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 22～24 年度　総合研究報告書

研究代表者　真田　昌

平成 25(2013)年　4 月

目　　次

Ⅰ　総合研究報告

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)
総合研究報告書

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
東京大学医学部附属病院　特任助教

研究要旨

骨髄異形成症候群（MDS）においてエピゲノム異常が生じていることが知られ、エピゲノム修飾を標的とした治療薬剤が臨床応用されている。さらに、近年、骨髄異形成症候群（MDS）においてエピゲノム修飾分子にゲノム異常が生じていることが明らかとなった。しかし、MDS におけるエピゲノム異常が、エピゲノム修飾分子の遺伝子変異に起因しているのか、明らかではなく、本研究ではゲノム・エピゲノム解析を統合し、両異常の関連を明らかとすることを目指した。MDS および関連疾患の全エクソン・シーケンスによる網羅的な変異解析により、我々が新たに見出した RNA スプライシング関連分子の変異と並んで、*TET2* などエピゲノム関連遺伝子の体細胞変異が MDS 症例の半数以上の症例において観察されることを明らかとした。さらには、MDS において変異の報告がある遺伝子について 192 例の MDS で target capture sequence を行い、エピゲノム関連遺伝子など既知の標的遺伝子を中心とした変異プロファイリングを行った。さらに、変異プロファイルが明らかとなった症例群についてイルミナ社の HumanMethylation 450 BeadChip を用いて、網羅的な DNA メチル化解析を行い、ゲノム異常との比較検討を行った。MDS において、正常末梢血に比しメチル化を受けている傾向にある 4000 遺伝子について教師なしクラスタリング解析を行い、メチル化パターンの異なる、いくつかのサブクラスに階層化された。192 例において変異頻度が高いエピゲノム修飾関連遺伝子、*TET2*（35%）、*ASXL1*（20%）、*DNMT3A*（12%）、*IDH1/2*（6%）および *EZH2*（5%）について変異の有無（内は 192 例での変異頻度）、および臨床情報とクラスタリング結果を比較検討した。骨髄芽球の増加を伴う RAEB-1 および 2 と診断されている症例は、増加を伴わない病型に比し、解析した遺伝子群においてはメチル化されている傾向が高かった。DNA メチルトランスフェラーゼとしてシトシンのメチル化に直接的に関わる *DNMT3A* 遺伝子においては、変異の有無とクラスタリング結果には関連は認められなかったが、脱メチル化過程で重要な働きを有することが明らかとなりつつある *TET2* および *IDH* 変異例は、特徴的なメチル化パターンを示すクラスター群に集中する傾向が認められ、ゲノムレベルでの変異がエピゲノム異常に関わっていることが示唆された。

A. 研究目的

シトシンのメチル化に代表されるエピゲノム修飾は遺伝子の発現調整に寄与し、細胞の発生・分化の過程において重要な分子機構である。エピゲノム修飾の異常は、遺伝子発現の異常を来たし、様々ながん種において、発がんに関わっていると考えられている。更には、近年の遺伝子解析技術の進歩に伴い、非常に広範な腫瘍において、エピゲノム修飾分子の遺伝子変異が生じていることが明らかとなっている。単純には、エピゲノム修飾分子に生じたゲノム異常が、エピゲノム異常を来たし、遺伝子の発現異常により、細胞の腫瘍化を招くと想定ができるが、腫瘍細胞で観察されるエピゲノム修飾異常が、ゲノムの異常に起因するのかは明らかではない。MDS は高齢者に多い造血器腫瘍であるが、古くからメチル化異常に関する報告がされてきた。造血幹細胞移植以外に治癒が期待できる治療法はない難治性腫瘍であるが、脱メチル化剤が生命予後を改善することが明らかとなり、移植が

困難な高齢者を含め広く使用されつつある。また、MDS において、*TET2* や *EZH2*、*ASXL1* など、メチル化修飾やポリコーム複合体に属するエピゲノム修飾分子に遺伝子変異が、しばしば生じていることが明らかとなっている。そこで、本研究では、MDS の臨床検体を用いて、ゲノム異常とエピゲノム異常の関連について明らかとする。

B. 研究方法

49 例の MDS 検体について、次世代シーケンサーを用いた全エクソーム解析を行い、MDS クローンに生じた体細胞変異を網羅的に同定した。また 192 例（低リスク MDS 145 例、高リスク MDS 47 例）の MDS 検体について、エクソーム解析で複数例に同定された遺伝子や MDS など骨髄系腫瘍で報告のある遺伝子を中心に、アジレント社の Sure Select を用いた Target-capture シーケンスを行った。既存および in house SNP データベースを用いて SNP を除外し、更にアリル頻度から SNP と推測される SNV は除外した後に、遺伝子変異のプロファイリングを行った。更に、同症例についてイルミナ社の Human Methylation 450 BeadChip を用いて、網羅的な DNA メチル化解析を行った。メチル化データは、教師なし階層クラスタリング解析を行った。メチル化修飾に関わる遺伝子を中心に変異とメチル化パターンの関連を解析した。

（倫理面への配慮）

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、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、事前に東京大学ならびに検体提供施設の倫理委員会の承認を得た。また、研究対象者からは文書による同意を得ている。

C. 研究結果

49 例の MDS の全エクソン・シーケンスにより、MDS 細胞特異的な変異候補が同定され、456 個（9.3 個/例）のタンパクの構造変化を伴う体細胞性変異が確認された。12 例で *TET2* 遺伝子変異を認めた他、*EZH2*、*ASXL1*、*DNMT3A* など半数近い症例でエピゲノム修飾に関わる遺伝子群に変異が観察された。192 例の MDS 検体において変異頻度が高いエピゲノム修飾関連遺伝子、*TET2*（35%）、*ASXL1*（20%）、*DNMT3A*（12%）、*IDH1/2*（6%）および *EZH2*（5%）について変異の有無（内は 192 例での変異頻度）を明らかとした。また、EZH2 とともに PRC2 複合体を形成する *EED* 遺伝子に、MDS において低頻度ながら変異が認められることを新たに見出した。*EED* に変異が生じることにより、PRC2 複合体形成が障害され、PRC2 の機能低下を招き、*EZH2* 変異と同様の分子病態を示すと推測された（Ueda et.al. Leukemia 2012）。

次に、この 192 例について、イルミナ社の Human Methylation 450 BeadChip を用いて、網羅的にシトシンのメチル化状態を観察した。本アレイ上には、480 万箇所、17,990 遺伝子のプローブが搭載をされているが、遺伝子のプロモーター領域に存在し、遺伝子の発現調整に関わりがあると推測されるプローブに限って解析を行った。正常末梢血における各プローブのシグナルと比較したメチル化の程度を 3 段階にスコア化し、MDS において、正常末梢血に比しメチル化を受けている傾向にある 4000 遺伝子について教師なしクラスタリング解析を行い、メチル化パターンの異なる、いくつかのサブクラスに階層化された。192 例において変異頻度が高いエピゲノム修飾関連遺伝子、*TET2*、*ASXL1*、*DNMT3A*、*IDH1/2* および *EZH2* について変異の有無と臨床情報とクラスタリング結果を比較検討した。骨髄芽球の増加を伴う RAEB-1 および 2 と診断されている症例は、増加を伴わない病型に比し、解析した遺伝子群においてはメチル化されている傾向が高かった。DNA メチルトランスフェラーゼとしてシトシンのメチル化に直接的に関わる *DNMT3A* 遺伝子においては、変異の有無とクラスタリング結果には関連は認められなかったが、脱メチル化過程で重要な働きを有することが明らかとなりつつある *TET2* および *IDH* 変異例は、特徴的なメチル化パターンを示すクラスター群に集中する傾向が認められ、ゲノムレベルでの変異がエピゲノム異常に関わっていることが示唆された。

D. 考察

エピゲノム修飾の異常は様々ながん種で観察され、がん抑制遺伝子のプロモーター領域のメチル化修飾を介した発現低下など、重要な発がんメカニズムの一つとして認識されている。一方、次世代シーケンス技術に代表される近年の遺伝子解析技術の進歩に伴い、多くの腫瘍性疾患の主要な遺伝子変異が須らく明らかになりつつあるが、エピゲノム修飾分子に遺伝子変異が広範ながん種において生じていることが報告されている。しかし、ゲノムの異常である遺伝子

変異が、患者細胞において、エピゲノム修飾に関わっているのか、さらには如何にして腫瘍化に寄与しているかなど不明な点が多い。
MDS は造血幹細胞に由来する腫瘍性疾患であるが、メチル化異常が生じていることが古くから報告をされ、また、脱メチル化剤などエピゲノム修飾を標的とした治療薬剤の臨床応用がされている。更にはエピゲノム修飾遺伝子の変異が高頻度に生じていることが近年明らかとなっており、エピゲノム異常が病態に大きく関与していると推測される代表的な腫瘍である。本研究では、MDS においてエピゲノム修飾関連遺伝子における遺伝子変異は高頻度に観察され、エピゲノム修飾異常を招いていると考えられることを、*TET2* および *IDH1/2* 変異とメチル化異常について臨床検体の解析を通じて明らかとした。
全エクソン解析の結果、1 例あたり約 10 個の体細胞性変異が同定されるが、複数の症例に観察される変異遺伝子は、49 例の解析でも 23 遺伝子に過ぎず、MDS がヘテロな疾患群であることを加味しても、主要な変異遺伝子は、ほぼ明らかとなったと考えられる。この中で、RNA スプライシング関連遺伝子とエピゲノム修飾関連遺伝子が多数を占めており、MDS の遺伝子病態を考える上で、これらは重要な遺伝子異常であると考えられる。
我々の解析結果においても最も高頻度に変異が認められた *TET2* は、近年の研究により、メチル化シトシンの脱メチル化過程で重要なメチル化シトシンヒドロキシラーゼ活性を有していることが明らかとなり、また IDH 変異体は α ケトグルタル酸から 2 ヒドロキシグルタル酸への変換を介して、TET2 の酵素活性を阻害することが示されている。すなわち、*TET2* の不活化変異および *IDH* 変異は、ともに脱メチル化過程が障害を受け、メチル化状態が維持されることが予測され、我々のメチル化解析結果とも合致する。*TET2* および *IDH1/2* 変異と関連の高いクラスターを特徴づけるメチル化されている領域・遺伝子を検索することを現在進めており、本変異に導かれるエピゲノム修飾異常のメカニズムならびに標的遺伝子を通じた MDS の分子病態の解明が進むことが期待される。
脱メチル化剤は、造血幹細胞移植が困難な高齢者 MDS の治療において重要な位置づけとなっているが、奏功率は必ずしも高くなく、作用機序も不明であり、耐性化の問題も生じている。今後、脱メチル化剤投与前後の検体の経時的な解析を行うことで、脱メチル化剤投与によるメチル化の変化、有効例と無効例の違いを明らかにすることで、脱メチル化剤の作用機序を解明し、治療反応性を事前に予測する臨床上有用なバイオマーカーの確立も望まれる。
MDS における RNA スプライシング関連分子の変異は排他的に生じているのに対し、エピゲノム修飾関連遺伝子の変異は、*TET2* と *IDH1/2* の変異の重複例は既報の通りに稀であるが、*TET2* 変異と *ASXL1* 変異など、しばしば重複して観察され、アレル頻度からも、同一の細胞に変異が生じていると推測される。すなわち、遺伝子異常が、エピゲノム修飾全体そして遺伝子発現に与える影響は単純ではないことが推測をされる。今後、遺伝子発現解析やヒストン修飾の解析も含めた、より多層的な解析が必要であると思われる。

E. 結論
MDS において、RNA スプライシング分子変異と並んで、エピゲノム修飾分子の異常は、最も頻度の高いゲノム異常である。ゲノムレベルの異常である *TET2* や *IDH1/2* 変異が、メチル化異常を来たしていることが推測された。

F. 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. **Rinshouketueki.** 2011;52(6):342-9
3) Sato-Otsubo A, Sanada M, Ogawa S. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? **Semin Oncol.** 2012; 39(1):13-25

4) 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; 13(11):1035-42
5) Ueda T, Sanada M, Matsui H, Yamasaki N, Honda ZI, Shih LY, Mori H, Inaba T, Ogawa S, Honda H. EED mutants impair polycomb repressive complex 2 in myelodysplastic syndrome and related neoplasms. **Leukemia.** 2012;26(12):2557-60.
6) Hosokawa K, Katagiri T, Sugimori N, Ishiyama K, Sasaki Y, Seiki Y, Sato-Otsubo A, Sanada M, Ogawa S, Nakao S. Favorable outcome of patients who have 13q deletion: a suggestion for revision of the WHO 'MDS-U' designation. **Haematologica.** 2012; 97(12): 1845-9.
7) 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; 26: 1879-81.
8) Sanada M, Ogawa S. Genome-wide analysis of myelodysplastic syndromes. **Curr Pharm Des.** 2012;18:3163-9.
9) Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, Shiba N, Hayashi Y, Kume H, Homma Y, Sanada M, Ogawa S, Miyano S. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. **Nucleic Acids Res.** 2013 in press
10) Kon S, Minegishi N, Tanabe K, Watanabe T, Funaki T, Wong WF, Sakamoto D, Higuchi Y, Kiyonari H, Asano K, Iwakura Y, Fukumoto M, Osato M, Sanada M, Ogawa S, Nakamura T, Satake M. Smap1 deficiency perturbs receptor trafficking and predisposes mice to myelodysplasia. **J Clin Invest.** 2013;123(3):1123-37.
11) Kunishima S, Okuno Y, Yoshida K, Shiraishi Y, Sanada M, Muramatsu H, Chiba K, Tanaka H, Miyazaki K, Sakai M, Ohtake M, Kobayashi R, Iguchi A, Niimi G, Otsu M, Takahashi Y, Miyano S, Saito H, Kojima S, Ogawa S. ACTN1 Mutations Cause Congenital Macrothrombocytopenia. **Am J Hum Genet.** 2013; 92(3):431-8.

2. 学会発表

1)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

2)真田 昌 Novel pathway mutations in myelodysplasia revealed by high-throughput sequencing technology.
第71回日本癌学会学術総会
International session（招待講演）
2012年9月20日 札幌

G．知的所有権の取得状況

1. 特許取得
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
真田　昌	正常核型急性骨髄性白血病のゲノム異常	高久史麿 小澤敬也 坂田洋一 金倉　譲 小島勢二	Annual Review 2013 血液	中外医学社	東京都新宿区	2012年	80-86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

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
Sanada M, Ogawa S	Genome-wide analysis of myelodysplastic syndromes.	Curr Pharm Des.	18	3163-9.	2012
Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, Shiba N, Hayashi Y, Kume H, Homma Y, Sanada M, Ogawa S, Miyano S.	An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data.	Nucleic Acids Research		In press	2013
Kon S, Minegishi N, Tanabe K, Watanabe T, Funaki T, Wong WF, Sakamoto D, Higuchi Y, Kiyonari H, Asano K, Iwakura Y, Fukumoto M, Osato M, Sanada M, Ogawa S, Nakamura T, Satake M.	Smap1 deficiency perturbs receptor trafficking and predisposes mice to myelodysplasia.	J Clin Invest.	123(3)	1123-37	2013
Ueda T, Sanada M, Matsui H, Yamasaki N, Honda ZI, Shih LY, Mori H, Inaba T, Ogawa S, Honda H.	EED mutants impair polycomb repressive complex 2 in myelodysplastic syndrome and related neoplasms.	Leukemia	26(12)	2557-60	2012
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	26(8)	1879-81	2012

Hosokawa K, Katagiri T, Sugimori N, Ishiyama K, Sasaki Y, Seiki Y, Sato-Otsubo A, Sanada M, Ogawa S, Nakao S.	Favorable outcome of patients who have 13q deletion: a suggestion for revision of the WHO 'MDS-U' designation.	Haematologica	97	1845-9	2012

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

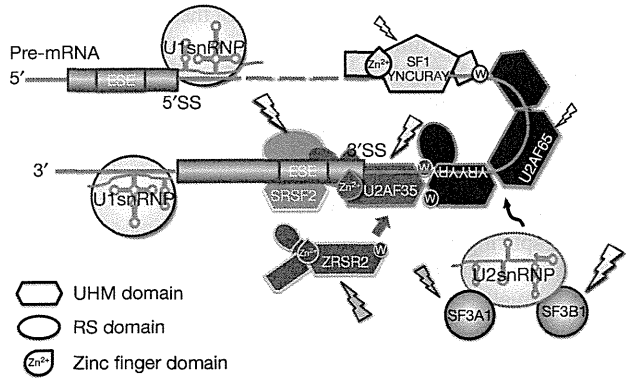

Figure 1 | Components of the splicing E/A complex mutated in myelodysplasia. RNA splicing is initiated by the recruitment of U1 snRNP to the 5′SS. SF1 and the larger subunit of the U2 auxiliary factor (U2AF), U2AF65, bind the branch point sequence (BPS) and its downstream polypyrimidine tract, respectively. 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.

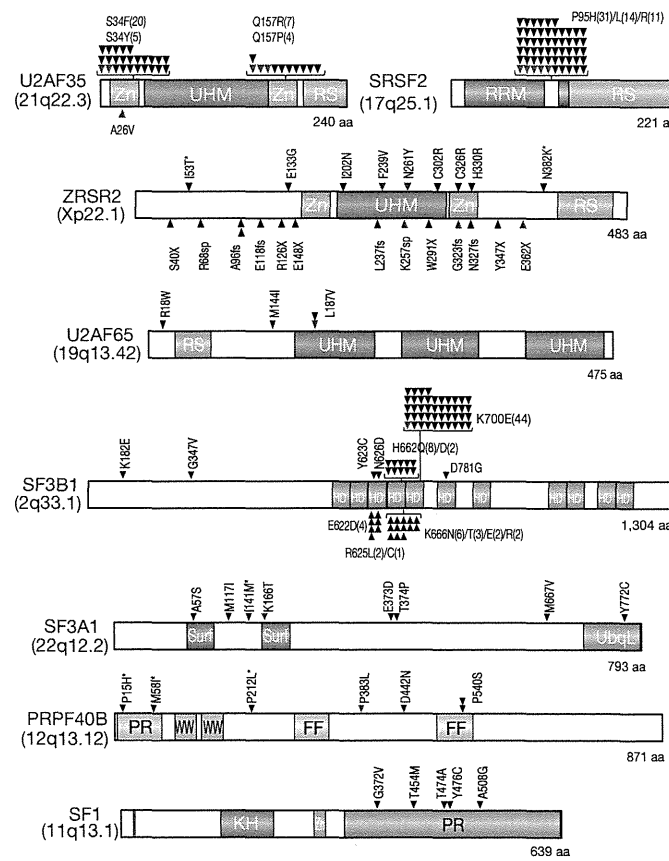

Figure 2 | Mutations of multiple components of the splicing machinery. Each mutation in the eight spliceosome components is shown with an arrowhead. Confirmed somatic mutations are discriminated by red arrows. Known domain structures are shown in coloured boxes as indicated. Mutations predicted as SNPs 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.

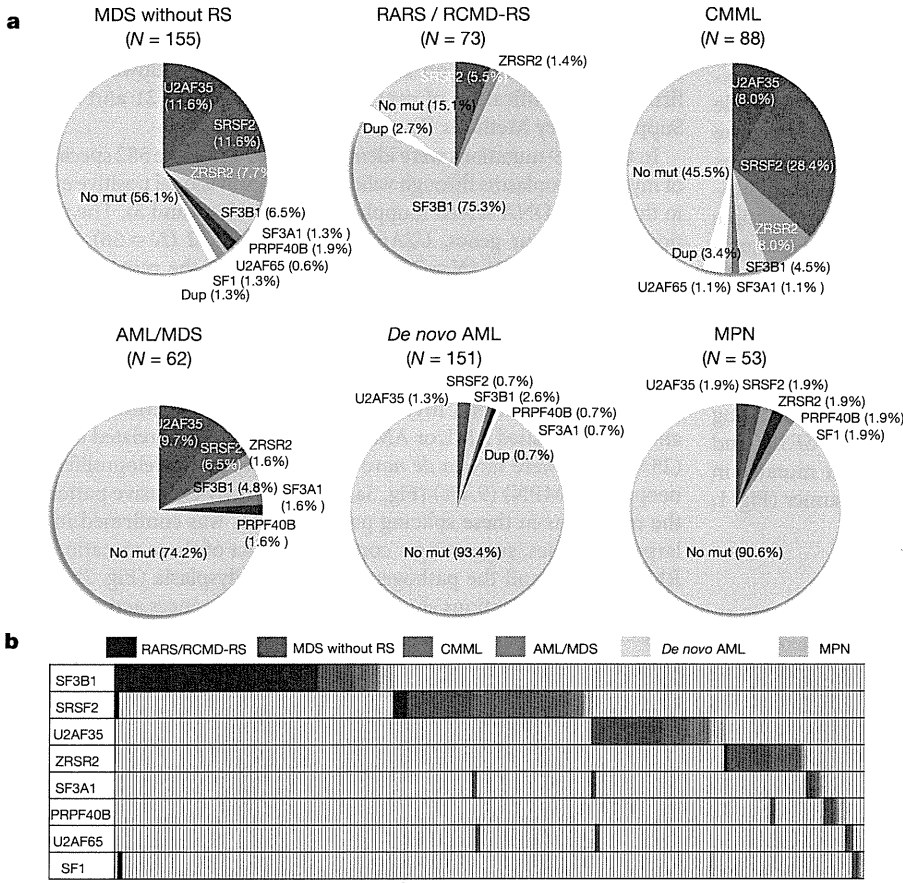

Figure 3 | Frequencies and distribution of spliceosome pathway gene mutations in myeloid neoplasms. **a**, Frequencies of spliceosome pathway mutations among 582 cases with various myeloid neoplasms. **b**, Distribution of mutations in eight spliceosome genes, where diagnosis of each sample is shown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

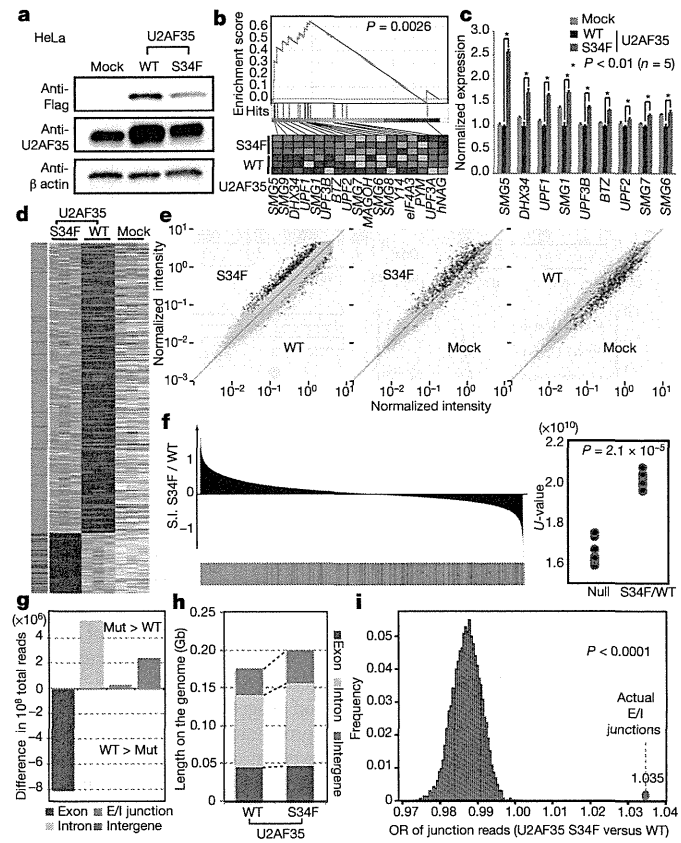

Figure 4 | Altered RNA splicing caused by a U2AF35 mutant. **a**, Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**, The GSEA demonstrating a significant enrichment of the set of 17 NMD 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

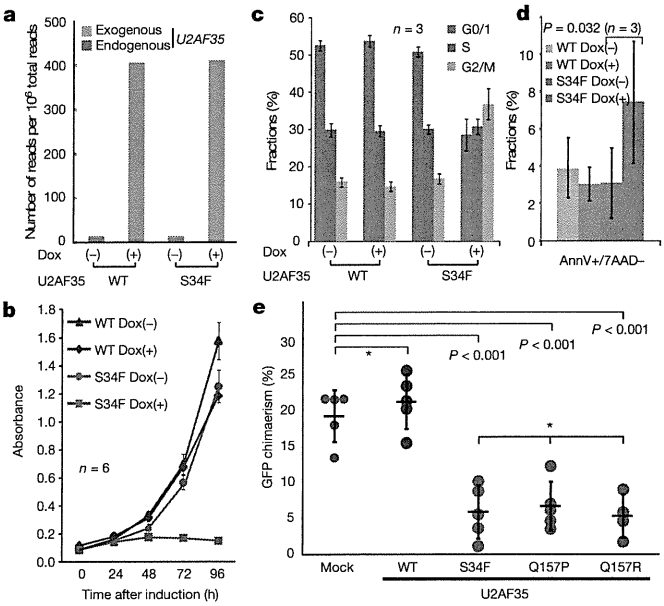

Figure 5 | Functional analysis of mutant U2AF35. **a**, Expression of endogenous and exogenous *U2AF35* transcripts in HeLa cells before and after induction determined by RNA sequencing. *U2AF35* transcripts were differentially enumerated for endogenous and exogenous species, which were discriminated 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-$Kit^+Scal^+Lin^-$, $CD34^-$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).

Single-Nucleotide Polymorphism Array Karyotyping in Clinical Practice: Where, When, and How?

Aiko Sato-Otsubo, Masashi Sanada, and Seishi Ogawa

Single-nucleotide polymorphism array (SNP-A) karyotyping is a new technology that has enabled genome-wide detection of genetic lesions in human cancers, including hematopoietic neoplasms. Taking advantage of very large numbers of allele-specific probes synthesized on microarrays at high density, copy number alterations as well as allelic imbalances can be sensitively detected in a genome-wide manner at unprecedented resolutions. Most importantly, SNP-A karyotyping represents the only platform currently available for genome-scale detection of copy neutral loss of heterozygosity (CN-LOH) or uniparental disomy (UPD), which is widely observed in cancer genomes. Although not applicable to detection of balanced translocations, which are commonly found in hematopoietic malignancies, SNP-A karyotyping technology complements and even outperforms conventional metaphase karyotyping, potentially allowing for more accurate genetic diagnosis of hematopoietic neoplasms in clinical practice. Here, we review the current status of SNP-A karyotyping and its application to hematopoietic neoplasms.
Semin Oncol 39:13-25

GENETIC ABNORMALITIES IN HEMATOLOGIC MALIGNANCIES

Since the discovery of the Philadelphia (Ph) chromosome in chronic myelogenous leukemia (CML) by Peter Nowell and David Hungerford in 1960,[1] hundreds of different genetic alterations/abnormalities have been identified and described in human cancers, including not only hematologic malignancies but also in a wide variety of solid cancers.[2] Human cancers show a diversity of genetic alterations, ranging from chromosome-scale lesions, such as translocations, gains/amplifications, and losses of large chromosomal segments, to small nucleotide substitutions, insertions, and deletions. Now it has been well established that the genetic alterations are central to the development of cancers, determining their biological or clinical behaviors. In fact, some genetic lesions, such as recurrent translocations, are highly specific to particular disease types or closely linked to tumor histologies, while others are commonly observed in a wide spectrum of cancer types, indicating more general roles of these genetic changes in carcinogenesis. Significantly, the information about these genetic lesions not only contributed to unmasking the underlying molecular pathogenesis of cancers but also enabled the development of novel diagnostics, therapeutics, and sensitive tumor monitoring that target these specific lesions.[3,4]

This has been best exemplified in hematologic cancers, in which underlying genetic changes have been most extensively studied."In particular, a number of disease-specific chromosomal translocations found in leukemias and lymphomas have been demonstrated to be critical genetic markers in clinical practice.[5-9] While several techniques have been developed to detect these genetic changes with different sensitivities and specificities for different purposes, probably the most widely used in clinical settings is metaphase karyotyping. The metaphase karyotyping technique was first developed during the 1960s and was soon introduced in both the experimental and clinical hematology fields. Since then, it has long been used as one of the indispensable clinical tests and research tools with which genetic alterations can be explored in a genome-

Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.
This work was supported in part by the Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, a Grant-in-Aid from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology.
Address correspondence to Seishi Ogawa, MD, PhD, Cancer Genomics Project, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. E-mail: sogawa-tky@umin.ac.jp
0270-9295/ - see front matter

doi:10.1053/j.seminoncol.2011.11.010

wide fashion, although resolutions are limited. Nevertheless, recent advances in genomics and microarray technologies have provided a new and, in a sense, more powerful alternative: SNP-array (SNP-A)-based analysis of cancer genomes or SNP-A karyotyping.[10-12] Here, we will focus on this novel genetic approach to hematologic oncology.

BASIC PRINCIPLES OF SNP-A–BASED COPY NUMBER ANALYSIS OF CANCER GENOMES

SNP-As were originally developed for large-scale SNP typing to enable genome-wide association studies (GWAS), in which more than hundreds of thousands of common SNPs across the entire genome are genotyped for thousands of specimens.[13,14] Currently, two SNP-A platforms are commercially available, which achieve highly paralleled genotyping of more than a million SNPs relying on hybridization to, and/or extension from, allele-specific oligonucleotide probes synthesized in high-density on array matrix (Affymetrix, Santa Clara, CA, GeneChip SNP Genotyping Arrays)[13] or numerous micro-beads (Illumina, San Diego, CA),[14] respectively. While making tremendous contributions to the recent achievements through a number of GWAS studies[15-17] both SNP-A technologies also have been applied with excellent results to genome-wide copy number analysis of cancer genomes.[18-21]

For the purpose of genotyping, the relative intensities of the two kinds of SNP-specific signals at individual SNP loci are evaluated to discriminate three possible genotypes, such as A/A, A/B, and B/B (Figure 1a). On the other hand, for copy number analysis, these signals are compared across all SNP loci to calculate genome-wide copy numbers, using "reference signal values" for diploid DNA (SNP-A karyotyping)[10-12] (Figure 1b). Note that like other DNA-based analyses, SNP-A karyotyping cannot determine cell ploidy precisely, which can only be enabled by cell-based analysis[22] (for more detail, see Ogawa et al[22]). The basic idea here is similar to array-based comparative genomic hybridization (array CGH), in that the hybridization signals from tumor DNA are compared to normal diploid signals at individual probe sites.[23,24] However, in SNP-A karyotyping, the comparisons are made between the corresponding two SNP-specific probes, which makes it possible to calculate allele-specific copy numbers (AsCNs) (Figure 1b). The AsCN analysis is a unique feature of SNP-A-based copy number analysis,[25] enabling sensitive detection of copy neutral loss of heterozygosity (CN-LOH) or uniparental disomies (UPD), which cannot be detected by metaphase karyotyping or array CGH (see below).[25] In addition, their high resolutions of analysis to precisely point out genetic targets with their positions, high-throughput sample processing with semi-automated experimental procedures, and cell division-independent nature of

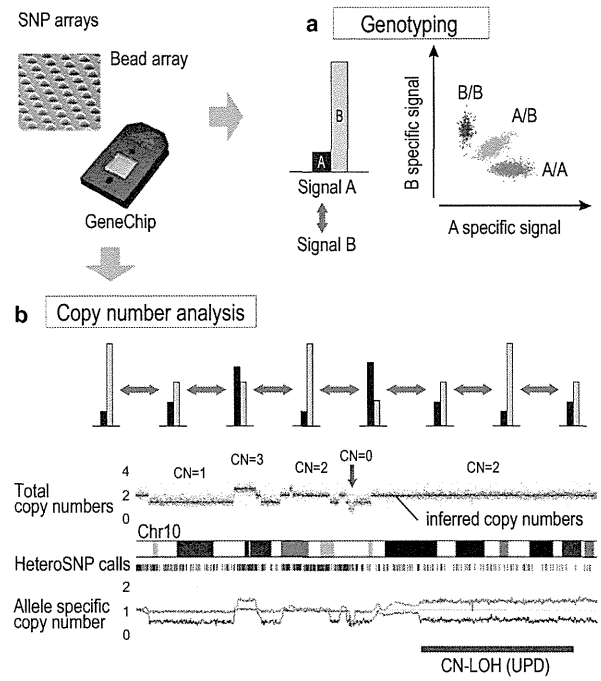

Figure 1. Principles of genotyping and copy number detection in SNP-A platforms. SNP arrays were originally designed for large-scale SNP typing, in which relative intensity of SNP-specific signals at each SNP site is compared to discriminate among three possible genotypes (a). On the other hand, SNP-specific signals can be used to calculate allele-specific copy numbers (AsCNs) by comparing them across the genomes in an allele-specific manner (SNP-A karyotyping). (b) A typical result of SNP-A karyotyping generated by CNAG software, where AsCNs are shown below the chromatogram (red and green lines), together with total genomic copy number plots on the top panel. The right end of the chromosome segment shows copy neutral LOH or UPD, as indicated by dissociated AsCN graphs with the normal total copy number (n = 2).

analysis, as well as computer-based detection of genetic lesions, are also among the outstanding features of SNP-A karyotyping platforms compared to conventional metaphase karyotyping and array CGH.

THE TARGET GENETIC LESIONS

In principle, the targets of SNP-A karyotyping are strictly limited to those genetic lesions that cause copy number alterations, such as numerical abnormalities of chromosomes and gains or losses of chromosomal segments. Balanced translocations, which are commonly found in hematopoietic malignancies and would be easily detected by metaphase karyotyping, such as t(8;21)(q22;q22) and t(15;17)(q22;q21), are not accompanied by copy number changes and therefore are out of scope of the SNP-A karyotyping. Moreover, SNP-A karyotyping does not provide any topological information about the copy number abnormalities it detects. For example, a high-level gene amplification is a very

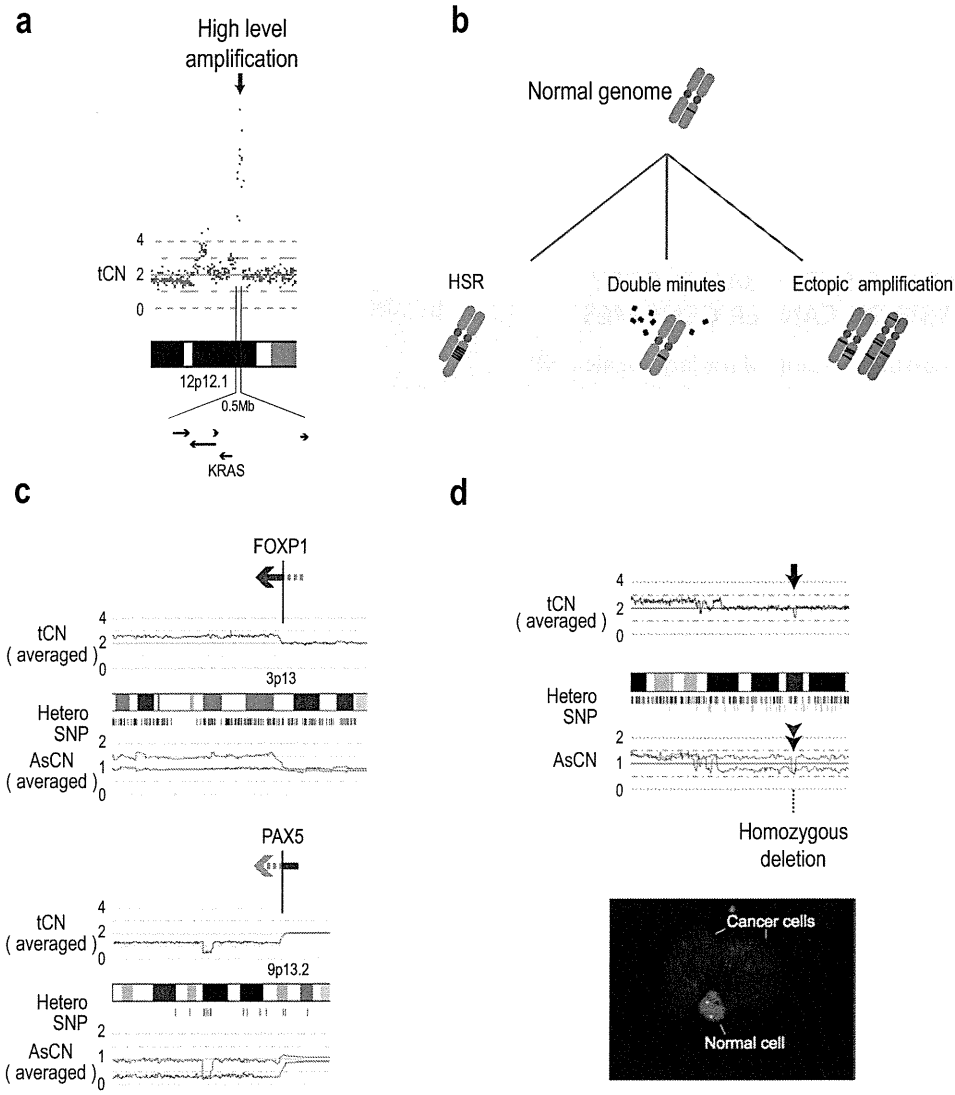

Figure 2. Detection of high-level amplifications and homozygous deletions in SNP-A karyotyping. (a) SNP-A karyotyping sensitively detects focal gene (chromosome) amplifications involving KRAS locus in a MDS case. (b) Gene amplifications may occur in situ as homogeneously staining regions (HSR), in episomal sites as double minutes (DMs), or at ectopic chromosomal sites. However, SNP-A karyotyping does not provide topological information of the amplifications. (c) SNP-A analysis of an unbalanced translocation between chromosomes 3 and 9 causing a *FOXP1/PAX5* fusion gene. The genomic positions of the breakpoints are determined at high precision. (d) Output of SNP-A karyotyping showing a homozygous deletion at 1p21.3 in a lymphoma specimen. The focal reduction found in total copy number plot (red arrows) is judged to represent a homozygous deletion, based on the fact that the region shows a biallelic reduction of AsCNs, as indicated by arrowheads.

nice target of SNP-A karyotyping, but it does not determine where and how it occurs within the genome, or whether it presents in a homogeneously stained region (HSR) or double minutes (DM), or represents episomal gene amplification (Figure 2a and b). On the other hand, in SNP-A karyotyping copy number change breakpoints can be precisely determined, relying on the method's high level of resolution (Figure 2a).

In terms of resolution, the SNP-A platforms far outperform metaphase karyotyping and typical bacterial artificial chromosome (BAC) array-based CGH, although oligonucleotide-based CGH arrays (Agilent, Santa Clara, CA) show even better performance in this regard.[26] On the currently available SNP arrays, genomic copy numbers are measured at approximately 10^5 to $>10^6$ SNP loci (Affymetrix GeneChip and Illumina BeadChip). Generally, the behavior of individual SNP-specific probes is not reliable enough to allow for precise single-point determination of copy number alterations at each SNP site, but the use of large numbers of probes enables the detection of genetic lesions less than 100k in size that would easily escape detection by metaphase karyotyping or even by BAC array CGH. SNP-A can detect the genes involved in the breakpoints of the unbalanced copy number changes[20,27] and precisely determine the genetic targets of amplifications and deletions[18,20,28-30] (Figure 2c).