations can cause activation of a nearby cryptic splice site (upstream or downstream of the normal site), omission of the whole exon, or intron retention, any of which can lead to frame shifts and premature stop codons that will trigger non-sense mediated decay and haploinsufficiency. Cohesin haploinsufficiency is thus the operative pathophysiologic mechanism, a notion further supported by the mutual exclusivity of cohesin mutations: haploinsufficiency for one cohesin complex member may be sufficient to initiate the myeloid transformation pathway, while a double hit would be deleterious. Recent studies of cohesin dynamics at super-enhancer sites are in line with these conclusions, cohesin localizes to sites of high mediator occupancy at the enhancer-promoter interaction sites in lymphoma and myeloma. Treatment with the BRD4 inhibitor JQ1 leads to a 2-fold decrease in cohesin occupancy at super-enhancer sites and normalization of the expression of oncogenic drivers. Taken together, these data suggest that cohesin function is integral to expression of key cell identity genes, and that reduction in gene dose of cohesin alters cellular identity.

Analysis of clonal dynamics in patients in whom serial samples at various time points along their disease course provides important insight into the timing and effect of cohesin loss-of-function on leukemia initiation and progression. Using deep sequencing and cross-sectional and serial analyses, we found that cohesin mutations were not commonly present in the founder clone but rather promoted clonal expansion and transformation to more aggressive disease. In 10 of 13 cohesin mutant patients with serial samples, the cohesin mutant clone achieved clonal dominance at the time of transformation. Cohesin mutations co-occur with other mutations known to be drivers of clonal evolution. The strong association with ASXL1 indicates that epigenetic dysregulation may be synergistic in the mechanism of action exerted by cohesin haploinsufficiency. Furthermore, activated Ras signaling may alter the expression of key identity genes, which are then further dysregulated by cohesin inactivity.

High numbers of informative patients (those with mutations) are needed to establish correlations with response to specific drugs. In exploring the effects of cohesin downmodulation/mutations on expression pattern in cohort, we noted a consistent downmodulation of DCK in patients with cohesin complex alterations. Expression of this enzyme has been reported to be reciprocal to the levels of UCK, a rate-limiting enzyme for azacitidine metabolism. Thus, azacitidine would be expected to have greater
efficacy in cohesin-deficient patients. Indeed, we observed a higher response rate in patients with cohesin mutations treated with azacitidine.

In summary, cohesin mutations, which likely result in loss-of-function, are recurrent molecular genetic events in both solid and hematopoietic tumors. Mechanistic studies to understand their functional and pathophysiologic role are currently underway, but our genetic data suggest that cohesin mutations do not contribute to hematopoietic transformation through altered chromosomal instability. The high frequency with which cohesin mutations occur in myeloid disease and their presence in expanding subclones during disease progression suggests that the cohesin complex may represent an attractive therapeutic target for future preclinical and clinical studies.
Acknowledgments:
Supported in part by NIH grants from R01 655365071402, U54 655365070704, K24 655365071503, (J.P.M.), R01 CA169784 (R.L.L.), a grant from the Aplastic Anemia and MDS International Foundation (J.P.M.); a grant from the Mortimer J. Lacher fellowship of the Lymphoma Foundation (A.D.V.), a grant from the Edward P Evan’s MDS CRC (M.S.), and a generous gift from the Trotter family

Authorship:
Designed research: AV, ST, HM, RL, JM
Performed research: AV, ST, BS, BP
Analyzed data: AV, ST, HM, BS, BP, TR
Wrote the paper: AV, ST, HM, BS, BP, TR, MS, RL, JM

Conflict of Interest Disclosure:
The authors disclose no potential conflicts of interest.
References:


### Table 1: Clinical characteristics of patients with cohesin family gene defects

<table>
<thead>
<tr>
<th></th>
<th>Mutants (N=93)</th>
<th>Wild type (N=687)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;60 years</td>
<td>72</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td>&lt;60 years</td>
<td>21</td>
<td>183</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>58</td>
<td>426</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>35</td>
<td>261</td>
<td>1</td>
</tr>
<tr>
<td><strong>Dx</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS-Low</td>
<td>22</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>MDS-High/sAML</td>
<td>52</td>
<td>219</td>
<td>0.0036</td>
</tr>
<tr>
<td>MDS/MPN</td>
<td>7</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>MPN</td>
<td>3</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>pAML</td>
<td>9</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>*<em>IPSS</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>8</td>
<td>99</td>
<td>0.05</td>
</tr>
<tr>
<td>INT-1</td>
<td>20</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>INT-2</td>
<td>10</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>HIGH</td>
<td>4</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>*<em>Karyotype</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>56 (48%)</td>
<td>351 (45%)</td>
<td>0.1</td>
</tr>
<tr>
<td>Abnormal</td>
<td>60 (53%)</td>
<td>425 (54%)</td>
<td>0.1</td>
</tr>
<tr>
<td>Complex</td>
<td>11 (10%)</td>
<td>104 (13%)</td>
<td>0.36</td>
</tr>
<tr>
<td>-5/del(5q)</td>
<td>4 (4%)</td>
<td>89 (11%)</td>
<td>0.007</td>
</tr>
<tr>
<td>-7/del(7q)</td>
<td>4 (4%)</td>
<td>99 (13%)</td>
<td>0.002</td>
</tr>
<tr>
<td>Trisomy 8</td>
<td>22 (19%)</td>
<td>69 (9%)</td>
<td>0.004</td>
</tr>
<tr>
<td>-20/del(20q)</td>
<td>6 (5%)</td>
<td>45 (6%)</td>
<td>1</td>
</tr>
<tr>
<td><strong>OS (months)</strong></td>
<td>27.2</td>
<td>39.9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*n=42 (Mutants), 281 (wild type)

*n =116 (Mutants), 776 (wild type)
<table>
<thead>
<tr>
<th></th>
<th>Cohesin Mutant</th>
<th>Cohesin WT</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pts treated with HMA and response assessed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=93</td>
<td>N=627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age at start of therapy (years)</td>
<td>71</td>
<td>69</td>
<td>0.98</td>
</tr>
<tr>
<td>Median IPSS Score at diagnosis</td>
<td>1.5</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Cytogenetics at treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>8</td>
<td>53</td>
<td>0.79</td>
</tr>
<tr>
<td>Abnormal Karyotype</td>
<td>8</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Complex Karyotype</td>
<td>2</td>
<td>28</td>
<td>0.52</td>
</tr>
<tr>
<td>-5/del(5q)</td>
<td>1</td>
<td>19</td>
<td>0.46</td>
</tr>
<tr>
<td>-7/del(7q)</td>
<td>0</td>
<td>27</td>
<td>0.08</td>
</tr>
<tr>
<td>Trisomy 8</td>
<td>3</td>
<td>8</td>
<td>0.22</td>
</tr>
<tr>
<td>-20/del(20q)</td>
<td>0</td>
<td>16</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Diagnosis at treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-risk MDS</td>
<td>1</td>
<td>38</td>
<td>0.04</td>
</tr>
<tr>
<td>High-risk MDS</td>
<td>7</td>
<td>38</td>
<td>0.39</td>
</tr>
<tr>
<td>sAML</td>
<td>6</td>
<td>20</td>
<td>0.07</td>
</tr>
<tr>
<td>MDS/MPN</td>
<td>2</td>
<td>24</td>
<td>0.76</td>
</tr>
<tr>
<td>MPN</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>pAML</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Types of response per IWG criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responder(CR,PR, HI-E)</td>
<td>13</td>
<td>59</td>
<td>0.02</td>
</tr>
<tr>
<td>Non-responder (SD &amp;Failure)</td>
<td>3</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td><strong>Other somatic mutations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET2</td>
<td>3</td>
<td>15</td>
<td>0.43</td>
</tr>
<tr>
<td>TP53</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>ASXL1</td>
<td>5</td>
<td>16</td>
<td>0.06</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>RUNX1</td>
<td>5</td>
<td>12</td>
<td>0.02</td>
</tr>
<tr>
<td>IDH1&amp;2</td>
<td>3</td>
<td>3</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Figure Legend

**Figure 1: Characterization of cohesin mutations in patients with myeloid disease**
A) Structure of cohesin ring
B) Position and disease subtype for identified cohesin mutations  
   i) SMC3  
   ii) STAG2  
   iii) RAD21.  
   No correlation was found with any structural motif, and no mutational hotspots were identified
C) Frequency of cohesin family gene mutations in each myeloid malignancy
D) Distribution of cohesin family gene mutations identified across the patient cohort

**Figure 2: Co-occurring somatic mutations in cohesin mutant patients**
A) Mutually exclusive cohesin gene mutations and co-occurring somatic genetic events.
B) Somatic mutational correlation in cohesin mutant/deletion cases

**Figure 3: Effect of cohesin mutational status on overall survival in MDS**
Kaplan-Meier curve for patients with MDS are shown for A) STAG2 mutant patients and cohesin WT patients and B) All cohesin mutant patients vs. WT cohesin patients.  
Median overall survival of MDS patients with STAG2 mutation was 35 months versus 50 months,  
p=0.17, HR: 1.8 CI [1.1-3.1], median survival for cohesin mutant MDS cases were inferior to  
wild type MDS cases, 34.2 versus 48.2, p=0.21, significant survival difference was pronounced  
in patients with MDS who survived for at least 12 months, p=0.01

**Figure 4: Analysis of clonal hierarchy points towards cohesin as a secondary mutation**
A&B) Exemplary serial samples illustrating clonal architecture of cohesin family genes.
C) VAFs of STAG2 mutations by deep sequencing in serial samples.  For each STAG2 mutant patient with available serial samples, the variant allele frequency in each sampling is shown.  
While only 2 patients were found to harbor a STAG2 mutation in the dominant clone at the first  
time point, 7/8 (87.5%) of patients who were initially found to have a subclonal STAG2 mutation  
underwent clonal expansion, becoming the dominant clone at the time of transformation  
(p=0.05).
D) VAFs of SMC3 & RAD21 mutations by deep sequencing in serial samples.  For 4 patients  
with SMC3 mutations and the 1 patient with a RAD21 mutation with available serial samples,  
the variant allele frequency in each sampling is shown.  All of these 5 patients harbored  
subclonal cohesin mutations at the first time point, with 2/4 (50%) of the SMC3 mutant patients  
and the 1/1 RAD21 mutant patient undergoing clonal expansion, becoming the dominant clone  
at the time of transformation.

**Figure 5: Analysis of cohesin expression identifies a discrete subset of low expressors**
A) Expression of STAG2 in MDS & primary AML subtypes.  A discrete subset of patients  
   exhibit low STAG2 expression defined as <2SD below the mean as compared to 17  
   healthy donors.  This was more prevalent in patients with high risk MDS and non-CBF  
   AML without complex cytogenetics.
B) Percent of low expressors in cohesin genes.  The frequency of cohesin low expressors  
   for each cohesin complex gene (STAG2, RAD21, SMC1A, SMC3, STAG1, and PDS5B)  
   is shown by disease subtype.  STAG2 was most commonly underexpressed in high risk  
   MDS whereas RAD21 was most commonly underexpressed in the remaining myeloid  
malignancies
Figure 6: Unsupervised hierarchical clustering of gene expression with annotation by cohesin mutational status and cohesin expression

A) Unsupervised hierarchical clustering of TCGA AML patients with annotation of STAG2 low expression and STAG2 mutation. Patients with STAG2 mutations and low STAG2 expression cluster show unique expression prolife, and cluster separately from patients with WT STAG2 and high expression.

B) Cohesin complex expression in i) STAG2 ii) RAD21 and iii) SMC3 low expressors. Within each group of cohesin low expressors, the entire cohesin family gene complex has similarly decreased expression.
Figure 1: Characterization of cohesin mutations in patients with myeloid disease

A

B i)

ii)

iii)

C

D

% of mutants

Low MDS (n=237)

High MDS (n=149)

sAML (n=149)

MDS/MPN (n=169)

MPN (n=55)

pAML (n=301)
Figure 2: Co-occurring somatic mutations in cohesin mutant patients
Figure 3: Effect of cohesin mutational status on overall survival in MDS
Figure 4: Analysis of clonal hierarchy points towards cohesin as a secondary mutation.
Figure 5: Analysis of cohesin expression identifies a discrete subset of low expressors
Figure 6: Unsupervised hierarchical clustering of gene expression with annotation by cohesin mutational status and cohesin expression.
Genetic alterations of the cohesin complex genes in myeloid malignancies

Swapna Thota, Aaron D. Viny, Hideki Makishima, Barbara Spitzer, Tomas Radivoyevitch, Bartlomiej Przychodzen, Mikkael A. Sekeres, Ross L. Levine and Jaroslaw P. Maciejewski