博士論文

骨髄系腫瘍におけるコヒーシン複合体の変異

昆　彩奈

骨髄系腫瘍におけるコヒーシン複合体の変異

東京大学大学院医学系研究科

医学博士課程

生体物理医学専攻

指導教員　宮川　清　教授

氏名　　昆　彩奈

Abstract

Cohesin is a multimeric protein complex involved in cohesion of sister chromatids, post-replicative DNA repair, and transcriptional regulation. Here I report recurrent mutations and deletions involving multiple components of the cohesin complex, including *STAG2, RAD21, SMC1A,* and *SMC3*, in different myeloid neoplasms. These mutations and deletions were largely mutually exclusive and occurred in 12.1% (19/157) of acute myeloid leukemia (AML), 8.0% (18/224) of myelodysplastic syndromes (MDS), 10.2% (9/88) of chronic myelomonocytic leukemia (CMML), 6.3% (4/64) of chronic myelogenous leukemia (CML), and 1.3% (1/77) of classical myeloproliferative neoplasms (MPN). Cohesin-mutated leukemic cells showed reduced amounts of chromatin-bound cohesin components, suggesting a substantial loss of cohesin binding sites on chromatin. The tumor suppressive role of cohesin was supported by growth suppression of leukemic cells having *RAD21*-mutation (Kasumi-1) or severely reduced RAD21/STAG2 expression (MOLM-13) induced by forced expression of wild-type RAD21 and/or STAG2, respectively. These findings highlight a role for compromised cohesin functions in myeloid leukemogenesis.

Introduction

Myelodysplastic syndromes (MDS) and related disorders comprise a group of myeloid neoplasms characterized by ineffective hematopoiesis leading to deregulated, dysplastic blood cell production and by progression to acute myeloid leukemia (AML) in a third of patients.[1] MDS are most common in elderly people.

Diagnosis of MDS is based on examination of the blood and bone marrow showing cytopenias and hypercellular marrow with dysplasia, with or without excess of blasts.[2] An optimized choice of treatment based on accurate diagnosis and risk stratification in individual patients is central to the current therapeutic strategy.[3] Many prognostic factors have been identified in MDS, of which the international prognostic scoring system (IPSS[4]) have been successfully used for the classification of patients, where the marrow blast percentage, number and extent of cytopenias, and cytogenetic abnormalities are taken into account.

Treatment of patients with lower-risk MDS includes transfusions, growth factors and lenalidomide, whereas treatment of higher-risk MDS patients includes hypomethylating agents and allogeneic stem-cell transplantation (allo-SCT).[5] Although allo-SCT remains the only curative treatment of higher-risk MDS, it is often difficult to choose this therapy because of higher rates of transplant-related mortality, especially in elderly patients. In this context, it is desired to develop novel therapeutic strategies that target the novel class of mutational targets.

The pathophysiology of MDS involves a multistep process of acquiring chromosomal abnormalities and/or somatic gene mutations in the multi-potent hematopoietic stem cells.[6,7] Our knowledge about the molecular pathogenesis of MDS has dramatically improved during the past 10 years through the identification of a

number of recurrent somatic mutations by array-based comparative genomic hybridization (aCGH), SNP-array copy number analysis and massively parallel sequencing technologies.[8-12] These mutations not only involved previously defined pathways such as signal transducing molecules (*NRAS*[13]) and transcription factors (*TP53*[14], *RUNX1*[15]), but also involved newly identified pathways including epigenetic regulators involved in the DNA methylation (*TET2*[16,17], *DNMT3A*[18], *IDH1/2*[19]) and chromatin modifications (*ASXL1*[20], *EZH2*[21,22]).

More recently, two landmark studies led to the discovery of mutations implicated in the RNA splicing machinery through whole exome or whole genome sequencing.[23,24] Ogawa and colleagues have performed whole exome sequencing of 29 paired samples of myelodysplasia and subsequent targeted sequencing in a large series of myeloid neoplasms and revealed that multiple components of the RNA splicing machinery, including *U2AF35*, *ZRSR2*, *SRSF2*, and *SF3B1* were frequently mutated (~45 to 85%) in a mutually exclusive pattern.[23] By contrast with mutations in epigenetic regulators, spliceosome mutations are highly specific to myeloid neoplasms showing features of myelodysplasia.

Thus, with the advent of high-throughput sequencing technologies, the catalog of gene mutations implicated in the pathophysiology of MDS is rapidly expanding. On the other hand, however, mutations of these sets of genes do not fully explain the pathogenesis of MDS because they are also commonly found in other myeloid malignancies and about 20 % of MDS cases have no known genetic alterations.[25] Therefore, more complete genetic characterization of MDS has great potential to deeply elucidate the molecular basis of MDS, to refine the risk stratification of MDS, and to develop novel targeted therapies.

Based on these perspectives, this study aimed to identify novel class of genetic lesions implicated in the pathogenesis of myeloid neoplasms using massively parallel sequencing technologies, and further aimed to understand the functional importance of such mutations and improve the diagnosis and treatment.

Materials and Method

Patients and samples

Whole exome sequencing were performed using bone marrow- or peripheral blood-derived DNAs from 29 patients with MDS or related myeloid neoplasms with paired CD3-positive T cells or buccal mucosa used as germline control. The detailed methodology of whole exome sequencing and WHO classification of the cases have been previously described.[23] The list of somatic mutations reported in the previous work,[23] where a total of 497 candidate SNVs and indels were identified through whole exome sequencing of 29 paired samples of myelodysplasia, were updated by validating the remaining 62 unconfirmed mutations by repeated Sanger sequencing and also by deep sequencing. Anonymized genomic DNA from 610 patients with myeloid neoplasms were collected from collaborating institutes, including the University of Tokyo, Chang Gung Memorial Hospital, University Hospital Mannheim, Tokyo Metropolitan Ohtsuka Hospital, Showa University Fujigaoka Hospital and University of Tsukuba, and analyzed using high-throughput DNA sequencing, and in part, SNP array-based copy number analysis, pyrosequencing, digital PCR and methylation analysis. All patients gave their written informed consent with genetic analyses. This study was approved by the ethics boards of the University of Tokyo (approved number: 948), University Hospital Mannheim, University of Tsukuba, Showa University, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital.

Cell lines

CMS, CMY, UTP-DSAL-1, MOLM-1, MOLM-7, HEL, SS9;22, and TS9;22 cell lines were provided from Dr. Yasuhide Hayashi (Gunma Children's Medical Centre).

P31FUJ and CMK-86 cells were purchased from the Health Science Research Resources Bank (Osaka, Japan). 293T, 293gp, KG-1, K562, and F-36P cells were obtained from RIKEN BRC Cell Bank (Tsukuba, Japan), and Kasumi-1, HL-60, MOLM-13, and TF-1 were from the American Type Culture Collection (ATCC).

Mutation screen enhanced by deep sequencing of pooled target exons or captured Target sequences

In total, 9 components of the cohesin complex, including *STAG1, STAG2, SMC1A, SMC3, RAD21*, *PDS5B*, *ESCO1, ESCO2 and NIPBL* were screened for mutations in 534 samples using high-throughput sequencing of pooled exons amplified from pooled genomic DNA samples according to the previously described method with some modifications in the algorithm for mutation call[23]. For 47 samples, mutations in *STAG2, SMC1A, SMC3,* and *RAD21* were investigated using SureSelect (Agilent) captured target sequences.

A) High-throughput sequencing enhanced by deep sequencing of pooled target exons

Screening of gene mutations in *STAG1, STAG2, SMC1A, SMC3, RAD21*, *PDS5B*, *ESCO1, ESCO2 and NIPBL* was performed using high-throughput sequencing of pooled exons amplified from pooled genomic DNA samples according to the following procedures.

1) In total, 48 DNA pools were generated from 576 DNA samples by equimolar mixture of 9μL each of whole genome-amplified genomic DNA from 12 different tumor samples or leukemia-derived cell lines, where concentrations of each DNA were adjusted to 75 ng/μL before mixture.

2) All target exons (N = 232) that encompass 89,323 nucleotides were PCR-amplified using a set of primers having a common NotI adaptor sequences on their 5' ends, in which a high-fidelity polymerase (LA Taq HS (TAKARA Ltd. Tokyo, Japan)) was used for PCR reactions.
3) After inspection by agarose gel electrophoresis, all the 232 amplicons in single pools were combined and purified using QIAquick PCR Purification Kit®(QIAGEN), followed by digestion with NotI.
4) After NotI digestion, DNA was re-purified in a similar manner and an aliquot of 2.5ug of purified DNA from each pool was subjected to DNA ligation using T4 DNA polymerase.
5) Ligated DNA was diluted to 120 μL volume, sonicated into ~200bp fragments on average using Covaris®, and used for generation of sequencing libraries, according to a modified pair-end library protocol from Illumina.
6) The libraries were then subjected to deep sequencing on Illumina HiSeq 2000® according to the standard protocol for 75 bp pair end reads, in which 3-12 DNA pools were analyzed in a single lane by using indexed sequencing primers. On average, 99.5% of the target bases were analyzed at the depth of 12,000 per pool or 1,000 per sample.

	Mean depth				
	≥ 8,000	≥ 4,000	≥ 2,000	≥ 1,000	≥ 500
Number of bases	32,544	78,451	87,762	88,910	89,100
Percentage *	36.4	87.8	98.3	99.5	99.8

* indicates a percentage of number of bases of the total number of bases included in the target region (89,323bp)

7) Sequencing data were analyzed by MAPDNA (Shiraishi Y. et al. manuscript in preparation) to identify candidate mutations.

8) Finally, each candidate mutation was confirmed by Sanger sequencing of each of the 12 individual DNA from the corresponding DNA pools, through which the individual or the cell line having the relevant mutation was finally determined.

Basically, the algorithms implemented in the MAPDNA (Shiraishi Y. et al. manuscript in preparation) were already described in the previously publication except for some modifications. Briefly, each short read from a given DNA pool was aligned to the set of target sequences having primer sequences, using blat (http://users.soe.ucsc.edu/~kent/src/), instead of BWA, with the -fine option. The mapping information in a .psl file was transformed into a .sam file using the my_psl2sam script, which was further converted into the .bam format using samtools. Among the successfully mapped reads, the following reads were removed from further analysis, which were mapped to multiple sites, mapped with more than 4 mismatched bases, or had more than 10 clipped bases. Next, the 'Estimation_CRME' script was run to eliminate strand-specific errors and exclude cycle-dependent errors. First, a strand-specific mismatch ratios was calculated for each nucleotide variation for both strands using those bases between 11 ~ 50 cycles. By excluding the top 5 cycles showing the highest mismatch rates, strand-specific mismatch rates were recalculated and the smaller value between both strands was adopted as the nominal mismatch ratio for that variation. In addition, those nucleotide variations that appeared across multiple pools were removed based on permutations across different pools, using the "Permut_Rm_com" script, because such variations are likely to result from systemic sequencing errors. Finally, after excluding variations found either in the dbSNP database or the 1000 genomes project database as well as our in-house SNP database,

those variants whose mismatch rate exceeded 0.009 were adopted as candidate mutations. In total, 178 positive pools were identified in deep sequencing, of which 58 mutations among 60 corresponding samples were finally validated by Sanger sequencing and/or deep sequencing.

B) Hybrid selection of targets and deep sequencing

In 47 cases, mutations in STAG2, RAD21, SMC1A and SMC3 were examined using deep sequencing. Genomic DNA (0.5–1.5 μg) from BM mononuclear cells or PB was enriched for target exons by liquid-phase hybridization using a SureSelect© custom kit (Agilent®) designed to capture all of the coding exons from the target genes according to the manufacturer's protocol and optimized for automated sample processing (Agilent®). After index sequences were attached to discriminate individual samples, the enriched target DNA was subjected to massively parallel sequencing using HiSeq 2500 with a standard 100-bp paired-end reads protocol. Sequence data were analyzed through our established pipeline to detect somatic mutations. For each sample, all sequencing reads were aligned to hg19 using BWA version 0.5.8 with default parameters. After all duplicated reads, low-quality reads and bases were removed, the allele frequencies of SNVs and insertions/deletions were calculated at each genomic position by enumerating the relevant reads using SAMtools. Initially, all variants showing allele frequencies >0.02 were extracted and annotated using ANNOVAR for further consideration, if they were found in >6 reads among >10 total reads and appeared in both positive- and negative-strand reads.

As for the cases for which no germline DNA was available, relevant somatic mutations were called by eliminating the following entries, unless they were registered in

COSMIC v60 (www.sanger.ac.uk/genetics/CGP/cosmic/) or reported as somatic mutations in PubMed (www.pubmed.gov).

1) All synonymous variants and those having ambiguous (unknown) annotations.
2) Known SNPs in public and private databases, including dbSNP131, the 1000 genomes project as of 2010/11/23, and our in-house database.
3) Sequencing/mapping errors.
4) All missense SNVs with 0.45–0.55 allele frequencies.

Mapping errors were removed by visual inspection on the Integrative Genomics Viewer (http://www.broadinstitute.org/igv/) browser.

Determination of variant allele frequencies

Variant allele frequencies were evaluated by deep sequencing of PCR-amplicons, pyro-sequencing,[26,27] and/or digital sequencing (Fluidigm CA, U.S.) [28-30] of the variants using non-amplified DNA. For amplicon sequencing, genomic fragments harboring the variants under interest were PCR-amplified using Not1-tagged primers. Randomly selected 92 SNP loci which do not contain repetitive sequences were amplified using normal genomic DNA as templates as control. Touch-down PCRs using high-fidelity DNA polymerase KOD-Plus-Neo (TOYOBO, Tokyo) were performed and an equimolar mixture of all PCR products was prepared for deep sequencing using HiSeq2000 or Miseq (Illumina) as described above with a 75bp or 100bp pair end read option. To calculate the frequency of each variant, all reads were mapped to the target reference sequence using BLAT[31] followed by differentially enumerating the dichotomic variant alleles. For indels, individual reads were first aligned to each of the wild-type and indel sequences and then assigned to the one to which better

alignment was obtained in terms of the number of matched bases.

Prediction of the functional impact of mutations

The functional impact of each amino acid substitutions was evaluated by the computer prediction using three open source programs, including SIFT (http://sift.jcvi.org/)[32], PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/index.shtml)[33], and Mutation Taster (http://www.mutationtaster.org/).[34]

Statistical analysis of significance of cohesin mutations

Significance of non-silent mutations in each cohesin component was evaluated, assuming the uniform distribution of the background mutations within the coding regions, which had been estimated to be ~ 0.3/Mb based on the previous whole exome sequencing of myelodysplasia[23].

Array-based copy-number and methylation analyses

Genomic DNA from 453 bone marrow samples with myeloid neoplasms were analyzed using GeneChip SNP-genotyping microarrays as previously described using CNAG/AsCNAR software.[35,36] The result of SNP array karyotyping for 284 of the 453 cases have been previously published.[23,25,36-38] Genomic DNA from the remaining 169 cases were newly analyzed using Affymetrix GeneChip® Human Mapping 250K Nsp array according to the previously described method.[35,36]

Promoter methylation of each cohesin gene was analyzed using Human Methylation 450 BeadChip (Illumina) as previously described ,[39,40] in which methylation status was evaluated by calculating the ratio of methylation-specific and

demethylation-specific fluorophores (β–value) at each CpG site using iScan software (Illumina).

RT-PCR analysis of *STAG2/RAD21* transcripts

To assess the allelic origin of the *STAG2* transcript in CMML-036, a 195bp *STAG2* sequence containing the mutation site (c.1840C>T) was amplified by PCR using archived tumor-derived cDNA as a template and subjected to direct sequencing. The primers for PCR-amplification were 5'- CGTGGCCCTTCCTCAGTTAT-3' (exon 18) and 5'-GAACATGCTTCCAAAACATCTG-3' (exon 20). To confirm the presence of abnormal splicing in MDS-12 caused by a splice donor site mutation c.1196+1G>A at the boundary of *STAG2* exon 13 and intron 14, RT-PCR was designed to amplify cDNA sequences containing exons 12-16, by which cDNA from MDS-12 was analyzed for aberrant transcript using cDNA from MDS-20 (having no known splice site mutations) as a control. The primers were 5'-GGAACTTTTTACCAGTCGGTTC-3' (exon 12; forward) and 5'-TCCCACATGCTATCCACAAG-3' (exon16; reverse). After separation through 4% agarose gel, an aberrant PCR band found in MDS-12 was excised for DNA extraction using NucleoSpin Extract II (MACHEREY-NAGEL) and subjected to Sanger sequencing. To evaluate the mutant allele frequency of the Kasumi-1 transcript, cDNA from Kasumi-1 cells was amplified by PCR using the following primers (5'-AGCATTTGCATTGGAGCCTA-3' and 5'-CTCCTCCTGTCTCTTTCCACA-3') and subjected to deep sequencing using HiSeq2000.

For cDNA synthesis, RNA was extracted from samples using TRIzol® Reagent (Invitrogen). After genomic DNA was depleted by DNase I using RNeasy Mini Kit

(QIAGEN), 1 μg of total RNA was subjected to reverse transcriptase reaction using ReverTra Ace qPCR RT Kit® (TOYOBO, Tokyo) according to the manufacturer's instructions and 50-250 ng of cDNA products were used for subsequent PCR.

Determination of the origin of two mutations found in a male case (MDS-176)

To determine the phase of the two *STAG2* mutations (c.1810C>T and c.2139C>G) identified in a male sample MDS-176, a 3488 bp genomic fragment of the *STAG2* gene containing both mutation sites was amplified from the genomic DNA from MDS-176 bone marrow cells by long PCR using primeSTAR Max DNA Polymerase (Takara), which was cloned into TOPO TA cloning vector (invitrogen) according to the manufacture's instruction. In total, 82 recombinant clones were isolated and each clone was analyzed for their mutation status by sequencing. The primers for long PCR were 5'-GAAGACACAGTTGGATGATAGGAC-3' (Forward) and 5'-AGCAAGTTGCCAAAGGATTAC-3' (Reverse), while those for sequencing were 5'-AAGCGGCCGCTTCCCTAAATGCCTCACAGAA-3' (for c.1810C>T mutation in + strand) and 5'-TGCAGTGCGTGAATAACAATC-3' (for c.2139C>G mutation in - strand).

Subcellular fractionation

Whole-cell extracts of myeloid cell lines were separated into soluble supernatant and chromatin-containing pellet fractions and analyzed by SDS-PAGE and immunoblotting as described.[41,42]

Western Blots

Cells were lysed in RIPA buffer and Western blots were performed following the standard protocol. Primary antibodies used were rabbit polyclonal anti-Rad21 (Bethyl Laboratories, A300-080A), rabbit polyclonal SMC1 (Abcam, ab21583), rabbit polyclonal anti-SMC3 (Abcam, ab9263), goat polyclonal SA1 (Abcam, ab4457), goat polyclonal SA2 (Abcam, ab4463), guineapig anti-Nipbl (Gift from Lena Ström, Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden), rabbit polyclonal Histone-H3 (Abcam, ab1791), goat polyclonal anti-Actin (I-19) (Santa Cruz Biotechnology, sc-1616), and rabbit polyclonal anti-FLAG (SIGMA-ALDRICH, F7425).

Chromosome spreads

Cells treated with Colcemid (100 μg/ml) were harvested by mitotic shake off and hypotonically swollen in 75 mM KCl for 20min at room temperature. Cells were fixed with Carnoy's solution (methanol:acetic acid=3:1), dropped on glass slides and dried. Slides were stained with 5% Giemsa (Merck, Tokyo, Japan), and washed with water, air-dried, mounted with Entellan (Merck, Tokyo, Japan).

Gene expression and cell proliferation assays

A full length *RAD21* cDNA (BC050381) was kindly provided by Dr. Sumio Sugano and a full length *STAG2* cDNA was obtained from total cDNA derived from bone marrow cells and cloned into pBluescript. The truncated mutant of *RAD21* was subcloned by PCR. FLAG-tagged *RAD21* or *STAG2* cDNAs were constructed into the retrovirus vector (pGCDNsamIRESEGFP) or tetracycline-inducible lentiviral vector (CS-TRE-Ubc-tTA-IRESPuro). The wild-type *RAD21*, mutant truncated *RAD21*

(*RAD21* K558X or E419X), or a mock vector plasmids were co-transfected with a VSV-G expression construct into 293GP cells to recover retrovirus-containing supernatant, which was then used to transduce 293GPG cells to establish stable cell lines. 293GP or 293GPG derived retrovirus was then transduced to Kasumi-1, K562, and TF1 cells, which were sorted by GFP-marking using a MoFlo® FACS cell sorter (Beckman Coulter) or BD FACSAria cell sorter (BD Biosciences) 48-96 hrs after retroviral transduction, and subjected to subsequent analysis. The wild-type *RAD21*, *STAG2* or mock-induced lentiviral vectors were generated as described previously (PNAS 103, 15933-938) and transduced to MOLM-13 cells and selected by 1μg/ml puromycin. Gene expression was introduced by 1μg/ml Doxcycline.

For cell growth assays, the cells were inoculated into 96 well culture plates and the cell growth was monitored in three independent experiments by MTT assay using Cell Counting Kit-8® (Dojindo Co.). For Quantitative RT-PCR analysis, 1ug of total RNA was subjected to reverse transcriptase reaction using ReverTra Ace qPCR RT Kit® (TOYOBO, Tokyo) and cDNA was amplified with Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies) using LightCycler 480 Real-time PCR System. Primer sequences were: Total (wild-type and mutant) *RAD21* forward, 5'-TGAAAATGGCATTACGGACA-3', reverse, 5'-GCAGTGGCTGATCAAAGTCA-3'; Full-length *RAD21* forward, 5'-ATGGTCTTCAGCGTGCTCTT-3', reverse, 5'-ACGGTTCTTCCTGTGTCAGC-3'; STAG2 forward, 5'-GGAACTTTTTACCAGTCGGTTC-3', reverse, 5'-TCCCACATGCTATCCACAAG-3'; and *GAPDH* forward, 5'-ACTGGCATGGCCTTCCGTGT-3', reverse, 5'-ATGCCAGCCCCAGCGTCAAA-3'.

Expression microarray analysis

RNA was extracted from Kasumi-1 cells retrovirally transduced with either mock or wild-type RAD21 and analyzed in triplicate using Human Genome U133 Plus 2.0 Array (Affymetrix) according to the manufacture's protocol. For data analysis, raw array signals were first extracted from .CEL files using dChip Software (www.dchip.org). After background correction and normalization across the 6 array data sets, the standardized signal value was obtained for each probe set on each of triplicate array experiments, which were compared between mock- and wild-type RAD21- transduced cells. Two independent gene experiments of transfection and subsequent microarray analysis were performed. To identify transcriptionally altered genes, we used the following criteria that satisfied both fold change > ±1.2 and P <0.05 (two-tailed paired t-test) in two independent experiments.

RNA sequencing

RNA sequencing of *RAD21*-transduced Kasumi-1 cells and subsequent data analysis were performed as previously described[23] with minor modifications. For quantifications of expression values from RNA sequencing data, slightly modified version of RKPM (reads per kilobase of exon per million mapped reads) measures were used.[43] After removing those sequencing reads inappropriately aligned or having low mapping quality (mapping quality below 60), the number of bases on each exonic regions for each RefSeq genes[49] were counted. Then, the number of bases was normalized as per kb of exon and per 100 million of aligned bases. Finally, expression value of each gene is determined by taking the maximum values among RefSeq genes

corresponding to the gene symbol.

RAD21 expression was measured by differentially enumerating endogenous and exogenous *RAD21* sequence reads, which were discriminated by the absence and presence of the FLAG sequence. After normalized by the number of total reads for each sample, the raw differential read counts were further calibrated against the read counts containing the stop codon in *RAD21*.

Statistical analysis

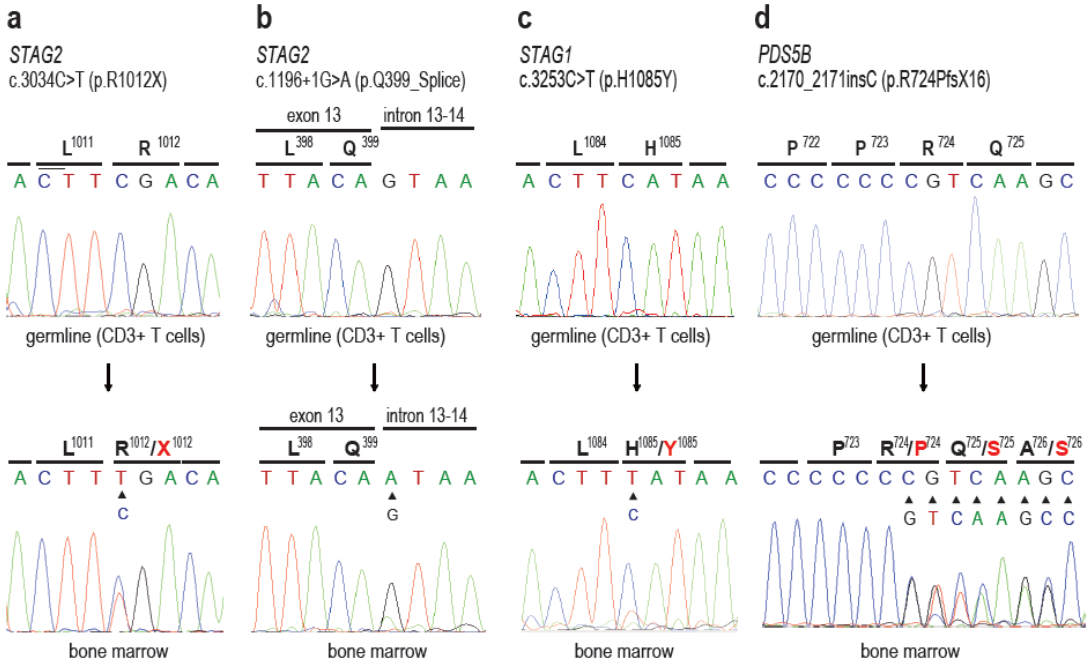
Significance of the difference in frequency of cohesin mutations between disease subtypes was tested by one-tailed Fisher's exact test. Coexistence of mutations was tested by two-tailed Fisher's direct method. The significance of the difference in the total number of somatic mutations between cohesin-mutated/deleted and non-mutated/deleted samples was tested by Mann-Whitney's *U*-test. Difference in the number of numerical abnormalities in cytogenetics between two groups with and without cohesin mutations/deletions was assessed by one-sided chi-square test.

Results

Whole exome sequencing revealed recurrent mutations in cohesin complex

In this study, the list of gene mutations of whole exome sequencing of 29 paired tumor and normal samples of myeloid neoplasms with myelodysplastic features[23] was closely investigated, where hundreds of unreported gene mutations were identified in addition to the frequent spliceosome mutations uniquely associated with myelodysplasia phenotypes. In that study, all the candidate nonsynonymous single nucleotide variations (SNVs) and small insertions/deletions (indels) (N = 509) detected by whole exome sequencing were validated by Sanger sequencing, and 248 somatic mutations (7.8 per sample), including 191 missense and 24 nonsense, 9 splice site mutations, and 24 frameshift-causing indels, were confirmed. In the present study, 60 out of the remaining 62 mutations were additionally validated, of which 14 and 46 SNVs/indels were validated as true positive and negative mutations, respectively. After all, 282 and 213 out of 497 candidate SNVs/indels were determined to be true positives and negatives, respectively. While most of them likely represented passenger changes affecting just single cases, additional recurrent mutations involving *STAG2*, a core component of the cohesin complex and two other functionally related cohesin components, *STAG1* and *PDS5B* were identified in single cases (Figure 1).

Figure 1. Mutations of *STAG1*, *STAG2* and *PDS5B* found in the discovery samples

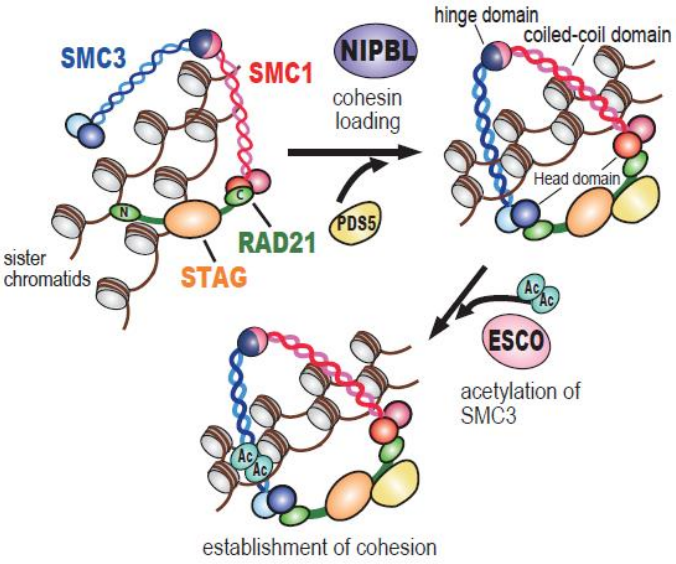


Somatic mutations of cohesin components identified by whole exome sequencing were validated by Sanger sequencing in bone marrow samples from cases MDS-19 (*STAG2*) (**a**), MDS-12 (*STAG2*) (**b**), MDS-11 (*STAG1*) (**c**) and tAML-02 (*PDS5B*) (**d**), using CD3+ T cells as germline control.

Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms

Cohesin is a multimeric protein complex conserved across species and composed of four core subunits, i.e., SMC1, SMC3, RAD21, and STAG proteins, together with a number of regulatory molecules, such as PDS5, NIPBL, and ESCO1/2 (Figure. 2).[44,45] Forming a ring-like structure, cohesin is engaged in cohesion of sister chromatids during cell division[45], post-replicative DNA repair,[46,47] and regulation of global gene expression through long-range *cis*-interactions.[42,48-51] Germline mutations in cohesin components lead to congenital multisystem malformation syndromes known as Cornelia de Lange syndrome and Roberts syndrome.[41,52,53]

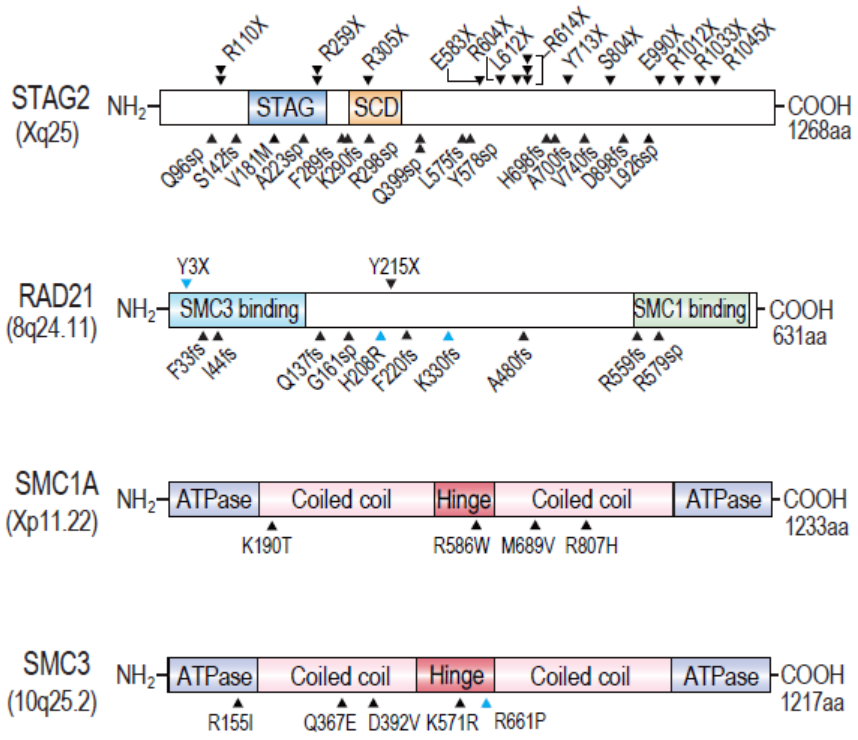
Figure 2. Multiple components of the cohesin complex



Cohesin holds chromatin strands within its ring-like structure composed of four core components, STAG, RAD21, SMC1, and SMC3. NIPBL promotes cohesin to be loaded on chromatin, while ESCO proteins stabilize the chromatin holding by acetylating N-terminal lysine residues on SMC3.

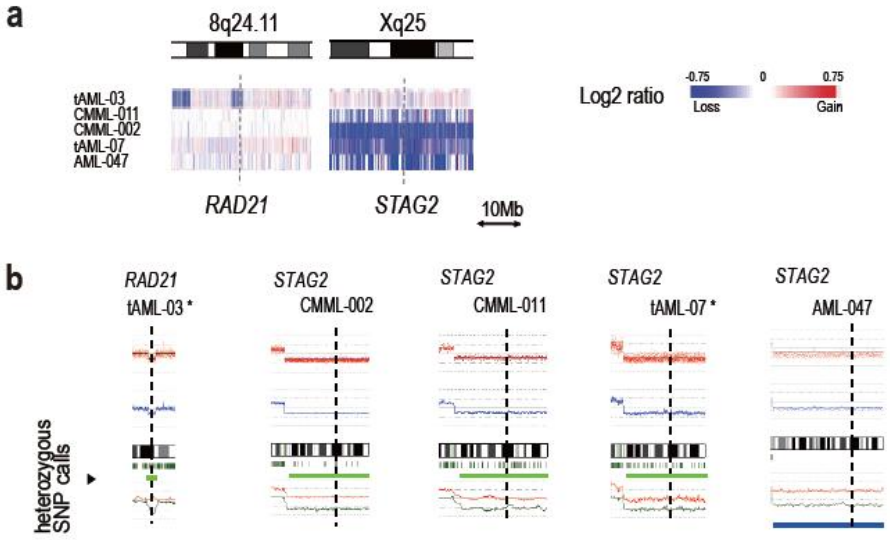
To investigate a possible role of cohesin mutations in myeloid leukemogenesis, an additional 581 primary specimens of various myeloid neoplasms was examined for mutations in nine cohesin genes implicated in mitosis[45] using high-throughput sequencing. Copy number alterations in cohesin loci were also investigated in 453 cases using SNP arrays. After excluding known/putative polymorphisms registered in the dbSNP database or the 1000 Genomes project or predicted from multiple computational imputations, a total of 60 non-synonymous mutations involving nine genes were identified in 610 primary samples and validated by Sanger sequencing (Figure 3 and Table 1). After conservative evaluation of the probability of random mutational events across these genes ($P<0.001$), only four genes were considered significantly mutated in the current cohort, including *STAG2*, *RAD21*, *SMC1A,* and *SMC3* (Table 2). Five deletions were also detected in *STAG2* (N=4) and *RAD21* (N=1) (Figure 4a, b). Mutations in these four genes were also found in four of the 34 myeloid leukemia cell lines (12%) (Table 3).

Figure 3. Mutations in the components of the cohesin complex



Mutations in the core components of the cohesin complex found in myeloid malignancies and myeloid leukemia cell lines are indicated by black and blue arrowheads, respectively.

Figure 4. Gene deletions of the cohesin and regulatory components



a. Gene deletions involving the *STAG2, and RAD21*, which are indicated by dotted lines. Copy numbers are indicated by color gradients as indicated. **b.** The somatic origin of the deletion was confirmed in 2 cases, in which germline control is available (asterisks). In 4 cases, the retention of substantial numbers of heterozygous SNP calls within the deletions (green bars) indicated residual normal cells that did not have LOH (i.e. deletions), supporting their somatic origin. In the remaining 1 case, there was almost complete loss of heterozygous SNP calls within the deletion (blue bar). However, the unusually large sizes (spanned >35Mbp) of these deletions are unusual for germline events, suggesting their somatic origin.

Table 1. Cohesin mutations identified in 610 cases with myeloid neoplasms

	Gene	Amino Acid Change	Allele Change	Sample Name[#]	WHO Classification	Gender	Karyotype
	STAG2	p.Q96_splice	c.288+1G>C	MDS-180	RAEB-2	F	46,XX,t(11;19)(q23;p13)
	STAG2	p.R110X	c.328C>T	CML-BC28	CML BC	M	46,XY,t(9;22)(q34;q11)
	STAG2	p.R110X	c.328C>T	CMML-055	CMML-2	M	ND
*	*STAG2*	p.S142fs	c.425_426insT	AML-154	AML M1	F	46,XX
	STAG2	p.V181M	c.541G>A	MPN-033	PMF	F	ND
	STAG2	p.A223_splice	c.668-1G>A	CMML-058	CMML-2	M	46,XY
*	*STAG2*	p.R259X	c.775C>T	AML-158	AML M2	M	46,XY,t(11;19)(q23;p13.11)
	STAG2	p.R259X	c.775C>T	CMML-064	CMML-2	F	47,XX,+8
	STAG2	p.F289fs	c.866_867delTinsAA	MDS-155	RAEB-1	M	47,XY,+8
*	*STAG2*	p.K290fs	c.870_871insG	MDS-230	RAEB-2	M	46,XY
*	*STAG2*	p.R298_splice	c.894-1G>A	AML-166	AML M6	F	46,XX,t(2;6)(q31;q25)
	STAG2	p.R305X	c.913C>T	CML-BC27	CML BC	M	51,XY,+8,+9,t(9;22)(q34;q11),+12,+19,+der(22)t(9;22)(q34;q11)
*	*STAG2*	p.Q399_splice	c.1196+1G>A	MDS-12	RCMD	F	46,XX
	STAG2	p.Q399_splice	c.1196+2T>G	MDS-139	RAEB-1	M	46,XY
*	*STAG2*	p.L575fs	c.1724_1725insA	MDS-208	RAEB-2	M	47,XY,+8
*	*STAG2*	p.Y578_splice	c.1732-2A>G	AML-178	AML M6	M	46,XY
*	*STAG2*	p.E583X	c.1747G>T	AML-177	AML M2	F	46,XX
	STAG2	p.R604X	c.1810C>T	MDS-176	RAEB-2	M	46,XY
	STAG2	p.L612X	c.1835T>A	MDS-091	RCMD	M	ND; 46,XY (at AML)
*	*STAG2*	p.R614X	c.1840C>T	MDS-234	RCMD	M	46,XY
	STAG2	p.R614X	c.1840C>T	AML-001	AML M0	M	47,XY,+8,t(9;22)(q34;q11)
	STAG2	p.R614X	c.1840C>T	CMML-036	CMML-1	F	ND
	STAG2	p.H698fs	c.2094_2095insAA	CMML-013	CMML-1	F	46,XX
*	*STAG2*	p.A700fs	c.2098_2099insC	MDS-240	RA	M	46,XY
	STAG2	p.Y713X	c.2139C>G	MDS-176	RAEB-2	M	46,XY
*	*STAG2*	p.V740fs	c.2218_2219insTAAT	AML-181	AML M6	F	46,XX
	STAG2	p.S804X	c.2411C>G	CMML-018	CMML-1	F	46,XX
	STAG2	p.D898fs	c.2694_2695insTC	CMML-070	CMML-2	M	46,XY,del(12)(p11)
	STAG2	p.L926_splice	c.2776-1G>T	MDS-096	RCMD	F	ND; 47,XX,-7,+2mar (at AML)
*	*STAG2*	p.E990X	c.2968G>T	AML-158	AML M2	M	46,XY,t(11;19)(q23;p13.11)
*	*STAG2*	p.R1012X	c.3034C>T	MDS-19	RAEB-2	F	46,XX
*	*STAG2*	p.R1033X	c.3097C>T	MDS-237	RA	M	46,XY,+1,der(1;7)(q10;p10),del(20)(q11)
	STAG2	p.R1045X	c.3133C>T	tAML-021	AML/MRC	M	46,XY
	RAD21	p.F33fs	c.99_100delCG	AML-109	AML M4	M	46,XY
	RAD21	p.I44fs	c.130_131insT	MDS-212	RAEB-2	M	47,XY,+9
*	*RAD21*	p.Q137fs	c.411_412insA	AML-190	AML M2	M	ND
	RAD21	p.G161_splice	c.482-1G>A	CML-AP1	CML AP	M	46,XY,t(9;22)(q34;q11)
	RAD21	p.Y215X	c.645T>A	AML-125	AML M5	M	46,XY
	RAD21	p.F220fs	c.660_661insGTAAGAAT	AML-033	AML M2	M	ND
	RAD21	p.A480fs	c.1438_1439insA	AML-011	AML M1	M	46,XY
*	*RAD21*	p.R559fs	c.1676_1677insA	MDS-243	RARS	M	46,XY
*	*RAD21*	p.R579_splice	c.1705-1G>A	AML-054	AML M2	F	46,XX,t(2;8)(q37;q22)
	SMC1A	p.K190T	c.569A>C	CML-BC28	CML BC	M	46,XY,t(9;22)(q34;q11)
	SMC1A	p.R586W	c.1756C>T	AML-049	AML M2	M	46,XY,t(8;21)(q22;q22)
	SMC1A	p.M689V	c.2065A>G	CML-CP30	CML CP	F	46,XX,t(9;22)(q34;q11)
	SMC1A	p.R807H	c.2420G>A	AML-007	AML M1	M	46,XY
	SMC3	p.R155I	c.464G>T	MDS-210	RAEB-2	M	46,XY
	SMC3	p.Q367E	c.1099C>G	MDS-060	RCMD-RS	M	46,XY
	SMC3	p.D392V	c.1175A>T	AML-039	AML M2	F	ND
	SMC3	p.K571R	c.1712A>G	MDS-191	RAEB-2	M	42,XY,del(7)(q34),-15,-16,-18,-21,-22,+mar
	STAG1	p.S741C	c.2222C>G	CMML-078	CMML-2	F	ND
*	*STAG1*	p.H1085Y	c.3253C>T	MDS-11	RCMD	M	46,XY
	PDS5B	p.G540A	c.1619G>C	CMML-052	CMML-1	M	46,XY
	PDS5B	p.S568C	c.1702A>T	MPN-061	PMF	F	46,XY,del(13)(q13q32)
*	*PDS5B*	p.R724fs	c.2170_2171insC	tAML-02	AML/MRC	F	47,XX,+1,der(1;7)(q10;p10),+8
	NIPBL	p.P1161S	c.3481C>T	CML-BC2	CML BC	M	46,XY,t(9;22)(q34;q12)
	NIPBL	p.M1485fs	c.4453_4454insT	AML-027	AML M2	M	45,X,-Y
	NIPBL	p.D1961N	c.5881G>A	MDS-141	RAEB-1	F	45,XX,del(5)(q23q32),der(11;12)(q10;q10),del(20)(q11)
	ESCO1	p.R779H	c.2336G>A	CMML-019	CMML-1	M	46,XY
	ESCO2	p.C263F	c.788G>T	MDS-129	RAEB-1	M	47,XY,+8

* Confirmed somatic mutations; # Sample names correspond to those defined in ref 3 except for newly analyzed cases

Abbreviations: RA, refractory anemia; RARS, refractory anemia with ringed sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RCMD-RS, refractory cytopenia with multilineage dysplasia and ringed sideroblasts; RAEB-1, refractory anemia with excess blasts-1; RAEB-2, refractory anemia with excess blasts-2; AML/MRC, acute myeloid leukemia with myelodysplasia related changes; CMML-1, chronic myelomonocytic leukemia-1; CMML-2, chronic myelomonocytic leukemia-2; CML-CP, CML in chronic phase; CML-AP, CML in accelerated phase; CML-BC, CML in blastic crisis; MPN, myeloproliferative neoplasm; PMF, primary myelofibrosis; ND,

Table 2. Significance of mutations in cohesin components

Gene	Target length (bp)	# of mutation in primary samples	P-value
STAG2	3935	33	2.39E-44
RAD21	1944	9	6.48E-12
SMC3	3766	4	7.33E-04
SMC1A	3798	4	7.61E-04
PDS5B	4476	3	7.49E-03
STAG1	3905	2	2.94E-02
NIPBL	8595	3	5.97E-02
ESCO2	1842	1	7.02E-02
ESCO1	2555	1	7.02E-02

Significance of non-silent mutations in each cohesin component was evaluated assuming the uniform distribution of the background mutations within the coding regions (~ 0.3/Mb).

Table 3. Cohesin mutations in myeloid leukemia derived cell lines

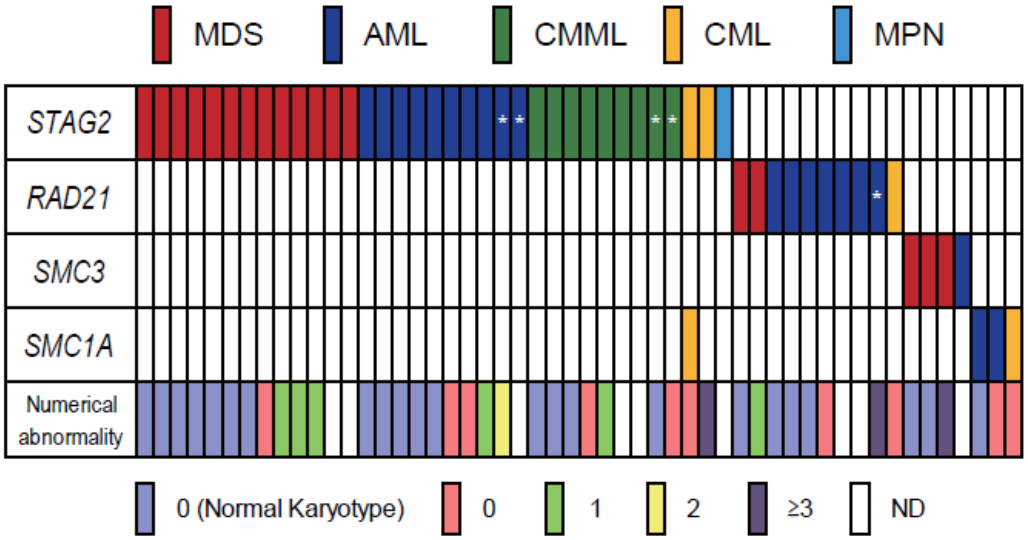
Cell line	Original tumor	Mutated gene	Mutation type	Allele change	Amino acid change
Mono-7	AML M0	–	–	–	–
Kasumi-3	AML M0	–	–	–	–
CTS	AML M1	–	–	–	–
YNH-1	AML M1	–	–	–	–
Kasumi-1	AML M2	*RAD21*	Frameshift	c.987insCCGG	p.K330PfsX6
KY821	AML M2	–	–	–	–
HL60	AML M2	–	–	–	–
NB4	AML M3	–	–	–	–
IMS-M1	AML M5a	–	–	–	–
THP-1	AML M5	–	–	–	–
P31FUJ	AML M5	*RAD21*	Missense	c.623A>G	p.H208R
CTV-1	AML M5	–	–	–	–
TF-1	AML M6	–	–	–	–
UT-7	AML M7	–	–	–	–
CMS	AML M7	–	–	–	–
INV-3	AML unspecified	–	–	–	–
ML-1	AML unspecified	–	–	–	–
P39TSU	sAML	–	–	–	–
MOLM-13	sAML	–	–	–	–
F-36P	sAML	–	–	–	–
GDM-1	sAML	–	–	–	–
HEL	sAML	–	–	–	–
CMK	DS-derived AMKL	–	–	–	–
CMY	DS-derived AMKL	*RAD21*	Nonsense	c.9C>G	p.Y3X
CMK-86	DS-derived AMKL	–	–	–	–
UTP-DSAL-1	DS-derived AMKL	–	–	–	–
KG-1	AML	–	–	–	–
EoL-3	Eosinophilic leukemia	–	–	–	–
SS9;22	CML	–	–	–	–
K-562	CML	–	–	–	–
MOLM-7	CML (BC)	*SMC3*	Missense	c.1982G>C	p.R661P
MOLM-1	CML (BC)	–	–	–	–
TS9;22	CML (BC)	–	–	–	–
W1	JMML	–	–	–	–

Abbreviations: CML, chronic myelogenous leukemia; BC, blastic crisis; JMML, juvenile myelomonocytic leukemia; DS-derived AMKL, Down syndrome-derived acute megakaryocytic leukemia; sAML, secondary AML.

– : No cohesin mutations detected

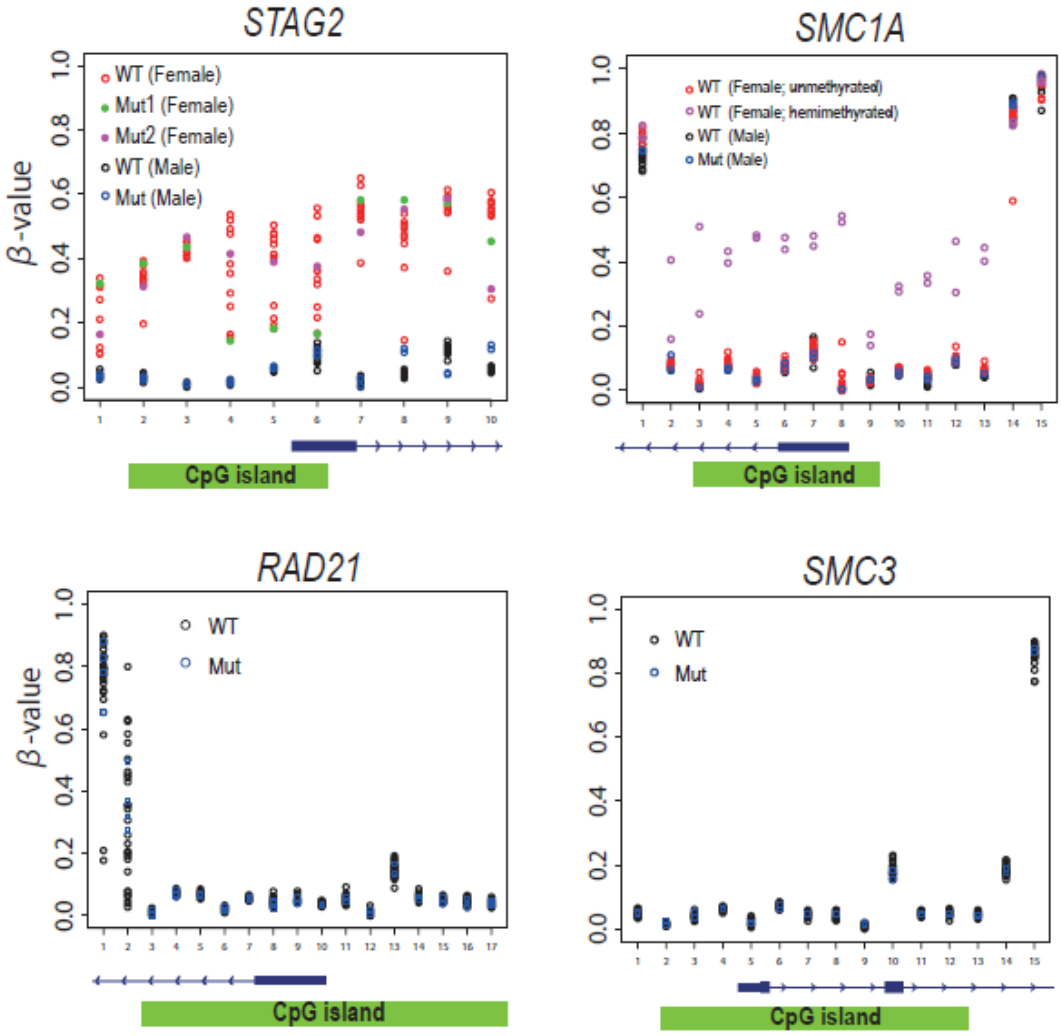
Mutations and deletions of these four genes were found in a mutually exclusive manner in a variety of myeloid neoplasms, including AML (19/157), CMML (9/88), MDS (18/224) and CML (4/64), and with a much lower mutation frequency in classical MPN (1/77) (Figure 5 and Table 4). In MDS, mutations were more frequent in refractory cytopenia with multilineage dysplasia (RCMD) and refractory anemia with excess blasts (RAEB) (11.4%), but rare in refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), RCMD and ringed sideroblasts (RCMD-RS), and 5q- syndrome (4.2%) (p=0.044). Promoter methylation was also evaluated in 33 cases either with (N=12) or without (N=21) cohesin mutations/deletions, using HumanMethylation450 BeadChip; however, no aberrant methylations in cohesin loci were found, except for hemimethylation of the *SMC1A* promoter in two female cases (Figure 6).

Figure 5. Distribution of cohesin mutations/deletions



Distribution of cohesin mutations/deletions shows a nearly mutually exclusive pattern among the different myeloid neoplasms. Gene deletions are indicated by asterisks.

Figure 6. Methylation status of promoter regions of 4 cohesin genes



Methylation of multiple sites within and around the promoter regions of 4 cohesin genes were measured using Human Methylation 450 BeadChip (Illumina) and plotted as β-values, where fully methylated, hemi-methylated and non-methylated sites are expected to show values close to 1.0, 0.5 and 0.0, respectively. The positions of CpG islands are indicated in green bars below each panel. In total, 33 samples with (N=12) or without (N=21) mutations of cohesin genes were analyzed. The samples having mutations of corresponding genes (Mut) are discriminated from the wild-type samples (Wt). For STAG2 and SMC1A, which are on X chromosome, male san female samples are shown in different color.

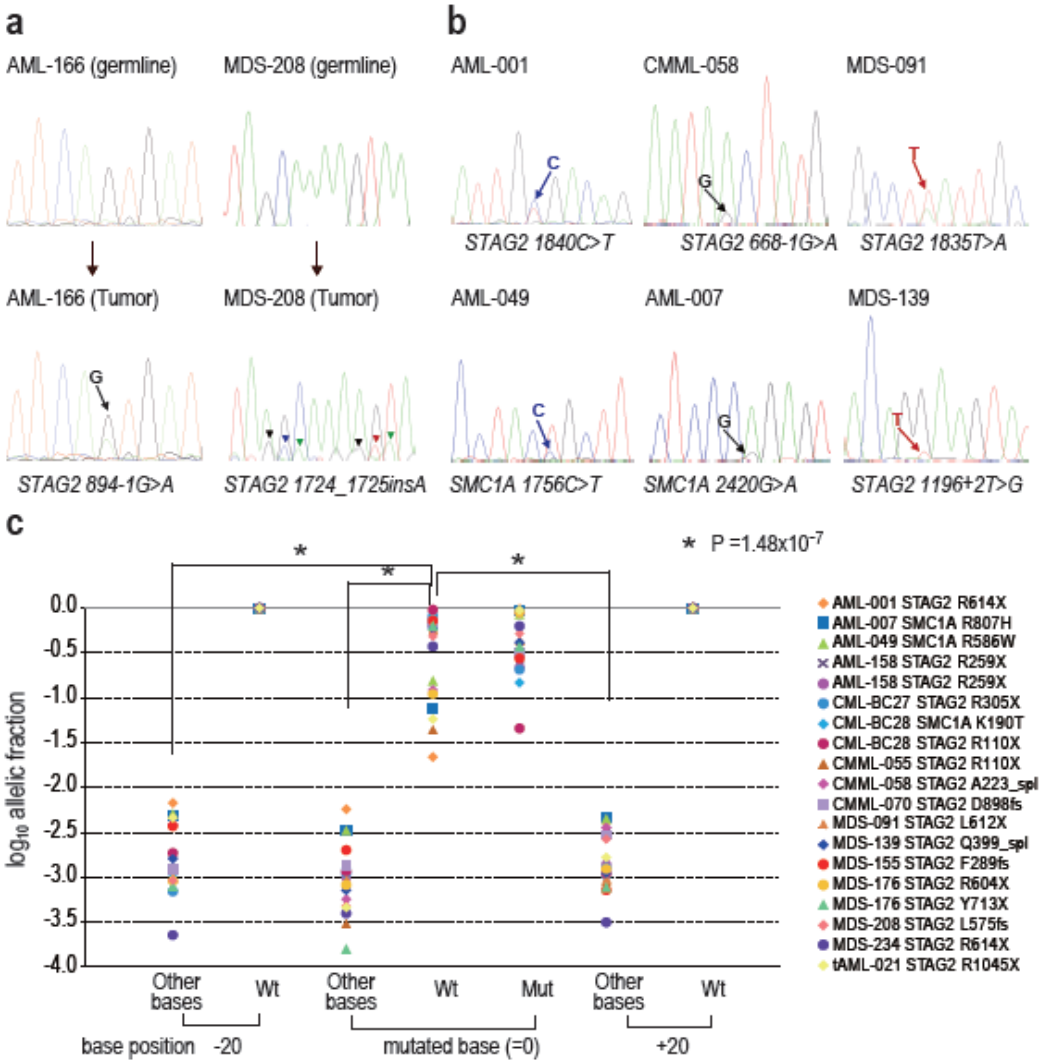
Table 4. Frequencies of mutations and deletions of cohesin components in 610 cases of myeloid neoplasms

Disease type	N	*STAG2*	*RAD21*	*SMC1A*	*SMC3*	Total	%
MDS	**224**	**13**	**2**	**0**	**3**	**18**	**8.0%**
RA	8	2	0	0	0	2	25.0%
RARS	28	0	1	0	0	1	3.6%
RCMD	34	4	0	0	0	4	11.8%
RCMD-RS	52	0	0	0	1	1	1.9%
RAEB–1	37	2	0	0	0	2	5.4%
RAEB–2	52	5	1	0	2	8	15.4%
5q- syndrome	8	0	0	0	0	0	0%
MDS-U	2	0	0	0	0	0	0%
Unknown	3	0	0	0	0	0	0%
MDS/MPN	**88**	**9 (2)**	**0**	**0**	**0**	**9(2)**	**10.2% (2.3%)**
CMML-1	53	5 (2)	0	0	0	5(2)	9.4% (3.8%)
CMML-2	35	4	0	0	0	4	11.4%
MPN	**77**	**1**	**0**	**0**	**0**	**1**	**1.3%**
PV	28	0	0	0	0	0	0%
ET	23	0	0	0	0	0	0%
PMF	26	1	0	0	0	1	3.8%
CML	**64**	**2**[#]	**1**	**2**[#]	**0**	**4**	**6.3%**
CML-CP	32	0	0	1	0	1	3.1%
CML-AP	10	0	1	0	0	1	10.0%
CML-BC	22	2[#]	0	1[#]	0	2	9.1%
AML	**157**	**10 (2)**	**7 (1)**	**2**	**1**	**19 (3)**	**12.1% (1.9%)**
de novo AML	120	8 (1)	6	2	1	16 (1)	13.3% (0.8%)
FAB M0	3	1	0	0	0	1	33.3%
FAB M1	15	1	1	1	0	3	20.0%
FAB M2	42	3 (1)	3	1	1	8 (1)	19.0% (2.4%)
FAB M3	17	0	0	0	0	0	0%
FAB M4	15	0	1	0	0	1	6.7%
FAB M5	13	0	1	0	0	1	7.7%
FAB M6	10	3	0	0	0	3	30.0%
FAB M7	2	0	0	0	0	0	0%
Unknown	3	0	0	0	0	0	0%
AML/MRC	37	2 (1)	1 (1)	0	0	3 (2)	8.1% (5.4%)
Total	**610**	**35 (4)**[#]	**10 (1)**	**4**[#]	**4**	**52 (5)**	**8.5% (0.8%)**

Diseases were classified according to World Health Organization 2008 classification. Number and percentage of cases with cohesin gene deletions are shown in parentheses. MDS, myelodysplastic syndromes; MDS/MPN, myelodysplastic-myeloproliferative neoplasms; MPN, myeloproliferative neoplasm; CML, chronic myelogenous leukemia; RA, refractory anemia; RARS, refractory anemia with ringed sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RCMD-RS, refractory cytopenia with multilineage dysplasia and ringed sideroblasts; RAEB-1, refractory anemia with excess blasts-1; RAEB-2, refractory anemia with excess blasts-2; MDS-U, MDS unclassified; CMML-1, chronic myelomonocytic leukemia-1; CMML-2, chronic myelomonocytic leukemia-2; PV, polycythemia vera; ET, essential thrombocythemia; PMF, primary myelofibrosis; CML-CP, CML in chronic phase; CML-AP, CML in accelerated phase; CML-BC, CML in blastic crisis; AML/MRC, AML with myelodysplasia-related changes; #; One CML case having both *STAG2* and *SMC1A* mutations are counted twice and counted as a single case in total.

A somatic origin was confirmed for 17 mutations in 16 cases for which matched normal DNA was available (Table 1). Somatic origins for 23 mutations in *STAG2* or *SMC1A* in 20 male cases were supported by the presence of reproducible wild-type signals/reads in Sanger and/or deep sequencing of tumor samples, which were considered to originate from the X chromosome of the residual normal cells (Figure 7). In addition, for 13 mutations, the observed allele frequencies determined by pyrosequencing, deep sequencing, and/or digital PCR showed significant deviations from the expected value for polymorphisms in the absence of apparent chromosomal alterations in SNP array analysis (Figure 8 and 9), suggesting their somatic origin. Finally, 32 of the 33 *STAG2* and all of the nine *RAD21* mutations were either nonsense (N=18), frameshift (N=14), or splice site (N=9) changes, which were predicted to cause premature truncation of the protein or abnormal exon skipping (Figure 10 and 11). Thus, the majority of the mutations were considered to represent functionally relevant changes, most likely of somatic origin (Table 5).

Figure 7. Residual wild-type *STAG2/SMC1A* alleles in male cases with *STAG2/SMC1A* mutations



Since *STAG2* and *SMC1A* are on X chromosome, the presence of the wild type allele within the tumor specimens in male cases should be interpreted as originated from the residual normal cell components and therefore, indicates the somatic origin of the mutations, except for a rare possibility of somatic chimera. This was actually the case with AML-166 and MDS-208 carrying a confirmed somatic mutation (a), in which the wild-type sequences were overlapped with mutated sequences in Sanger sequencing of the tumor specimen (arrow heads). Similarly, the wild-type signals were confirmed in the 11 male cases with an *STAG2* or *SMC1A* mutation, of which 6 mutations are illustrated (b), which are indicated by arrows. **c.** The presence of the wild-type allele was more sensitively and explicitly detected in deep sequencing, where the genomic region containing each mutation was PCR amplified and subjected to deep sequencing.

The number of reads having mutated (Mut) or wild-type (Wt) (or other (Other bases)) base calls at indicated base positions were enumerated and are plotted as logarithm of allelic fractions. The fraction of the wild-type alleles detected at the mutated positions (position 0) are significantly higher than the background error rates (Other bases) that were evaluated at the mutated position and the position 20 base-pair upstream and downstream from the mutated position (=0), respectively (Mann-Whitney's U test).

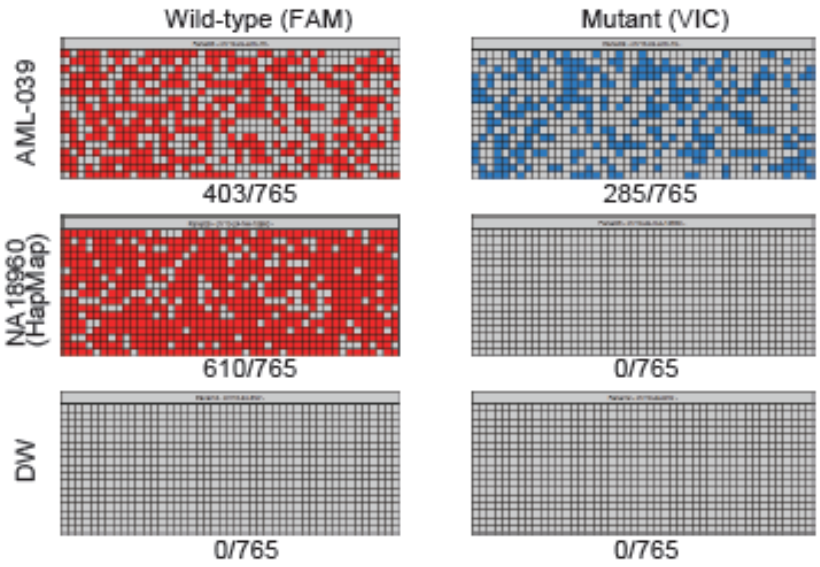
Figure 8. Measurements of frequencies of mutant alleles by digital PCR

a

Mutations	Sample	Alleles	Labelling	Forward (5'->3')	Reverse (5'->3')
SMC3 (1099C>G)	MD5-060	wild-type	FAM	TGGCTCAAGCTAC	TCCCATTTGCTTCTATTTTGTGTAGA
		mutant	VIC	TGGCTGAAGCTAC	CTGCTTTGCATAAAGATCCGTTCT
SMC3 (1175A>T)	AML-039	wild-type	FAM	AGAAGAAAGGGATAAGTG	GGGTCGAGGAAGCCAGTTTAC
		mutant	VIC	AGAAGAAAGGGTTAAGTG	GCAATCTGTCTTTTCTTGTCATTAATAGC

b

c



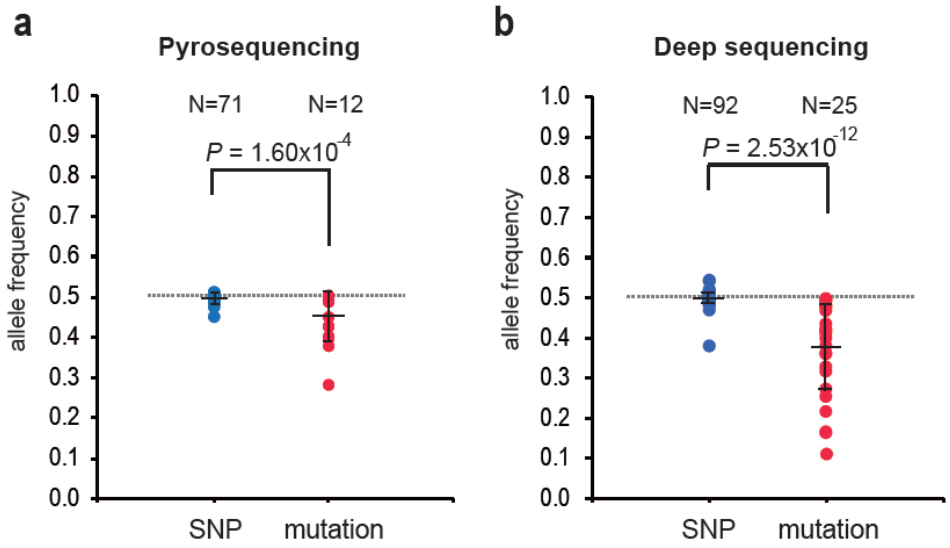
d

Mutations		WT			Mut			WT/Mut (95%CI)
		DW	NA18960	Tumor	DW	NA18960	Tumor	
SMC3 (1099C>G)	# Positive wells	0	616	4793	0	0	4489	1.100 (1.055-1.146)
	# Total wells	1530	765	9180	1530	765	9180	
SMC3 (1175A>T)	# Positive wells	0	610	788	0	0	588	1.492 (1.340-1.664)
	# Total wells	765	765	1530	765	765	1530	

In digital PCR on the Fluidigm system, the presence or absence of the wild-type (WT) and mutant (Mut) is determined by allele-specific PCR using differentially labeled (FAM- or VIC-) primers (**a**) in each of a large number of micro-wells, in which samples are properly diluted so that each well contains less than one molecule on average to allow for determination of the concentration of each allele according to the numbers of

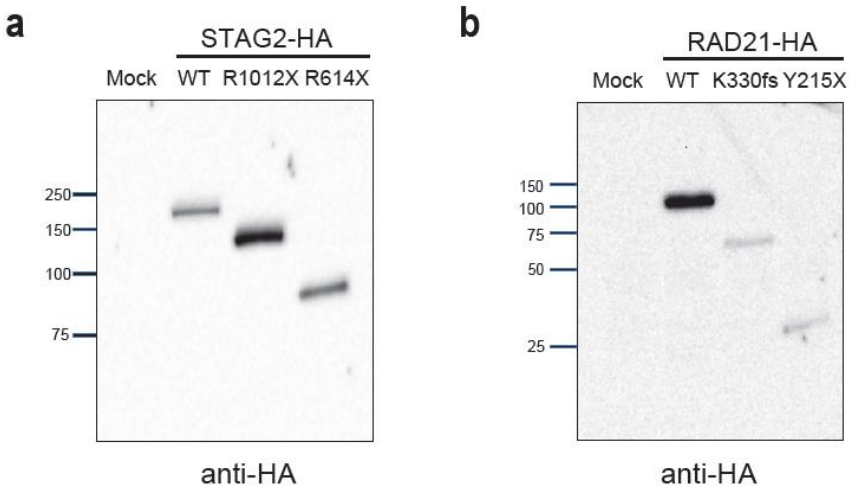
observed positive wells, assuming a Poisson's distribution. Representative assays are shown in **b** and **c**. Each PCR showed discrete distributions of signals between productive and non-productive (**b** and **c**, top and middle panels) and no false positive amplifications were observed in samples with no DNA (DW) for both wild-type and mutant-specific PCRs (**c**, bottom panels) as well in normal DNA (NA18960 from the HapMap panel) for mutant-specific PCRs (**c**, right middle panel). The results of measurements are summarized in **d**, where total number of positive and negative wells for each allele-specific PCR for control and test samples, together with observed WT/Mut allelic ratios with their 95% confidential intervals.

Figure 9. Measurements of frequencies of mutant alleles by pyrosequencing and deep sequencing



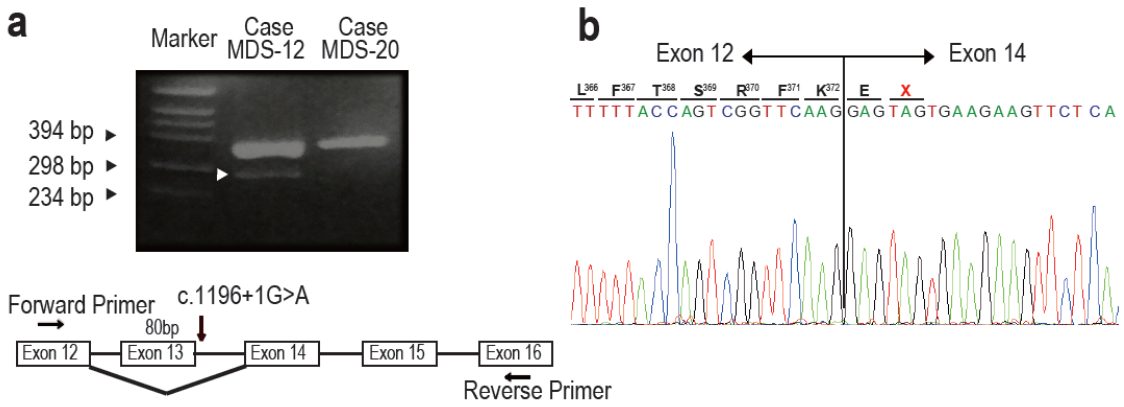
Frequencies of mutant alleles together with those of SNP alleles were measured by pyrosequencing (a) and deep sequencing (b). The mean ± S.D. value of measured allele frequencies are also indicated. All mutant and SNP loci did not show copy number alteration or copy neutral loss-of-heterozygosity in SNP array karyotyping. In both assays, the mutant alleles show significantly lower allele frequencies compared to known SNP alleles (Mann-Whitney's U test), supporting the hypothesis that many, if not all, of the mutations are not SNPs but represent somatic changes.

Figure 10. Expression of truncating *STAG2* and *RAD21* alleles in HeLa cells



Anti-HA immunoblots of total lysates from HeLa cells transduced with HA-tagged wild-type and truncating alleles of *STAG2* (**a**) and *RAD21* (**b**) showing truncated STAG2 and RAD21 proteins having expected molecular weights, respectively. Transduced truncating mutants are indicated on top of each panel.

Figure 11. Aberrant splicing of STAG2 transcript caused by a splice site mutation in MDS-12



a. RT PCR analysis of transcripts from MDS-12 shows an aberrant PCR product having a smaller fragment length (indicated by an while arrowhead) in addition to a major product, which is also present in a control case (MDS-20). The difference in intensity of both bands is thought to reflect the low tumor content in this sample, which was also indicated from the lower mutation peak in the genomic sequencing (Figure 1b, lower panel). The primer design is shown at the bottom. The position of the splice site mutation is shown by a vertical arrow. **b.** Sanger sequencing of the aberrant RT-PCR product confirmed the aberrant splicing that cause a skipping of exon 13, leading to a premature stop codon.

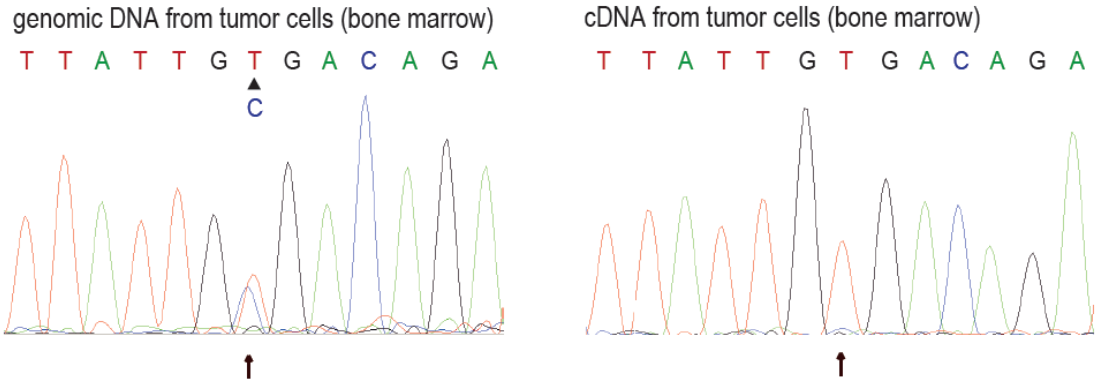
Table 5. Summary of characteristics of 50 cohesin mutations

Gene	Mutation	Case	Gender	Chr	Type of mutation	Tumor	Mutant allele frequency		
							DS	PS	DigPCR
STAG2	R110X	CML-BC28	M#	chrX	Nonsense	primary	0.358	–	–
STAG2	R110X	CMML-055	M#	chrX	Nonsense	primary	0.956	–	–
STAG2	R259X	AML-158	M#	chrX	Nonsense	primary	0.499	–	–
STAG2	R305X	CML-BC27	M#	chrX	Nonsense	primary	0.612	–	–
STAG2	R604X	MDS-176	M#	chrX	Nonsense	primary	0.580	–	–
STAG2	L612X	MDS-091	M#	chrX	Nonsense	primary	0.310	–	–
STAG2	R614X	AML-001	M#	chrX	Nonsense	primary	0.425	–	–
STAG2	R614X	MDS-234	M#	chrX	Nonsense	primary	0.934	–	–
STAG2	Y713X	MDS-176	M#	chrX	Nonsense	primary	0.170	–	–
STAG2	E990X	AML-158	M#	chrX	Nonsense	primary	0.392	–	–
STAG2	R1033X	MDS-237	M#	chrX	Nonsense	primary	-	–	–
STAG2	R1045X	tAML-021	M#	chrX	Nonsense	primary	0.931	–	–
STAG2	F289fs	MDS-155	M#	chrX	Frameshift	primary	0.913	–	–
STAG2	K290fs	MDS-230	M#	chrX	Frameshift	primary	0.776	–	–
STAG2	L575fs	MDS-208	M#	chrX	Frameshift	primary	0.834	–	–
STAG2	A700fs	MDS-240	M#	chrX	Frameshift	primary	-	–	–
STAG2	D898fs	CMML-070	M#	chrX	Frameshift	primary	0.975	–	–
STAG2	A223_spl	CMML-058	M#	chrX	Splice site	primary	0.874	–	–
STAG2	Q399_spl	MDS-139	M#	chrX	Splice site	primary	0.887	–	–
STAG2	Y578_spl	AML-178	M#	chrX	Splice site	primary	0.819	–	–
STAG2	R259X	CMML-064	F	chrX	Nonsense	primary	0.493	0.489	–
STAG2	E583X	AML-177	F	chrX	Nonsense	primary	0.440*	–	–
STAG2	R614X	CMML-036	F	chrX	Nonsense	primary	0.466	0.505	–
STAG2	S804X	CMML-018	F	chrX	Nonsense	primary	0.411*	0.400*	–
STAG2	R1012X	MDS-19	F	chrX	Nonsense	primary	0.417*	–	–
STAG2	S142fs	AML-154	F	chrX	Frameshift	primary	0.398*	–	–
STAG2	H698fs	CMML-013	F	chrX	Frameshift	primary	0.496	–	–
STAG2	V740fs	AML-181	F	chrX	Frameshift	primary	0.215*	–	–
STAG2	Q96_spl	MDS-180	F	chrX	Splice site	primary	0.359*	0.381*	–
STAG2	R298_spl	AML-166	F	chrX	Splice site	primary	0.252*	–	–
STAG2	Q399_spl	MDS-12	F	chrX	Splice site	primary	0.161*	–	–
STAG2	L926_spl	MDS-096	F	chrX	Splice site	primary	0.477	0.452*	–
STAG2	V181M	MPN-033	F	chrX	Missense	primary	–	–	–
RAD21	Y215X	AML-125	M	chr8	Nonsense	primary	0.433*	0.430*	–
RAD21	F33fs	AML-109	M	chr8	Frameshift	primary	0.315*	–	–
RAD21	I44fs	MDS-212	M	chr8	Frameshift	primary	0.164*	–	–
RAD21	Q137fs	AML-190	M	chr8	Frameshift	primary	0.319*	–	–
RAD21	F220fs	AML-033	M	chr8	Frameshift	primary	0.106*	–	–
RAD21	A480fs	AML-011	M	chr8	Frameshift	primary	0.270*	–	–
RAD21	R559fs	MDS-243	M	chr8	Frameshift	primary	-	–	–
RAD21	G161_spl	CML-AP1	M	chr8	Splice site	primary	0.358*	0.405*	–
RAD21	R579_spl	AML-054	F	chr8	Splice site	primary	0.326*	–	–
SMC1A	K190T	CML-BC28	M#	chrX	Missense	primary	0.258	–	–
SMC1A	R586W	AML-049	M#	chrX	Missense	primary	0.876	–	–
SMC1A	R807H	AML-007	M#	chrX	Missense	primary	0.810	–	–
SMC1A	M689V	CML-CP30	F	chrX	Missense	primary	0.492	0.429*	–
SMC3	R155I	MDS-210	M	chr10	Missense	primary	0.495	0.495	–
SMC3	Q367E	MDS-060	M	chr10	Missense	primary	0.420*	0.492	0.476*
SMC3	D392V	AML-039	F	chr10	Missense	primary	0.379*	0.286*	0.340*
SMC3	K571R	MDS-191	M	chr10	Missense	primary	0.481	0.449*	–

Somatic orign of the mutation are supported by one or more of the following observations in 48 mutations:
1) confirmed wild-type allele in a germline control (in blue),
2) presence of an additional wild-type signals in addition to the mutation signals in male cases for STAG2 or SMC1A (on X cheomosome) measured by sanger sequencing and/or deep sequencing (# in the Gendar columns),
3) significant deviation of candidate allele from 50%, which was determined by pyrosequencing (PS), digital PCR (DigPCR) and/or deep sequencing (DS) (* in the corresponding colums), and
4) nonsense (X), frameshift (fs), or splice site (spl) changes (in the Mutation column).

Most of the cohesin mutations/deletions were heterozygous, except for *STAG2* and *SMC1A* mutations on the single X chromosome in male cases (N=23). In female samples, the *STAG2* promoter was hemimethylated through X-inactivation regardless of the mutation status, and a heterozygous mutation of the unmethylated *STAG2* allele would lead to biallelic *STAG2* inactivation, as previously documented in a female case of Ewing's sarcoma[54] and also confirmed in a single case (CMML-036) in our cohort (Figure 12).

Figure 12. Expression of *STAG2* from the mutated allele in case CMML-036

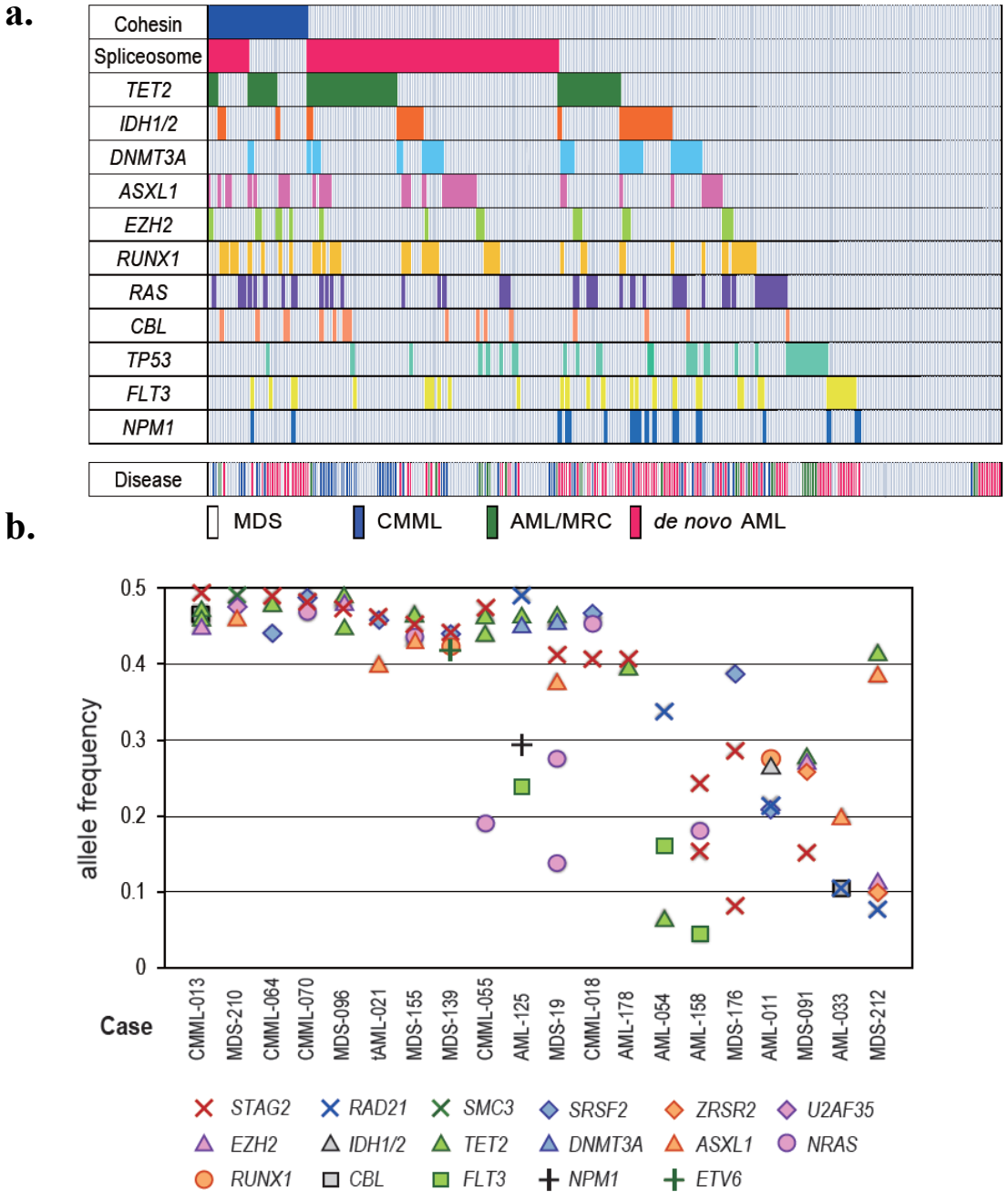


Results of Sanger sequencing of genomic DNA and cDNA from a female patient (CMML-036) carrying a heterozygous *STAG2* mutation (c.1840C>T). While genomic DNA from tumor cells shows comparable C (wild-type) and T (mutated) signals (left) at the nucleotide position 1840, the tumor-derived cDNA displays a T signal (right), indicating that the *STAG2* message was expressed exclusively from the mutated rather than the wild-type allele on the inactivated X chromosome.

Relationship between cohesin mutations and other common mutations in myeloid malignancies

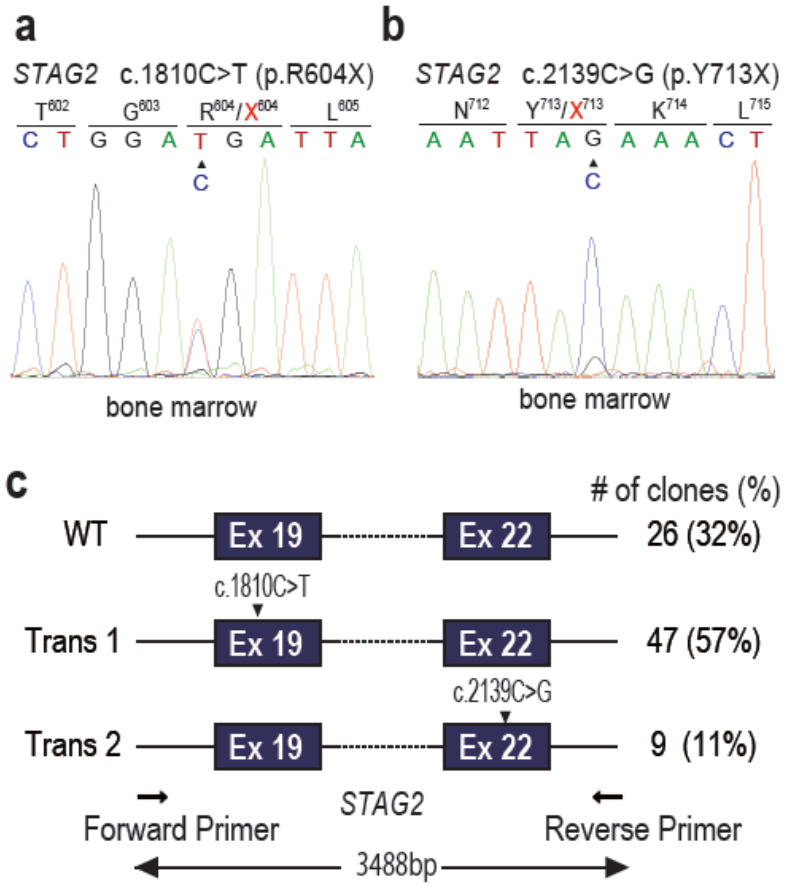
Cohesin mutations frequently coexisted with other common mutations in myeloid neoplasms and were significantly associated with mutations in *TET2* (p=0.027), *ASXL1* (p=0.045) and *EZH2* (p=0.011) (Fig. 13a).[23] Deep sequencing of these mutant alleles was performed in 20 cases of cohesin mutations, which allowed for accurate determination of their allele frequencies. The majority of the cohesin mutations (15/20) existed in the major tumor populations, indicating their early origin during leukemogenesis. In the remaining five cases, cohesin mutations were only found in a subpopulation of leukemic clones, indicating that the mutations were relatively late events (Fig. 13b). Two male cases (MDS-176 and AML-158) harbored two independent subclones with different *STAG2* mutations, indicating that these *STAG2* mutations could confer a strong advantage during clonal evolution (Figure 14). The number of mutations determined by whole exome sequencing[23] was significantly higher in four cohesin-mutated/deleted cases compared to non-mutated/deleted cases (p=0.049) (Figure 15).

Figure 13. Relationship between cohesin mutations and other common mutations in myeloid malignancies



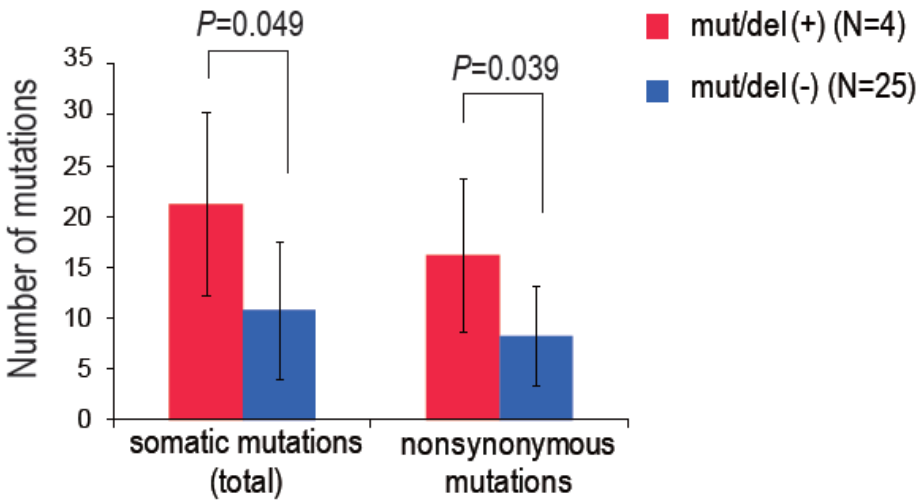
a. Mutations in cohesin and other common targets in 310 cases with different myeloid neoplasms are shown. Disease types are shown in the bottom lane. **b.** Allele frequencies of cohesin and coexisting mutations in 20 myeloid neoplasms were determined by deep sequencing.

Figure 14. Two independent *STAG2* mutations in a male case



Sanger sequencing showing two *STAG2* mutations, 1810C>T (a) and 2139C>G (b), found in a male case (MDS-176). The presence of the wild-type signals in both chromatogram indicates that these mutations were somatic. To determine the phase of these two mutations, a 3488 bp *STAG2* gene fragment was amplified, subcloned into plasmids and individually sequenced for the mutation sites (c). The distribution of mutations supports the model that two mutations are harbored in different subclones, one major (1810C>T) and the other minor (2139C>G) populations. This is consistent with the a smaller peak for 2139C>G (b) compared to 1810C>T. The presence of two independent STAG2 mutant alleles in a single case underscores the important driver role of these mutations.

Figure 15. Number of mutations in the discovery cohort

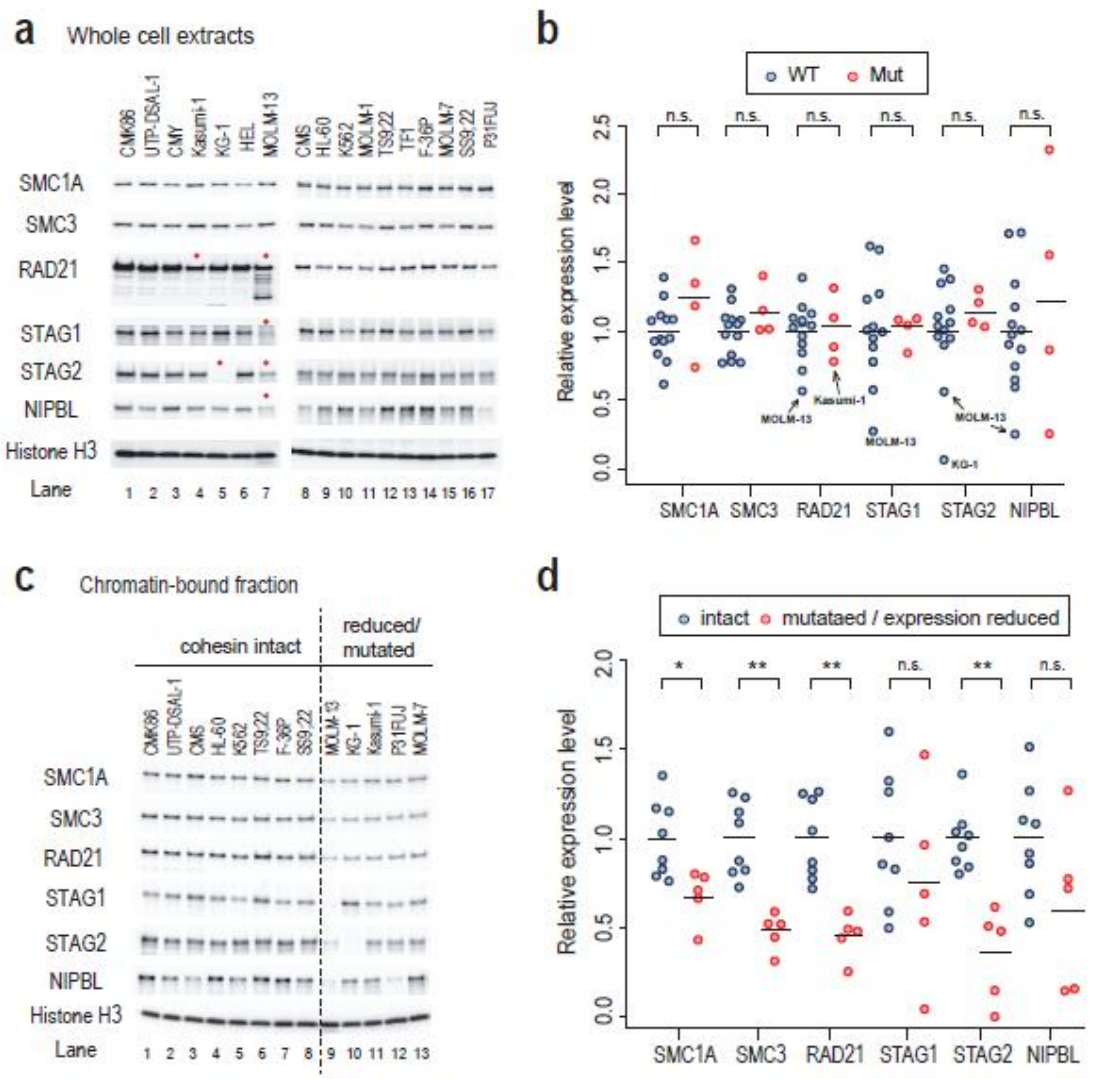


The impact of cohesin mutations on the number of coexisting somatic mutations was tested by Mann-Whitney U test based on the result of whole exome study. Cohesin-mutated/deleted cases (red) show significantly higher numbers of total somatic mutations and non-silent mutations as determined by whole exome sequencing, compared to those having no cohesin mutations/deletions (blue). Means ± SD are shown.

Reduced amounts of chromatin-bound cohesin components in cohesin-mutated leukemic cell lines

To investigate the possible impact of mutations on cohesin function, I first examined the expression of STAG1, STAG2, RAD21, SMC3, SMC1A, and NIPBL in 17 myeloid leukemia cell lines with (N=4) or without (N=13) known cohesin mutations, as well as in the chromatin-bound fractions of 13 cell lines (Figure 16a–d and Table 6).[41,55-57] Although a significant reduction in RAD21 was observed in Kasumi-1 harboring RAD21 p.K330PfsX6 (Figure 16a, lane 4), P31FUJ (RAD21 p.H208R), CMY (RAD21 p.Y3X), and MOLM-7 (SMC3 p.R661P) were not accompanied by significant decreases in the corresponding proteins compared with the wild-type cell lines. In contrast, severely reduced expression of one or more cohesin components was observed in KG1 (STAG2)[54] and MOLM-13 (STAG1/2, RAD21, and NIPBL) without any accompanying mutations in the relevant genes (Figure 16a, lanes 5 and 7). No significant differences in protein expression of the cohesin components in cohesin-mutated and non-mutated cell lines were noted in whole cell lysates (Figure 16b). However, one or more cohesin components, including SMC1, SMC3, RAD21, and STAG2, in the chromatin-bound fractions were substantially reduced in cell lines with mutated or reduced cohesin components, including Kasumi-1, KG1, P31FUJ, MOLM-7, and MOLM-13, compared with the cell lines with no known cohesin mutations or abnormal cohesin expression, suggesting a substantial loss of cohesin-bound sites on chromatin (Figure 16c, d and Table 6).[41]

Figure 16. Abnormal cohesin expression and chromatin binding of various cohesin components in myeloid leukemic cell lines



a. Western blotting of various cohesin components was performed on total lysates from 17 myeloid leukemia cell lines. Significantly reduced components are indicated by red dots. **b.** Differences in cohesin components between cohesin mutated and non-mutated cell lines. Normalized expression of each cohesin component, with histone H3 signal as a control, is plotted after standardization to the mean value across all non-mutated cell lines. The significantly reduced cohesin components in Kasumi-1, MOLM-13, and KG-1 are indicated. **c.** Western blotting of cohesin components were performed on the chromatin-bound fractions of 13 myeloid leukemia cell lines, including eight with intact (lanes 1–8) and five with abnormal (lanes 9–13) cohesin components. The

reduced chromatin-bound cohesin fractions were confirmed in two independent experiments (data not shown). **d.** Differences in chromatin-bound cohesin components between cohesin intact and abnormal (mutated and expression reduced) cell lines. For each cell line, the expression of each chromatin-bound cohesin component was normalized with the histone H3 signal and standardized to the mean value across all intact cell lines. *, $p < 0.05$; **, $p < 0.005$; n.s., not significant (Mann–Whitney U test). Horizontal bars indicate the mean values.

Table 6 Western blot quantification of each cohesin component of several myeloid cell lines

a. whole cell extracts

Lane	Cell line	SMC1A	SMC3	RAD21	STAG1	STAG2	NIPBL
1	CMK86	0.78	0.97	1.10	1.00	1.16	1.72
2	UTP-DSAL-1	1.12	1.23	1.39	1.59	1.35	1.35
3	CMY	0.74	1.01	1.10	1.05	1.21	2.32
4	Kasumi-1	1.19	1.41	0.78	1.08	1.06	1.56
5	KG-1	0.83	1.09	0.97	1.62	0.06	1.18
6	HEL	0.62	0.83	1.06	1.23	1.38	1.71
7	MOLM-13	0.94	1.05	0.57	0.27	0.56	0.74
8	CMS	0.93	1.08	1.04	0.95	0.90	0.25
9	HL-60	1.09	0.98	0.71	1.03	0.96	0.65
10	K562	0.95	0.78	0.84	0.57	1.06	0.87
11	MOLM-1	1.09	0.77	0.91	0.89	1.01	0.59
12	TS9;22	1.08	1.05	1.08	1.01	1.04	0.90
13	TF1	0.93	0.77	1.03	0.78	0.95	0.98
14	F36P	1.26	1.10	1.13	0.78	1.45	1.05
15	MOLM-7	1.35	1.02	1.32	1.09	1.31	0.86
16	SS9;22	1.39	1.31	1.17	1.27	1.11	1.01
17	P31FUJ	1.66	1.15	0.89	0.84	1.03	0.25

b. Chromatin-bound fraction

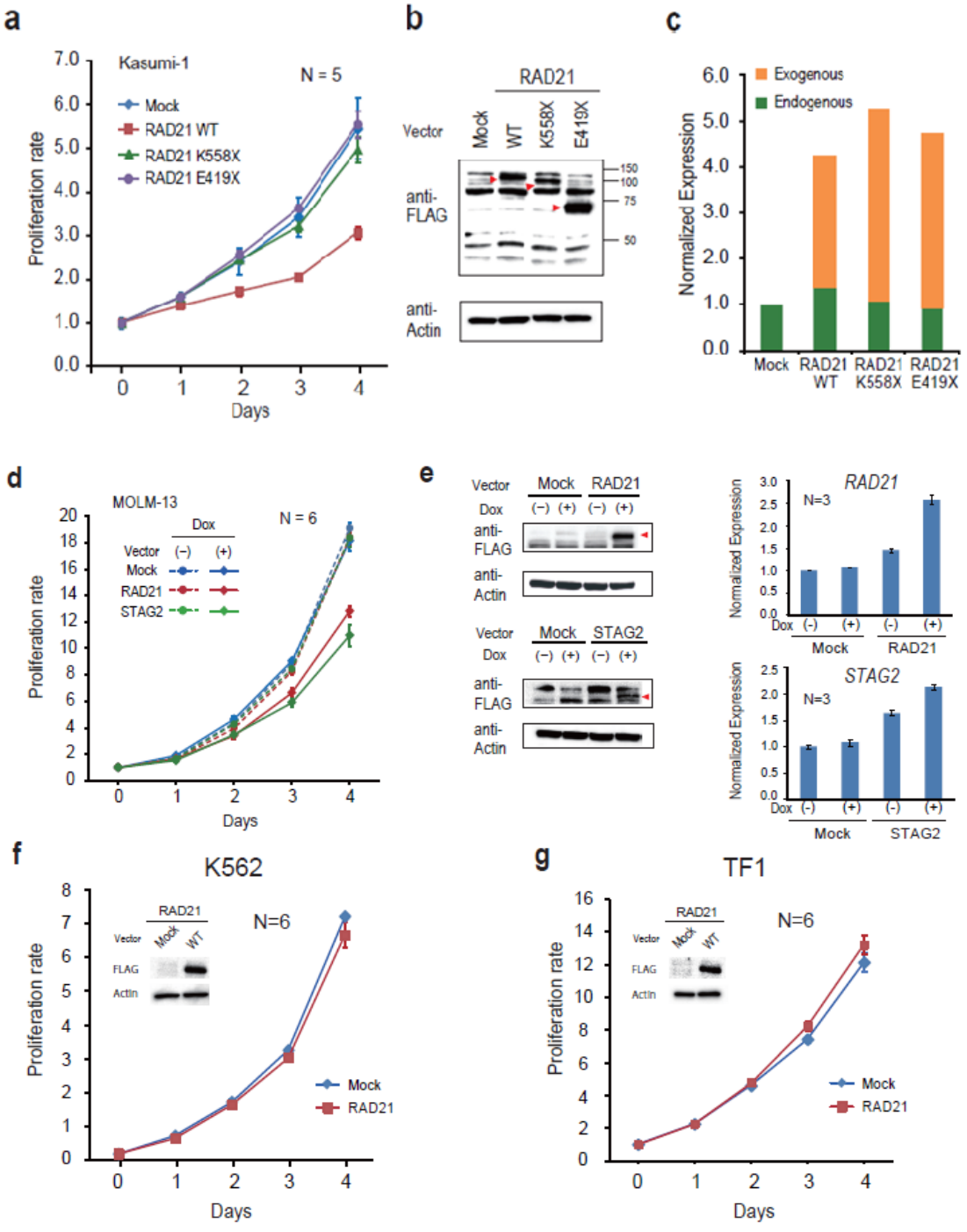
Lane	Cell line	SMC1A	SMC3	RAD21	STAG1	STAG2	NIPBL
1	CMK86	1.16	1.24	1.26	1.01	1.37	1.52
2	UTP-DSAL-1	1.18	1.26	1.23	1.33	1.08	0.92
3	CMS	1.36	1.15	1.05	1.61	1.04	0.53
4	HL-60	0.77	0.83	0.83	0.86	0.81	1.11
5	K562	0.79	0.82	0.78	0.50	0.96	0.69
6	TS9;22	1.03	1.09	1.27	1.27	1.02	1.27
7	F36P	0.83	0.73	0.72	0.59	0.88	1.09
8	SS9;22	0.88	0.88	0.87	0.83	0.85	0.86
9	MOLM-13	0.43	0.32	0.26	0.04	0.15	0.15
10	KG-1	0.71	0.53	0.49	1.48	0.00	0.78
11	Kasumi-1	0.79	0.52	0.48	0.69	0.51	0.72
12	P31FUJ	0.67	0.45	0.44	0.54	0.48	0.16
13	MOLM-7	0.81	0.59	0.60	0.97	0.62	1.27

a. Densitometoric data corresponding to **Figs. 16a and 16b,** which shows the expression level of each cohesin component in whole cell extracts of 17 cell lines after normalized with histone H3 signal and standardized to the mean value across all non-mutated cell lines. **b.** Densitometric data corresponding to **Figs. 16c and 16d,** which shows the expression of chromatin-bound form of each cohesin component in 13 cell lines normalized with histone H3 signal, which was then standardized to the mean value across all the intact cell lines.

The effect of forced expression of wild-type cohesin on cell proliferation of cohesin-defective cell lines

Next, I examined the effect of forced expression of wild-type cohesin on cell proliferation of cohesin-defective cell lines. Forced expression of the wild-type 52*RAD21* and/or *STAG2*, but not a truncated *RAD21* allele, induced significant growth suppression of the Kasumi-1 (*RAD21*-mutated) and MOLM-13 (RAD21/STAG2 expression severely reduced) cell lines, but not of the K562 and TF1 (wild-type *RAD21*) cell lines, supporting a leukemogenic role for compromised RAD21 function (Figure 17a–g). To explore the effect of forced expression of RAD21 on global gene expression, I performed expression microarray analysis of *RAD21*- and mock-transduced Kasumi-1 cells. In agreement with previous experiments with other cohesin and cohesin-related components, magnitude of transcriptional changes by RAD21 expression was generally small.[41,54,58] However, a total of 63 genes reproducibly and significantly showed more than 1.2 fold increase (N=35) or decrease (N=28) in gene expression, which was confirmed by quantitative PCR and/or RNA sequencing for 59 genes (Figure 18a–c).

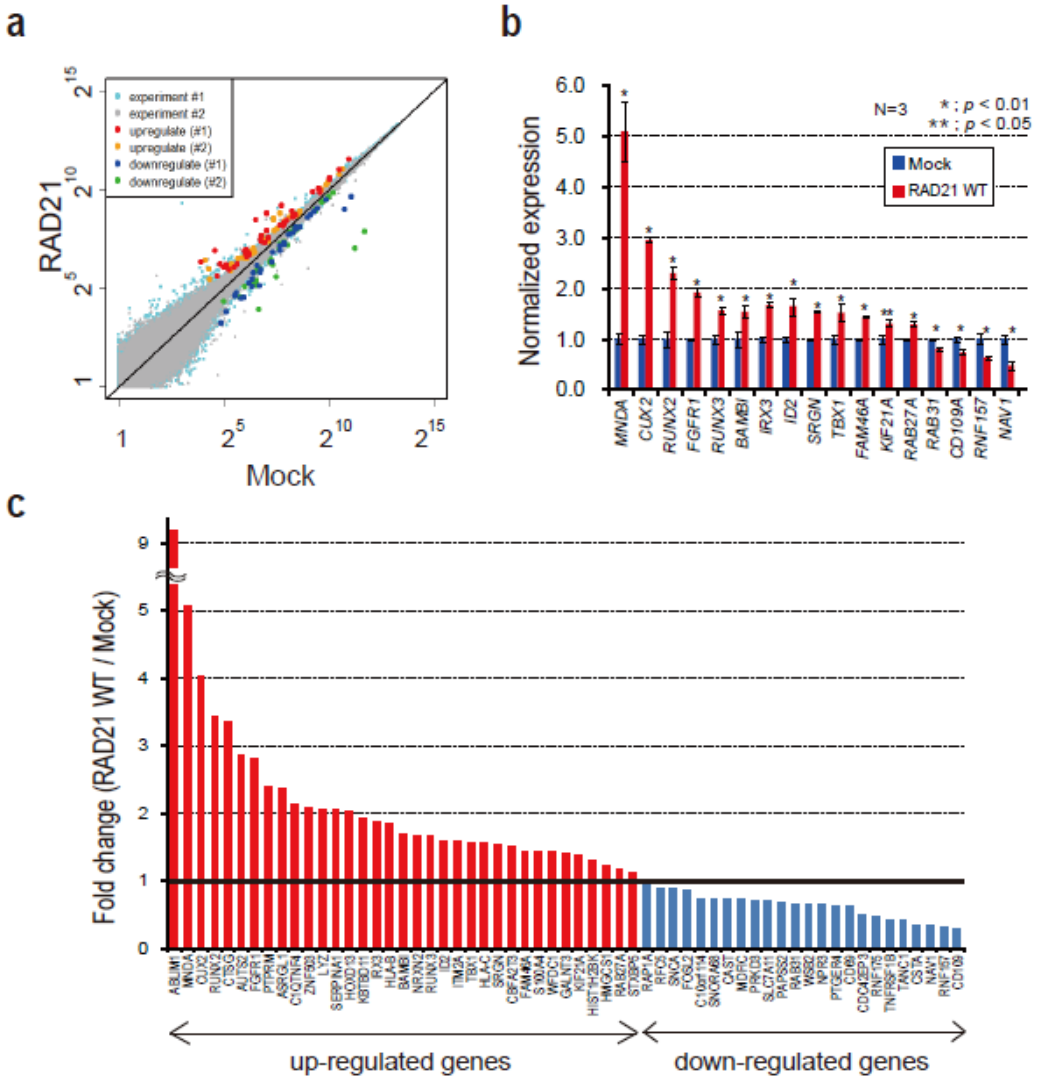
Figure 17. Impact of cohesin mutations on cell proliferation



a. Proliferation of the Kasumi-1 cell line stably transduced with the wild-type *RAD21* allele, mutant truncated *RAD21* alleles (*RAD21* K558X or E419X), or a mock construct was measured by MTT assays (N=5). Means ± SD of the absorbance at 450 nm relative to the value at day 0 are plotted. Representative results of three independent

experiments are shown. **b.** Western blotting showing the expression of the transduced wild-type/mutant *RAD21* allele. **c.** Expression of endogenous and exogenous *RAD21* transcripts in Kasumi-1 cells transduced with indicated mutants was measured by RNA sequencing. **d.** Effects of induced expression of wild-type RAD21 or STAG2 alleles on the proliferation of MOLM-13 with abnormal RAD21/STAG2 expression were determined (N=6). Means ± SD of the absorbance at 450 nm relative to the value at day 0 are plotted. A representative result from three independent experiments is shown. The induced expression of RAD21 and STAG2 are shown in **(e)**. **e.** RT-PCR analysis of RAD21 and STAG2 expression (N=3, means ± SD). Proliferation of K562 **(f)** and TF1 **(g)** cell lines stably transduced with the wild-type *RAD21* allele or a mock vector was measured (N=6). Means ± SD of the absorbance at 450 nm relative to the value at day 0 are plotted. A representative result from three independent experiments is shown. Western blotting of the transduced wild-type RAD21 allele are also shown.

Figure 18. Expression microarray analysis of RAD21- and mock- transduced Kasumi-1 cells



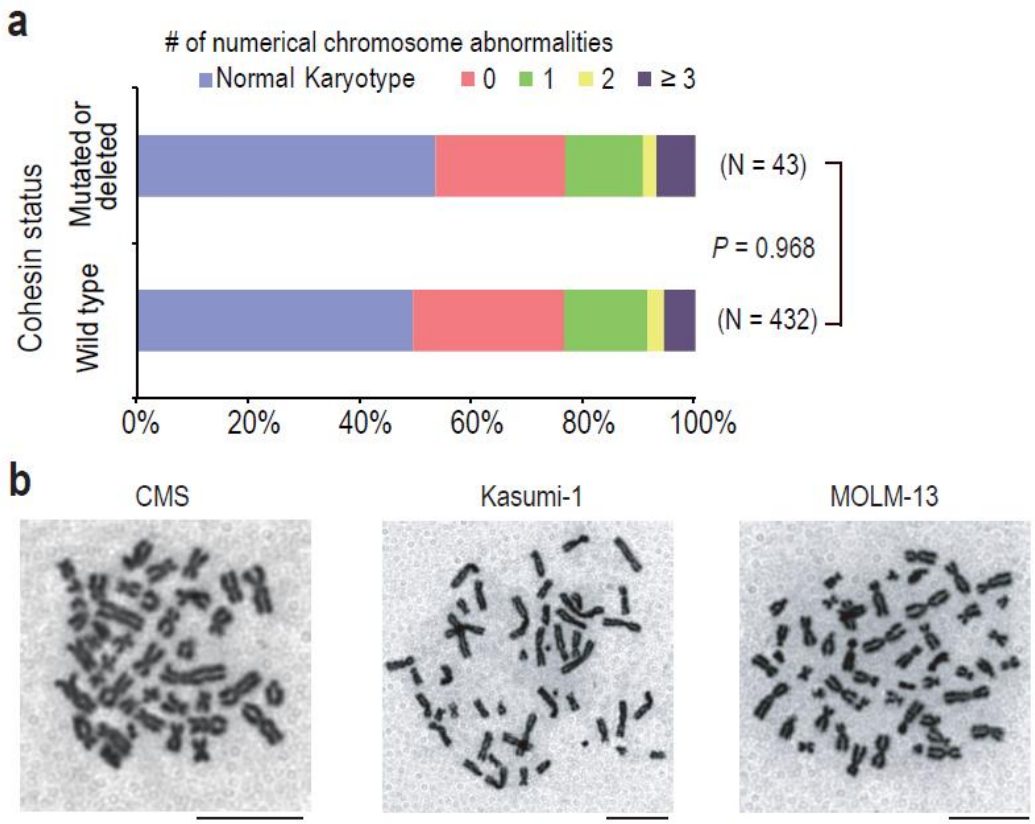
a. Comparison of gene expression profiles between RAD21- and mock-transduced Kasumi-1 cells. Mean expression values in triplicate array measurements were plotted for two independent experiments. Significantly and reproducibly up-regulated (>1.2-fold) or down-regulated (<0.83-fold) genes are shown in the indicated colors. **b.** The validation of the microarray analysis for the expression of 17 genes by qPCR. *GAPDH* normalized data are plotted (N=3, Means ± SD). *, $p < 0.01$; **, $p < 0.05$ (Student's t-test, one-sided). **c.** The validation of the microarray data by RNA sequencing, in which altered gene expression was confirmed for 59 out of 63 genes that showed significant alteration in gene expression in two independent microarray

experiments. Fold changes in gene expression in RNA sequencing is shown in up-regulated (N=35, red) and down-regulated (N=29, blue) genes.

Impact of cohesin mutations on chromosomal instability

It has been recently reported that *STAG2* mutations cause aneuploidy in human cancers.[54] Therefore, I investigated the relationship of cohesin mutations and aneuploidy in the present cohort of myeloid neoplasms. Contrary to my expectations, no significant differences were found in the number of chromosome abnormalities between cohesin-mutated and non-mutated cases, and the 43 cases with cohesin mutations/deletions showed diploid or near-diploid karyotypes, including 23 completely normal karyotypes (Figure 19a). Therefore, in these euploid cases, cohesin-mutated cells were not clonally selected as a result of aneuploidy. Furthermore, Kasumi-1 and MOLM-13 showed almost normal cohesion of sister chromatids, even though Kasumi-1 has a truncated *RAD21* allele, and MOLM-13 has substantially reduced expression of multiple cohesin components (Figure 19b).

Figure 19. Impact of cohesin mutations on chromosomal stability.



a. Number of numerical abnormalities in cytogenetics was compared between two groups with and without cohesin mutations/deletions (p=0.968, chi-square test). The numbers of numerical abnormalities in each cohesin mutated/deleted case are also shown at the bottom of panel Figure 5. **b.** Representative metaphases of Kasumi-1, MOLM-13, and CMS cells, showing almost normal sister chromatid cohesion.

Discussion

In this study, I have reported the identification of novel recurrent mutations and deletions in four core components of the cohesin complex, including *STAG2*, *RAD21*, *SMC1A* and *SMC3* in various myeloid neoplasms, which are novel class of genetic lesions, making them attractive targets for therapeutic intervention. These mutations were first identified through whole exome sequencing of 29 cases with myelodysplasia[23] and further validated in a large set of additional 581 cases by amplicon sequencing of pooled DNA or Sureselect captured sequencing of the targeted genes. In addition, genetic deletions were analyzed by SNP array based copy number analysis in 453 cases.

In analyzing the genetics data, some points were further investigated in order not to make a misleading interpretation. First, in the cases where germline control samples were not available, I studied the likelihood whether they were somatic mutations rather than SNPs by multiple methods. In addition to predicting the functional impact of each amino acid substitutions, the allele frequencies of the mutant alleles were exactly determined by deep sequencing, pyrosequencing and/or digital PCR and the deviations from those of SNPs were evaluated (Figure 7-9, Table5). Second, since MDS and related neoplasms are highly heterogeneous diseases, there might exist complex clonal architectures within the tumor samples and could affect the sensitivity of sequencing data. The percentages of tumor cells in the samples could also affect the sensitivity. However, Walter et al have previously performed whole genome sequencing of seven paired samples from the MDS stage and the secondary AML stage, and reported that nearly all the bone marrow cells in patients with MDS and secondary AML were clonally derived, even with a myloblast count of less than 5%.[59] So the

sensitivity of the whole exome sequencing could be considered to be considerably high.

Interestingly, mutations in four cohesin genes occurred in a mutually exclusive pattern, indicating that mutation in one component of the cohesin is sufficient for the loss of cohesin functions as a whole. The observation that multiple cohesin components in the chromatin-bound fractions were substantially reduced in cell lines with mutated or reduced cohesin components compared with the cell lines with no known cohesin mutations or abnormal cohesin expression (Figure 16) may support the idea. Therefore, there is a possibility that cohesin mutated cells might be synthetic lethal with mutations in other cohesin components or other related molecules. Given the integral functions of cohesin for cell viability, genetic defects in cohesin could be exploited as potential targets for myeloid neoplasms.[41,60] In fact, one intriguing study has demonstrated that inhibition of replication fork stability mediators such as poly (ADP-ribose) polymerase (PARP) could result in the specific killing of tumors with cohesin mutations through the mechanisms of synthetic lethality.[61]

Among the four recurrently mutated genes, *STAG2* was most frequently mutated and all of them were truncating mutations. It is unknown why the mutations rates of *STAG2* are higher than other genes, but it might be that *STAG1* partially compensate the function of *STAG2*, or STAG2 has different functions from other cohesin components. On the other hand, all of the mutations found in *SMC1A* and *SMC3* were heterozygous missense mutations. It has not yet elucidated whether these mutations might reduce the amount of functional cohesin (haploinsufficiency), or they could function in a dominant negative fashion. Further functional analysis would be required to address these questions.

In MDS, cohesin mutations were more frequent in advanced stages (RCMD

and RAEB) than lower-risk stages (RA, RARS, RCMD, RCMD-RS, and 5q- syndrome). Recently, comprehensive analysis of driver mutations in large cohort of MDS also demonstrated that cohesin mutations are more frequent in progressed stages[62,63] and associated with poorer prognosis.[62] Therefore, cohesin mutations are important driver mutations involved in the disease progression, and could be used as predictive indicators for prognosis.

Previously, less frequent mutations of cohesin components have been described in other cancers, including *STAG2* mutations in glioblastoma (4/68), melanoma (1/48), and Ewing's sarcoma (1/24).[54] Solomon et al have demonstrated that targeted inactivation of *STAG2* led to chromatid cohesion defects and aneuploidy, whereas in aneuploid human glioblastoma cell lines, targeted correction of the endogenous mutant alleles of *STAG2* led to enhanced chromosomal stability[54]. In primary colon cancer samples, *SMC1A* (4/132), *NIPBL* (4/132), *STAG3* (1/130), and *SMC3* (1/130) mutations have been reported, in which impaired cohesion and consequent aneuploidy were implicated in oncogenesis.[64] In contrast, in our cohort of myeloid neoplasms, no significant differences were found in the number of chromosome abnormalities between cohesin-mutated and non-mutated cases. Therefore, in these euploid cases, cohesin-mutated cells were not clonally selected as a result of aneuploidy. Supporting this is the observation that only 13% of the normal expression level of *scc1p*, a *RAD21* homologue, was sufficient for normal cohesion in yeast.[65] Other groups also recently reported recurrent mutations in the cohesin complex in a cohort of *de novo* AML and MDS, in which four major cohesin components were mutated in 6.0–9.0 % of the cases, where more than half cases with cohesin mutations showed normal karyotype.[59,66-69]

In the meanwhile, a growing body of evidence suggests that cohesin mediates long-range chromosomal *cis*-interactions[70] and regulates global gene expression. For example, two cohesin subunits, Rad21 and Smc3, have been implicated in the expression of the hematopoietic transcription factor Runx1 in zebrafish.[50] Furthermore, up to 80% downregulation of *Nipped-B*, a *NIPBL* homologue in *Drosophila,* does not affect chromosomal segregation but causes impaired regulation of gene expression.[58] It has also been demonstrated that only mild loss (17–28%) of cohesin binding sites within the genome resulted in deregulated global gene expression.[41,56,57] These observations suggest the possibility that cohesin mutations participate in leukemogenesis through deregulated expression of genes involved in myeloid development and differentiation.

In this study, the tumor suppressive role of cohesin was supported by growth suppression of leukemic cells having *RAD21*-mutation (Kasumi-1) or severely reduced RAD21/STAG2 expression (MOLM-13) induced by forced expression of wild-type RAD21 and/or STAG2, respectively (Figure17). Since most of the identified *RAD21* mutations in tumor samples were nonsense, frameshift, or splice-site changes causing premature truncation, Kasumi-1, which harbors a truncation mutation (K330PfsX6), was considered to be a good model for *RAD21*-mutated leukemia cells. As shown in Figure 17c, the wild-type RAD21-transduced Kasumi-1 cells had about four-fold increased RAD21 transcript expression compared to mock transduced cells. These results indicate that haploinsufficiency could be a relevant leukemogenic mechanism, at least for *RAD21* mutations. Further, expression microarray analysis of *RAD21*- and mock-transduced Kasumi-1 cells revealed that a total of 63 genes reproducibly and significantly showed more than 1.2 fold increase or decrease in gene expression,

supporting the possibility that cohesin mutations participate in leukemogenesis through deregulated expression of genes. Nevertheless, however, the molecular mechanisms of how cohesin mutations contribute to leukemogenesis largely remained unknown.

In conclusion, I report frequent mutations in cohesin components involving a wide variety of myeloid neoplasms. Genetic evidence suggests that aneuploidy alone may not be the leukemogenic mechanism, at least *in vivo,* and that deregulated gene expression and/or other mechanisms, such as DNA hypermutability, might also operate in leukemogenesis. Given the integral functions of cohesin for cell viability, genetic defects in cohesin could be exploited as potential targets for myeloid neoplasms. In the future project, further functional studies using a conditional knockout mouse model would be helpful to elucidate the mechanisms through which cohesin mutations play a role in myeloid leukemogenesis.

This work was published in *Nature Genetics* on 18th August 2013, entitled “Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms” (doi:10.1038/ng.2731).

Acknowledgments

I thank my supervisor Seishi Ogawa (Kyoto University) and Kiyoshi Miyagawa (The University of Tokyo) for continuous support and encouragement in completing the present study.

I would also like to thank the following collaborators for all the help and cooperation: Lee-Yung Shih, Masashi Minamino, Masashi Sanada, Yuichi Shiraishi, Yasunobu Nagata, Kenichi Yoshida, Yusuke Okuno, Masashige Bando, Ryuichiro Nakato, Shunpei Ishikawa, Aiko Sato-Otsubo, Genta Nagae, Aiko Nishimoto, Claudia Haferlach, Daniel Nowak, Yusuke Sato, Tamara Alpermann, Masao Nagasaki, Teppei Shimamura, Hiroko Tanaka, Kenichi Chiba, Ryo Yamamoto, Tomoyuki Yamaguchi, Makoto Otsu, Naoshi Obara, Mamiko Sakata-Yanagimoto, Tsuyoshi Nakamaki, Ken Ishiyama, Florian Nolte, Wolf-Karsten Hofmann, Shuichi Miyawaki, Shigeru Chiba, Hiraku Mori, Hiromitsu Nakauchi, H. Phillip Koeffler, Hiroyuki Aburatani, Torsten Haferlach, Katsuhiko Shirahige, and Satoru Miyano.

I am also grateful to Yuji Yamazaki for cell sorting, and Yuka Mori, Maki Nakamura, Noriko Mizota, and Shizue Ichimura for their technical assistance.

References

1 Tefferi, A. & Vardiman, J. W. Myelodysplastic syndromes. *N Engl J Med* **361**, 1872-1885, 2009

2 Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellstrom-Lindberg, E., Tefferi, A. & Bloomfield, C. D. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951, 2009

3 Cazzola, M., Della Porta, M. G., Travaglino, E. & Malcovati, L. Classification and prognostic evaluation of myelodysplastic syndromes. *Semin Oncol* **38**, 627-634, 2011

4 Greenberg, P., Cox, C., LeBeau, M. M., Fenaux, P., Morel, P., Sanz, G., Sanz, M., Vallespi, T., Hamblin, T., Oscier, D., Ohyashiki, K., Toyama, K., Aul, C., Mufti, G. & Bennett, J. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* **89**, 2079-2088, 1997

5 Ades, L., Itzykson, R. & Fenaux, P. Myelodysplastic syndromes. *Lancet* **383**, 2239-2252, 2014

6 Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 504-515, 2011

7 Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R. L., Neuberg, D. & Ebert, B. L. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* **364**, 2496-2506, 2011

8 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46, 2010

9 Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., McGrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D. E., Koboldt, D. C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L. W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J. R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W. D., Nagarajan, R., Walter, M. J., Link, D. C., Graubert, T. A., DiPersio, J. F. & Wilson, R. K. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, 2008

10 Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., Stebbings, L. A., Leroy, C., Edkins, S., Hardy, C., Teague, J. W., Menzies, A., Goodhead, I., Turner, D. J., Clee, C. M., Quail, M. A., Cox, A., Brown, C., Durbin, R., Hurles, M. E., Edwards, P. A., Bignell, G. R., Stratton, M. R. & Futreal, P. A. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729, 2008

11 Lee, W., Jiang, Z., Liu, J., Haverty, P. M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K. P., Bhatt, D., Ha, C., Johnson, S., Kennemer, M. I., Mohan, S., Nazarenko, I., Watanabe, C., Sparks, A. B., Shames, D. S., Gentleman, R., de Sauvage, F. J., Stern, H., Pandita, A., Ballinger, D. G., Drmanac, R., Modrusan, Z., Seshagiri, S. & Zhang, Z. The mutation spectrum revealed by paired genome

sequences from a lung cancer patient. *Nature* **465**, 473-477, 2010

12 Chapman, M. A., Lawrence, M. S., Keats, J. J., Cibulskis, K., Sougnez, C., Schinzel, A. C., Harview, C. L., Brunet, J. P., Ahmann, G. J., Adli, M., Anderson, K. C., Ardlie, K. G., Auclair, D., Baker, A., Bergsagel, P. L., Bernstein, B. E., Drier, Y., Fonseca, R., Gabriel, S. B., Hofmeister, C. C., Jagannath, S., Jakubowiak, A. J., Krishnan, A., Levy, J., Liefeld, T., Lonial, S., Mahan, S., Mfuko, B., Monti, S., Perkins, L. M., Onofrio, R., Pugh, T. J., Rajkumar, S. V., Ramos, A. H., Siegel, D. S., Sivachenko, A., Stewart, A. K., Trudel, S., Vij, R., Voet, D., Winckler, W., Zimmerman, T., Carpten, J., Trent, J., Hahn, W. C., Garraway, L. A., Meyerson, M., Lander, E. S., Getz, G. & Golub, T. R. Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472, 2011

13 Hirai, H., Kobayashi, Y., Mano, H., Hagiwara, K., Maru, Y., Omine, M., Mizoguchi, H., Nishida, J. & Takaku, F. A point mutation at codon 13 of the N-ras oncogene in myelodysplastic syndrome. *Nature* **327**, 430-432, 1987

14 Sugimoto, K., Hirano, N., Toyoshima, H., Chiba, S., Mano, H., Takaku, F., Yazaki, Y. & Hirai, H. Mutations of the p53 gene in myelodysplastic syndrome (MDS) and MDS-derived leukemia. *Blood* **81**, 3022-3026, 1993

15 Osato, M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene* **23**, 4284-4296, 2004

16 Langemeijer, S. M., Kuiper, R. P., Berends, M., Knops, R., Aslanyan, M. G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A. G., Raymakers, R. A., Kamping, E. J., Verhoef, G. E., Verburgh, E., Hagemeijer, A., Vandenberghe, P., de Witte, T., van der Reijden, B. A. & Jansen, J. H. Acquired

mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-842, 2009

17 Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lecluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguie, F., Fontenay, M., Vainchenker, W. & Bernard, O. A. Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301, 2009

18 Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., Lin, L., O'Laughlin, M., McMichael, J. F., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Magrini, V. J., Vickery, T. L., Hundal, J., Cook, L. L., Conyers, J. J., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T., Walker, J., Kalicki, J., Watson, M. A., Heath, S., Shannon, W. D., Varghese, N., Nagarajan, R., Westervelt, P., Tomasson, M. H., Link, D. C., Graubert, T. A., DiPersio, J. F., Mardis, E. R. & Wilson, R. K. DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine* **363**, 2424-2433, 2010

19 Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Locke, D. P., Magrini, V. J., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., McMichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X., Lin, L., Schmidt, H., Tang, Y., Haipek, C., Wiechert, M. E., Ivy, J.

V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath, S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., DiPersio, J. F., Wilson, R. K. & Ley, T. J. Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine* **361**, 1058-1066, 2009

20 Gelsi-Boyer, V., Trouplin, V., Adelaide, J., Bonansea, J., Cervera, N., Carbuccia, N., Lagarde, A., Prebet, T., Nezri, M., Sainty, D., Olschwang, S., Xerri, L., Chaffanet, M., Mozziconacci, M. J., Vey, N. & Birnbaum, D. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* **145**, 788-800, 2009

21 Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V., Waghorn, K., Zoi, K., Ross, F. M., Reiter, A., Hochhaus, A., Drexler, H. G., Duncombe, A., Cervantes, F., Oscier, D., Boultwood, J., Grand, F. H. & Cross, N. C. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722-726, 2010

22 Nikoloski, G., Langemeijer, S. M., Kuiper, R. P., Knops, R., Massop, M., Tonnissen, E. R., van der Heijden, A., Scheele, T. N., Vandenberghe, P., de Witte, T., van der Reijden, B. A. & Jansen, J. H. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42**, 665-667, 2010

23 Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K.,

Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S. & Ogawa, S. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69, 2011

24 Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J. S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., Godfrey, A. L., Rapado, I., Cvejic, A., Rance, R., McGee, C., Ellis, P., Mudie, L. J., Stephens, P. J., McLaren, S., Massie, C. E., Tarpey, P. S., Varela, I., Nik-Zainal, S., Davies, H. R., Shlien, A., Jones, D., Raine, K., Hinton, J., Butler, A. P., Teague, J. W., Baxter, E. J., Score, J., Galli, A., Della Porta, M. G., Travaglino, E., Groves, M., Tauro, S., Munshi, N. C., Anderson, K. C., El-Naggar, A., Fischer, A., Mustonen, V., Warren, A. J., Cross, N. C., Green, A. R., Futreal, P. A., Stratton, M. R. & Campbell, P. J. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* **365**, 1384-1395, 2011

25 Sanada, M., Suzuki, T., Shih, L. Y., Otsu, M., Kato, M., Yamazaki, S., Tamura, A., Honda, H., Sakata-Yanagimoto, M., Kumano, K., Oda, H., Yamagata, T., Takita, J., Gotoh, N., Nakazaki, K., Kawamata, N., Onodera, M., Nobuyoshi, M., Hayashi, Y., Harada, H., Kurokawa, M., Chiba, S., Mori, H., Ozawa, K., Omine, M., Hirai, H., Nakauchi, H., Koeffler, H. P. & Ogawa, S. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904-908, 2009

26 Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome research* **11**, 3-11, 2001

27 Shih, L. Y., Kuo, M. C., Kuo, C. Y., Lin, T. H., Bai, L. Y., Chen, T. Y., Wang, M.

C., Lin, T. L., Lan, Y. J., Chen, C. C., Yang, Y., Hsiao, P. C., Lai, C. L. & Chang, C. H. Emerging kinetics of BCR-ABL1 mutations and their effect on disease outcomes in chronic myeloid leukemia patients with imatinib failure. *Leuk Res* **37**, 43-49, 2013

28 Qin, J., Jones, R. C. & Ramakrishnan, R. Studying copy number variations using a nanofluidic platform. *Nucleic Acids Res* **36**, e116, 2008

29 Dube, S., Qin, J. & Ramakrishnan, R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS One* **3**, e2876, 2008

30 Totoki, Y., Tatsuno, K., Yamamoto, S., Arai, Y., Hosoda, F., Ishikawa, S., Tsutsumi, S., Sonoda, K., Totsuka, H., Shirakihara, T., Sakamoto, H., Wang, L., Ojima, H., Shimada, K., Kosuge, T., Okusaka, T., Kato, K., Kusuda, J., Yoshida, T., Aburatani, H. & Shibata, T. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43**, 464-469, 2011

31 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664, 2002

32 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**, 1073-1081, 2009

33 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, 2010

34 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**,

575-576, 2010

35 Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D. K., Kennedy, G. C. & Ogawa, S. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer research* **65**, 6071-6079, 2005

36 Yamamoto, G., Nannya, Y., Kato, M., Sanada, M., Levine, R. L., Kawamata, N., Hangaishi, A., Kurokawa, M., Chiba, S., Gilliland, D. G., Koeffler, H. P. & Ogawa, S. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *American journal of human genetics* **81**, 114-126, 2007

37 Hosoya, N., Sanada, M., Nannya, Y., Nakazaki, K., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S. & Ogawa, S. Genomewide screening of DNA copy number changes in chronic myelogenous leukemia with the use of high-resolution array-based comparative genomic hybridization. *Genes, chromosomes & cancer* **45**, 482-494, 2006

38 Kawamata, N., Ogawa, S., Yamamoto, G., Lehmann, S., Levine, R. L., Pikman, Y., Nannya, Y., Sanada, M., Miller, C. W., Gilliland, D. G. & Koeffler, H. P. Genetic profiling of myeloproliferative disorders by single-nucleotide polymorphism oligonucleotide microarray. *Experimental hematology* **36**, 1471-1479, 2008

39 Nagae, G., Isagawa, T., Shiraki, N., Fujita, T., Yamamoto, S., Tsutsumi, S., Nonaka, A., Yoshiba, S., Matsusaka, K., Midorikawa, Y., Ishikawa, S., Soejima,

H., Fukayama, M., Suemori, H., Nakatsuji, N., Kume, S. & Aburatani, H. Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. *Hum Mol Genet* **20**, 2710-2721, 2011

40 Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H., Nagata, Y., Yoshida, K., Kon, A., Suzuki, Y., Chiba, K., Tanaka, H., Niida, A., Fujimoto, A., Tsunoda, T., Morikawa, T., Maeda, D., Kume, H., Sugano, S., Fukayama, M., Aburatani, H., Sanada, M., Miyano, S., Homma, Y. & Ogawa, S. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* **45**, 860-867, 2013

41 Deardorff, M. A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K. E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L. J., Gyftodimou, Y., Hirashima, K., Hullings, M., Ishikawa, Y., Jaulin, C., Kaur, M., Kiyono, T., Lombardi, P. M., Magnaghi-Jaulin, L., Mortier, G. R., Nozaki, N., Petersen, M. B., Seimiya, H., Siu, V. M., Suzuki, Y., Takagaki, K., Wilde, J. J., Willems, P. J., Prigent, C., Gillessen-Kaesbach, G., Christianson, D. W., Kaiser, F. J., Jackson, L. G., Hirota, T., Krantz, I. D. & Shirahige, K. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* **489**, 313-317, 2012

42 Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. & Peters, J. M. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801, 2008

43 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping

and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, 2008

44 Gruber, S., Haering, C. H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765-777, 2003

45 Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annual review of genetics* **43**, 525-558, 2009

46 Strom, L., Karlsson, C., Lindroos, H. B., Wedahl, S., Katou, Y., Shirahige, K. & Sjogren, C. Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317**, 242-245, 2007

47 Watrin, E. & Peters, J. M. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *The EMBO journal* **28**, 2625-2635, 2009

48 Dorsett, D. Cohesin, gene expression and development: lessons from Drosophila. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**, 185-200, 2009

49 Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B. & McKim, K. Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. *Development* **132**, 4743-4753, 2005

50 Horsfield, J. A., Anagnostou, S. H., Hu, J. K., Cho, K. H., Geisler, R., Lieschke, G., Crosier, K. E. & Crosier, P. S. Cohesin-dependent regulation of Runx genes. *Development* **134**, 2639-2649, 2007

51 Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori,

K., Dillon, N., Aragon, L., Fisher, A. G. & Merkenschlager, M. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433, 2008

52 Bose, T. & Gerton, J. L. Cohesinopathies, gene expression, and chromatin organization. *The Journal of cell biology* **189**, 201-210, 2010

53 Deardorff, M. A., Wilde, J. J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Monnich, M., Yan, Y., Xu, W., Gil-Rodriguez, M. C., Clark, D., Hakonarson, H., Halbach, S., Michelis, L. D., Rampuria, A., Rossier, E., Spranger, S., Van Maldergem, L., Lynch, S. A., Gillessen-Kaesbach, G., Ludecke, H. J., Ramsay, R. G., McKay, M. J., Krantz, I. D., Xu, H., Horsfield, J. A. & Kaiser, F. J. RAD21 Mutations Cause a Human Cohesinopathy. *American journal of human genetics* **90**, 1014-1027, 2012

54 Solomon, D. A., Kim, T., Diaz-Martinez, L. A., Fair, J., Elkahloun, A. G., Harris, B. T., Toretsky, J. A., Rosenberg, S. A., Shukla, N., Ladanyi, M., Samuels, Y., James, C. D., Yu, H., Kim, J. S. & Waldman, T. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**, 1039-1043, 2011

55 Beckouet, F., Hu, B., Roig, M. B., Sutani, T., Komata, M., Uluocak, P., Katis, V. L., Shirahige, K. & Nasmyth, K. An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Molecular cell* **39**, 689-699, 2010

56 Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N. B., Vega, H., Jackson, L. G., Shirahige, K. & Krantz, I. D. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS biology* **7**, e1000119, 2009

57 Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Li, J. R., Clark, D.,

Kaur, M., Tatsuro, K., Kline, A. D., Chang, C., Vega, H., Jackson, L. G., Spinner, N. B., Shirahige, K. & Krantz, I. D. Genome-wide DNA methylation analysis in cohesin mutant human cell lines. *Nucleic acids research* **38**, 5657-5671, 2010

58 Schaaf, C. A., Misulovin, Z., Sahota, G., Siddiqui, A. M., Schwartz, Y. B., Kahn, T. G., Pirrotta, V., Gause, M. & Dorsett, D. Regulation of the Drosophila Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PLoS One* **4**, e6202, 2009

59 Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., Larson, D. E., McLellan, M. D., Dooling, D., Abbott, R., Fulton, R., Magrini, V., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Fan, X., Grillot, M., Witowski, S., Heath, S., Frater, J. L., Eades, W., Tomasson, M., Westervelt, P., DiPersio, J. F., Link, D. C., Mardis, E. R., Ley, T. J., Wilson, R. K. & Graubert, T. A. Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine* **366**, 1090-1098, 2012

60 Chan, D. A. & Giaccia, A. J. Harnessing synthetic lethal interactions in anticancer drug discovery. *Nature reviews. Drug discovery* **10**, 351-364, 2011

61 McLellan, J. L., O'Neil, N. J., Barrett, I., Ferree, E., van Pel, D. M., Ushey, K., Sipahimalani, P., Bryan, J., Rose, A. M. & Hieter, P. Synthetic lethality of cohesins with PARPs and replication fork mediators. *PLoS Genet* **8**, e1002574, 2012

62 Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., Yoshida, K., Roller, A., Nadarajah, N., Shiraishi, Y., Shiozawa, Y., Chiba, K., Tanaka, H., Koeffler, H. P., Klein, H. U., Dugas, M., Aburatani, H., Kohlmann, A., Miyano, S., Haferlach,

C., Kern, W. & Ogawa, S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241-247, 2014

63 Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C. J., Ellis, P., Wedge, D. C., Pellagatti, A., Shlien, A., Groves, M. J., Forbes, S. A., Raine, K., Hinton, J., Mudie, L. J., McLaren, S., Hardy, C., Latimer, C., Della Porta, M. G., O'Meara, S., Ambaglio, I., Galli, A., Butler, A. P., Walldin, G., Teague, J. W., Quek, L., Sternberg, A., Gambacorti-Passerini, C., Cross, N. C., Green, A. R., Boultwood, J., Vyas, P., Hellstrom-Lindberg, E., Bowen, D., Cazzola, M., Stratton, M. R. & Campbell, P. J. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616-3627; quiz 3699, 2013

64 Barber, T. D., McManus, K., Yuen, K. W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Lengauer, C. & Hieter, P. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3443-3448, 2008

65 Heidinger-Pauli, J. M., Mert, O., Davenport, C., Guacci, V. & Koshland, D. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Current biology : CB* **20**, 957-963, 2010

66 Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L.,

McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K. & DiPersio, J. F. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510, 2012

67 Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klco, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., DiPersio, J. F., Ding, L., Mardis, E. R. & Wilson, R. K. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264-278, 2012

68 The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074, 2013

69 Walter, M. J., Shen, D., Shao, J., Ding, L., White, B. S., Kandoth, C., Miller, C. A., Niu, B., McLellan, M. D., Dees, N. D., Fulton, R., Elliot, K., Heath, S., Grillot, M., Westervelt, P., Link, D. C., Dipersio, J. F., Mardis, E., Ley, T. J., Wilson, R. K. & Graubert, T. A. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia*, 2013

70 Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G. & Merkenschlager, M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**, 410-413, 2009