xenograft models of primary patient samples from the preleukemic phase have been rarely reported, and the TAM xenograft model would be an attractive method to investigate leukemogenesis.

We previously described the development of novel immunodeficient NOD/Shi-*scid*, interleukin (IL)-2Rγ^{null} (NOG) mice with a superior capacity for the engraftment of human hematopoietic and neoplastic cells.[23-26] In contrast to a previous study in which TAM cells showed a limited ability to expand in immunodeficient mice,[27] we established a xenograft model where TAM cells were transplanted into NOG mice to recapitulate the pathophysiology of TAM/ML-DS. This xenograft model in combination with high-throughput genomic technology was used to show that genetically heterogeneous minor subclones with leukemia-initiating potential already exist in the neonatal TAM phase and could serve as initiating clones evolving to ML-DS in a patient. Our TAM xenograft model may be of value to gain insight into the evolutionary process of leukemia.

Materials and methods

Patients and sample collection

Peripheral blood (PB) samples were obtained from patients diagnosed with TAM associated with DS in acute and complete remission phases. Mononuclear cells were separated by Ficoll-Hypaque (Pharmacia, Uppsala, Sweden) density gradient centrifugation, as previously described.[23] Informed consent was obtained from the patients' parents in accordance with the Declaration of Helsinki, and the research was approved by the institutional ethics committee of Kyoto University Hospital.

Mice

NOG mice were developed at the Central Institute of Experimental Animals (Kawasaki, Japan) as previously described[28] and were maintained in our pathogen-free facility and cared for in accordance with the institutional guidelines for animal welfare.

Primary and serial xenogeneic transplantation into NOG mice

Xenotransplantation and analysis of TAM cells were performed using a previously reported method with some modifications.[26] In brief, PB mononuclear cells (PBMCs) obtained from TAM patients ($1–3 \times 10^6$ cells) were injected into 2.4 Gy–irradiated 8- to 12-week-old NOG mice through the tail vein. To screen for the proliferation of TAM-derived cells, bone marrow (BM) cells were aspirated from the tibia every 4 weeks. Engraftment was defined as >1% of cells staining positive for human CD7 (hCD7), hCD33, hCD41a, hCD45, and hCD117 at 12 weeks after transplantation. For serial transplantation, recipient BM cells were collected 12 to 18 weeks after transplantation; the equivalent of 1×10^6 hCD45$^+$ cells was intravenously transplanted into new mice. For a detailed determination of chromosomal and genetic alterations in TAM-derived cells, serial transplantation experiments using preserved PBMC samples were performed.

Flow cytometric analysis of transplanted TAM-derived cells

For analysis of TAM-derived cells in murine BM, mice were euthanized, and the BM was removed and mechanically dispersed. Mononuclear cells were purified from the BM and stained with antibodies. Dead cells were excluded according to 4′,6-diamidino-2-phenylindole staining. Blast cells were identified by classical CD45/SSC blast gating.[29] See supplemental Methods on the *Blood* Web site for details.

Human cell sorting

Human cell isolation was performed according to a previously described method with some modifications.[23,24] See supplemental Methods for details.

Colony assay

Leukemic colony formation was assessed according to a previously described method with some modifications.[30] See supplemental Methods for details.

GATA1 genomic sequencing analysis

The *GATA1* gene was amplified using polymerase chain reaction (PCR) as previously described[8] and sequenced by an ABI 3130xl Genetic analyzer (Applied Biosystems, Foster City, CA).

DNA copy number analysis

DNA copy number analysis was performed using GeneChip Human Mapping 250K Nsp arrays (Affymetrix, Inc., Santa Clara, CA) according to the manufacturer's standard protocols. Genomic copy numbers including allele-specific copy numbers were calculated using CNAG/AsCNAR software (http://www.genome.umin.jp), and genomic DNA obtained from PB of patients in the remission phase was used as a control. Copy number abnormalities and other allelic imbalances were detected using a hidden Marcov model–based algorithm.

Statistical analysis

Data are presented as the mean ± standard deviation. The 2-sided P value was determined by testing the null hypothesis that the 2 population medians are equal. P values <0.05 were considered to be significant.

Results

Establishment of a TAM xenograft model using NOG mice

To determine whether NOG mice provide a TAM xenograft model, Ficoll-purified PB samples from 11 TAM patients were transplanted into irradiated NOG mice. Patient characteristics are shown in Table 1. Patients' ages at sample collection, percentage of blast cells, number of cells injected, and number of engrafted recipients for each PB sample are shown in supplemental Table 1. Of 11 patient samples, 3 (patients 1, 2, and 9) were engrafted successfully in the recipient mice. Engraftment was maintained ≥12 weeks in all cases (Figure 1A). The spleen and liver of the recipients were also infiltrated with hCD45$^+$ blast cells (data not shown). These TAM-derived cells were morphologically similar to the primary TAM cells obtained from the patients (Figure 1B). Flow cytometric analysis of surface antigens detected the expression of CD117, CD34, CD33, and CD41a on hCD45$^+$ cells, which was consistent with the pattern observed in primary cells of TAM patients (Figure 1C). The presence of the same *GATA1* mutation was confirmed in the primary TAM cells and the engrafted cells in NOG mice (Figure 1D; supplemental Table 1). Chromosomal analysis of engrafted cells showed no abnormalities other than trisomy 21 (Figure 1E). These TAM-derived cells were detectable in the recipient's BM for >24 weeks (data not shown).

NOG mice can support self-renewal of TAM-derived cells

To examine the self-renewal capacity of TAM-derived cells, we performed serial transplantation of engrafted cells in the BM of recipient mice. Only the TAM-derived cells from patient 1 were successfully engrafted into the secondary (2°) and tertiary (3°) recipients. The morphology and surface antigen expression of these engrafted cells remained unchanged throughout the serial transplantation (Figure 2A-B). Interestingly, the TAM-derived cells

Regular Article

MYELOID NEOPLASIA

Clonal selection in xenografted TAM recapitulates the evolutionary process of myeloid leukemia in Down syndrome

Satoshi Saida,[1] Ken-ichiro Watanabe,[1] Aiko Sato-Otsubo,[2] Kiminori Terui,[3] Kenichi Yoshida,[2] Yusuke Okuno,[2] Tsutomu Toki,[3] RuNan Wang,[3] Yuichi Shiraishi,[4] Satoru Miyano,[4] Itaru Kato,[1] Tatsuya Morishima,[1] Hisanori Fujino,[1] Katsutsugu Umeda,[1] Hidefumi Hiramatsu,[1] Souichi Adachi,[5] Etsuro Ito,[3] Seishi Ogawa,[2] Mamoru Ito,[6] Tatsutoshi Nakahata,[7] and Toshio Heike[1]

[1]Department of Pediatrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [2]Cancer Genomics Project, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [3]Department of Pediatrics, Graduate School of Medicine, Hirosaki University, Hirosaki, Japan; [4]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; [5]Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [6]Laboratory Animal Research Department, Central Institute for Experimental Animals, Kawasaki, Japan; and [7]Department of Clinical Application, Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan

Key Points

- Genetically heterogeneous subclones with varying leukemia-initiating potential exist in neonatal transient abnormal myelopoiesis.
- This novel xenograft model of transient abnormal myelopoiesis may provide unique insight into the evolutionary process of leukemia.

Transient abnormal myelopoiesis (TAM) is a clonal preleukemic disorder that progresses to myeloid leukemia of Down syndrome (ML-DS) through the accumulation of genetic alterations. To investigate the mechanism of leukemogenesis in this disorder, a xenograft model of TAM was established using NOD/Shi-*scid*, interleukin (IL)-2Rγ^{null} mice. Serial engraftment after transplantation of cells from a TAM patient who developed ML-DS a year later demonstrated their self-renewal capacity. A *GATA1* mutation and no copy number alterations (CNAs) were detected in the primary patient sample by conventional genomic sequencing and CNA profiling. However, in serial transplantations, engrafted TAM-derived cells showed the emergence of divergent subclones with another *GATA1* mutation and various CNAs, including a 16q deletion and 1q gain, which are clinically associated with ML-DS. Detailed genomic analysis identified minor subclones with a 16q deletion or this distinct *GATA1* mutation in the primary patient sample. These results suggest that genetically heterogeneous subclones with varying leukemia-initiating potential already exist in the neonatal TAM phase, and ML-DS may develop from a pool of such minor clones through clonal selection. Our xenograft model of TAM may provide unique insight into the evolutionary process of leukemia. (*Blood*. 2013;121(21):4377-4387)

Introduction

Neonates with Down syndrome (DS) are at high risk of developing a unique hematologic disorder referred to as transient abnormal myelopoiesis (TAM), transient myeloproliferative disorder, or transient leukemia. In most cases, TAM resolves spontaneously within 3 months.[1,2] However, after spontaneous remission, 20% of TAM patients develop myelodysplastic syndrome and acute megakaryocytic leukemia referred to as myeloid leukemia of DS (ML-DS) within 4 years.[3,4] Blast cells in most patients with TAM and ML-DS have mutations in exon 2 of the gene coding for the transcription factor *GATA1*,[5-8] which is essential for the normal development of erythroid and megakaryocytic cells.[9,10] Although blast cells in most TAM and ML-DS patients share the identical *GATA1* mutation, recurrent additional cytogenetic abnormalities are commonly observed during disease progression.[2,5,11,12] In fact, a ML-DS case derived from a minor clone with a distinct *GATA1* mutation in the TAM phase was previously reported by our group.[13] These clinical findings suggest that although most TAM cells disappear in the early neonatal phase, a few clones persist during apparent remission to develop ML-DS later. Because only one fifth of TAM cases progress to ML-DS, additional genetic events besides *GATA1* mutation are likely to be involved in the progression of TAM to ML-DS.[14] As mentioned above, the development of ML-DS is significantly correlated with karyotypic abnormalities such as duplication (dup)(1q), deletion (del)(6q), del(7p), dup(7q), +8, +11, and del(16q),[2,11,12] which are rarely observed in the TAM phase. These clinical findings have led many physicians to consider TAM as preleukemia and the progression of TAM to ML-DS as an attractive model to investigate multistep leukemogenesis.

Animal models have contributed to our understanding of the pathogenesis of TAM/ML-DS and other leukemias.[15-21] Mice models in which primary human leukemic cells were transplanted into immunodeficient hosts provided significant clues to advance our understanding of the pathogenesis of human leukemia.[19-22] However,

Submitted December 18, 2012; accepted February 4, 2013. Prepublished online as *Blood* First Edition paper, March 12, 2013; DOI 10.1182/blood-2012-12-474387.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

mutations and patients with these mutations have significantly shorter telomeres than those with other dyskeratosis congenita subtypes [3]. It has also been reported that most patients with dyskeratosis congenita with *TINF2* mutations have severe disease, and, compared with other dyskeratosis congenita genes, patients with *TINF2* mutations have a high incidence of aplastic anaemia before the age of 10 years [3].

Aberrant repair process by enhanced apoptosis of alveolar epithelial cells plays a critical role in the pathogenesis of pulmonary fibrosis such as idiopathic pulmonary fibrosis, although the precise mechanism is still unclear. The mechanism(s) of pulmonary fibrosis in dyskeratosis congenita has also not yet been clarified. However, because mutations in dyskeratosis congenita genes cause short telomere length with functional deficits in telomere maintenance, telomeres in alveolar epithelial cells may be short. In patients with dyskeratosis congenita, we speculate that aberrant lung repair by enhanced cell death causes pulmonary fibrosis, although the short telomere length in alveolar epithelial cells has not been directly demonstrated.

Herein, we describe the first case report of dyskeratosis congenita with pulmonary fibrosis associated with *TINF2* mutation. This report proved that mutations not only in *TERC*, *TERT* and *DKC1*, but also *TINF2*, cause pulmonary fibrosis in dyskeratosis congenita. However, we do not know why mutations in *TERC*, *TERT* and *DKC1* are frequently found in dyskeratosis congenita patients with pulmonary fibrosis in contrast to the other five genes. In addition, sex hormones, which can increase telomerase activity, are potential therapeutic drugs; however, no standard treatment has been established for pulmonary fibrosis in dyskeratosis congenita patients. Because the clinical characteristics and pathogenesis of pulmonary fibrosis in dyskeratosis congenita is not clear, the accumulation of case-based reports sheds light on the understanding of this devastating disease.

@ERSpublications
The first reported case of a dyskeratosis congenita patient with pulmonary fibrosis and *TINF2* mutation http://ow.ly/pheRW

Atsuro Fukuhara[1], Yoshinori Tanino[1], Taeko Ishii[1], Yayoi Inokoshi[1], Kazue Saito[1], Naoko Fukuhara[1], Suguru Sato[1], Junpei Saito[1], Takashi Ishida[1], Hiroki Yamaguchi[2] and Mitsuru Munakata[1]
[1]Dept of Pulmonary Medicine, Fukushima Medical University School of Medicine, Fukushima, and [2]Division of Hematology, Dept of Internal Medicine, Nippon Medical School, Tokyo, Japan.

Correspondence: Y. Tanino, Dept of Pulmonary Medicine, Fukushima Medical University, 1 Hikarigaoka, Fukushima-City, Fukushima 960-8157, Japan. E-mail: ytanino@fmu.ac.jp

Received: Aug 27 2013 | Accepted after revision: Sept 04 2013 | First published online: Sept 26 2013

Conflict of interest: None declared.

References

1 Dokal I. Dyskeratosis congenita. *Hematology Am Soc Hematol Educ Program* 2011; 2011: 480–486.
2 Vulliamy TJ, Marrone A, Knight SW, *et al.* Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. *Blood* 2006; 107: 2680–2685.
3 Walne AJ, Vulliamy T, Beswick R, *et al.* *TINF2* mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood* 2008; 112: 3594–3600.
4 Vulliamy T, Marrone A, Goldman F, *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 2001; 413: 432–435.
5 Marrone A, Sokhal P, Walne A, *et al.* Functional characterization of novel telomerase RNA (*TERC*) mutations in patients with diverse clinical and pathological presentations. *Haematologica* 2007; 92: 1013–1020.
6 Armanios M, Chen JL, Chang YP, *et al.* Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci USA* 2005; 102: 15960–15964.
7 Yamaguchi H, Calado RT, Ly H, *et al.* Mutations in *TERT*, the gene for telomerase reverse transcriptase, in aplastic anemia. *N Engl J Med* 2005; 352: 1413–1424.
8 Safa WF, Lestringant GG, Frossard PM. X-linked dyskeratosis congenita: restrictive pulmonary disease and a novel mutation. *Thorax* 2001; 56: 891–894.
9 Savage SA, Giri N, Baerlocher GM, *et al.* TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. *Am J Hum Genet* 2008; 82: 501–509.
10 de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005; 19: 2100–2110.

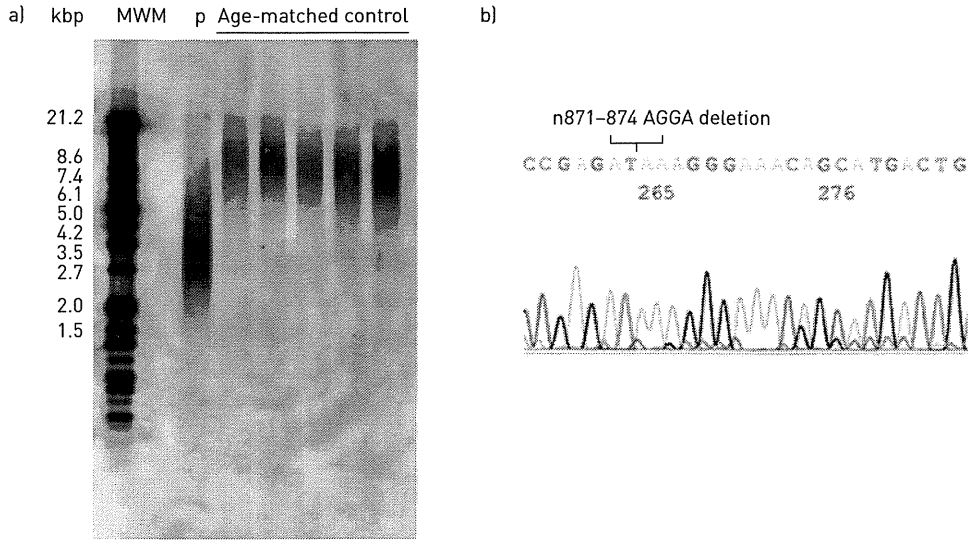


FIGURE 1 a) Southern blot analysis showed shorter telomere length of the patient (P) compared to age-matched healthy controls. MWM: molecular weight marker. b) Gene mutation analysis by direct sequencing showed n871–874 tetranucleotide AGGA deletion in *TINF2* gene.

At this point, we strongly suspected that she had dyskeratosis congenita. To make a definite diagnosis, we first examined the *TERC* and *TERT* genes by direct sequencing. However, no mutations were found in either gene. Southern blot analysis showed short telomere length (fig. 1a), therefore mutations in *TINF2* were next explored. As shown in figure 1b, because direct sequencing showed a n871–874 tetranucleotide AGGA deletion in *TINF2*, she was diagnosed as having dyskeratosis congenita with pulmonary fibrosis associated with *TINF2* mutation. As her respiratory condition progressed, steroid pulse therapy followed by oral prednisolone was conducted. However, no improvement of her symptoms was observed, and bilateral pneumothorax with mediastinal and subcutaneous emphysemas developed. She died of respiratory failure 1 year after starting the treatment.

Dyskeratosis congenita is a rare genetic ectodermal disorder characterised by skin hyperpigmentation, nail dystrophy and leukoplakia of the mucous membranes. Bone marrow failure is a frequent finding and a predisposition to malignancy has been noted. Although pulmonary manifestations of dyskeratosis congenita were believed to be uncommon, DOKAL [1] reported that abnormal pulmonary features may be seen in as many as 10–15% of patients.

Genetically, dyskeratosis congenita is heterogeneous, with three forms having been identified: X-linked recessive, autosomal dominant and autosomal recessive. In the present case, the patient's father had suffered from the same disease; therefore, we suspected that the form of dyskeratosis of this patient was autosomal dominant. The autosomal dominant form of dyskeratosis congenita is caused by heterozygous mutations in the core components of telomerase, *TERC* [4, 5] and *TERT* [6, 7], as well as in the component of the shelterin telomere protection complex, *TINF2* [3]. In this patient, mutation of *TINF2*, but not *TERC* and *TERT*, was confirmed by gene mutation analysis. It has previously been reported that mutations in *DKC1* [8], *TERC* [5] and *TERT* [6] were associated with pulmonary fibrosis in dyskeratosis congenita patients. *DKC1* was not analysed in this patient, because mutation in *DKC1* causes the X-linked form of dyskeratosis congenita. Regarding the relationship between pulmonary fibrosis and *TINF2* mutation in dyskeratosis congenita, WALNE *et al.* [3] have reported that only one patient had pulmonary fibrosis among other clinical features in 33 dyskeratosis congenita patients with *TINF2* mutations. However, they did not describe the patient in detail. To the best of our knowledge, this is the first case report showing pulmonary fibrosis in dyskeratosis congenita with *TINF2* mutation.

TINF2 mutations were reported to be heterozygous mutations in the sixth-found dyskeratosis congenita gene by SAVAGE *et al.* [9] in 2008. *TINF2* encodes TIN2, and is a component of the shelterin telomere-protection complex. The shelterin complex has at least three effects on telomeres: it determines the structure of the telomeric terminus, is implicated in the generation of t-loops and controls the synthesis of telomeric DNA by telomerase [1, 10]. Without the protective activity of shelterin, telomeres are no longer hidden from DNA repair mechanisms and chromosome ends are therefore incorrectly processed by the DNA repair pathways. Approximately 11% of all dyskeratosis congenita has been reported to be accounted for by *TINF2*

2 Martineau AR, Newton SM, Wilkinson KA, *et al.* Neutrophil-mediated innate immune resistance to mycobacteria. *J Clin Invest* 2007; 117: 1988–1994.
3 Sugawara I, Udagawa T, Yamada H. Rat neutrophils prevent the development of tuberculosis. *Infect Immun* 2004; 72: 1804–1806.
4 Nandi B, Behar SM. Regulation of neutrophils by interferon-γ limits lung inflammation during tuberculosis infection. *J Exp Med* 2011; 208: 2251–2262.
5 Brahmbhatt S, Black GF, Carroll NM, *et al.* Immune markers measured before treatment predict outcome of intensive phase tuberculosis therapy. *Clin Exp Immunol* 2006; 146: 243–252.
6 Martineau AR, Timms PM, Bothamley GH, *et al.* High-dose vitamin D_3 during intensive-phase antimicrobial treatment of pulmonary tuberculosis: a double-blind randomised controlled trial. *Lancet* 2011; 377: 242–250.
7 Lawn SD, Kerkhoff AD, Vogt M, *et al.* Characteristics and early outcomes of patients with Xpert MTB/RIF-negative pulmonary tuberculosis diagnosed during screening before antiretroviral therapy. *Clin Infect Dis* 2012; 54: 1071–1079.
8 Lefebvre N, Falzon D. Risk factors for death among tuberculosis cases: analysis of European surveillance data. *Eur Respir J* 2008; 31: 1256–1260.
9 Sloand E. Hematologic complications of HIV infection. *AIDS Rev* 2005; 7: 187–196.
10 Ferrand RA, Herman J, Elgalib A, *et al.* Septic shock and multi-organ failure in HIV infection – 'sepsis tuberculosa gravissima'. *Int J STD AIDS* 2006; 17: 562–564.

Eur Respir J 2013; 42: 1752–1757 | DOI: 10.1183/09031936.00140913 | Copyright ©ERS 2013

Pulmonary fibrosis in dyskeratosis congenita with *TINF2* gene mutation

To the Editor:

Dyskeratosis congenita is a rare inherited disorder of ectodermal dysplasia characterised by the classical mucocutaneous triad of abnormal skin pigmentation, nail dystrophy and leukoplakia [1–3], at least one of which is present in around 80–90% of dyskeratosis congenita cases. Bone marrow failure is another common feature, and a variety of other abnormalities (*e.g.* dental, gastrointestinal, neurological, ophthalmic, pulmonary and skeletal) have been also described [1–3]. The main causes of mortality in dyskeratosis congenita are bone marrow failure, pulmonary disease and malignancy [1]. Three modes of inheritance have been recognised: X-linked recessive, autosomal dominant and autosomal recessive [1, 3]. Eight dyskeratosis congenita genes (*DKC1* (dyskeratosis congenita 1), *TERC* (telomerase RNA component), *TERT* (telomerase reverse transcriptase), *NOP10* (nucleolar protein 10), *NHP2*, *TINF2* (TERF1-interacting nuclear factor 2), *TCAB1* and *RTEL1* (regulation of telomere elongation helicase 1)) have already been identified, and their mutations account for ~60% of all dyskeratosis congenita cases [1]. Among the dyskeratosis congenita genes, mutations in *TERC*, *TERT* and *DKC1* have recently been reported to be associated with familial pulmonary fibrosis and idiopathic pulmonary fibrosis, and pulmonary fibrosis is recognised as one of the features of dyskeratosis congenita. However, the relationship between mutations in the other dyskeratosis congenita genes and pulmonary fibrosis has not yet been clarified. To the best of our knowledge, this is the first case report describing a dyskeratosis congenita patient with pulmonary fibrosis who had a *TINF2* mutation.

A 43-year-old female visited our hospital with cough and progressive dyspnoea. She had never smoked, and had a history of aplastic anaemia, ocular pemphigoid, erythroplasia of Queyrat and infertility. Her father had been diagnosed as having aplastic anaemia and his whole body was pigmented. About 2 years ago, she complained of cough and consulted her personal doctor. Her chest radiographs showed diffuse reticular shadows in the bilateral lung fields. She was referred to a general hospital and was diagnosed with idiopathic interstitial pneumonia. Because her general condition was stable at that time, she was followed up without any specific therapy for 1 year. She was referred to our hospital due to gradual worsening of dyspnoea and admitted for further examinations. Her physical examination was remarkable for skin pigmentation on her whole body, ocular pemphigoid in the left eye and fine crackles in both lung fields. Her fingertip skin was rough but her nails were not dystrophic. Although no leukoplakia was found in the oral mucosa, she had erythroplasia of Queyrat of the vulva. Laboratory data showed elevated lactate dehydrogenase, transaminases, erythrocyte sedimentation rate and sialylated carbohydrate antigen KL-6 with thrombocytopenia. Chest radiographs demonstrated consolidation and reticular shadows in the bilateral lung fields. Furthermore, chest computed tomography revealed consolidation and reticular shadows in both lung fields, as well as bronchiectasis and cystic shadows in the left lung.

Natl Acad Sci U S A. 2012;109(15):5803–5808.
40. Yagi T, et al. Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Hum Mol Genet.* 2011;20(23):4530–4539.
41. Isken O, Maquat LE. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet.* 2008;9(9):699–712.
42. Olnes MJ, et al. Eltrombopag and improved hematopoiesis in refractory aplastic anemia. *N Engl J Med.* 2012;367(1):11–19.
43. Fielder PJ, et al. Regulation of thrombopoietin levels by c-mpl-mediated binding to platelets. *Blood.* 1996;87(6):2154–2161.
44. Stoffel R, Wiestner A, Skoda RC. Thrombopoietin in thrombocytopenic mice: evidence against regulation at the mRNA level and for a direct regulatory role of platelets. *Blood.* 1996;87(2):567–573.
45. Hitchcock IS, Chen MM, King JR, Kaushansky K. YRRL motifs in the cytoplasmic domain of the thrombopoietin receptor regulate receptor internalization and degradation. *Blood.* 2008;112(6):2222–2231.
46. Patel SR, Hartwig JH, Italiano JE Jr. The biogenesis of platelets from megakaryocyte proplatelets. *J Clin Invest.* 2005;115(12):3348–3354.
47. Pikman Y, et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med.* 2006;3(7):e270.
48. Guglielmelli P, et al. Anaemia characterises patients with myelofibrosis harbouring Mpl mutation. *Br J Haematol.* 2007;137(3):244–247.
49. Tefferi A. JAK and MPL mutations in myeloid malignancies. *Leuk Lymphoma.* 2008;49(3):388–397.
50. Kohn AD, Summers SA, Birnbaum MJ, Roth RA. Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *J Biol Chem.* 1996;271(49):31372–31378.
51. Kato Y, et al. Selective activation of STAT5 unveils its role in stem cell self-renewal in normal and leukemic hematopoiesis. *J Exp Med.* 2005;202(1):169–179.
52. Olthof SG, Fatrai S, Drayer AL, Tyl MR, Vellenga E, Schuringa JJ. Downregulation of signal transducer and activator of transcription 5 (STAT5) in CD34+ cells promotes megakaryocytic development, whereas activation of STAT5 drives erythropoiesis. *Stem Cells.* 2008;26(7):1732–1742.
53. Amabile G, et al. In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. *Blood.* 2013;121(8):1255–1264.
54. Suzuki N, et al. Generation of engraftable hematopoietic stem cells from induced pluripotent stem cells by way of teratoma formation molecular therapy: the Journal Of The American Society Of Gene Therapy [published online ahead of print May 14, 2013]. *Mol Ther.* doi:10.1038/mt.2013.71.
55. Kitamura T, et al. Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp Hematol.* 2003;31(11):1007–1014.
56. Eto K, et al. The WAVE2/Abi1 complex differentially regulates megakaryocyte development and spreading: implications for platelet biogenesis and spreading machinery. *Blood.* 2007;110(10):3637–3647.

To the editor:

ANKRD26-related thrombocytopenia and myeloid malignancies

Since the discovery that mutations in the 5′ untranslated region (UTR) of *ANKRD26* are responsible for an autosomal-dominant form of thrombocytopenia (*ANKRD26*-RT),[1] 21 affected families were reported.[2] A study analyzing this series of patients suggested that *ANKRD26*-RT is characterized by normal platelet size, moderate thrombocytopenia, and absent or mild bleeding tendency.[2] The study also found that the number of hematologic malignancies in affected families was higher than expected, but the relatively small cohort of patients precluded firm conclusions.

To gain further information on this matter, we screened for mutations in the 5′ UTR of *ANKRD26* 215 subjects with an inherited thrombocytopenia of unknown origin and found 11 mutations (3 not previously described) in 23 cases (Table 1[3]). Analysis of family members identified another 52 affected subjects. Moreover, we found 43 additional subjects, first- or second-degree relatives of *ANKRD26*-RT patients, who were known to be thrombocytopenic from an early age but were not available for genetic investigation (they were dead or not willing to perform the analysis). Analysis of the new series of 75 patients with *ANKRD26* mutations confirmed that *ANKRD26*-RT is characterized by moderate thrombocytopenia with normal platelet size and a mild bleeding phenotype.

In the extended series of 118 subjects certainly or very likely affected, we identified 10 patients who had developed myeloid malignancies: 4 acute myeloid leukemias (age at onset, 40-60 years), 4 myelodysplastic syndromes (MDS) (age at onset, 35-70 years), and 2 chronic myeloid leukemias (CML) (age at onset, 30-65 years). The total observation time was 4741 years, and the incidence of acute myeloid leukemia, MDS, and CML was 84 (confidence interval [CI], 23-216), 84 (CI, 23-216), and 42 (CI, 5-152) per 100 000, respectively. Putting together the 118 cases examined in this study and the 104 cases already reported, 4.9% of patients had acute leukemias, 2.2% MDS, and 1.3% CML. The total observation time was 8915 years, with an incidence of acute leukemias, MDS, and CML of 123 (CI, 62-221), 56 (CI, 18-131), and 34 (CI, 7-98) per 100 000, respectively, thus higher than expected in the general population (5.2, 4.5, and 1.6 per 100 000, respectively, according to the National Cancer Institute[4]). Unlike myeloid malignancies, the incidence of lymphoproliferative disorders and nonhematologic cancers in *ANKRD26*-RT pedigrees was not higher than expected (data not reported). Available data are therefore compatible with the hypothesis that *ANKRD26*-RT, as the inherited thrombocytopenia deriving from *RUNX1* mutations,[5] predisposes to myeloid malignancies, in particular acute leukemias. The observation that this unfavorable outcome occurred in a limited proportion of subjects and has been observed in only 12 of 44 families indicates that the penetrance for malignancies was incomplete and other genetic and/or environmental factors contributed to the development of these disorders. Of note, the the case series described in this paper confirmed the already reported finding of unexplained high leukocyte counts in a large

Table 1. Main clinical and laboratory features of patients with *ANKRD26*-RT grouped by family

Family/No. of patients/Country	*ANKRD26* 5′ UTR mutation	Mean age, y (range)	WHO bleeding score[3] (no. of patients)	Mean platelet count, $\times 10^9$/L (range)	Mean MPV, fL (range)	Mean hemoglobin, g/dL (range)	Mean WBC, $\times 10^9$/L (range)
1/3/Canada	c.-116C>G*	18 (6-37)	1 (2), 2 (1)	70.6 (45-107)	9.55 (9.4-9.7)†	12.85 (12.1-13.6)†	11.7 (10.6-12.8)†
2/2/France	c.-118C>A	34 (17-51)	2 (2)	63.5 (45-82)	10.2 (9.9-10.5)	15.5 (14.8-16.3)	7.25 (6.8-7.7)
3/1/US	c.-118C>T	57	0 (1)	38	7.7	14.3	10.5
4/3/France	c.-119C>A	32.6 (2-65)	2 (2), 3	55.6 (36-81)	10.1 (9-11.3)	13.7 (12.8-14.2)	7.5 (4.7-9.6)
5/1/Italy	c.-126T>G	28	1 (1)	14	8.1	16.6	7.3
6/9/Argentina	c.-127A>G	37.7 (6-74)	0 (2), 1 (4), 2 (3)	96.2 (68-147)	8.6 (7.7-9.5)‡	15.3 (12.9-17.7)	8.47 (6.1-11.1)
7/2/Italy	c.-127A>G	32 (17-47)	0 (1), 2 (1)	79.5 (68-91)	8.8 (8.6-9)	14.15 (12.4-15.9)	8.59 (7.92-9.26)
8/3/France	c.-127A>T	19 (6-38)	0 (1), 1 (2)	65.6 (47-85)	7.5 (7.1-7.8)	14.8 (12.8-17.2)	7 (5.3-8.7)
9/8/France	c.-127A>T	39.25 (1-68)	0 (5), 1 (3)	39.37 (24-84)	10.8 (9.3-11.7)	14.15 (11-16.8)	9.55 (6.7-12.1)
10/6/France	c.-127delAT*	34.6 (1-66)	0 (6)	54.6 (26-96)	10.2 (8.7-11)	14.2 (10.4-16.6)	7.86 (5.8-13.4)
11/3/France	c.-128G>A	63 (21-97)	2 (2)	19 (12-30)	8.1 (7.7-8.5)†	13.6 (13.6-13.6)†	7.75 (7.54-7.96)†
12/2/France	c.-128G>A	36.5 (26-47)	1, 2	71.5 (68-75)	8.4§	14.8§	10§
13/3/Italy	c.-128G>A	20 (12-26)	0 (3)	34 (14-70)	7.6 (6.3-8.5)	15.3 (14-16.5)	14.6 (9.21-21)
14/1/Japan	c.-128G>A	25	0 (1)	54	8.9	15.8	12.5
15/2/Italy	c.-128G>A	52 (50-54)	1 (2)	20 (19-21)	8.4§	14.2§	5.1§
16/10/US	c.-128G>A	23.2 (1-62)	0 (4), 1 (6)	35.5 (19-65)	8.7 (7.5-10.2)	14.43 (12.7-16.1)	9.93 (7-12.4)
17/8/Italy	c.-128G>A	40.2 (6-93)	0 (1), 1 (1), 2 (3), 3 (19), 4 (1)	18.7 (5-34)	8.5 (7-10.3)‖	13.47 (11.8-16.3)	7.64 (5.73-11.31)
18/1/Italy	c.-128G>A	38	2 (1)	48	9.04	15.9	8.51
19/2/France	c.-128G>C*	20 (1-39)	0 (2)	52.5 (24-81)	12 (10-14)	13.15 (12.3-4)	9.35 (6.6-12.1)
20/1/Japan	c.-134G>A	13	0 (1)	81	9.2	14.2	7.64
21/2/US	c.-134G>A	58.5 (50-67)	2 (2)	10 (7-13)	8.2 (5.7-10.7)	11.7 (10.7-12.7)	7.85 (5-10-75)
22/1/The Netherlands	c.-134G>A	34	2 (1)	24	11.1	15.5	12.5
23/1/US	c.-134G>A	64	1 (1)	20	8.8	13.4	8.2

MPV, mean platelet volume; US, United States; WBC, white blood cell; WHO, World Health Organization.
*New mutations.
†Data from 2 of 3 patients.
‡Data from 7 of 9 patients.
§Data from 1 of 2 patients.
‖Data from 7 of 8 patients.

proportion of subjects. Further observation is required to ascertain whether the higher leukocyte counts are related to eventual development of leukemia.

In conclusion, *ANKRD26*-RT is an insidious form of inherited thrombocytopenias that exposes patients to a low risk of bleeding but predisposes them to hematologic myeloid malignancies. Recognizing this disorder and its attendant risks is important for proper management of affected subjects.[6]

Patrizia Noris
Department of Internal Medicine, University of Pavia–Istituto Di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo Foundation, Pavia, Italy

Remi Favier
French Reference Center for Inherited Platelet Disorders, Armand Trousseau Children's Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France
Institut National de la Santé et de la Recherche Médicale, Unités Mixtes de Recherche 1009, Villejuif, France

Marie-Christine Alessi
Aix-Marseille University, Faculty of Medicine, Institut National de la Santé et de la Recherche Médicale, Unités Mixtes de Recherche 1062, Marseille, France

Amy E. Geddis
Division of Pediatric Hematology Oncology, University of California at San Diego, Rady Children's Hospital San Diego, San Diego, CA

Shinji Kunishima
Department of Advanced Diagnosis, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan

Paula G. Heller
Instituto de Investigaciones Médicas Alfredo Lanari, Consejo Nacional de Investigaciones Científicas y Técnicas, University of Buenos Aires, Buenos Aires, Argentina

Paola Giordano
Dipartimento di Scienze Biomediche e Oncologia Umana, Sezione di Pediatria, Università di Bari, Bari, Italy

Karen Y. Niederhoffer
Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

James B. Bussel
Division of Pediatric Hematology-Oncology, Weill Medical College of Cornell University, New York, NY

Gian Marco Podda
UO Medicina III, Ospedale San Paolo, Milan, Italy

Nicola Vianelli
Department of Haematology and Clinical Oncology "L. and A. Seràgnoli," S. Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy

Rogier Kersseboom
Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands

Alessandro Pecci
Department of Internal Medicine, University of Pavia–Istituto Di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo Foundation, Pavia, Italy

Chiara Gnan
Institute for Maternal and Child Health, Istituto Di Ricovero e Cura a Carattere Scientifico "Burlo Garofolo," Trieste, Italy

Caterina Marconi
Medical Genetics Unit, Policlinico S. Orsola-Malpighi, University of Bologna, Bologna, Italy

Anne Auvrignon
French Reference Center for Inherited Platelet Disorders, Armand Trousseau Children's Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

William Cohen
Aix-Marseille University, Faculty of Medicine, Institut National de la Santé et de la Recherche Médicale, Unités Mixtes de Recherche 1062, Marseille, France

Jennifer C. Yu
Division of Pediatric Hematology Oncology, University of California at San Diego, Rady Children's Hospital San Diego, San Diego, CA

Akihiro Iguchi
Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Allison Miller Imahiyerobo
Division of Pediatric Hematology-Oncology, Weill Medical College of Cornell University, New York, NY

Francoise Boehlen
Division of Angiology and Hemostasis, University Hospital of Geneva and Faculty of Medicine, Geneva, Switzerland

Dorsaf Ghalloussi
Aix-Marseille University, Faculty of Medicine, Institut National de la Santé et de la Recherche Médicale, Unités Mixtes de Recherche 1062, Marseille, France

Daniela De Rocco
Institute for Maternal and Child Health, Istituto Di Ricovero e Cura a Carattere Scientifico "Burlo Garofolo," Trieste, Italy

Pamela Magini
Medical Genetics Unit, Policlinico S. Orsola-Malpighi, University of Bologna, Bologna, Italy

Elisa Civaschi
Department of Internal Medicine, University of Pavia–Istituto Di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo Foundation, Pavia, Italy

Ginevra Biino
Institute of Molecular Genetics, and Pavia and Institution of Population Genetics, Consiglio Nazionale delle Ricerche, Sassari, Italy

Marco Seri
Medical Genetics Unit, Policlinico S. Orsola-Malpighi, University of Bologna, Bologna, Italy

Anna Savoia
Institute for Maternal and Child Health, Istituto Di Ricovero e Cura a Carattere Scientifico "Burlo Garofolo" and Department of Medical Sciences, University of Trieste, Trieste, Italy

Carlo L. Balduini
Department of Internal Medicine, University of Pavia–Istituto Di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo Foundation, Pavia, Italy

P.N. and R.F. contributed equally to this manuscript.

Acknowledgments: This work was supported by grants from Telethon Foundation, Italy (no. GGP10089) and the Cariplo Foundation (2012-0529), Italy. C.G. is recipient of a fellowship from Italian Association for Cancer Research, Italy. R.F. is recipient of a research fellowship from Institut National de la Santé et de la Recherche Médicale–Assistance Publique-Hôpitaux de Paris (contrat d'interface 2012-2014), France.

Contribution: P.N., R.F., M.-C.A., M.S., A.S., and C.L.B. designed the research, interpreted results, and wrote the manuscript; A.E.G., P.G.H., P.G., K.Y.N., J.B.B., G.M.P., N.V., R.K., A.P., J.C.Y., A.I., A.M.I., and E.C. acquired clinical and laboratory data; S.K., C.G., C.M., A.A., W.C., F.B., D.G., D.D.R., and P.M. performed mutation screening; G.B. performed statistical analysis; and all authors had access to primary clinical trial data and discussed and approved the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Carlo L. Balduini, Medicina Interna 3, Fondazione IRCCS Policlinico San Matteo, piazzale Golgi, 27100 Pavia, Italy; e-mail: c.balduini@smatteo.pv.it.

References

1. Pippucci T, Savoia A, Perrotta S, et al. Mutations in the 5′ UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am J Hum Genet.* 2011;88(1):115-120.
2. Noris P, Perrotta S, Seri M, et al. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood.* 2011;117(24):6673-6680.
3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer.* 1981;47(1):207-214.
4. National Cancer Institute. Cancer statistics. Available at: www.seer.cancer.gov/statistics/. Accessed May 20, 2013.
5. Liew E, Owen C. Familial myelodysplastic syndromes: a review of the literature. *Haematologica.* 2011;96(10):1536-1542.
6. Balduini CL, Pecci A, Noris P. Diagnosis and management of inherited thrombocytopenias. *Semin Thromb Hemost.* 2013;39(2):161-171.

To the editor:

Plasma hepcidin of Ethiopian highlanders with steady-state hypoxia

Hepcidin impedes iron absorption and is suppressed when erythropoietic iron requirements are increased. Recent studies show that during acute exposure to high-altitude hypoxia, plasma hepcidin concentrations drop when iron demands for erythropoiesis and hemoglobin synthesis are sharply increased.[1,2] However, the effects of chronic exposure to high-altitude hypoxia with stable erythropoietic iron requirements have not been examined. We hypothesized that plasma hepcidin would not be suppressed in iron-replete individuals chronically adapted to high altitude.

People of Amhara and Oromo ethnicity have been living at high altitude in Ethiopia for more than 5000 years and about 500 years, respectively, and have been shown to differ from one another in hemoglobin and oxyhemoglobin percentage.[3] Healthy volunteers from 3700 to 4000 m (high altitude) and 1200 to 1500 m (low altitude) were recruited, and they provided blood samples for analyses (see the supplemental Video(s)/Data Set(s) link at the top of the online article for genetic analysis methodology). The sample reported here had normal calculated body iron stores[4] (Figure 1D) and did not have infection or inflammation, assessed with C-reactive protein levels and malarial plasmodium DNA.

In contrast to acutely exposed Europeans[2] (see also the online supplement), high-altitude Amhara had higher plasma hepcidin, and high-altitude Oromo had similar hepcidin, compared with their respective lowland counterparts (Figure 1E). Furthermore, Amhara had higher plasma hepcidin and oxyhemoglobin percentage as well as lower hemoglobin and erythropoietin than Oromo at high altitudes (Figure 1A-C,E). Within Ethiopian subsamples, age, sex, BMI, erythropoietin, hemoglobin, oxyhemoglobin percentage, and transferrin receptor were not correlated with hepcidin (natural-log transformed for normality). Like Europeans at low altitude,[5] serum ferritin ($r = 0.35$ to 0.77) and body iron stores ($r = 0.39$ to 0.85) correlated with ln(hepcidin) (all $P < .05$, except in the small sample of female low-altitude Oromo [$n = 6$]). An intronic SNP in *GRAMD3* was associated with plasma hepcidin among Amhara at genome-wide significance ($P = 4.94*10^{-8}$), accounting for 22% of the variation in covariate-adjusted hepcidin level. Allele A of rs7700582 increased hepcidin levels by 8.1 ng/mL only among Amhara, although allele frequency was similar in all 4 subsamples.

Hepcidin was not suppressed in Amhara or Oromo highland samples under steady-state hypoxia, likely because erythropoietic drive was stable.[6] It is interesting to speculate that the higher plasma hepcidin of highlander Amhara, compared with Oromo, is due to lower iron demand indicated by lower hemoglobin and erythropoietin concentrations and higher body iron stores. Variants in *GRAMD3* are associated with macular degeneration, a retinal disease that has been related to abnormalities in hepcidin and iron accumulation.[7,8] Another variant near *GRAMD3* (rs1366100) has been associated with erythrocyte counts,[9] consistent with the idea that this region plays a role in iron metabolism. Thus, the genetic results also support the idea that iron stores are primary regulators of hepcidin levels in hypoxic populations without increased erythropoietic drive. Previous work has shown that various highlander populations demonstrate different responses to hypoxia,[10] which may also be the case with iron regulation.

Erika L. Lundgrin
Cleveland Clinic Lerner College of Medicine, Cleveland, OH

Allison J. Janocha
Department of Pathobiology, Lerner Research Institute, Cleveland, OH

Carl D. Koch
Cleveland Clinic Lerner College of Medicine, Cleveland, OH

Amha Gebremedhin
Faculty of Medicine, Addis Ababa University, Addis Ababa, Ethiopia

Anna Di Rienzo
Department of Human Genetics, University of Chicago, Chicago, IL

Gorka Alkorta-Aranburu
Department of Human Genetics, University of Chicago, Chicago, IL

Gary M. Brittenham
Department of Pediatrics, Columbia University College of Physicians and Surgeons, New York, NY

nature genetics

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

Ayana Kon[1], Lee-Yung Shih[2], Masashi Minamino[3], Masashi Sanada[1,4], Yuichi Shiraishi[5], Yasunobu Nagata[1], Kenichi Yoshida[1], Yusuke Okuno[1], Masashige Bando[3], Ryuichiro Nakato[3], Shumpei Ishikawa[6,7], Aiko Sato-Otsubo[1], Genta Nagae[8], Aiko Nishimoto[6], Claudia Haferlach[9], Daniel Nowak[10], Yusuke Sato[1], Tamara Alpermann[9], Masao Nagasaki[11], Teppei Shimamura[5], Hiroko Tanaka[12], Kenichi Chiba[5], Ryo Yamamoto[13], Tomoyuki Yamaguchi[13,14], Makoto Otsu[15], Naoshi Obara[16], Mamiko Sakata-Yanagimoto[16], Tsuyoshi Nakamaki[17], Ken Ishiyama[18], Florian Nolte[10], Wolf-Karsten Hofmann[10], Shuichi Miyawaki[18], Shigeru Chiba[16], Hiraku Mori[17], Hiromitsu Nakauchi[13,14], H Phillip Koeffler[19,20], Hiroyuki Aburatani[8], Torsten Haferlach[9], Katsuhiko Shirahige[3], Satoru Miyano[5,12] & Seishi Ogawa[1,4]

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

Recent genetic studies have led to the discovery of a number of new mutational targets in myeloid malignancies, unmasking unexpected roles for deregulated histone modification and DNA methylation in both acute and chronic myeloid neoplasms[1,2]. However, knowledge of the spectrum of gene mutations in myeloid neoplasms remains incomplete. We previously reported a whole-exome sequencing study of 29 paired tumor and normal samples of myeloid neoplasms with myelodysplastic features[3]. Although our major discovery was that frequent spliceosome mutations are uniquely associated with myelodysplasia phenotypes, we also identified hundreds of previously unreported gene mutations[3]. Most of those mutations affected single individuals only and are probably passenger changes. Therefore, their importance in leukemogenesis remains undetermined. However, through closer inspection of an updated list of mutations, including newly validated single-nucleotide variants, we identified additional recurrent mutations involving *STAG2*, a core component of the cohesin complex (Online Methods and **Supplementary Table 1**). In addition, we found that two other functionally related cohesin components, *STAG1* and *PDS5B*, were mutated in single specimens (**Supplementary Fig. 1**).

Cohesin is a multimeric protein complex that is conserved across species and is composed of four core subunits, SMC1, SMC3, RAD21

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [2]Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, Taipei, Taiwan. [3]Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [4]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Kyoto-shi Sakyo-ku, Kyoto, Japan. [5]Laboratory of DNA Information Analysis, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [6]Department of Pathology, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [7]Department of Genomic Pathology, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo, Japan. [8]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Meguro-ku, Tokyo, Japan. [9]Munich Leukemia Laboratory, Munich, Germany. [10]Department of Hematology and Oncology, University Hospital Mannheim, Mannheim, Germany. [11]Division of Biomedical Information Analysis, Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Aoba-ku, Sendai, Japan. [12]Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [13]Division of Stem Cell Therapy, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [14]Stem Cell and Organ Regeneration Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency (JST), Chiyoda-ku, Tokyo, Japan. [15]Stem Cell Bank, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [16]Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan. [17]Division of Hematology, Department of Medicine, Showa University School of Medicine, Shinagawa-ku, Tokyo, Japan. [18]Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, Toshima-ku, Tokyo, Japan. [19]Hematology/Oncology, Cedars-Sinai Medical Center, Los Angeles, California, USA. [20]National University of Singapore, Cancer Science Institute of Singapore, Singapore. Correspondence should be addressed to S.O. (sogawa-tky@umin.ac.jp).

Received 15 February; accepted 24 July; published online 18 August 2013; doi:10.1038/ng.2731

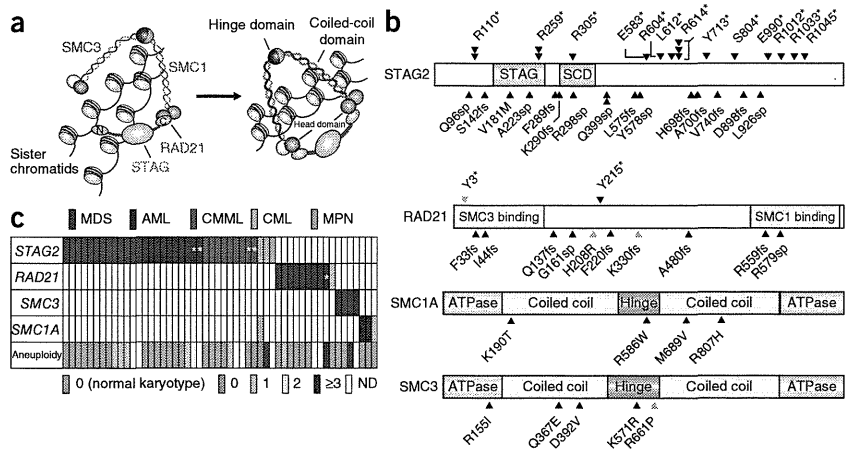


Figure 1 Genetic alterations of the cohesin complex in myeloid neoplasms. (**a**) Cohesin holds chromatin strands within a ring-like structure that is composed of four core components STAG, RAD21, SMC1 and SMC3. (**b**) Mutations in the core components of the cohesin complex found in myeloid malignancies (black arrowheads) and myeloid leukemia-derived cell lines (blue arrowheads). The amino acids in the alterations are referred to using their one-letter abbreviations (for example, R110* represents p.Arg110*). (**c**) Distribution of cohesin mutations and deletions showing a nearly mutually exclusive pattern among different myeloid neoplasms. Gene deletions are indicated by asterisks. The number of numerical chromosome abnormalities in each cohesin-mutated or -deleted case is shown at the bottom. ND, not determined.

and STAG proteins, together with a number of regulatory molecules such as PDS5, NIPBL and ESCO proteins (**Fig. 1a**)[4,5]. Forming a ring-like structure, cohesin is thought to be engaged in the cohesion of sister chromatids during cell division[5], post-replicative DNA repair[6,7] and the regulation of global gene expression through long-range *cis* interactions[8–12]. Germline mutations in cohesin components lead to the congenital multisystem malformation syndromes known as Cornelia de Lange syndrome and Roberts syndrome[13–15].

To investigate a possible role of cohesin mutations in myeloid leukemogenesis, we examined an additional 581 primary specimens of various myeloid neoplasms for mutations in nine cohesin or cohesin-related genes that have been implicated in mitosis[5] using high-throughput sequencing (**Supplementary Table 2**). We also investigated copy-number alterations in cohesin loci in 453 samples using SNP arrays (**Supplementary Table 3**). After excluding known and putative polymorphisms that are registered in the dbSNP or the 1000 Genomes project databases or that were predicted from multiple computational imputations, we identified a total of 60 nonsynonymous mutations involving nine genes in a total of 610 primary samples, which we validated by Sanger sequencing (**Fig. 1b** and **Supplementary Table 4**). After conservative evaluation of the probability of random mutational events across these genes, only four genes remained significantly mutated: *STAG2*, *RAD21*, *SMC1A* and *SMC3* ($P < 0.001$) (**Supplementary Table 5** and Online Methods). In addition, we detected five deletions in *STAG2* ($n = 4$) and *RAD21* ($n = 1$) (**Supplementary Fig. 2a,b** and **Supplementary Table 6**). We also found mutations in these four genes in four of the 34 myeloid leukemia cell lines studied (12%) (**Supplementary Table 7**).

We found mutations and deletions of these four genes in a mostly mutually exclusive manner in a variety of myeloid neoplasms, including acute myeloid leukemia (AML) (19/157), chronic myelomonocytic leukemia (CMML) (9/88), myelodysplastic syndromes (MDS) (18/224) and chronic myelogenous leukemia (CML) (4/64). Mutations were rate in classical myeloproliferative neoplasms (MPN) (1/77) (**Fig. 1c**, **Table 1** and **Supplementary Table 8**). In MDS, mutations were more frequent in refractory cytopenia with multilineage dysplasia and refractory anemia with excess blasts (11.4%) but were rare in refractory anemia, refractory anemia with ring sideroblasts, refractory cytopenia with multilineage dysplasia and ring sideroblasts and MDS with isolated del(5q) (4.2%) ($P = 0.044$). We also evaluated promoter methylation in 33 cases either with ($n = 12$) or without ($n = 21$) cohesin mutations or deletions for which sufficient nonamplified DNA was available using the HumanMethylation450 BeadChip; however, we found no aberrant methylations in cohesin loci, with the exception of hemimethylation of the *SMC1A* promoter that we found in two female cases (**Supplementary Fig. 3**).

We confirmed somatic origins for 17 mutations detected in 16 cases for which matched normal DNA was available (**Supplementary Table 4**). The somatic origins of an additional 23 mutations in *STAG2* or *SMC1A* found in 20 male cases were supported by the presence of reproducible wild-type signals or reads in Sanger and/or deep sequencing of the tumor samples, which were considered to originate from the X chromosome of the residual normal cells (**Supplementary Fig. 4**). In addition, for 20 mutations, the observed allele frequencies determined by pyrosequencing, deep sequencing or digital PCR showed significant deviations from the expected value for polymorphisms in the absence of apparent chromosomal alterations in a SNP array analysis ($P < 0.01$) (**Supplementary Figs. 5** and **6** and **Supplementary Tables 9–12**), suggesting their somatic origins. In addition, 32 of the 33 *STAG2* mutations and all of the nine *RAD21* mutations were either nonsense ($n = 18$), frameshift ($n = 14$) or splice-site ($n = 9$) changes, which were predicted to cause premature truncation of the protein or abnormal exon skipping (**Fig. 1b** and **Supplementary Figs. 7** and **8**). Thus, we considered the majority of the mutations to represent functionally relevant changes, probably of somatic origins (**Supplementary Table 13**).

Most of the cohesin mutations and deletions were heterozygous, except for the *STAG2* and *SMC1A* mutations on the single X chromosome in male cases ($n = 23$). In female samples, the *STAG2* promoter

Table 1 Frequencies of mutations and deletions of cohesin components in 610 myeloid neoplasms

Disease type	*n*	*STAG2*	*RAD21*	*SMC1A*	*SMC3*	Total	Percentage
MDS	224	13	2	0	3	18	8.0
CMML	88	9[a]	0	0	0	9	10.2
AML	157	10	7	2	1	19	12.1
de novo AML	120	8[a]	6	2	1	16	13.3
AML/MRC	37	2[a]	1[a]	0	0	3	8.1
CML	64	2[b]	1	2[b]	0	4	6.3
MPN	77	1	0	0	0	1	1.3
Total	610	35[b]	10	4[b]	4	52	8.5

Diseases are classified according to the World Health Organization 2008 classification. AML/MRC, AML with myelodysplasia-related changes.
[a]Two of the nine cases with *STAG2* alterations in CMML, one of the eight cases with *STAG2* alterations in *de novo* AML, one of the two cases with *STAG2* alterations in AML/MRC cases and one case with *RAD21* alteration in AML/MRC case involved genetic deletions. [b]One CML case having mutations in both *STAG2* and *SMC1A* was counted as a single case. A more detailed list is available in **Supplementary Table 8**.

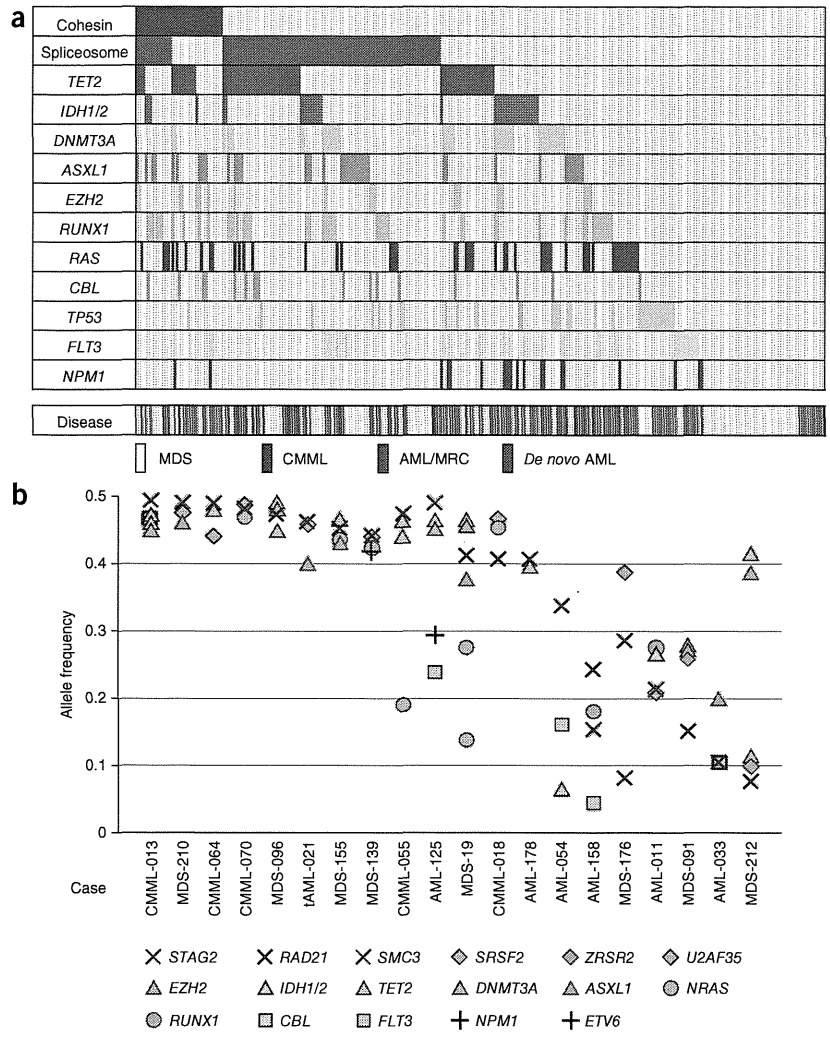


Figure 2 Relationship between cohesin mutations and other common mutations in myeloid malignancies. (**a**) Mutations in the cohesin complex and other common targets in 310 cases with different myeloid neoplasms. The corresponding disease types are shown in the bottom lane. *IDH1/2*, either *IDH1* or *IDH2*. AML/MRC, AML with myelodysplasia-related changes. (**b**) Allele frequencies of mutations in cohesin components and other coexisting mutations in 20 myeloid neoplasms determined by deep sequencing.

was hemimethylated through X inactivation regardless of mutation status (**Supplementary Fig. 3**), and a heterozygous mutation of the unmethylated *STAG2* allele would lead to biallelic *STAG2* inactivation, as has been previously documented in a female case with Ewing's sarcoma[16] and was also confirmed in a single case (CMML-036) in our cohort (**Supplementary Fig. 9**).

Cohesin mutations frequently coexisted with other mutations that are common in myeloid neoplasms and significantly associated with mutations in *TET2* ($P = 0.027$), *ASXL1* ($P = 0.045$) and *EZH2* ($P = 0.011$) (**Fig. 2a**). We performed deep sequencing of the mutant alleles in 20 available samples with cohesin mutations, which allowed for accurate determination of their allele frequencies. The majority of the cohesin mutations (15/20) existed in the major tumor populations, indicating their early origin during leukemogenesis. In the remaining five samples, we found cohesin mutations only in a tumor subpopulation, indicating that the mutations were relatively late events (**Fig. 2b**). Two male cases (MDS-176 and AML-158) harbored two independent subclones with different *STAG2* mutations, indicating that *STAG2* mutation could confer a strong advantage to pre-existing leukemic cells during clonal evolution (**Supplementary Fig. 10**). The number of mutations determined by whole-exome sequencing[3] was significantly higher in four cases with cohesin mutation or deletion compared to cases with no mutation or deletion of cohesin ($P = 0.049$) (**Supplementary Fig. 11**).

Next we investigated the possible impact of mutations on cohesin function. We examined the expression of STAG1, STAG2, RAD21, SMC3, SMC1A and NIPBL in 17 myeloid leukemia cell lines with ($n = 4$) or without ($n = 13$) known cohesin mutations, as well as in the chromatin-bound fractions of 13 cell lines (**Fig. 3a–d** and **Supplementary Table 14**)[14,17–19]. Although we observed an evaluable reduction in RAD21 expression in Kasumi-1 cells that harbored a frameshift alteration in RAD21 (p.Lys330ProfsX6) (**Fig. 3a**), alterations in P31FUJ (RAD21 p.His208Arg), CMY (RAD21 p.Tyr3X) and MOLM-7 (SMC3 p.Arg661Pro) cells were not accompanied by measurable decreases in the corresponding mutated proteins compared to wild-type cell lines. In contrast, we observed severely reduced expression of one or more cohesin components in KG-1 (STAG2)[16] and MOLM-13 (STAG1, STAG2, RAD21 and NIPBL) cells without any accompanying mutations in the relevant genes (**Fig. 3a**). We found no significant differences in protein expression of the cohesin components in cohesin-mutated and non-mutated cell lines in whole-cell extracts (**Fig. 3b**). However, expression of one or more cohesin components, including SMC1, SMC3, RAD21 and STAG2, was significantly reduced in the chromatin-bound fractions of cell lines with mutated or reduced expression of cohesin components, including Kasumi-1, KG-1, P31FUJ, MOLM-7 and MOLM-13 cells, compared with the cell lines with no known cohesin mutations or abnormal cohesin expression ($P < 0.05$), suggesting a substantial loss of cohesin-bound sites on chromatin (**Fig. 3c,d** and **Supplementary Table 14**)[14].

We next examined the effect of forced expression of wild-type cohesin components on the proliferation of a cohesin-mutated cell line (Kasumi-1) or a cell line with reduced expression of cohesin components (MOLM-13). Forced expression of wild-type *RAD21* and/or *STAG2*, but not of a truncated *RAD21* allele, induced significant growth suppression of the Kasumi-1 (with mutated *RAD21*) and MOLM-13 (with severe reduction of RAD21 and STAG2 expression) cell lines but not the K562 and TF1 (with wild-type *RAD21*) cell lines, supporting a leukemogenic role for compromised cohesin functions (**Fig. 4a–c** and **Supplementary Fig. 12a–g**). To explore the effect of forced expression of RAD21 on global gene expression, we performed expression microarray analysis of *RAD21*- and mock-transduced Kasumi-1 cells. In agreement with previous experiments with other cohesin and cohesin-related components, the magnitudes of the

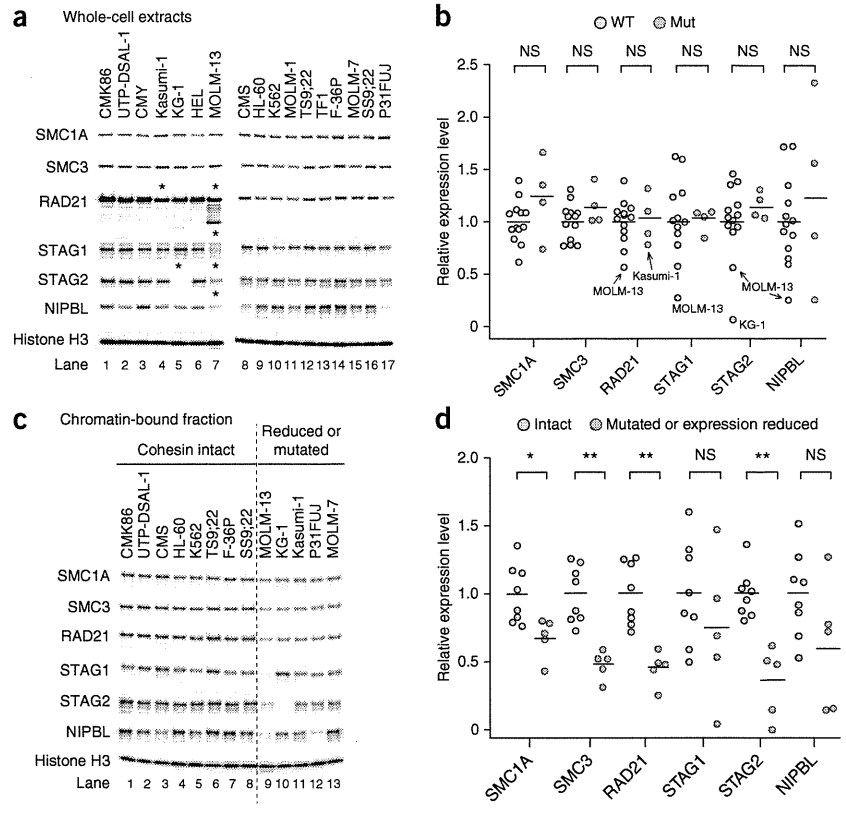
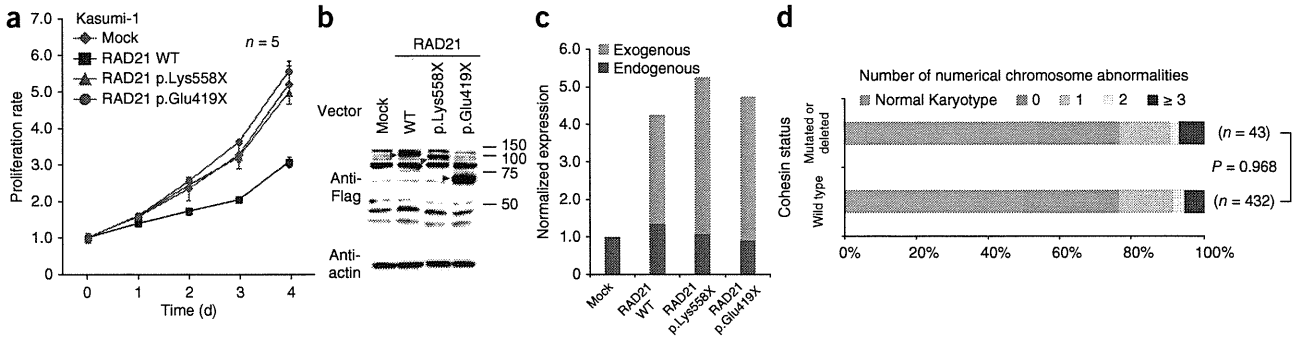


Figure 3 Abnormal cohesin expression and chromatin binding of various cohesin components in myeloid leukemic cell lines. (**a**) Protein blot analysis of the expression of various cohesin components in whole-cell extracts in 17 myeloid leukemia cell lines. Cohesin components showing evaluable reduction in expression are indicated by asterisks, which were reproducible in two independent experiments. (**b**) Expression levels of each cohesin component measured by densitometry after normalization for the mean value across all non-mutated cell lines, with histone H3 signals serving as controls. Evaluably reduced RAD21 expression in Kasumi-1 cells and severely reduced expression of cohesin components in MOLM-13 and KG-1 cells are indicated within the plots. No significant differences (NS) in the expression of the cohesin components were observed between cohesin-mutated and non-mutated cell lines (Mann-Whitney *U* test). Each circle represents a single cell line. (**c**) Protein blot analysis of cohesin components in the chromatin-bound fractions of 13 myeloid leukemia cell lines having intact cohesin (lanes 1–8), cohesin mutations and/or reduced expression of cohesin in whole-cell extracts (lanes 9–13). A representative result of two independent experiments reproducibly showing reduced chromatin-bound cohesin fractions in the cell lines in lanes 9–13 is presented. (**d**) Expression levels of cohesin components in the chromatin-bound fractions measured by densitometry after normalization for the mean value across cell lines with intact cohesin components, with histone H3 signals serving as controls. $*P < 0.05$, $**P < 0.005$ (Mann-Whitney *U* test). Horizontal bars in **b** and **d** indicate the mean values. The densitometric data are presented in **Supplementary Table 14**.

transcriptional changes induced by forced RAD21 expression were generally small[14,16,20]. However, 63 genes reproducibly and significantly showed a more than 1.2-fold increase ($n = 35$) or decrease ($n = 28$) in gene expression ($P < 0.05$), which was validated by quantitative PCR and/or RNA sequencing for 59 of the 63 genes (**Supplementary Fig. 13a–c** and **Supplementary Tables 15** and **16**).



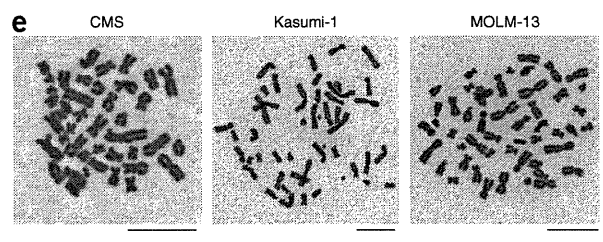


Figure 4 Impact of cohesin mutations on cell proliferation and karyotypes. (**a**) Proliferation of the Kasumi-1 cell line stably transduced with either wild-type RAD21, a truncated allele of *RAD21* (RAD21 p.Lys558X or p.Glu419X) or a mock construct measured by MTT assays ($n = 5$ wells per group). The data are shown as the means ± s.d. of the absorbance at 450 nm relative to the value at day 0. Representative results of three independent experiments are shown. (**b**) Protein blot analysis showing expression of the transduced wild-type and mutant RAD21 alleles. (**c**) Expression of endogenous and exogenous *RAD21* transcripts in Kasumi-1 cells transduced with indicated constructs measured using RNA sequencing by enumerating the corresponding reads. (**d**) The numbers of cases with numerical cytogenetic abnormalities were compared between two groups, those with and those without cohesin mutations or deletions ($P = 0.968$, χ^2). The numbers of numerical chromosome abnormalities are shown at the top. (**e**) Representative metaphases of cell lines with intact (CMS) or abnormal (Kasumi-1 and MOLM-13) cohesin components showing almost normal sister chromatid cohesion. Scale bars, 10 μm.

Mutations in the cohesin complex have recently been reported in a cohort of *de novo* AML and MDS in which four major cohesin components were mutated in 6.0–13.0% of cases[21–25]. Less frequent mutations of cohesin components have been described in other cancers, including *STAG2* mutations in glioblastoma (4/68), melanoma (1/48) and Ewing's sarcoma (1/24)[16]. In primary colon cancer samples, in which impaired cohesion and consequent aneuploidy have been implicated in oncogenesis, mutations in *SMC1A* (4/132), *NIPBL* (4/132), *STAG3* (1/130) and *SMC3* (1/130) have been reported[26]. In contrast, in our cohort of myeloid neoplasms, we found no significant differences in the number of numerical chromosome abnormalities between cohesin-mutated and non-mutated cases, and the 43 cases with cohesin mutations or deletions showed diploid or near-diploid karyotypes, including 23 cases with completely normal karyotypes (**Fig. 4d**). Therefore, in these euploid cases, cohesin-mutated cells were not clonally selected as a result of aneuploidy. Supporting this finding is the observation that expression of *scc1p*, a *RAD21* homolog, at only 13% of its normal level was sufficient for normal cohesion in yeast[27]. Furthermore, Kasumi-1 and MOLM-13 cells showed almost normal cohesion of sister chromatids, even though Kasumi-1 cells have a truncated *RAD21* allele and MOLM-13 cells have substantially reduced expression of multiple cohesin components (**Fig. 4e**).

A growing body of evidence has suggested that cohesin mediates long-range chromosomal *cis* interactions[28] and regulates global gene expression[11,12]. For example, two cohesin subunits, Rad21 and Smc3, have been implicated in the transcriptional regulation of the hematopoietic transcription factor Runx1 in zebrafish[10]. Furthermore, an up to 80% downregulation of *Nipped-B*, a *NIPBL* homolog in *Drosophila*, does not affect chromosomal segregation but does cause impaired regulation of gene expression[20]. We also previously demonstrated that only mild loss (17–28%) of cohesin binding sites within the genome results in deregulated global gene expression[14,18,19]. These observations suggest the possibility that cohesin mutations participate in leukemogenesis through the deregulated expression of genes that are involved in myeloid development and differentiation.

In conclusion, we report frequent mutations in cohesin components that involve a wide variety of myeloid neoplasms. Genetic evidence suggests that aneuploidy may not be the only leukemogenic mechanism, at least *in vivo*, and that deregulated gene expression and/or other mechanisms, such as DNA hypermutability, might also operate in leukemogenesis. Given the integral functions of cohesin for cell viability, genetic defects in cohesin might be potential targets in myeloid neoplasms[14,29].

URLs. dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; the 1000 Genomes Project, http://www.1000genomes.org/; the UCSC Genome Browser; http://genome.ucsc.edu/cgi-bin/hgGateway/; hg19, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/; RefSeq genes, http://www.ncbi.nlm.nih.gov/RefSeq/; CNAG/AsCNAR, http://www.genome.umin.jp/; dChip, http://www.dchip.org/; the Integrative Genomics Viewer, http://www.broadinstitute.org/igv/; SIFT, http://sift.jcvi.org/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/; Mutation Taster, http://www.mutationtaster.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Whole-exome sequence data have been deposited in the DNA Data Bank of Japan (DDBJ) repository under accession number DRA000433. RNA sequencing data have been deposited in the DDBJ repository under accession number DRA001013. Microarray data have been deposited in the Gene Expression Omnibus under accession number GSE47684.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006 and 21790907; S.O.), the Industrial Technology Research Grant Program from the New Energy and Industrial Technology Development Organization (NEDO; S.O.) (08C46598a), NHRI-EX100-10003NI Taiwan (L.-Y.S.), the project for development of innovative research on cancer therapies (p-direct; S.O.) and the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science and Technology, initiated by the Council for Science and Technology Policy (CSTP; S.O.). We thank Y. Hayashi (Gunma Children's Medical Centre), R.C. Mulligan (Harvard Medical School), S. Sugano (The University of Tokyo), M. Onodera (National Center for Child Health and Development, Japan) and L. Ström (Karolinska Institute) for providing materials. We thank Y. Yamazaki for cell sorting. We also thank Y. Mori, M. Nakamura, N. Mizota and S. Ichimura for their technical assistance and M. Ueda for encouragement.

AUTHOR CONTRIBUTIONS

A.K., Y.N., K.Y., A.S.-O., Y. Sato and M.S. processed and analyzed genetic materials and performed sequencing and SNP array analysis. Y. Shiraishi, Y.O., R.N., A.S.-O., H.T., T.S., K.C., M.N. and S. Miyano performed bioinformatics analyses of the sequencing data. L.-Y.S. performed pyrosequencing analysis, and A.N. and S.I. performed digital PCR. G.N. and H.A. performed methylation analysis. M.M., M.B. and K.S. performed studies on protein expression of cohesin components. A.K., M.S., T.Y., R.Y., M.O. and H.N. were involved in the functional studies. A.K. and A.S.-O. performed expression microarray experiments and their analyses. L.-Y.S., D.N., T.A., C.H., F.N., W.-K.H., T.H., H.P.K., T.N., H.M., S. Miyawaki, M.S.-Y., K.I., N.O. and S.C. collected specimens and were involved in project planning. A.K., L.-Y.S., M.M., A.S.-O. and S.O. generated figures and tables. S.O. led the entire project, and A.K. and S.O. wrote the manuscript. All authors participated in the discussion and interpretation of the data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Bejar, R., Levine, R. & Ebert, B.L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29**, 504–515 (2011).
2. Marcucci, G., Haferlach, T. & Dohner, H. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J. Clin. Oncol.* **29**, 475–486 (2011).
3. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
4. Gruber, S., Haering, C.H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765–777 (2003).
5. Nasmyth, K. & Haering, C.H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
6. Ström, L. *et al.* Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317**, 242–245 (2007).
7. Watrin, E. & Peters, J.M. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *EMBO J.* **28**, 2625–2635 (2009).
8. Dorsett, D. Cohesin, gene expression and development: lessons from *Drosophila*. *Chromosome Res.* **17**, 185–200 (2009).
9. Dorsett, D. *et al.* Effects of sister chromatid cohesion proteins on cut gene expression during wing development in *Drosophila*. *Development* **132**, 4743–4753 (2005).
10. Horsfield, J.A. *et al.* Cohesin-dependent regulation of Runx genes. *Development* **134**, 2639–2649 (2007).
11. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
12. Wendt, K.S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
13. Bose, T. & Gerton, J.L. Cohesinopathies, gene expression, and chromatin organization. *J. Cell Biol.* **189**, 201–210 (2010).
14. Deardorff, M.A. *et al.* HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* **489**, 313–317 (2012).
15. Deardorff, M.A. *et al.* RAD21 mutations cause a human cohesinopathy. *Am. J. Hum. Genet.* **90**, 1014–1027 (2012).

16. Solomon, D.A. *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
17. Beckouët, F. *et al.* An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Mol. Cell* **39**, 689–699 (2010).
18. Liu, J. *et al.* Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.* **7**, e1000119 (2009).
19. Liu, J. *et al.* Genome-wide DNA methylation analysis in cohesin mutant human cell lines. *Nucleic Acids Res.* **38**, 5657–5671 (2010).
20. Schaaf, C.A. *et al.* Regulation of the *Drosophila* enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PLoS ONE* **4**, e6202 (2009).
21. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
22. Walter, M.J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098 (2012).
23. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
24. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
25. Walter, M.J. *et al.* Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* **27**, 12785–1282 (2013).
26. Barber, T.D. *et al.* Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl. Acad. Sci. USA* **105**, 3443–3448 (2008).
27. Heidinger-Pauli, J.M., Mert, O., Davenport, C., Guacci, V. & Koshland, D. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Curr. Biol.* **20**, 957–963 (2010).
28. Hadjur, S. *et al.* Cohesins form chromosomal *cis*-interactions at the developmentally regulated IFNG locus. *Nature* **460**, 410–413 (2009).
29. Chan, D.A. & Giaccia, A.J. Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat. Rev. Drug Discov.* **10**, 351–364 (2011).

ONLINE METHODS

Patients and samples. Twenty-nine cases analyzed by whole-exome sequencing were described previously[3]. Anonymized genomic DNA from an additional 581 patients with different myeloid neoplasms were collected from collaborating institutes and used for the analyses described below. All the analyses were performed after written informed consent was obtained. This study was approved by the ethics boards of the University of Tokyo, University Hospital Mannheim, University of Tsukuba, the Munich Leukemia Laboratory, Showa University, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital.

Cell lines. The CMS, CMY, UTP-DSAL-1, MOLM-1, MOLM-7, HEL, SS9;22 and TS9;22 cell lines were provided by Y. Hayashi. 293gp and 293gpg cells were provided by R.C. Mulligan. P31FUJ and CMK-86 cells were purchased from the Health Science Research Resources Bank (Osaka, Japan). 293T, KG-1, K562 and F-36P cells were obtained from RIKEN BioResource Center Cell Bank (Tsukuba, Japan), and Kasumi-1, HL-60, MOLM-13 and TF-1 cells were from the American Type Culture Collection. Chromosome spreads were performed for the CMS, Kasumi-1 and MOLM-13 cell lines as previously described[14], except that cells were treated with colcemid (100 μg/ml) and hypotonically swollen in 75 mM KCl for 20 min.

Whole-exome sequencing. The whole-exome sequencing of the 29 paired samples of myelodysplasia was previously described[3], through which we identified a total of 497 candidate single-nucleotide variants and insertions/deletions (indels), of which 268 and 167 were determined by Sanger sequencing as true positives and negatives, respectively, with 62 mutations unconfirmed. In the present study, we updated the list of somatic mutations by rigorously validating the remaining 62 unconfirmed mutations by Sanger sequencing and also by deep sequencing (**Supplementary Table 1**).

Mutation analysis of cohesin components. In total, 534 tumor DNA samples from a variety of myeloid neoplasms were analyzed for possible mutations in nine components of the cohesin complex, *STAG1*, *STAG2*, *SMC1A*, *SMC3*, *RAD21*, *PDS5B*, *ESCO1*, *ESCO2* and *NIPBL*, using high-throughput sequencing of pooled exons amplified from pooled genomic DNA samples. In an additional 47 samples, mutations in *STAG2*, *RAD21*, *SMC1A* and *SMC3* were examined by deep sequencing after enrichment for these targets using a SureSelect custom kit (Agilent) designed to capture all of the coding exons from the target genes, performed as previously described with minor modifications in the algorithm for mutation call[30].

For pooled-DNA sequencing, all target exons (n = 232) encompassing 89,323 nucleotides were PCR amplified using a set of primers having common NotI adaptor sequences on their 5′ ends, digested with NotI, ligated using T4 ligase and sonicated to approximately 200-bp fragments using an ultrasonicator (Covaris); these fragments were used for the generation of sequencing libraries according to a modified pair-end protocol from Illumina. The libraries were then sequenced using HiSeq 2000 (Illumina) with a standard 100-bp paired end–reads protocol. On average, 99.5% of the target bases were analyzed at the depth of 12,000 per pool or 1,000 per sample. Data processing and variant calling were performed as previously described[3] with minor modifications. First, each read from a given DNA pool was aligned to the set of target sequences using BLAT[31] with the -fine option. The mapping information in a .psl format was transformed into a .sam format using the my_psl2sam script, which was further converted into the .bam format using SAMtools[32]. Among the successfully mapped reads, reads were removed from further analysis that either mapped to multiple sites, mapped with more than four mismatched bases or had more than ten clipped bases. Next, the Estimation_CRME script was run to eliminate strand-specific errors and exclude PCR-derived errors. Then, a strand-specific mismatch ratio was calculated for each nucleotide variation for both strands using the bases corresponding to 11–50 cycles. By excluding the top five cycles showing the highest mismatch rates, strand-specific mismatch rates were recalculated, and the smaller value between both strands was adopted as the nominal mismatch ratio. In addition, the nucleotide variations that were present across multiple pools were removed based on permutations across different pools using the Permut_Rm_com script because it is probable that such variations result from systemic sequencing errors. Finally, after excluding variations found in the dbSNP database, the database from the 1000 Genomes project or our in-house SNP database, the variants whose mismatch rate exceeded 0.009 were adopted as candidate mutations. Each candidate mutation was validated by Sanger sequencing of the 12 original individual DNAs from the corresponding DNA pools.

The functional impact of each amino acid substitution was evaluated by computer prediction using SIFT[33], PolyPhen-2 (ref. 34) and Mutation Taster[35]. The significance of nonsilent mutations in each cohesin component was evaluated assuming a uniform distribution of the background mutations within the coding regions, which was estimated to be ~0.3 Mb^{-1} on the basis of a previous whole-exome sequencing of myelodysplasia[3].

Determination of variant allele frequencies. Variant allele frequencies were evaluated by deep sequencing of PCR amplicons, pyrosequencing[36,37] and/or digital PCR (Fluidigm CA, US)[38–40] of the variants using nonamplified DNA. For amplicon sequencing, genomic fragments harboring the variants of interest were PCR amplified using Not1-tagged primers. Ninety-two randomly selected SNP loci that do not contain repetitive sequences were amplified using normal genomic DNA as a template, which served as the control. Touch-down PCRs using high-fidelity DNA polymerase KOD-Plus-Neo (TOYOBO, Tokyo) were performed, and an equimolar mixture of all PCR products was prepared for deep sequencing using HiSeq2000 or Miseq (Illumina), as described above, with a 75-bp or 100-bp pair end–read option. To calculate the allele frequency of each variant, all reads were mapped to the target reference sequence using BLAT[31], followed by differential enumeration of the dichotomic variant alleles. For indels, individual reads were first aligned to each of the wild-type and altered sequences and then assigned to the one with better alignment in terms of the number of matched bases.

Array-based copy-number and methylation analyses. Genomic DNA from 453 bone marrow samples with myeloid neoplasms was analyzed using GeneChip SNP genotyping microarrays as previously described using CNAG/AsCNAR software[41,42]. The results of the SNP array karyotyping for 290 of the 453 cases have been previously published[3,41–44]. The promoter methylation of each cohesin component gene was analyzed using the HumanMethylation450 BeadChip (Illumina), as previously described[30,45], in which methylation status was evaluated by calculating the ratio of methylation-specific and demethylation-specific fluorophores (β value) at each CpG site using iScan software (Illumina).

RT-PCR. Complementary DNA synthesis and quantitative RT-PCR analyses were performed as previously described[3]. The primer sequences used are listed in **Supplementary Tables 16** and **17**.

Protein expression of cohesin components in whole-cell extracts and chromatin-enriched fractions. Whole-cell extracts of myeloid cell lines were separated into soluble supernatant and chromatin-containing pellet fractions and analyzed by SDS-PAGE and protein blot analysis for the expression of different cohesin components as previously described[12,14]. Antibodies used for protein blot analysis are described in **Supplementary Table 18**.

Gene expression and cell proliferation assays. A full-length *RAD21* cDNA (BC050381) was provided by S. Sugano. A full-length *STAG2* cDNA was obtained from total cDNA derived from bone marrow cells and cloned into pBluescript. The truncated mutant of *RAD21* was subcloned by PCR. Flag-tagged *RAD21* or *STAG2* cDNAs were constructed into the retrovirus vector pGCDNsamIRESEGFP (provided by M. Onodera)[46] or a tetracycline-inducible lentiviral vector, CS-TRE-Ubc-tTA-IRESPuro. The wild-type *RAD21*, the mutant *RAD21* and/or a mock-induced retroviral vector were generated as previously described[3] and transduced into Kasumi-1, K562 and TF1 cells, which were sorted by GFP marking using a MoFlo FACS cell sorter (Beckman Coulter) or a BD FACSAria cell sorter (BD Biosciences) 48–96 h after retroviral transduction. The wild-type *RAD21*, the wild-type *STAG2* and a mock-induced lentiviral vector were generated as described previously[47], transduced into MOLM-13 cells and selected by 1 μg/ml puromycin. Gene expression was induced by 1 μg/ml doxycycline. For cell growth assays, the cells were inoculated into 96-well culture plates in RPMI 1640 medium supplemented

with 5% FCS (and 5 ng/ml GM-CSF for TF1 cells), and cell growth was monitored in three independent experiments by MTT assay using the Cell Counting Kit-8 (Dojindo Co.).

Expression microarray analysis. RNA was extracted from Kasumi-1 cells that were either mock transduced or transduced with wild-type RAD21 and analyzed in triplicate using the Human Genome U133 Plus 2.0 Array (Affymetrix) according to the manufacturer's protocol. For data analysis, raw array signals were first extracted from .CEL files using dChip Software[48]. After background correction and normalization across the six array data sets, the standardized signal value was obtained for each probe set in each of triplicate array experiments, which were compared between mock-transduced and wild-type *RAD21*–transduced cells. Two independent microarray experiments were performed. To identify transcriptionally altered genes, we used the criteria of fold change greater than ±1.2 and $P < 0.05$ (two-tailed paired t test) in two independent experiments.

RNA sequencing. RNA sequencing of *RAD21*-transduced Kasumi-1 cells and subsequent data analyses were performed as previously described[3] with minor modifications. For quantifications of expression values from the RNA sequencing data, we used a slightly modified version of RKPM (reads per kb of exon per million mapped reads) measures[49]. After removing the sequencing reads that were inappropriately aligned or that had low mapping quality, the number of bases on each exonic region for each RefSeq gene[50] was counted. Then the number of bases was normalized per kb of exon and per 100 million aligned bases. Finally, the expression value of each gene was determined by taking the maximum values among the RefSeq genes corresponding to the gene symbol.

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

Statistical analyses. The significance of the difference in frequency of cohesin component mutations between disease subtypes was tested by one-tailed Fisher's exact test. The coexistence of mutations was tested by two-tailed Fisher's direct method. The significance of the difference in the total number of somatic mutations between cohesin-mutated or -deleted and non-mutated or -deleted samples was tested by Mann-Whitney U test. Differences in the number of numerical abnormalities in cytogenetics between two groups with and without cohesin mutations or deletions was assessed by one-sided χ^2 test.

30. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* doi:10.1038/ng.2699 (24 June 2013).
31. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
32. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
33. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
34. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
35. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
36. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3–11 (2001).
37. Shih, L.Y. *et al.* Emerging kinetics of BCR-ABL1 mutations and their effect on disease outcomes in chronic myeloid leukemia patients with imatinib failure. *Leuk. Res.* **37**, 43–49 (2013).
38. Qin, J., Jones, R.C. & Ramakrishnan, R. Studying copy number variations using a nanofluidic platform. *Nucleic Acids Res.* **36**, e116 (2008).
39. Dube, S., Qin, J. & Ramakrishnan, R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS ONE* **3**, e2876 (2008).
40. Totoki, Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nat. Genet.* **43**, 464–469 (2011).
41. 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).
42. 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).
43. Hosoya, N. *et al.* Genomewide screening of DNA copy number changes in chronic myelogenous leukemia with the use of high-resolution array-based comparative genomic hybridization. *Genes Chromosom. Cancer* **45**, 482–494 (2006).
44. Sanada, M. *et al.* Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904–908 (2009).
45. Nagae, G. *et al.* Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. *Hum. Mol. Genet.* **20**, 2710–2721 (2011).
46. Nabekura, T., Otsu, M., Nagasawa, T., Nakauchi, H. & Onodera, M. Potent vaccine therapy with dendritic cells genetically modified by the gene-silencing–resistant retroviral vector GCDNsap. *Mol. Ther.* **13**, 301–309 (2006).
47. Agarwal, S. *et al.* Isolation, characterization, and genetic complementation of a cellular mutant resistant to retroviral infection. *Proc. Natl. Acad. Sci. USA* **103**, 15933–15938 (2006).
48. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98**, 31–36 (2001).
49. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
50. Pruitt, K.D., Tatusova, T., Brown, G.R. & Maglott, D.R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012).

nature genetics

Integrated molecular analysis of clear-cell renal cell carcinoma

Yusuke Sato[1,2,11], Tetsuichi Yoshizato[1,11], Yuichi Shiraishi[3,11], Shigekatsu Maekawa[1,2,11], Yusuke Okuno[1,11], Takumi Kamura[4], Teppei Shimamura[3], Aiko Sato-Otsubo[1], Genta Nagae[5], Hiromichi Suzuki[1], Yasunobu Nagata[1], Kenichi Yoshida[1], Ayana Kon[1], Yutaka Suzuki[6], Kenichi Chiba[3], Hiroko Tanaka[7], Atsushi Niida[3], Akihiro Fujimoto[8], Tatsuhiko Tsunoda[8], Teppei Morikawa[9], Daichi Maeda[9], Haruki Kume[2], Sumio Sugano[6], Masashi Fukayama[9], Hiroyuki Aburatani[5], Masashi Sanada[1,10], Satoru Miyano[3,7], Yukio Homma[2] & Seishi Ogawa[1,10]

Clear-cell renal cell carcinoma (ccRCC) is the most prevalent kidney cancer and its molecular pathogenesis is incompletely understood. Here we report an integrated molecular study of ccRCC in which ≥100 ccRCC cases were fully analyzed by whole-genome and/or whole-exome and RNA sequencing as well as by array-based gene expression, copy number and/or methylation analyses. We identified a full spectrum of genetic lesions and analyzed gene expression and DNA methylation signatures and determined their impact on tumor behavior. Defective VHL-mediated proteolysis was a common feature of ccRCC, which was caused not only by *VHL* inactivation but also by new hotspot *TCEB1* mutations, which abolished Elongin C–VHL binding, leading to HIF accumulation. Other newly identified pathways and components recurrently mutated in ccRCC included PI3K-AKT-mTOR signaling, the KEAP1-NRF2-CUL3 apparatus, DNA methylation, p53-related pathways and mRNA processing. This integrated molecular analysis unmasked new correlations between DNA methylation, gene mutation and/or gene expression and copy number profiles, enabling the stratification of clinical risks for patients with ccRCC.

Renal cell carcinomas (RCCs) constitute 2–3% of all adult malignancies, with 271,000 new cases and 116,000 related deaths estimated worldwide in 2008 (ref. 1). RCC can be histologically classified into several subtypes, among which ccRCC is the most common, accounting for 70–80% of all kidney cancers[2]. Although immunomodulation using interferon-α and VEGF and/or mTOR inhibition has been applied as systemic therapy for patients with locally advanced or metastatic disease[3], complete surgical resection remains the only curative treatment for ccRCC, except for high-dose interleukin-2, which is used for only limited cases[3]. Genetically, ccRCC is characterized by a very high frequency of biallelic *VHL* inactivation caused by allelic deletion or loss of heterozygosity (LOH) on chromosome 3p (>90%)[4] along with gene mutation (~50%)[5,6] or promoter hypermethylation (5–10%)[7]. In addition, recent whole-exome and targeted sequencing studies have identified frequent recurrent mutations in genes involved in chromatin modification, such as *PBRM1* (ref. 8), *SETD2* (ref. 9), *KDM5C*[9], *KDM6A*[9] and *BAP1* (refs. 10,11), as well as in those involved in the ubiquitin-mediated proteolysis pathway[11]. However, in previous studies, gene mutations were comprehensively investigated for entire coding sequences in only a limited number of cases, and other genetic or epigenetic lesions, including structural abnormalities and DNA methylation, have not been addressed in a comprehensive manner. Thus, knowledge about genetic and/or epigenetic alterations in ccRCC is most likely still incomplete. For example, a subset of ccRCC cases has no detectable *VHL* alterations, and pathogenesis in this subset is poorly characterized compared to that in *VHL*-mutated ccRCC cases in which accumulated hypoxia-inducible factors (HIFs) have a critical role.

Here we performed an integrated molecular study of ccRCC in which ≥100 ccRCC specimens were simultaneously analyzed by whole-genome and/or whole-exome and RNA sequencing in conjunction with microarray-based gene expression, DNA methylation and genomic copy number analyses and immunohistochemistry (Online Methods and **Supplementary Table 1**). An extended cohort of 240 ccRCC specimens, including 106 discovery specimens, was analyzed by SureSelect-based targeted deep sequencing to validate and clarify the effects of major genetic lesions. In addition to previously described common

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [2]Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [4]Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya, Japan. [5]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [6]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [7]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [8]Center for Genomic Medicine, RIKEN, Yokohama, Japan. [9]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [10]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [11]These authors contributed equally to this work. Correspondence should be addressed to S.O. (sogawa-tky@umin.ac.jp).

Received 15 April; accepted 18 June; published online 24 June 2013; doi:10.1038/ng.2699

mutational targets, we identified new mutated genes and pathways that are involved in the pathogenesis of ccRCC, including potentially druggable molecular targets. We also identified unique correlations between mutations, gene expression, DNA methylation and copy number profiles. Our study highlights the role of integrated genome, transcriptome and methylome analyses in clarifying tumor biology and identifying potential therapeutic targets in human cancers.

RESULTS

Whole-genome and whole-exome sequencing

The mean coverage by whole-genome sequencing for paired tumor-normal DNA from 14 ccRCC specimens was 47.2× and 33.6× with 95% and 90% of the entire genome analyzed with ≥20 independent reads on average, respectively (**Supplementary Fig. 1**). A total of 71,424 somatic changes, including 68,273 single-nucleotide variants (SNVs) and 3,151 insertion and/or deletion polymorphisms (indels), were detected in 14 cases (1.7 per megabase per sample) with a true positive rate (TPR) of 99% (630 of the 634 tested were confirmed) (**Supplementary Fig. 2a** and **Supplementary Table 2**). The spectrum of SNVs was over-represented by T>C/A>G transitions followed by C>T/G>A transitions and C>A/G>T transversions. C>T/G>A transitions are predominant in most cancer types[12], whereas T>C/A>G transitions and C>A/G>T transversions were characteristic of ccRCC and have also been reported in hepatocellular carcinoma[13–16] (**Supplementary Fig. 2c**). The mean number of structural variations per case was 12 (range of 0–35) with no apparent breakpoint cluster regions (**Supplementary Fig. 3**).

Table 1 Significantly mutated genes in whole-exome analysis of 106 ccRCCs

Gene	Missense mutations	Nonsense, indel or splicing mutations	Total mutations	Samples	Passenger probability (*P* value)	*q* value
VHL	19	23	42	42	1.32×10^{-102}	1.03×10^{-99}
PBRM1	4	24	28	28	2.63×10^{-36}	1.02×10^{-33}
BAP1	3	5	8	8	1.82×10^{-9}	4.71×10^{-7}
TCEB1	5	0	5	5	7.07×10^{-9}	1.37×10^{-6}
SETD2	5	7	12	12	2.06×10^{-8}	3.20×10^{-6}
FPGT	4	1	5	3	1.13×10^{-7}	1.46×10^{-5}
MUDENG	6	1	7	2	3.38×10^{-7}	3.75×10^{-5}
KEAP1	3	2	5	5	5.95×10^{-5}	5.78×10^{-3}
TET2	7	1	8	6	5.59×10^{-5}	4.83×10^{-3}
MUC4	6	0	6	6	1.02×10^{-4}	7.91×10^{-3}
MLLT10	3	0	3	3	2.30×10^{-4}	1.62×10^{-2}
MSGN1	3	0	3	2	2.85×10^{-4}	1.85×10^{-2}
KRT32	3	1	4	4	2.21×10^{-4}	1.32×10^{-2}
M6PR	1	2	3	3	2.77×10^{-4}	1.54×10^{-2}
RPL14	3	0	3	2	3.90×10^{-4}	2.02×10^{-2}
GRB7	4	0	4	4	4.20×10^{-4}	2.04×10^{-2}
TP53	1	2	3	3	3.85×10^{-4}	1.76×10^{-2}
CSMD3	8	1	9	8	7.08×10^{-4}	3.06×10^{-2}
DNHD1	3	1	4	3	6.44×10^{-4}	2.64×10^{-2}
PIK3CA	5	0	5	5	6.90×10^{-4}	2.68×10^{-2}
NLRP12	3	0	3	3	8.93×10^{-4}	3.31×10^{-2}
VMO1	2	0	2	2	9.89×10^{-4}	3.49×10^{-2}
OR4C13	2	1	3	3	1.10×10^{-3}	3.72×10^{-2}
KCNMA1	4	1	5	5	1.24×10^{-3}	4.00×10^{-2}
LMAN2L	1	2	3	2	1.69×10^{-3}	5.24×10^{-2}
MTOR	7	0	7	6	1.44×10^{-3}	4.31×10^{-2}
ZNF536	5	0	5	5	1.63×10^{-3}	4.70×10^{-2}
YIPF3	2	1	3	2	1.57×10^{-3}	4.36×10^{-2}

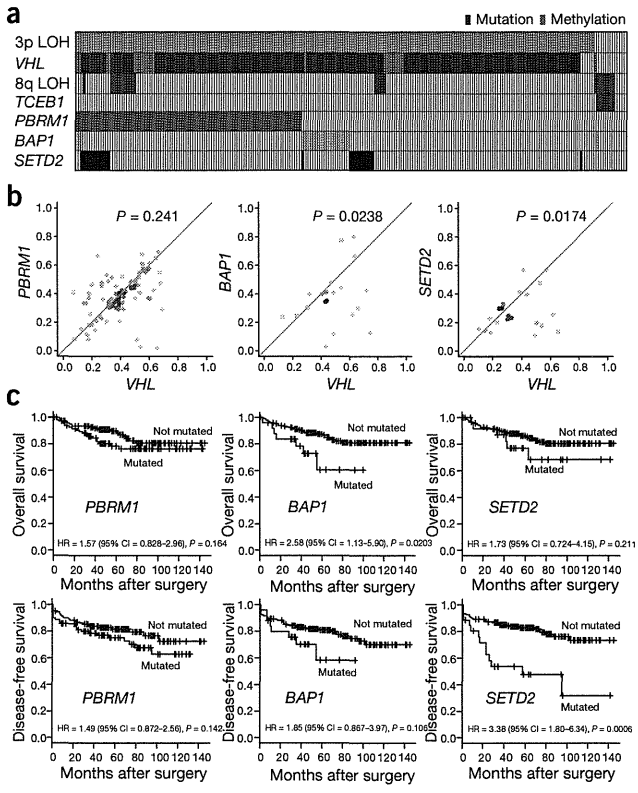


Figure 1 Mutations in 3p target genes and their impact on survival. (**a**) Distribution of common gene mutations and LOH in 240 ccRCC specimens. (**b**) Diagonal plots of observed mutant allele frequencies for *VHL* (*x* axes) and *PBRM1*, *BAP1* and *SETD2* (*y* axes). *P* values were calculated using the paired *t* test (two-sided). (**c**) Effects of common gene mutations on overall survival (top) and disease-free survival (bottom) for 240 ccRCC cases. *P* values were calculated using the log-rank test.

On average, 47 non-silent mutations were identified per case (**Supplementary Fig. 2b**), which accounted for approximately 0.92% of all somatic mutations. The numbers of mutations within coding, intronic (regulatory) and intergenic sequences were roughly proportional across the 14 cases, indicating that mutations were largely random events. To identify the complete spectrum of driver gene targets, we analyzed a total of 106 paired ccRCC specimens by whole-exome sequencing (SureSelect v.4, Agilent Technologies) in which approximately 89% of the target sequences were covered by ≥20 independent reads (**Supplementary Fig. 4**). A total of 5,171 non-silent somatic mutations (48.8 per tumor) were detected with a TPR of 96% (559 of the 582 tested were confirmed). These consisted of 4,234 missense mutations, 232 nonsense mutations, 140 splice-site mutations, 557 indels and 8 read-through changes (**Supplementary Table 3**).

In the 14 specimens that were analyzed by both whole-genome and whole-exome sequencing, 539 of the 839 non-silent mutations (64%) were identified with both platforms. However, reflecting its higher coverage (129×), whole-exome sequencing more efficiently captured the subclonal mutations harbored by a subset of the tumor population, which consequently had lower allele frequencies (**Supplementary Figs. 5** and **6**). Only whole-genome sequencing captured 117 mutations, for which coverage depths were lower in whole-exome than in whole-genome sequencing in most cases (n = 96), even though the mean

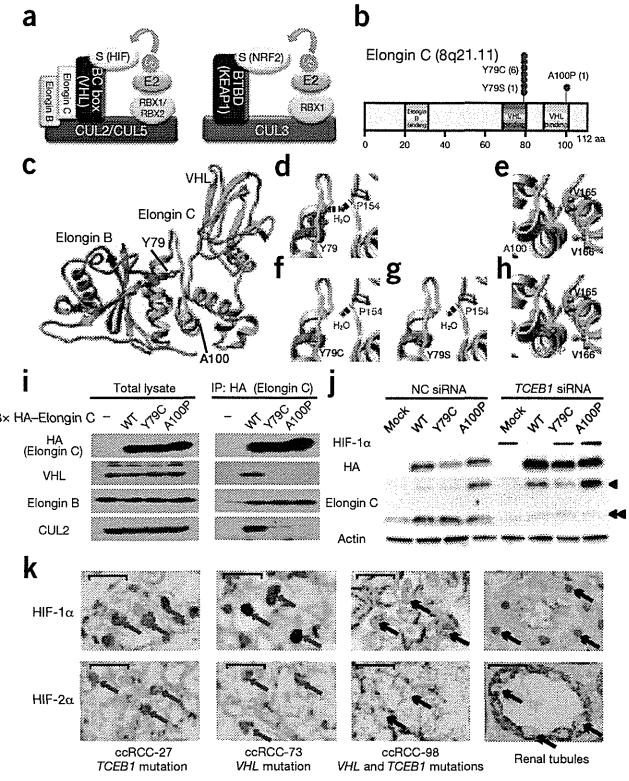


Figure 2 New *TCEB1* mutations and HIF accumulation. (**a**) Two examples of cullin–RING ubiquitin ligase system molecular assemblies using CUL2 or CUL5 (left) and CUL3 (right) that interact with the BC-box protein–Elongin C–Elongin B complex and BTB protein, respectively, to recruit substrate for ubiquitination and subsequent degradation. VHL and KEAP1 are examples of BC-box and BTB proteins, respectively, that recruit HIF and NRF2 proteins for ubiquitin-mediated degradation. (**b**) *TCEB1* mutations (8 of 240 tumors) affect the domains for binding to VHL in Elongin C. (**c–e**) Structure of the VHL complex comprising Elongin B, Elongin C and VHL (**c**), with the positions of mutated amino acids Tyr79 (**d**) and Ala100 (**e**) in Elongin C indicated. (**f,g**) A critical hydrogen bond between Tyr79 in Elongin C and Pro154 in VHL (**d**) was predicted to be abolished in the Tyr79Cys (**f**) and Tyr79Ser (**g**) Elongin C mutants. (**h**) Hydrophobic binding around Ala100 in Elongin C and Val165 and Val166 in VHL (**g**) could be compromised in the Ala100Pro Elongin C mutant. (**i**) Protein blotting for the indicated components of the VHL-CUL2 complex in total cell lysates (left) and in lysates after immunoprecipitation (IP) with antibody to HA (Elongin C) (right). Lysates were from HEK 293T cells transduced with mock, wild-type (WT) or mutant *TCEB1* constructs (encoding Tyr79Cys and Ala100Pro Elongin C). (**j**) Protein blotting for the effect of *TCEB1* (Elongin C) mutations on HIF accumulation using non-specific siRNA (left) or siRNA specific for endogenous *TCEB1* (right). Endogenous and exogenous Elongin C and HIF-1α were examined. Exogenous 3′ HA-tagged Elongin C was detected with an antibody to HA. Note that antibody to Elongin C could discriminate slower migrating exogenous protein (single arrowhead) from faster migrating endogenous protein (double arrowheads). (**k**) Immunohistochemical analysis of HIF-1α and HIF-2α expression in representative cases from primary ccRCC specimens with *TCEB1* mutations ($n = 5$), *VHL* mutations ($n = 92$) or without *TCEB1* or *VHL* mutations ($n = 9$) as well as in normal kidney tissue ($n = 1$). Red and black arrows indicate positive and negative nuclear immunoreactivity, respectively. Scale bars, 20 μm.

depths of coverage of the entire targeted regions were much higher in whole-exome sequencing (129×) than in whole-genome sequencing (~50×). The coverage in whole-exome sequencing was especially low (<8×) for 45 variants owing to high GC content (for 30 variants) or to no bait designing in targeted exome capture at all (for 15 variants). Subsequent deep sequencing of mutations identified intratumoral heterogeneity in 12 of the 14 ccRCC specimens (**Supplementary Fig. 7**), with the presence of heterogeneity was more explicitly demonstrated in a previous study using whole-exome and targeted deep sequencing combined with extensive multisite sampling from the same tumors and/or metastasized tumor blocks[17].

Recurrent mutations in 3p targets

In whole-genome and/or whole-exome sequencing of the 106 ccRCC specimens, recurrent mutations were observed in 777 genes, of which 28 were considered to be significantly mutated ($q < 0.05$) compared to background mutation rates (**Table 1** and **Supplementary Table 4**). Of the top five significantly mutated genes, *VHL*, *PBRM1*, *BAP1* and *SETD2* were all located within the common site of LOH at 3p between the 3p25 and 3p21 segments (**Supplementary Fig. 8**) and were considered to be the targets of the LOH at 3p found in more than 90% of ccRCC specimens.

Mutations of these common targets were further investigated in detail by deep and/or Sanger sequencing of the relevant genes in combination with assays for DNA methylation status and SNP array–based allele-specific copy number analysis in the extended cohort of paired tumor-normal DNA samples from 240 ccRCC cases (**Fig. 1a** and **Supplementary Fig. 9**). LOH at 3p was found in 226 specimens (94%), which was caused either by simple 3p loss ($n = 175$) or copy-neutral LOH (uniparental disomy, UPD; $n = 51$). There were no significant differences in the mutation rate of 3p target genes in cases with 3p loss and those with 3p UPD. Mutation or promoter hypermethylation was rarely found in cases without LOH at 3p. The vast majority of the 226 cases with LOH at 3p ($n = 221$; 97.8%) had the remaining *VHL* allele affected either by somatic mutation ($n = 197$; 16 nonsense mutations, 70 missense mutations, 100 indels and 11 splice-site mutations) or promoter methylation ($n = 24$). Inactivation of other genes was exclusively caused by gene mutation. Almost all mutations (147/149) involving *PBRM1* (98/98), *SETD2* (25/26) and *BAP1* (24/25) were found in a subset of *VHL*-inactivated cases (**Fig. 1a**) in which *SETD2* and *BAP1* mutations tended to show significantly lower allelic burdens than coexisting *VHL* mutations (**Fig. 1b**). This finding indicates that *SETD2* and *BAP1* mutations are likely to be acquired and selected from within pre-existing *VHL*- and/or *PBRM1*-mutated clones and contribute to tumor progression, as suggested in a recent report.

We next investigated the impact of these mutations on survival and tumor recurrence. In accordance with recent reports[18–20], there was no significant impact of *PBRM1* mutations on overall survival or disease-free survival in univariate analysis (**Fig. 1c**). In contrast, *BAP1* mutations, which were mutually exclusive with *PBRM1* mutations[10] ($P = 2.05 \times 10^{-3}$, Fisher's exact test), were significantly associated with shorter overall survival time (hazards ratio (HR) = 2.58, 95% confidence interval (CI) = 1.13–5.90; $P = 0.0203$), although their impact on relapsed disease was less prominent ($P = 0.106$). This effect could be partly owing to the effects of *SETD2* mutations in *BAP1* mutation–negative cases, as *SETD2*-mutated cases showed a very high relapse rate (HR = 3.38, 95% CI =1.80–6.34; $P = 6.00 \times 10^{-4}$) but did not necessarily have shorter overall survival times. Equivalent results were obtained in multivariate analysis in which mutations in all three 3p target genes were included (**Supplementary Table 5**).

TCEB1 mutations in ccRCC

Another highly significant mutational target was *TCEB1*, which encodes Elongin C, a 112-residue protein[21]. Elongin C was originally identified as a subunit of the heterotrimeric RNA polymerase II elongation factor complex (Elongin) that potently induces mRNA elongation but is also known to be a vital component of the VHL complex. In the latter complex,