厚生労働科学研究費補助金
第3次対がん総合戦略研究事業

骨髄異形成症候群におけるエピゲノム修飾分子異常の解明

平成23年度　　総括研究報告書

研究代表者　　真田　昌

平成24(2012)年　3月

目　　次

Ⅰ．総括研究報告

厚生労働科学研究費補助金（第3次対がん総合戦略研究事業）
総括研究報告書

骨髄異形成症候群におけるエピゲノム修飾分子異常の解明

研究代表者　真田　昌　　東京大学キャンサーボード がんゲノミクスプロジェクト 特任助教

研究要旨

骨髄異形成症候群(MDS)は高齢者に好発する難治性造血器疾患であるが、高齢者に適した根治的治療がなく、急速な少子高齢化による患者数の増加も危惧される。MDS におけるDNA のメチル化などのエピゲノム異常が生じていることは1990 年代から報告され、近年脱メチル化剤やヒストン修飾酵素阻害剤などが欧米で臨床応用され、一定の臨床効果が得られ、従来型の殺細胞性抗腫瘍剤では治療効果が期待できないMDS において重要な治療薬として認識され、H23 年度より本邦でも臨床応用が本格的に開始された。一方で、最近になり、MDS においてエピゲノム関連分子のゲノム異常が生じていることが明らかとなってきている。しかし、MDS において観察され、治療標的と考えられているエピゲノム異常が、これらのエピゲノム関連遺伝子のゲノムレベルでの変異に起因しているのか、さらにはエピゲノム修飾を治療標的とした薬剤の反応性の予測マーカーとなり得るのかなど、不明な点は多い。そこで本研究では、研究計画に基づき、平成23 年度は151 例のMDS 臨床検体を対象に、エピゲノム修飾に関連する81 遺伝子のコーディング領域について、次世代シークエンサーを活用したターゲットシークエンスによる変異解析を行った。本解析で得られた変異候補については、サンガーシークエンス法を用いてバリデーションを行い、151 例中 112 例（71%）の症例において、本遺伝子群にアミノ酸変化を伴う変異が1 つ以上同定された。*TET2*、*ASXL1*、*EZH2*、*DNMT3A* など既に変異の報告のある遺伝子に多数例で変異が観察された。また、頻度は低いものの、これまで変異が報告されていないエピゲノム関連遺伝子にも機能喪失が考えられる変異が同定され、MDS においては大半の症例でエピゲノム修飾が遺伝子（ゲノム）レベルで異常を来していることが明らかとなった。これらエピゲノム関連遺伝子群のゲノムレベルでの変異とエピゲノム異常との関連について検討する目的で、平成24 年度はメチル化解析を行う予定であるが、既存の網羅的なメチル化解析法では、サンプルの必要量が多く、またエピゲノム解析では増幅後検体を使用できないことは、臨床検体の解析においては大きな問題である。そこで、少ないサンプル量での網羅的な解析法を検討した。メチル化感受性制限酵素 HpaII もしくは非感受性の isoschizomer である MspI 処理後にアフィメトリクス社のSNP アレイを用いて解析をすることにより、少量のサンプル量で、簡便なプロトコールのもと、61,428 サイトのメチル化状態を網羅的に解析が可能であり、バイサルファイトシークエンスにより再現性が確認された。またSNP アレイの特性を活かし、ヘテロ SNP の近傍の CCGG 配列に着目することで、アレルによるメチル化修飾の差も検出が可能であった。

A. 研究目的

骨髄異形成症候群(MDS)は高齢者に好発する難治性造血器腫瘍であるが、高齢者に適した根治的治療がなく、急速な少子高齢化による患者数の増加も危惧される。MDSにおけるDNAのメチル化などのエピゲノム異常が生じていることは1990年代から報告され、近年脱メチル化剤やヒストン修飾酵素阻害剤などが欧米で臨床応用され、一定の臨床効果が得られ、従来の抗腫瘍剤では効果が期待できないMDSにおいて重要な治療薬として認識され、本邦でも臨床応用が開始された。一方で、最近になり、MDSにおいて*EZH2*、*TET2*、*ASXL1*などのエピゲノム関連分子の後天的変異が報告されたが、他にも多くの分子がエピゲノム修飾に関与しており、それらの分子の変異の有無は明らかではない。しかしMDSにおけるエピゲノム修飾異常が、同修飾に関わる分子のゲノム異常に起因しているのか、多数存在する他のエピゲノム修飾分子の変異の有無など不明な点は多く、MDSにおけるエピゲノム関連分子異常の全体像は明らかではない。本研究では、最新のゲノム解析技術を駆使し、MDS検体における計81個のエピゲノム修飾関連分子の変異プロファイルを明らかとし(H23)、網羅的なメチル化プロファイル(H24)、先行研究に基づくSNPアレイを用いたゲノム異常、既知の遺伝子変異、生命予後を含む臨床データとの関連を明らかとする。

B. 研究方法

(1) 実施経過

151例のMDS臨床検体(CMML 44例、MDSから移行したAML 18例を含む)を対象に、エピゲノム修飾に関連する81遺伝子のコーディング領域について、次世代シークエンサーを活用したターゲットシークエンスによる変異解析を行った。MDS患者の骨髄細胞から抽出したゲノムDNAを超音波処理により断片化した後に、解析予定領域に対し相補的なRNAベイトを設計し(アジレント社SureSelect®)、断片化したDNAと液相でハイブリダイズさせ、ビーズ回収することにより解析領域を濃縮し、濃縮後サンプルに検体識別用の配列を付加した後に、illumina社の次世代シークエンサーを用いてシークエンスを行った。本解析で得られた変異候補については、サンガーシークエンス法を用いてバリデーションを行った。また、次年度に予定している網羅的なメチル化解析に向け、新たな解析方法の検討として、SNPアレイを用いたアレル特異性も含めたメチル化解析についても細胞株を用いて、有用性を検証した。

(2) 倫理面の配慮

本研究で実施される患者検体を用いた遺伝子解析研究は、原則としてMDS細胞の体細胞突然変異を扱うものであるが、平成16年(平成20年改訂)文部科学省、厚生労働省および経済産業省告示第1号「ヒトゲノム・遺伝子研究に関する倫理指針」を遵守し、事前に検体提供施設の倫理委員会の承認を得ている。東京大学における遺伝子解析研究については、学内のヒトゲノム遺伝子解析研究倫理委員会の審査・承認済み(「造血器腫瘍における遺伝子異常の網羅的解析」948-7)である。なお、承認済みの研究計画書に基づき、研究対象(検体提供)者から文書による同意を得た上で検体を採取し、匿名化作業を行った上で、遺伝子解析研究に用いた。

C. 研究結果

本研究で採用したターゲットキャプチャシークエンスにより、限られたRUN数で多数検体の変異解析を、多数の遺伝子セットについて効率良く行うことが可能であった。本解析で得られた変異候補については、サンガーシークエンス法を用いてバリデーションを行い、151例中112例(71%)の症例において、本遺伝子群にアミノ酸変化を伴う変異が1つ以上同定された。*TET2*、*ASXL1*、*EZH2*、*DNMT3A*など既に変異の報告のある遺伝子に多数例で変異が観察された。また、頻度は低いものの、

これまで変異が報告されていないエピゲノム関連遺伝子にもナンセンス変異やフレームシフトなど機能喪失が考えられる変異が同定され、MDSにおいては広範な症例でエピゲノム修飾が遺伝子（ゲノム）レベルで異常を来していることが明らかとなった。またMDS例における全エクソン解析においても、エピゲノム関連分子異常は、RNAスプライシング関連分子異常と並んで、MDSに共通して観察される代表的な遺伝子異常であることが示された(Nature 2011)。

また、既存の網羅的なメチル化解析法では、サンプルの必要量が多く、またエピゲノム解析では増幅後検体を使用できないことは、臨床検体の解析においては大きな問題である。そこで、少ないサンプル量での網羅的な解析法を検討した。メチル化感受性制限酵素HpaIIもしくは非感受性のisoschizomerであるMspI処理後にアフィメトリクス社のSNPアレイを用いて解析をすることにより、少量のサンプル量で、簡便なプロトコールのもと、61,428サイトのメチル化状態を網羅的に解析が可能であり、バイサルファイトシークエンスにより再現性が確認された。またSNPアレイの特性を活かし、ヘテロSNPの近傍のCCGG配列に着目することで、アレルによるメチル化修飾の差も検出が可能であった。

D. 考察

エピゲノム修飾は遺伝子発現調節において重要な機構であるが、メチル化に代表されるエピゲノム修飾異常が発がんと関わっていることが、多くのがんで示されている。従来の抗腫瘍剤の有効性が期待できないMDSでは、脱メチル化剤やヒストン脱アセチル化酵素阻害剤が、有効性が評価され、H23年度より日本においても脱メチル化剤の臨床応用が開始された。しかしながら、奏功率は高くなく、有効性を予測できる指標もない。一方で近年の遺伝子解析研究により、MDSにおいて、*TET2*や*ASXL1*などエピゲノム修飾に関わるとされる遺伝子の変異が明らかとなってきた。さらにMDSで変異頻度が高いTET2が、脱メチル化に重要な役割を担っていることが最近になり示され(Guo et al. Cell 2011)、MDSを含む骨髄系腫瘍で観察されるIDH変異は、TET2の不活化を導くことも明らかとなっている(Figueroa et al. Cancer Cell 2010)。しかしMDS症例で観察されるエピゲノム修飾の異常が、同修飾に関わる分子のゲノム異常に起因しているのか、多数存在する他のエピゲノム修飾分子の変異の有無など不明な点は多く、MDSにおけるエピゲノム関連分子異常の全体像は明らかではない。本年度行った次世代シークエンサーを活用したエピゲノム関連遺伝子変異の解析から、MDSにおけるエピゲノム関連分子の変異プロファイルが明らかとなり、多くのMDS症例において、エピゲノム修飾に関わる分子に異常が生じていることが明らかとなった。また全エクソンシークエンスの結果からも、エピゲノム関連遺伝子変異は、MDSの分子病態に強く関与していることが推測された。次年度以降、ゲノム解析研究結果などと比較検討することにより、ゲノム異常・臨床情報との関連が明らかとなり、MDSの分子病態の解明が進むことが期待される。またDNAのメチル化やヒストン修飾状態のパターンと変異の有無を調べることにより、変異の意義、すなわち腫瘍性疾患におけるエピゲノム異常の分子メカニズム解明に寄与することが期待される。これらの分子基盤の解明は、脱メチル化剤などのエピゲノム標的治療薬の有効性を予測する重要な分子マーカーとなり、これらの薬剤の臨床応用に向けて大きな資産となることが期待され、MDSの治療成績の向上に直接寄与し得ると考える。本貢献は患者個人の生命予後やQOLのみならず、急速な高齢化社会を迎え、MDS患者数の増加が懸念される我が国においては医療経済や輸血行政においても重要である。

またエピゲノム異常は、MDSに限られるものではなく、多くの腫瘍性疾患に共通して観察

される異常である。本アプローチは、他の「がん」においても応用可能であり、MDS 以外の難治性腫瘍における新規治療法の開発において有用な基盤が構築されることが期待される。

E. 結論

MDS 症例においては、エピゲノム修飾分子にゲノムレベルでの異常が既知の遺伝子のみならず、高頻度に生じていることが本研究を通じて明らかとなった。更にエピゲノム解析を行うことにより、MDS におけるエピゲノム修飾分子のゲノム異常とエピゲノム異常の関わりが明らかになることが期待される。

F. 健康危険情報

なし

G. 研究発表

1 論文発表

1) 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 WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S.: Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478:64-9

2) Sanada M. Acquired uniparental disomy and c-CBL mutation in myelodysplastic syndromes. Rinsho Ketsueki. 2011 Jun;52(6):342-9.

3) Takita J, Yoshida K, Sanada M, Nishimura R, Okubo J, Motomura A, Hiwatari M, Oki K, Igarashi T, Hayashi Y, Ogawa S. Novel splicing-factor mutations in juvenile myelomonocytic leukemia. Leukemia. 2012 Feb 20

4) Sato-Otsubo A, Sanada M, Ogawa S. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? Semin Oncol. 2012 Feb; 39(1):13-25

5) Kao HW, Sanada M, Liang DC, Lai DC, Lee EH, Kuo MC, Lin TL, Shih YS, Wu JH, Huang CF, Ogawa S, Shih LY. A high occurrence of acquisition and/or expansion of C-CBL mutant clones in the progression of high-risk myelodysplastic syndrome to acute myeloid leukemia. Neoplasia. 2011 Nov; 13(11):1035-42

2 学会発表

真田 昌 高速シークエンサーを活用した骨髄異形成症候群における新規標的遺伝子の探索 第56回日本人類遺伝学会（シンポジウム・招待講演）2011年11月 千葉市

H. 知的財産権の出願・登録状況

1. 特許出願中
   なし
2. 実用新案登録
   なし
3. その他
   なし

Ⅱ．研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書　籍　名	出版社名	出版地	出版年	ページ
真田　昌	ゲノム異常	松田　晃	骨髄異形成症候群（MDS）のマネジメント	医薬ジャーナル	大阪市	2011年	42-48ページ

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
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 WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S.	Frequent pathway mutations of splicing machinery in myelodysplasia	Nature	478	64-9	2011
Sanada M	Acquired uniparental disomy and c-CBL mutation in myelodysplastic syndromes	Rinshou ketueki	52	342-9	2011
Takita J, Yoshida K, Sanada M, Nishimura R, Okubo J, Motomura A, Hiwatari M, Oki K, Igarashi T, Hayashi Y, Ogawa S	Novel splicing-factor mutations in juvenile myelomonocytic leukemia. Leukemia	Leukemia			2012
Sato-Otsubo A, Sanada M, Ogawa S.	Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how?	Semin Oncol	39	13-25	2012

Kao HW, Sanada M, Liang DC, Lai DC, Lee EH, Kuo MC, Lin TL, Shih YS, Wu JH, Huang CF, Ogawa S, Shih LY.	A high occurrence of acquisition and/or expansion of C-CBL mutant clones in the progression of high-risk myelodysplastic syndrome to acute myeloid leukemia.	Neoplasia	13	1035-42	2011

Ⅲ. 研究成果の刊行物・別刷

ARTICLE

doi:10.1038/nature10496

Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida[1]*, Masashi Sanada[1]*, Yuichi Shiraishi[2]*, Daniel Nowak[3]*, Yasunobu Nagata[1]*, Ryo Yamamoto[4], Yusuke Sato[1], Aiko Sato-Otsubo[1], Ayana Kon[1], Masao Nagasaki[5], George Chalkidis[6], Yutaka Suzuki[7], Masashi Shiosaka[1], Ryoichiro Kawahata[1], Tomoyuki Yamaguchi[8], Makoto Otsu[4], Naoshi Obara[9], Mamiko Sakata-Yanagimoto[9], Ken Ishiyama[10], Hiraku Mori[11], Florian Nolte[3], Wolf-Karsten Hofmann[3], Shuichi Miyawaki[10], Sumio Sugano[7], Claudia Haferlach[12], H. Phillip Koeffler[13,14], Lee-Yung Shih[15], Torsten Haferlach[12], Shigeru Chiba[9], Hiromitsu Nakauchi[4,8], Satoru Miyano[2,6] & Seishi Ogawa[1]

Myelodysplastic syndromes and related disorders (myelodysplasia) are a heterogeneous group of myeloid neoplasms showing deregulated blood cell production with evidence of myeloid dysplasia and a predisposition to acute myeloid leukaemia, whose pathogenesis is only incompletely understood. Here we report whole-exome sequencing of 29 myelodysplasia specimens, which unexpectedly revealed novel pathway mutations involving multiple components of the RNA splicing machinery, including *U2AF35*, *ZRSR2*, *SRSF2* and *SF3B1*. In a large series analysis, these splicing pathway mutations were frequent (~45 to ~85%) in, and highly specific to, myeloid neoplasms showing features of myelodysplasia. Conspicuously, most of the mutations, which occurred in a mutually exclusive manner, affected genes involved in the 3′-splice site recognition during pre-mRNA processing, inducing abnormal RNA splicing and compromised haematopoiesis. Our results provide the first evidence indicating that genetic alterations of the major splicing components could be involved in human pathogenesis, also implicating a novel therapeutic possibility for myelodysplasia.

Myelodysplastic syndromes (MDS) and related disorders (myelodysplasia) comprise a group of myeloid neoplasms characterized by deregulated, dysplastic blood cell production and a predisposition to acute myeloid leukaemia (AML)[1]. Although the prevalence of MDS has not been determined precisely, more than 10,000 people are estimated to develop myelodysplasia annually in the United States[2]. Their indolent clinical course before leukaemic transformation and ineffective haematopoiesis with evidence of myeloid dysplasia indicate a pathogenesis distinct from that involved in *de novo* AML. Currently, a number of gene mutations and cytogenetic changes have been implicated in the pathogenesis of MDS, including mutations of *RAS*, *TP53* and *RUNX1*, and more recently *ASXL1*, *c-CBL*, *DNMT3A*, *IDH1/2*, *TET2* and *EZH2* (ref. 3). Nevertheless, mutations of this set of genes do not fully explain the pathogenesis of MDS because they are also commonly found in other myeloid malignancies and roughly 20% of MDS cases have no known genetic changes (ref. 4 and unpublished data). In particular, the genetic alterations responsible for the dysplastic phenotypes and ineffective haematopoiesis of myelodysplasia are poorly understood. Meanwhile, the recent development of massively parallel sequencing technologies has provided an expanded opportunity to discover genetic changes across the entire genomes or protein-coding sequences in human cancers at a single-nucleotide level[5–10], which could be successfully applied to the genetic analysis of myelodysplasia to obtain a better understanding of its pathogenesis.

Overview of genetic alterations

In this study, we performed whole-exome sequencing of paired tumour/control DNA from 29 patients with myelodysplasia (Supplementary Table 1). Although incapable of detecting non-coding mutations and gene rearrangements, the whole-exome approach is a well-established strategy for obtaining comprehensive registries of protein-coding mutations at low cost and high performance. With a mean coverage of 133.8, 80.4% of the target sequences were analysed at more than ×20 depth on average (Supplementary Fig. 1). All the candidates for somatic mutations ($N = 497$) generated through our data analysis pipeline were subjected to validation using Sanger sequencing (Supplementary Methods I and Supplementary Fig. 2). Finally, 268 non-synonymous somatic mutations were confirmed with an overall true positive rate of 53.9% (Supplementary Fig. 3), including 206 missense, 25 nonsense, and 10 splice site mutations, and 27 frameshift-causing insertions/deletions (indels) (Supplementary Fig. 4). The mutation rate of 9.2 (0–21) per sample was significantly lower than that in solid tumours (16.2–302)[7,11,12] and multiple myeloma (32.4)[6], but was comparable to that in AML (7.3–13)[13–15] and chronic lymphocytic leukaemia (11.5)[16]. Combined with the genomic copy number profile obtained by single nucleotide polymorphism (SNP) array karyotyping, this array of somatic mutations provided a landscape of myelodysplasia genomes (Supplementary Fig. 5)[17,18].

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. [2]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [3]Department of Hematology and Oncology, Medical Faculty Manheim of the University of Heidelberg, 1–3 Theodor-Kutzer-Ufer, Mannheim 68167, Germany. [4]Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [5]Laboratory of Functional Genomics, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [6]Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [7]Division of Systems Biomedical Technology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [8]Nakauchi Stem Cell and Organ Regeneration Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [9]Department of Hematology, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8571, Japan. [10]Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, 2-8-1 Minami-Ohtsuka, Toshima-ku, Tokyo 170-0005, Japan. [11]Division of Hematology, Internal Medicine, Showa University Fujigaoka Hospital, 1-30 Fujigaoka, Aoba-ku, Yokohama, Kanagawa 227-8501, Japan. [12]Munich Leukemia Laboratory, Max-Lebsche-Platz 31, Munich 81377, Germany. [13]Hematology/Oncology, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. [14]National University of Singapore, Cancer Science Institute of Singapore, 28 Medical Drive, Singapore 117456, Singapore. [15]Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, 199 Tung Hwa North Rd, Taipei 105, Taiwan.
*These authors contributed equally to this work.

Novel gene targets in myelodysplasia

The list of the somatic mutations (Supplementary Table 2) included most of the known gene targets in myelodysplasia with similar mutation frequencies to those previously reported, indicating an acceptable sensitivity of the current study. The mutations of the known gene targets, however, accounted for only 12.3% of all detected mutations ($N = 33$), and the remaining 235 mutations involved previously unreported genes. Among these, recurrently mutated genes in multiple cases are candidate targets of particular interest, for which high mutation rates are expected in general populations. In fact, 8 of the 12 recurrently mutated genes were among the well-described gene targets in myelodysplasia (Supplementary Table 3). However, what immediately drew our attention were the recurrent mutations involving *U2AF35* (also known as *U2AF1*), *ZRSR2* and *SRSF2* (*SC35*), because they belong to the common pathway known as RNA splicing. Including an additional three genes mutated in single cases (*SF3A1*, *SF3B1* and *PRPF40B*), six components of the splicing machinery were mutated in 16 out of the 29 cases (55.2%) in a mutually exclusive manner (Fig. 1, Supplementary Fig. 6 and Supplementary Table 2).

Frequent mutations in splicing machinery

RNA splicing is accomplished by a well-ordered recruitment, rearrangement and/or disengagement of a set of small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, and either U4/5/6 or U11/12), as well as many other protein components onto the pre-mRNAs. Notably, the mutated components of the spliceosome were all engaged in the initial steps of RNA splicing, except for PRPF40B, whose functions in RNA splicing are poorly defined. Making physical interactions with SF1 and a serine/arginine-rich (SR) protein, such as SRSF1 or SRSF2, the U2 auxiliary factor (U2AF) that consists of the U2AF65 (U2AF2)–U2AF35 heterodimer, is involved in the recognition of the 3′ splice site (3′SS) and its nearby polypyrimidine tract, which is thought to be required for the subsequent recruitment of the U2 snRNP, containing SF3A1 as well as SF3B1, to establish the splicing A complex (Fig. 1)[19]. ZRSR2 (or Urp), is another essential component of the splicing machinery. Showing a close structural similarity to U2AF35, ZRSR2 physically interacts with U2AF65, as well as SRSF1 and SRSF2, with a distinct function from its homologue, U2AF35 (ref. 20).

To confirm and extend the initial findings in the whole-exome sequencing, we studied mutations of the above six genes together with three additional spliceosome-related genes, including *U2AF65*, *SF1* and *SRSF1*, in a large series of myeloid neoplasms ($N = 582$) using a high-throughput mutation screen of pooled DNA followed by confirmation/identification of candidate mutations (refs 21 and 22 and Supplementary Methods II).

In total, 219 mutations were identified in 209 out of the 582 specimens of myeloid neoplasms through validating 313 provisional positive events in the pooled DNA screen (Supplementary Tables 4 and 5). The mutations among four genes, *U2AF35* ($N = 37$), *SRSF2* ($N = 56$), *ZRSR2* ($N = 23$) and *SF3B1* ($N = 79$), explained most of the mutations with much lower mutational rates for *SF3A1* ($N = 8$), *PRPF40B* ($N = 7$), *U2AF65* ($N = 4$) and *SF1* ($N = 5$) (Fig. 2). Mutations of the splicing machinery were highly specific to diseases showing myelodysplastic features, including MDS either with (84.9%) or without (43.9%) increased ring sideroblasts, chronic myelomonocytic leukaemia (CMML) (54.5%), and therapy-related AML or AML with myelodysplasia-related changes (25.8%), but were rare in *de novo* AML (6.6%) and myeloproliferative neoplasms (MPN) (9.4%) (Fig. 3a). The mutually exclusive pattern of the mutations in these splicing pathway genes was confirmed in this large case series, suggesting a common impact of these mutations on RNA splicing and the pathogenesis of myelodysplasia (Fig. 3b). The frequencies of mutations showed significant differences across disease types. Surprisingly, *SF3B1* mutations were found in the majority of the cases with MDS characterized by increased ring sideroblasts, that is, refractory anaemia with ring sideroblasts (RARS) (19/23 or 82.6%) and refractory cytopenia with multilineage dysplasia with ≥ 15% ring sideroblasts (RCMD-RS) (38/50 or 76%) with much lower mutation frequencies in other myeloid neoplasms. RARS and RCMD-RS account

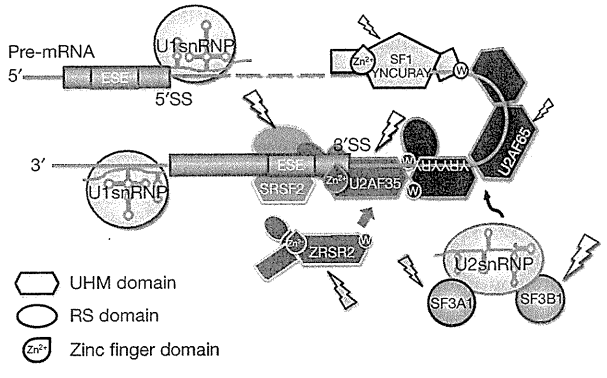

Figure 1 | Components of the splicing E/A complex mutated in myelodysplasia. RNA splicing is initiated by the recruitment of U1 snRNP to the 5′SS. SF1 and the larger subunit of the U2 auxiliary factor (U2AF), U2AF65, bind the branch point sequence (BPS) and its downstream polypyrimidine tract, respectively. The smaller subunit of U2AF (U2AF35) binds to the AG dinucleotide of the 3′SS, interacting with both U2AF65 and a SR protein, such as SRSF2, through its UHM and RS domain, comprising the earliest splicing complex (E complex). ZRSR2 also interacts with U2AF and SR proteins to perform essential functions in RNA splicing. After the recognition of the 3′SS, U2 snRNP, together with SF3A1 and SF3B1, is recruited to the 3′SS to generate the splicing complex A. The mutated components in myelodysplasia are indicated by arrows.

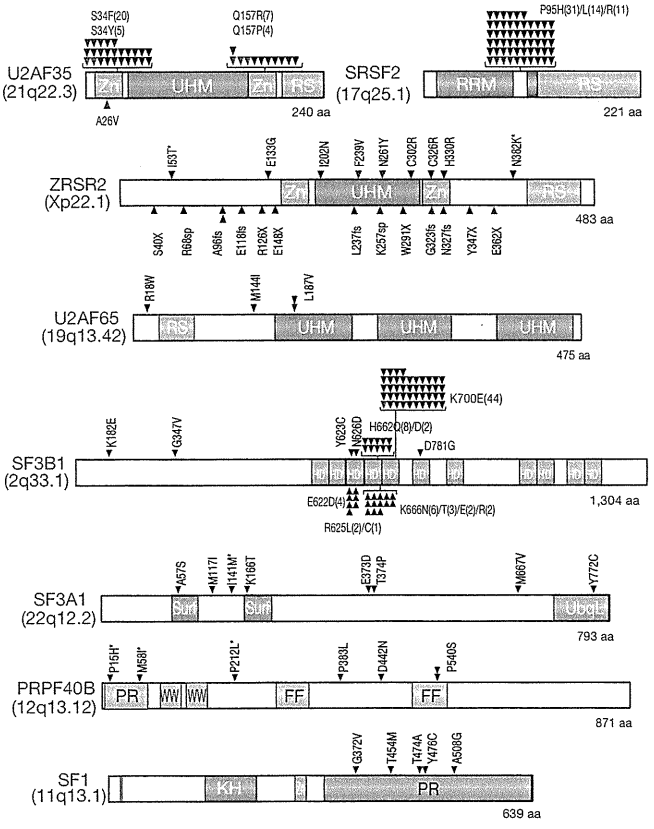

Figure 2 | Mutations of multiple components of the splicing machinery. Each mutation in the eight spliceosome components is shown with an arrowhead. Confirmed somatic mutations are discriminated by red arrows. Known domain structures are shown in coloured boxes as indicated. Mutations predicted as SNPs by MutationTaster (http://www.mutationtaster.org/) are indicated by asterisks. The number of each mutation is indicated in parenthesis. ZRSR2 mutations in females are shown in blue.

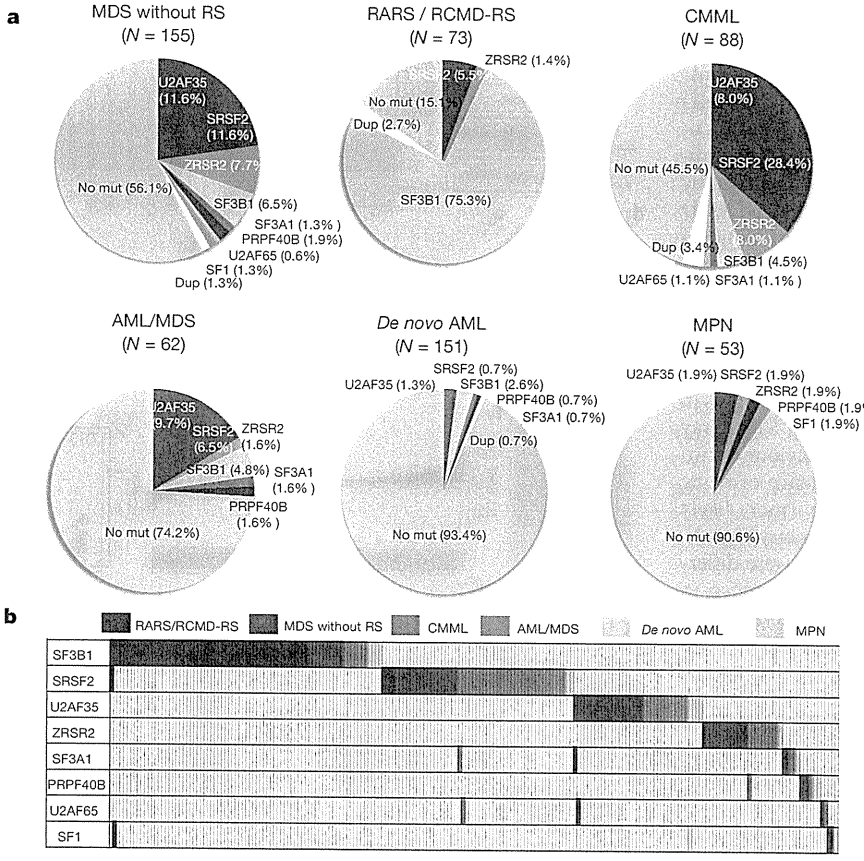

Figure 3 | Frequencies and distribution of spliceosome pathway gene mutations in myeloid neoplasms. a, Frequencies of spliceosome pathway mutations among 582 cases with various myeloid neoplasms. **b**, Distribution of mutations in eight spliceosome genes, where diagnosis of each sample is shown by indicated colours.

for 4.3% and 12.9% of MDS cases, respectively, where deregulated iron metabolism has been implicated in the development of refractory anaemia[23]. With such high mutation frequencies and specificity, the *SF3B1* mutations were thought to be almost pathognomonic to these MDS subtypes characterized by increased ring sideroblasts, and strongly implicated in the pathogenesis of MDS in these categories. Less conspicuously but significantly, *SRSF2* mutations were more frequent in CMML cases (Fig. 3 and Supplementary Table 4). Thus, although commonly involving the E/A splicing complexes, different mutations may still have different impacts on cell functions, contributing to the determination of discrete disease phenotypes. For example, studies have demonstrated that *SRSF2* was also involved in the regulation of DNA stability and that depletion of SRSF2 can lead to genomic instability[24]. Of interest in this context, regardless of disease subtypes, samples with *SRSF2* mutations were shown to have significantly more mutations of other genes compared with *U2AF35* mutations ($P = 0.001$, multiple regression analysis) (Supplementary Table 6 and Supplementary Fig. 7).

Notably, with a rare exception of A26V in a single case, the mutations of U2AF35 exclusively involved two highly conserved amino acid positions (S34 or Q157) within the amino- and the carboxyl-terminal zinc finger motifs flanking the U2AF homology motif (UHM) domain. SRSF2 mutations exclusively occurred at P95 within an intervening sequence between the RNA recognition motif (RRM) and arginine/serine-rich (RS) domains (Fig. 2 and Supplementary Figs 8 and 9). Similarly, SF3B1 mutations predominantly involved K700 and, to a lesser extent, K666, H662 and E622, which are also conserved across species (Fig. 2 and Supplementary Fig. 10). The involvement of recurrent amino acid positions in these spliceosome genes strongly indicated a gain-of-function nature of these mutations, which has been a well-documented scenario in other oncogenic mutations[25]. On the other hand, the 23 mutations in *ZRSR2* (Xp22.1) were widely distributed along the entire coding region (Fig. 2). Among these, 14 mutations were nonsense or frameshift changes, or involved splicing donor/acceptor sites that caused either a premature truncation or a large structural change of the protein, leading to loss-of-function. Combined with their strong male preference for the mutation (14/14 cases), *ZRSR2* most likely acts as a tumour suppressor gene with an X-linked recessive mode of genetic action. The remaining nine *ZRSR2* mutations were missense changes and found in both males (six cases) and females (three cases), whose somatic origin was only confirmed in two cases. However, neither the dbSNP database (build131 and 132) nor the 1000 Genomes database (May 2011 snp calls) contained these missense nucleotides, suggesting that many, if not all, of these missense changes are likely to represent functional somatic changes, especially those found in males. Interrogation of these hot spots for mutations in *U2AF35* and *SRSF2* found no mutations among lymphoid neoplasms, including acute lymphoblastic leukaemia ($N = 24$) or non-Hodgkin's lymphoma ($N = 87$) (data not shown).

RNA splicing and spliceosome mutations

Because the splicing pathway mutations in myelodysplasia widely and specifically affect the major components of the splicing complexes E/A in a mutually exclusive manner, the common consequence of these mutations is logically the impaired recognition of 3′SSs that would lead to the production of aberrantly spliced mRNA species. To appreciate this and also to gain an insight into the biological/biochemical impact of these splicing mutations, we expressed the wild-type and the mutant (S34F) U2AF35 in HeLa cells using retrovirus-mediated gene transfer with enhanced green fluorescent protein (EGFP) marking (Fig. 4a and Supplementary Methods III) and examined their effects on gene expression in these cells using GeneChip Human genome U133 plus 2.0 arrays (Affymetrix), followed by gene set enrichment analysis (GSEA) (Supplementary Methods IV)[26]. Intriguingly, the GSEA disclosed a significant enrichment of the genes on the nonsense-mediated mRNA decay (NMD) pathway among the significantly upregulated genes in mutant *U2AF35*-transduced HeLa cells (Fig. 4b, Supplementary Fig. 11a and Supplementary Table 7), which was

confirmed by quantitative polymerase chain reactions (qPCR) (Fig. 4c and Supplementary Methods 5V). A similar result was also observed for the gene expression profile of an MDS-derived cell line (TF-1) transduced with the S34F mutant (Supplementary Figs 11b, c). The NMD activation by the mutant U2AF35 was suppressed significantly by the co-overexpression of the wild-type protein (Supplementary Fig. 11d), indicating that the effect of the mutant protein was likely to be mediated by inhibition of the functions of the wild-type protein. Given that the NMD pathway, known as mRNA surveillance, provides a post-transcriptional mechanism for recognizing and eliminating abnormal transcripts that prematurely terminate translation[27], the result of the GSEA analyses indicated that the mutant U2AF35 induced abnormal RNA splicing in HeLa and TF-1 cells, leading to the generation of unspliced RNA species having a premature stop codon and induction of the NMD activity.

To confirm this, we next performed whole transcriptome analysis in these cells using the GeneChip Human exon 1.0 ST Array (Affymetrix), in which we differentially tracked the behaviour of two discrete sets of probes showing different level of evidence of being exons, that is, 'Core' (authentic exons) and 'non-Core' (more likely introns) sets (Supplementary Methods IV and Supplementary Fig. 12). As shown in Fig. 4d, the Core and non-Core set probes were differentially enriched among probes showing significant difference in expression between wild-type and mutant-transduced cells (false discovery rate (FDR) = 0.01). The Core set probes were significantly enriched in those probes significantly downregulated in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells, whereas the non-Core set probes were enriched in those probes significantly upregulated in mutant *U2AF35*-transduced cells (Fig. 4e). The significant differential enrichment was also demonstrated, even when all probe sets were included (Fig. 4f). Moreover, the significantly differentially expressed Core set probes tended to be up- and downregulated in wild-type and mutant *U2AF35*-transduced cells compared with mock-transduced cells, respectively, and vice versa for the differentially expressed non-Core set probes (Fig. 4e). Combined, these exon array results indicated that the wild-type U2AF35 correctly promoted authentic RNA splicing, whereas the mutant U2AF35 inhibited this processes, rendering non-Core and therefore, more likely intronic sequences to remain unspliced.

The abnormal splicing in mutant *U2AF35*-transduced cells was more directly demonstrated by sequencing mRNAs extracted from HeLa cells, in which expression of the wild-type and mutant (S34F) U2AF35 were induced by doxycycline. First, after adjusting by the total number of mapped reads, the wild-type *U2AF35*-transduced cells showed an increased read counts in the exon fraction, but reduced counts in other fractions, compared with mutant *U2AF35*-transduced cells (Fig. 4g). The reads from the mutant-transduced cells were mapped to broader genomic regions compared with those from the wild-type *U2AF35*-transduced cells, which were largely explained by non-exon reads (Fig. 4h). Finally, the number of those reads that encompassed the authentic exon/intron junctions was significantly increased in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells (Fig. 4i and Supplementary Methods VI). These results clearly demonstrated that failure of splicing ubiquitously occurred in mutant *U2AF35*-transduced cells. A typical example of abnormal splicing in mutant-transduced cells and the list of significantly unspliced exons are shown in Supplementary Fig. 13 and Supplementary Table 8, respectively.

Biological consequence of U2AF35 mutations

Finally, we examined the biological effects of compromised functions of the E/A splicing complexes. First, TF-1 and HeLa cells were transduced with lentivirus constructs expressing either the S34F U2AF35 mutant or wild-type U2AF35 under a tetracycline-inducible promoter (Fig. 5a and Supplementary Figs 14a and 15a), and cell proliferation was examined after the induction of their expression. Unexpectedly, after the induction of gene expression with

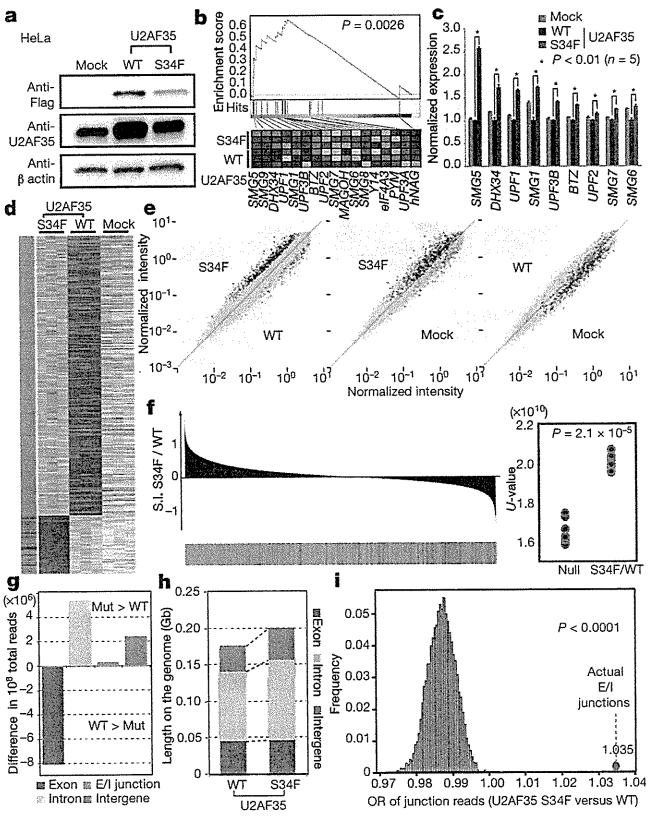

Figure 4 | Altered RNA splicing caused by a U2AF35 mutant. **a**, Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**, The GSEA demonstrating a significant enrichment of the set of 17 NMD pathway genes among significantly differentially expressed genes between wild-type and mutant *U2AF35*-transduced HeLa cells. The significance of the gene set was empirically determined by 1,000 gene-set permutations. **c**, The confirmation of the microarray analysis for the expression of nine genes that contributed to the core enrichment in the NMD gene set. Means ± s.e. are provided for the indicated NMD genes. *P* values were determined by the Mann–Whitney *U* test. **d**, Significantly upregulated and downregulated probe sets (FDR = 0.01) in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells in triplicate exon array experiments are shown in a heat map. The origin of each probe set is depicted in the left lane, where red and green bars indicate the Core and non-Core sets, respectively. **e**, Pair-wise scatter plots of the normalized intensities of entire probe sets (grey) across different experiments. The Core and non-Core set probes that were significantly differentially expressed between the wild-type and mutant *U2AF35*-transduced cells are plotted in red and green, respectively. **f**, Distribution of the Core (red) and non-Core (green) probe sets within the entire probe sets ordered by splicing index (S.I.; Supplementary Methods IV), calculated between wild-type and mutant *U2AF35*-transduced cells. In the right panel, the differential enrichment of both probe sets was confirmed by Mann–Whitney *U* test. **g**, Difference in read counts for the indicated fractions per 10^8 total reads in RNA sequencing between wild-type and mutant U2AF35-expressing HeLa cells analysis. Increased/decreased read counts in mutant U2AF35-expressing cells are plotted upward/downward, respectively. **h**, Comparison of the genome coverage by the indicated fractions in wild-type- and mutant-U2AF35-expressing cells. The genome coverage was calculated for each fraction within the 10^8 reads randomly selected from the total reads and averaged for ten independent selections. **i**, The odds ratio of the junction reads within the total mapped reads was calculated between the two experiments (red circle), which was evaluated against the 10,000 simulated values under the null hypothesis (histogram in blue).

doxycycline, the mutant *U2AF35*-transduced cells, but not the wild-type *U2AF35*-transduced cells, showed reduced cell proliferation (Fig. 5b and Supplementary Fig. 15b) with a marked increase in the G2/M fraction (G2/M arrest) together with enhanced apoptosis as

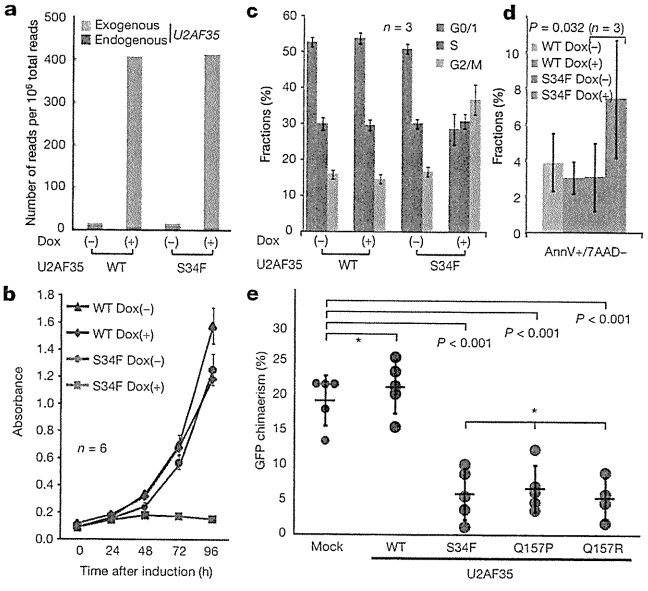

Figure 5 | Functional analysis of mutant U2AF35. **a**, Expression of endogenous and exogenous *U2AF35* transcripts in HeLa cells before and after induction determined by RNA sequencing. *U2AF35* transcripts were differentially enumerated for endogenous and exogenous species, which were discriminated by the Flag sequence. **b**, Cell proliferation assays of *U2AF35*-transduced HeLa cells, where cell numbers were measured using cell-counting apparatus and are plotted as mean absorbance ± s.d. **c**, The flow cytometry analysis of propidium iodide (PI)-stained HeLa cells transduced with the different U2AF35 constructs. Mean fractions ± s.d. in G0/G1, S and G2/M populations after the induction of U2AF35 expression are plotted. **d**, Fractions of the annexin V-positive (AnnV+) populations among the 7-amino-actinomycin D (7AAD)-negative population before and after the induction of U2AF35 expression are plotted as mean ± s.d. for indicated samples. The significance of difference was determined by paired *t*-test. **e**, Competitive reconstitution assays for CD34-negative KSL cells transduced with indicated U2AF35 mutants. Chimaerism in the peripheral blood 6 weeks after transplantation are plotted as mean %EGFP-positive Ly5.1 cells ± s.d., where outliers were excluded from the analysis. The significance of differences was evaluated by the Grubbs test with Bonferroni's correction for multiple testing. *not significant.

indicated by the increased sub-G1 fraction and annexin V-positive cells (Fig. 5c, d, Supplementary Fig. 14b and Supplementary Methods VI). To confirm the growth-suppressive effect of U2AF35 mutants *in vitro*, a highly purified haematopoietic stem cell population ($CD34^-$c-$Kit^+Sca1^+Lin^-$, $CD34^-KSL$) prepared from C57BL/6 (B6)-Ly5.1 mouse bone marrow[28] was retrovirally transduced with either the mutant (S34F, Q157P and Q157R) or wild-type *U2AF35*, or the mock constructs, each harbouring the EGFP marker gene (Supplementary Fig. 16). The ability of these transduced cells to reconstitute the haematopoietic system was tested in a competitive reconstitution assay. The transduced cells were mixed with whole bone marrow cells from B6-Ly5.1/5.2 F1 mice, transplanted into lethally irradiated B6-Ly5.2 recipients, and peripheral blood chimaerism derived from EGFP-positive cells was assessed 6 weeks after transplantation by flow cytometry. We confirmed that each recipient mouse received comparable numbers of EGFP-positive cells among the different retrovirus groups by estimating the percentage of EGFP-positive cells and overall proliferation in transduced cells by *ex vivo* tracking. Also no significant difference was observed in their homing capacity to bone marrow as assessed by transwell migration assays (Supplementary Fig. 17). As shown in Fig. 5e, the wild-type *U2AF35*-transduced cells showed a slightly higher reconstitution capacity than the mock-transduced cells. On the other hand, the recipients of the cells transduced with the various *U2AF35* mutants showed significantly lower EGFP-positive cell chimaerism than those of either the mock- or the wild-type *U2AF35*-transduced cells, indicating a compromised reconstitution capacity of the haematopoietic stem/progenitor cells expressing the U2AF35 mutants. In summary, these mutants lead to loss-of-function of U2AF35 most probably by acting in a dominant-negative fashion to the wild-type protein.

Discussion

Our whole-exome sequencing study unexpectedly unmasked a complexity of novel pathway mutations found in approximately 45% to 85% of myelodysplasia patients depending on the disease subtypes, which affected multiple but distinctive components of the splicing machinery and, as such, demonstrated the unquestionable power of massively parallel sequencing technologies in cancer research.

The RNA splicing system comprises essential cellular machinery, through which eukaryotes can achieve successful transcription and guarantee the functional diversity of their protein species using alternative splicing in the face of a limited number of genes[29]. Accordingly, the meticulous regulation of this machinery should be indispensable for the maintenance of cellular homeostasis[30], deregulation of which causes severe developmental abnormalities[31,32]. The current discovery of frequent mutations of the splicing pathway in myelodysplasia, therefore, represents another remarkable example that illustrates how cancer develops by targeting critical cellular functions. It also provides an intriguing insight into the mechanism of 'cancer specific' alternative splicing, which have long been implicated in the development of cancer, including MDS and other haematopoietic neoplasms[33,34].

In myelodysplasia, the major targets of spliceosome mutations seemed to be largely confined to the components of the E/A splicing complex, among others to *SF3B1*, *SRSF2*, *U2AF35* and *ZRSR2*, and to a lesser extent, to *SF3A1*, *SF1*, *U2AF65* and *PRPF40B*. The broad coverage of the wide spectrum of spliceosome components in our exome sequencing was likely to preclude frequent involvement of other components on this pathway (Supplementary Fig. 18). The surprising frequency and specificity of these mutations in this complex, together with the mutually exclusive manner they occurred, unequivocally indicate that the compromised function of the E/A complex is a hallmark of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS.

Given the critical functions of the E/A splicing complex on the precise 3′SS recognition, the logical consequence of these relevant mutations would be the impaired splicing involving diverse RNA species. In fact, when expressed in HeLa cells, the mutant U2AF35 induced global abnormalities of RNA splicing, leading to increased production of transcripts having unspliced intronic sequences. On the other hand, the functional link between the abnormal splicing of RNA species and the phenotype of myelodysplasia is still unclear. Mutant U2AF35 seemed to suppress cell growth/proliferation and induce apoptosis rather than confer a growth advantage or promote clonal selection. *ZRSR2* knockdown in HeLa cells has been reported to also result in reduced viability, arguing for the common consequence of these pathway mutations[35]. These observations suggested that the oncogenic actions of these splicing pathway mutations are distinct from what is expected for classical oncogenes, such as mutated kinases and signal transducers, but could be more related to cell differentiation. Of note in this regard, the commonest clinical presentation of MDS is severe cytopenia in multiple cell lineages due to ineffective haematopoiesis with increased apoptosis rather than unlimited cell proliferation[1]. In this regard, lessons may be learned from the recent findings on the pathogenesis of the 5q− syndrome, where haploinsufficiency of *RPS14* leads to increased apoptosis of erythroid progenitors, but not myeloproliferation[36,37].

A lot of issues remain to be answered, however, to establish the functional link between these splicing pathway mutations and the

pathogenesis of MDS, where the broad spectrum of RNA species affected by impaired splicing hampers identification of responsible gene targets. Moreover, the mutated components of the splicing machinery have distinct function of their own other than direct regulation of RNA splicing, involved in elongation and DNA stability, which may be important to determine specific disease phenotypes. Clearly, more studies are required to answer these questions through understanding of the molecular basis of their oncogenic actions.

METHODS SUMMARY

Whole-exome sequencing of paired tumour/normal DNA samples from the 29 patients was performed after informed consent was obtained. SNP array-based copy number analysis was performed as previously described[17,18]. Mutation analysis of the splicing pathway genes in a set of 582 myeloid neoplasms were performed by first screening mutations in PCR-amplified pooled targets from 12 individuals, followed by validation/identification of the candidate mutations within the corresponding 12 individuals by Sanger sequencing. Flag-tagged cDNAs of the wild-type and mutant *U2AF35* were generated by *in vitro* mutagenesis, constructed into a murine stem cell virus-based retroviral vector as well as a tetracycline-inducible lentivirus-based expression vector, and used for gene transfer to $CD34^-$KSL cells and cultured cell lines, with EGFP marking, respectively. Total RNA was extracted from wild-type or mutant *U2AF35*-transduced HeLa and TF-1 cells, and analysed on microarrays. RNA sequencing was performed according to the manufacturer's instructions (Illumina). Cell proliferation assays (MTT assays) on HeLa and TF-1 cells stably transduced with lentivirus U2AF35 constructs were performed in the presence or absence of doxycycline. For competitive reconstitution assays, $CD34^-$KSL cells collected from C57BL/6 (B6)-Ly5.1 mice were retrovirally transduced with various *U2AF35* constructs with EGFP marking, and transplanted with competitor cells (B6-Ly5.1/5.2 F1 mouse origin) into lethally irradiated B6-Ly5.2 mice 48 h after gene transduction. Frequency of EGFP-positive cells was assessed in peripheral blood by flow cytometry 6 weeks after the transplantation (Supplementary Methods VII). The primer sets used for validation of gene mutations and qPCR of NMD gene expression are listed in Supplementary Tables 9–11. A complete description of the materials and methods is provided in the Supplementary Information. This study was approved by the ethics boards of the University of Tokyo, Munich Leukaemia Laboratory, University Hospital Mannheim, University of Tsukuba, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital. Animal experiments were performed with approval of the Animal Experiment Committee of the University of Tokyo.

Received 7 June; accepted 24 August 2011.
Published online 11 September 2011.

1. Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature Rev. Cancer* **7,** 118–129 (2007).
2. Ma, X., Does, M., Raza, A. & Mayne, S. T. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* **109,** 1536–1542 (2007).
3. Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29,** 504–515 (2011).
4. Sanada, M. *et al.* Gain-of-function of mutated *C-CBL* tumour suppressor in myeloid neoplasms. *Nature* **460,** 904–908 (2009).
5. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40,** 722–729 (2008).
6. Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471,** 467–472 (2011).
7. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465,** 473–477 (2010).
8. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
9. Metzker, M. L. Sequencing technologies — the next generation. *Nature Rev. Genet.* **11,** 31–46 (2010).
10. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26,** 1135–1145 (2008).
11. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
12. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469,** 539–542 (2011).
13. Ley, T. J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).
14. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
15. Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43,** 309–315 (2011).
16. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
17. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65,** 6071–6079 (2005).
18. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81,** 114–126 (2007).
19. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718 (2009).
20. Tronchère, H., Wang, J. & Fu, X. D. A protein related to splicing factor $U2AF^{35}$ that interacts with $U2AF^{65}$ and SR proteins in splicing of pre-mRNA. *Nature* **388,** 397–400 (1997).
21. Bevilacqua, L. *et al.* A population-specific *HTR2B* stop codon predisposes to severe impulsivity. *Nature* **468,** 1061–1066 (2010).
22. Calvo, S. E. *et al.* High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nature Genet.* **42,** 851–858 (2010).
23. Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110,** 4385–4395 (2007).
24. Xiao, R. *et al.* Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol. Cell. Biol.* **27,** 5393–5402 (2007).
25. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42,** 181–185 (2010).
26. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
27. Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5,** 89–99 (2004).
28. Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nature Protocols* **1,** 2979–2987 (2007).
29. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10,** 741–754 (2009).
30. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21,** 708–718 (2007).
31. He, H. *et al.* Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332,** 238–240 (2011).
32. Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332,** 240–243 (2011).
33. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24,** 2343–2364 (2010).
34. Pajares, M. J. *et al.* Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol.* **8,** 349–357 (2007).
35. Shen, H., Zheng, X., Luecke, S. & Green, M. R. The U2AF35-related protein Urp contacts the 3′ splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* **24,** 2389–2394 (2010).
36. Barlow, J. L. *et al.* A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q− syndrome. *Nature Med.* **16,** 59–66 (2010).
37. Ebert, B. L. *et al.* Identification of *RPS14* as a $5q^-$ syndrome gene by RNA interference screen. *Nature* **451,** 335–339 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology, and also by the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)', initiated by the Council for Science and Technology Policy (CSTP). pGCDNsamIRESEGFP vector was a gift from M. Onodera. We thank Y. Mori, O. Hagiwara, M. Nakamura and N. Mizota for their technical assistance. We are also grateful to K. Ikeuchi and M. Ueda for their continuous encouragement throughout the study.

Author Contributions Y.Sh., Y.Sa., A.S.-O., Y.N., M.N., G.C., R.K. and S.Miyano were committed to bioinformatics analyses of resequencing data. M.Sa., A.S.-O. and Y.Sa. performed microarray experiments and their analyses. R.Y., T.Y., M.O., M.Sa., A.K., M.Sh. and H.N. were involved in the functional analyses of U2AF35 mutants. N.O., M.S.-Y., K.I., H.M., W.-K.H., F.N., D.N., T.H., C.H., S.Miyawaki, S.C., H.P.K. and L.-Y.S. collected specimens and were also involved in planning the project. K.Y., Y.N., Y.Su., A.S.-O. and S.S. processed and analysed genetic materials, library preparation and sequencing. K.Y., M.Sa., Y.Sh., A.S.-O., Y. Sa. and S.O. generated figures and tables. S.O. led the entire project and wrote the manuscript. All authors participated in the discussion and interpretation of the data and the results.

Author Information Sequence data have been deposited in the DDBJ repository under accession number DRA000433. Microarray data have been deposited in the GEO database under accession numbers GSE31174 (for SNP arrays), GSE31171 (for exon arrays) and GSE31172 (for expression arrays). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.ac.jp).

第71回日本血液学会学術集会
学会奨励賞受賞論文

MDSにおけるacquired uniparental disomyとc-CBL変異

真田　昌

Key words : SNP array, UPD, c-CBL

骨髄異形成症候群（Myelodysplastic syndromes, MDS）においては染色体コピー数の欠失や増加などのゲノム異常が約半数の症例で観察され，重要な予後指標となっている[1,2]。これらのゲノム異常が観察される領域にはMDSの病態に関連する標的遺伝子が存在すると考えられるが，その多くは同定には至っていない。近年，ゲノム解析技術は急速な進歩を遂げ，多くの腫瘍性疾患で遺伝子病態が明らかになりつつある。本稿では高密度SNPアレイを用いたMDSの網羅的なゲノム解析と11番染色体長腕aUPDの標的として同定された*c-CBL*変異について概説する。

高密度SNPアレイを用いたゲノム解析

高密度SNPアレイは大規模SNP（一塩基多型，single nucleotide polymorphism）タイピング用に開発されたが，マイクロアレイに搭載された各SNPプローブのシグナル強度とその領域のコピー数には高い相関がみられることから，ゲノムのコピー数解析に広く応用されている。マイクロアレイ技術の進歩により，非常に多数のプローブを搭載することが可能となっており，染色体分析に比し高解像度なコピー数解析を網羅的に行うことができる。コピー数解析についてはアレイCGH（comparative genomic hybridization）法によっても同様の解析結果が得られるが，SNPアレイにおいてはアレル別のコピー数解析が可能であり，網羅的なLOH（loss of heterozygosity）解析も同時に行うことができる。我々が開発したCNAG・AsCNARプログラム[3,4]は，ノイズの軽減と感度の向上が得られ，腫瘍細胞比率が10～20%程度しかない検体においてもゲノム異常の同定が可能である。また自己正常対照がない検体においても，他人由来の最適な正常対照プールを自動的に選択させることで，同等の解析結果を得ることも可能である。これらの利点は正常細胞の混入を避けることが困難，かつ自己正常対照検体を得ることも難しいMDS検体のゲノム解析において非常に有用な点である。染色体分析とは異なり，分裂核が得られない検体においてもゲノム解析が可能である。

MDSにおけるLOH

LOHは，古典的ながん抑制遺伝子の不活化メカニズムとして知られ[5]，新規がん抑制遺伝子の探索に用いられてきた。MDSにおいては，5番染色体長腕（5q）や7番染色体長腕（7q），20番染色体長腕（20q）などに代表される片アレルの欠失によるLOHの最小共通欠失領域から同領域の標的遺伝子探索がなされ，いくつかの候補遺伝子が報告されてきたが，未だ不明な点は多い。高密度SNPアレイによるゲノム解析では，5qや7qの欠失など染色体分析で同定される異常の他にも多くのゲノム異常が認められ，比較的狭い領域の欠失の他にも，コピー数の変化を伴わないLOH領域（Copy number neutral LOH（CN-LOH），片親性2倍体，Acquired uniparental disomy（aUPD））が新たに同定される（図1）。aUPDは，狭義のMDSにおいて10～25%程度の症例に観察され，慢性骨髄単球性白血病（chronic myelomonocytic leukemia, CMML）において，他病型に比し高頻度（35～40%）に認められる[6,7]。初期に報告されたSNPアレイを用いたMDSのゲノム解析ではSNPコールに基づく低感度なLOH判定にも関わらず，高頻度にaUPDが検出されたとの報告[8]もあるが，両親から同じアレルを引き継いでいる領域，すなわち先天的なホモ接合もカウントされている可能性が指摘されている。MDSクローンのみに生じた後天的な変化かは，自己正

受付：2011年2月2日
受理：2011年4月7日
東京大学医学部附属病院がんゲノミクスプロジェクト

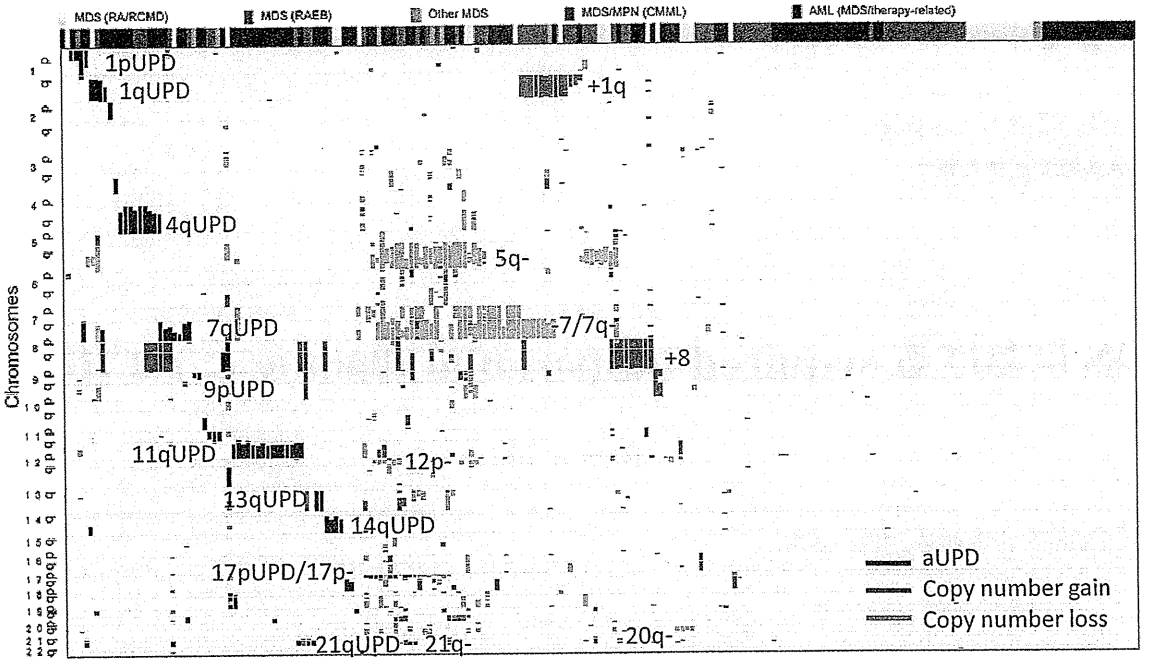

図 1 SNP アレイによる MDS および類縁疾患のゲノム解析結果（文献 7 を改変）
MDS108 例，CMML86 例，MDS から移行した AML28 例（WHO 分類による）計 222 例の臨床検体を用いた解析結果。縦に 1 番染色体短腕～22 番染色体長腕が順に並び，横に各症例の解析結果を示す。薄灰で表示しているのが欠失領域，濃灰が増加領域，黒が aUPD 領域。MDS に特徴的とされる 5 番染色体長腕，7 番染色体長腕，17 番染色体短腕，20 番染色体長腕の欠失，1 番染色体長腕，8 番染色体の増加などが認められる。aUPD は約 30％症例で観察され，CMML で頻度が高い。4 番染色体長腕，7 番染色体長腕，9 番染色体短腕，11 番染色体長腕，13 番染色体，17 番染色体短腕，21 番染色体長腕 aUPD の標的は，それぞれ *TET2*，*EZH2*，*JAK2*，*c-CBL*，*FLT3*，*TP53*，*RUNX1/AML1* 変異と考えられる（表 1）。

常細胞と比較した時のみ結論付けることが可能であるが，MDS においては正常細胞の取得は困難である。Heirichs ら[9]は同一の患者から採取した頬粘膜細胞を正常対照に解析を行っているが，SNP アレイで観察されるゲノム変化の一部は，頬粘膜においても観察されており，十分な注意が必要と考えられる。これまでの観察結果から造血器腫瘍における aUPD の多くは，somatic recombination により生じ[10]，染色体断端を含む形で認められる。一方で，コピー数の変化を伴わず，かつ染色体断端を含まない比較的短い領域のホモ接合は正常細胞の解析においてもしばしば観察され，腫瘍細胞特異的に認められることは稀である。また，臨床検体においては正常細胞の混入が少なからずあり，変化したアレルとともに正常細胞由来のアレルも同時に解析されるので，多くの場合は区別をすることが可能である。

MDS における aUPD

MDS における aUPD は 1 番染色体，4 番染色体長腕（4q），7q，11 番染色体長腕（11q），14 番染色体長腕（14q），17 番染色体短腕（17p），21 番染色体長腕（21q）等いくつかの染色体領域に集積をする[6, 7, 11, 12]（図 1・表 1）。7q，17p，21q は，欠失による LOH と共通して観察されるが，4q，11q，14q については，MDS において欠失例は稀である。一方，5q や 20q は欠失による LOH がしばしば認められるが，aUPD は稀である。17p，21q に aUPD を有する症例は，欠失例と同様に，それぞれ *TP53* および *RUNX1* 遺伝子変異を高頻度に伴っており，共通した分子病態が存在すると考えられる[7]。一方，7q-aUPD の標的遺伝子変異として 7q36.1 に存在する *EZH2* 遺伝子変異が報告された[13]が，本変異は aUPD 例においては高頻度に変異が認められるが，7q 欠失例における変異の頻度は稀であり，これまで報告されてきた 7q における最小共通欠失領域には含まれない。5q 欠失については，最近の研究結果からは *RPS14* もしくは *miR-145* と *miR-146a* のハプロ不全により MDS 様の病態がマウスモデルで再現されており[14, 15]，単一の遺伝子が標的ではないことも推測され，aUPD 例が稀であることとも一致すると考えられる。MDS および類縁疾患において高頻度に認められる aUPD 領域である 11q-aUPD の標的遺伝子として *C-CBL* 遺伝子変異は同定されたが，

表 1　骨髄系腫瘍における LOH と標的遺伝子変異

position	Diseases	LOH	Mutated Gene Target
1p13. 1	MDS	UPD	*Nras*
1p34	MPN, RARSt	UPD	*cMPL*
4q24	MDS, MPN	loss/UPD	*TET2*
5q	AML, MDS	loss	
7q35	MDS, MDS/MPN	UPD	*EZH2*
7q	AML, MDS	loss	
9p24	MPN	UPD	*JAK2*
11p13	AML	UPD	*WT1*
11q23. 3	MDS/MPN	UPD	*c-CBL*
13q12	AML	UPD	*FLT3*
14q	MDS, MDS/MPN	UPD	
17p13. 1	AML, MDS	loss/UPD	*TP53*
17q11. 2	JMML	UPD	*NF1*
19q13. 1	AML	UPD	*CEBPA*
20q	MDS, MPN	loss	
21q22. 3	AML, MDS	loss/UPD	*RUNX1*

C-CBL 変異は 11q-aUPD 例では高頻度に観察されるのに対し，非 aUPD 例においては頻度が低く，11q-aUPD と強い相関が認められる[7, 16, 17]。4q-aUPD の標的遺伝子として同定された *TET2* 変異[12, 18]や 7q-aUPD の標的である *EZH2* 変異[13, 19]は，aUPD が認められない症例においても，しばしば観察されるのとは対照的である。

c-CBL

c-CBL は，マウスにリンパ腫や白血病を発症させる Casitas-NS-リンパ腫ウィルスから単離された v-Cbl の相同遺伝子として同定された。c-CBL は，主として RING 型 E3 ユビキチンリガーゼとして作用することが知られている[20, 21]。ユビキチン化は細胞内のタンパク質の分解・機能制御において重要な翻訳後修飾であるが，c-CBL は，サイトカインなどのリガンド刺激で活性化された EGFR や FLT3，KIT などのチロシンキナーゼに TKB ドメインで結合し，E2 ユビキチン結合酵素を介してチロシンキナーゼをユビキチン化する。ユビキチン化されたチロシンキナーゼはプロテオソームなどで速やかに分解され，チロシンキナーゼを介したシグナルは終息する，すなわち c-CBL は同シグナルの重要な負の調整分子として機能すると考えられている。また c-CBL はチロシンキナーゼのみならず非常に多くのシグナル伝達分子と結合することが知られており，負の制御のみならずシグナル伝達においても重要な役割を担っていると推測される[20, 21]。

骨髄系腫瘍における *c-CBL* 変異

急性骨髄性白血病 AML で *c-CBL* 変異例の報告がされ[22, 23]，続いて MDS[7, 16]，骨髄増殖性疾患 MPN[17]での報告もされているが，変異例の多くは MDS/MPN (myelodysplastic/myeloproliferative neoplasms) 例である。MDS/MPN は，FAB 分類においては MDS の一亜型として扱われてきたが，白血球増多や肝脾腫を認めら