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 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

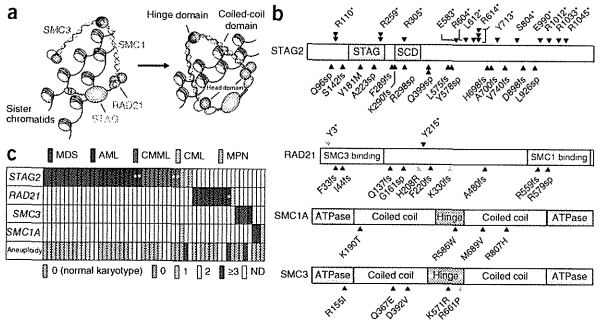

Figure 1 Genetic alterations of the cohesin complex in myeloid neoplasms. (**a**) Cohesin holds chromatin strands within a ring-like structure that is composed of four core components STAG, RAD21, SMC1 and SMC3. (**b**) Mutations in the core components of the cohesin complex found in myeloid malignancies (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.

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**.

[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

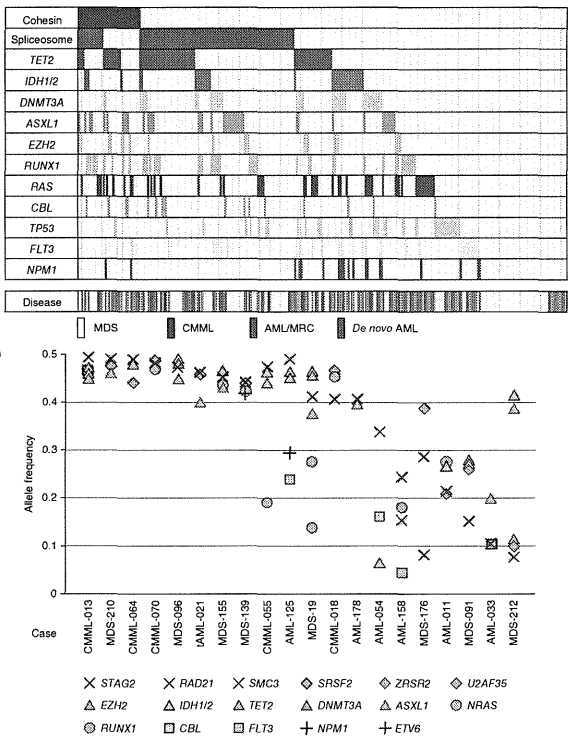

Figure 2 Relationship between cohesin *mutations and other common mutations in* myeloid malignancies. **(a)** Mutations in the cohesin complex and other common targets in 310 cases with different myeloid neoplasms. The corresponding disease types are shown in the bottom lane. *IDH1/2*, either *IDH1* 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

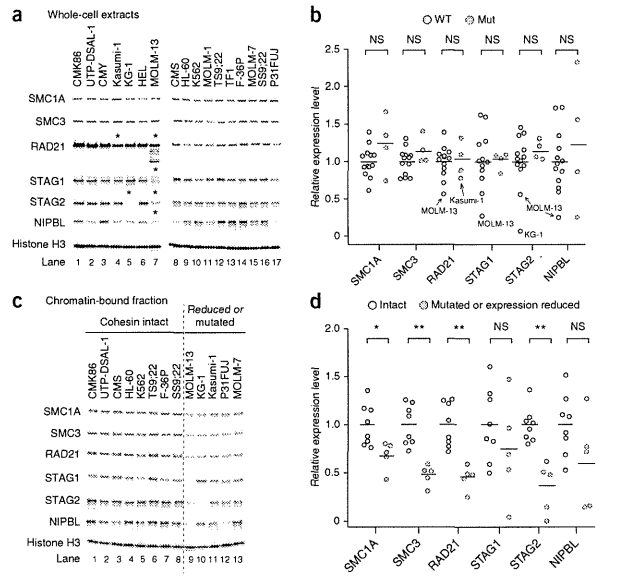

Figure 3 Abnormal cohesin expression *and chromatin binding of various cohesin* components in myeloid leukemic cell lines. **(a)** Protein blot analysis of the expression of various cohesin components in whole-cell extracts in 17 myeloid leukemia cell lines. Cohesin components showing evaluable reduction 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**).

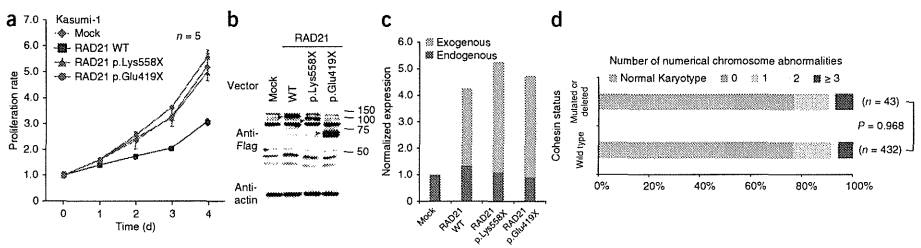

Figure 4 Impact of cohesin mutations on cell proliferation and karyotypes. **(a)** Proliferation of the Kasumi-1 cell line stably transduced with either wild-type RAD21, a truncated allele of *RAD21* (RAD21 p.Lys558X or p.Glu419X) or a mock construct measured by MTT assays ($n = 5$ wells per group). 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.

LETTERS

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).

© 2013 Nature America, Inc. All rights reserved.

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 NotI-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/*A*sCNAR 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).

doi:10.1038/ng.2731 NATURE GENETICS

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 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.

[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

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**).

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

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}

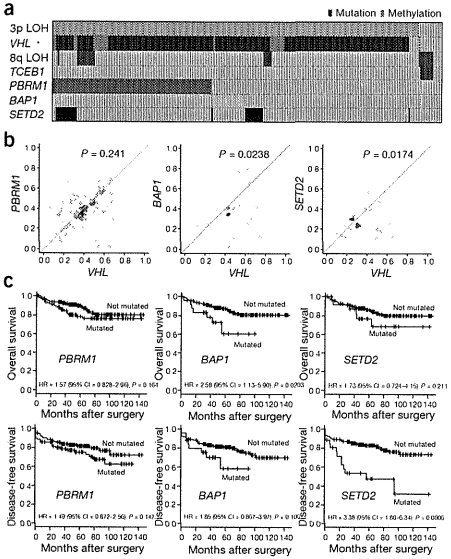

Figure 1 Mutations in 3p target genes and their impact on survival. (**a**) Distribution of common gene mutations and LOH in 240 ccRCC specimens. (**b**) Diagonal plots of observed mutant allele frequencies for *VHL* (x axes) and *PBRM1*, *BAP1* and *SETD2* (y axes). P values were calculated using the 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.

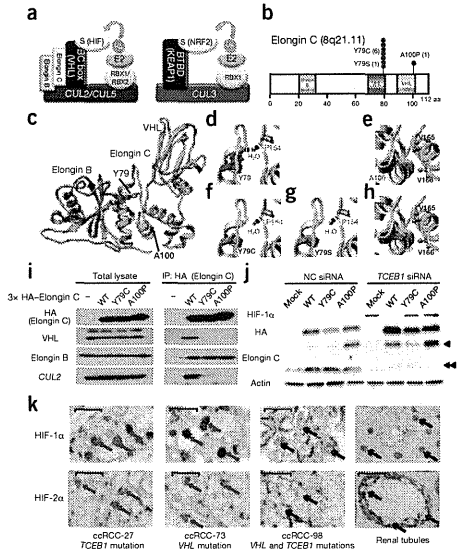

Figure 2 New *TCEB1* mutations and HIF accumulation. (**a**) Two examples of cullin–RING ubiquitin ligase system molecular assemblies using CUL2 or CUL5 (left) and CUL3 (right) that interact with the BC-box protein–Elongin C–Elongin B complex and BTB protein, respectively, to recruit substrate for ubiquitination 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, VHL, Elongin B, Elongin C and a catalytic RING subunit (RBX1), which binds ubiquitin-conjugated E2 component, are organized on a cullin scaffold protein (CUL2) to accomplish ubiquitination of VHL-bound HIF proteins (**Fig. 2a**, left)[22]. Targeted deep sequencing and methylation analysis for the entire cohort identified 8 mutations (3.3%) in *TCEB1*. No mutations were detected in other complex components, including *TCEB2* (encoding Elongin B), *CUL2* and *RBX1*. Together with *VHL* lesions, genetic and epigenetic alterations in the *VHL* complex accounted for 229 of the 240 ccRCC specimens (95.4%)in which *TCEB1* mutations and *VHL* lesions were completely mutually exclusive ($P = 2.50 \times 10^{-14}$, Fisher's exact test), further underscoring the critical role of VHL complex inactivation in the pathogenesis of ccRCC (**Fig. 1a**). There were no significant differences in the clinicopathological characteristics of cases with mutated *VHL* and those with mutated *TCEB1* (**Supplementary Table 6**).

Whereas *VHL* mutations and methylation were closely associated with LOH at 3p, *TCEB1* mutations were always accompanied by loss of chromosome 8 ($P = 3.03 \times 10^{-9}$, Fisher's exact test), leading to complete loss of wild-type *TCEB1* alleles (on 8q21) (**Fig. 1a**). However, in contrast to *VHL* mutations of which nonsense or frameshift alterations frequently result in complete loss of protein function, *TCEB1* mutations exclusively involved two conserved amino acids, Tyr79 ($n = 7$) and Ala100 ($n = 1$), with the former being a mutational hotspot (**Fig. 2b** and **Supplementary Fig. 10**). Notably, these two amino acids are positioned close together within the binding interface for the VHL protein (**Fig. 2c**): Tyr79 mediates a hydrogen bond with Pro154 of VHL via a water molecule (**Fig. 2d**), and Ala100 participates in hydrophobic interactions with Val165 and Val166 of VHL (**Fig. 2e**)[23,24]. Thus, mutations that affect these two amino acids (p.Tyr79Cys, p.Tyr79Ser and p.Ala100Pro) are predicted to abolish the interaction between Elongin C and VHL and to result in compromised ubiquitination and subsequent accumulation of HIF (**Fig. 2f–h**).

Indeed, when expressed in HEK 293T cells, wild-type Elongin C effectively coprecipitated with VHL and CUL2, whereas the interaction with VHL and CUL2 was markedly reduced with mutant Elongin C (Tyr79Cys and Ala100Pro; **Fig. 2i** and **Supplementary Fig. 11a**). We also examined the effects of mutant Elongin C on HIF-1α accumulation by exogenously expressing various Elongin C proteins in HeLa cells. As expected, HIF-1α accumulation was not observed after simply expressing these putative loss-of-function Elongin C mutants or the wild-type protein (**Fig. 2j**, left). However, when endogenous wild-type Elongin C expression was suppressed by small interfering RNA (siRNA) specific for the endogenous *TCEB1* transcripts (Online Methods and **Supplementary Table 7a**), both mock-transduced cells and mutant *TCEB1*–transduced cells but not wild-type *TCEB1*–transduced cells showed recognizable HIF-1α accumulation (**Fig. 2j**, right). These results suggest that the two *TCEB1* mutants (encoding Tyr79Cys and Ala100Pro) actually represent loss-of-function alleles with regard to VHL complex function and that biallelic inactivation is required for HIF-1α accumulation, explaining why *TCEB1* mutations were always accompanied by a deletion of the intact *TCEB1* allele on chromosome 8.

Finally, to confirm the oncogenic role of these *TCEB1* mutations in primary ccRCC, we investigated HIF-1α and HIF-2α expression in primary surgical ccRCC specimens with *TCEB1* mutations by immunohistochemistry. As with *VHL*-mutated tumors, all five tumors with *TCEB1* mutation exhibited increased HIF-1α expression in immunohistochemistry compared to normal kidney samples or tumors that lacked *TCEB1* and *VHL* mutations (**Fig. 2k**, **Supplementary Fig. 12** and **Supplementary Table 8**).

Other recurrent mutations

Other newly identified recurrent mutational targets included *TET2*, *KEAP1* and *MTOR* (**Fig. 3**). *TET2* encodes an α-ketoglutarate–dependent oxigenase frequently inactivated in myeloid malignancies by gene

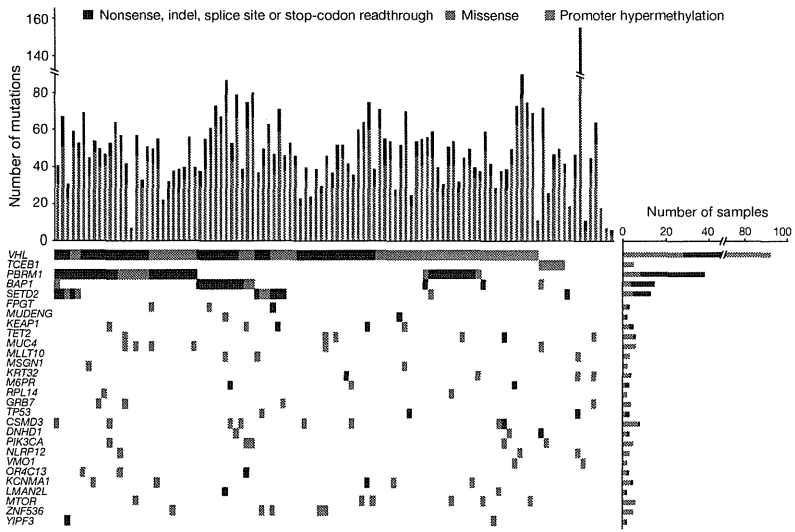

Figure 3 Significantly mutated genes and pathways for 106 ccRCC specimens. The number of somatic mutations in each case (top) and the number of cases that had alterations in significantly mutated genes (bottom right) are shown in a bar plot.

mutation[25,26]. TET2 catalyzes the conversion of 5-methylcytosine to 5-hydroxymethlcytosine, which is now believed to be a critical step in DNA demethylation. A recent study indicated that TET2 also mediates histone O-GlcNAcylation during gene transcription[27]. No *TET2* mutations have been reported in non-hematopoietic tumors, except for rare mutations in colorectal cancers (5/214 examined; 2.3%)[28]. *TET2* was mutated in 6 of 106 ccRCC cases (5.7%). Except for one frameshift mutation, five were missense mutations, of which four affected the cysteine-rich or catalytic domain (**Supplementary Fig. 13a**). In copy number analysis using SNP arrays (**Fig. 4**), *TET2* was also located within the significantly deleted regions at 4q24 ($n = 11$; 10.4%) (**Fig. 4f**). In combination, *TET2* mutations and deletions accounted for 17 ccRCC cases (16.0%), with no case having biallelic inactivation that indicated a haploinsufficiency effect of *TET2* on the pathogenesis of ccRCC. KEAP1 is a key component of another cullin–RING ubiquitin ligase complex that is involved in oxidative stress responses by regulating the ubiquitination of the KEAP1-bound NRF2 transcription factor (also known as NEF2L2) (**Fig. 2a,** right)[29]. Frequent *KEAP1* and *NRF2* mutations that abrogate their physical interaction were originally reported in squamous cell carcinoma of the lung and in other solid cancers[30–32], with KEAP1-mediated NRF2 degradation compromised, resulting in deregulated transcriptional activity of the abnormally accumulated NRF2. Of note, compromised KEAP1-mediated NRF2 degradation is also caused by abnormally accumulated fumarate in congenital fumarate hydratase deficiency[33,34], which predisposes to type 2 papillary renal cell carcinoma (pRCC2), and also by somatic mutations in *NRF2* and *CUL3* in sporadic cases with pRCC2 (ref. 35). The current study confirmed that mutually exclusive mutations in *KEAP1* ($n = 5$), *NRF2* ($n = 1$) and *CUL3* ($n = 1$) are also found in the clear-cell subtypes of RCC (6.6%) (**Supplementary Fig. 14**), together with deletions in the *CUL3* locus at 2q36 ($n = 11$; 10.4%) (**Fig. 4d**), with no case having biallelic inactivation in this pathway. *MTOR* was also a newly identified recurrent mutational target and was mutated in 6 of 106 ccRCC cases (5.7%), although a single case with an activating *MTOR* mutation was previously reported[17]. Together with mutations in *PTEN* ($n = 2$), *PIK3CA* ($n = 5$), *PIK3CG* ($n = 2$), *RPS6KA2* ($n = 3$), *TSC1* and *TSC2* (ref. 36) ($n = 2$), and other genes, a total of 28 cases (26%) had mutations that involved phosphoinositide 3-kinase (PI3K)-AKT-mTOR signaling. Except for 3 known tumor suppressor genes—*PTEN*, *TSC1* and *TSC2*—27 mutations were found in 13 genes that are thought to functionally act as oncogenes. In fact, none of the 27 mutations were nonsense, frameshift or splice-site changes, which was highly unexpected from the observed overall frequencies of these types of mutations in ccRCC ($P = 0.00946$, Fisher's exact test), suggesting that these mutations largely act as oncogenes. These mutations were mutually exclusive, except for in two cases that had both *PTEN* and *AKT2* mutations. *FGFR4* was within the significantly amplified region at 5q35 ($n = 69$), and, in total, 81 cases (76%) had genetic alterations in this pathway (**Fig. 5a**). These findings provide additional rationale for the effectiveness of mTOR inhibitors in ccRCC.

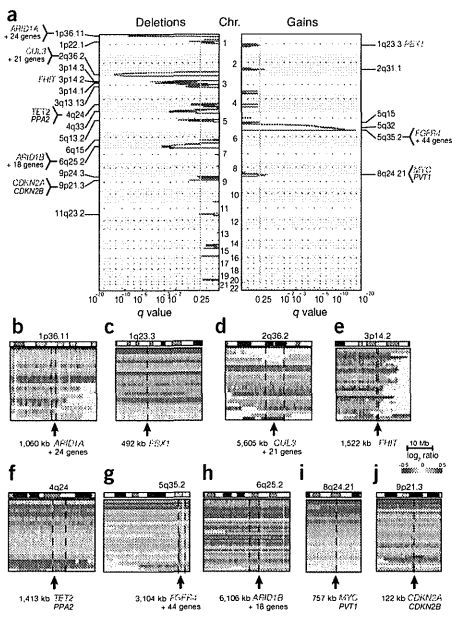

Figure 4 Significant copy number alterations in 240 ccRCC specimens. (**a**) Regions showing statistically significant increase or decrease in genomic copy number were detected using the GISTIC algorithm based on SNP array analysis. For each *q*-value peak, putative gene targets are listed. A dashed line represents the centromere of each chromosome. Red and blue lines indicate *q* value for gains and deletions, respectively. (**b**–**j**) Log-ratio copy number heatmaps are shown for gene targets at 1p36.11 (**b**), 1q23.3 (**c**), 2q36.2 (**d**), 3p14.2 (**e**), 4q24 (**f**), 5q35.2 (**g**), 6q25.2 (**h**), 8q24.21 (**i**) and 9p21.3 (**j**).

Copy number lesions and significantly affected pathways

We also performed SNP array–based copy number analysis for the 240 ccRCC specimens to identify candidate target genes involved in pathogenesis. Most copy number lesions involved large chromosomal segments, as found in LOH at 3p (94%), gain of 5q (65%), gain of 7q (41%), loss of 8p with or without loss of 8q (20%), LOH at 9p (15%), LOH at 14q (27%) and LOH at 18q (11%) (**Supplementary Fig. 15a**). Even though there was a close correlation between monosomy 8 and *TCEB1* mutation, total or partial loss of chromosome 8 was also found in cases with wild-type *TCEB1*, in which common loss of 8p seemed to be relevant to ccRCC pathogenesis. Hyperploid tumors (defined by ploidy of >2.5) accounted for 17.5% ($n = 42$) of the cases and had a significantly higher rate of metastasis ($P = 2.98 \times 10^{-5}$, Cox proportional hazards model) and poor prognosis ($P = 3.93 \times 10^{-2}$, Cox proportional hazards model). In hyperploid tumors, copy numbers in LOH involving 3p, 9p and 14q were largely neutral, suggesting that these tumors had evolved from diploid tumors with typical deletions of the relevant chromosome segments (**Supplementary Fig. 15b,c**). Supporting this notion was the fact that mutations in the 3p target genes in cases with UPD at 3p showed higher allele frequencies than those for mutations within 2*N* regions (**Supplementary Fig. 16**). Using GISTIC 2.0 analysis, significant focal gains and deletions ($q < 0.25$) were found at 20 loci (5 gains and 15 losses) that involved known tumor suppressors and oncogenes, including *ARID1A* (1p36.11), *CUL3* (2q36.2), *FHIT* (3p14.2), *TET2* (4q24), *ARID1B* (6q25.2), *CDKN2A* and *CDKN2B* (9p21.3), *PBX1* (1q23.3), *FGFR4* (5q35.2) and *MYC* (8q24) (**Fig. 4**).

Significantly affected pathways in ccRCC were further investigated by searching for statistically overrepresented gene families that were somatically mutated and expressed and/or showed copy number abnormalities (Online Methods). In addition to PI3K-AKT-mTOR signaling, significantly affected pathways (false discovery rate (FDR) < 0.01) included p53 signaling and cell cycle checkpoints, mRNA processing and the SWI/SNF complex (**Supplementary Fig. 13b** and **Supplementary Table 9**). *TP53* was mutated in only 3 cases (2.8%), but, considering these cases together with those with mutations in *ATM* ($n = 2$), *CHEK2* ($n = 1$), *MDM2* ($n = 1$) and *E2F3* ($n = 1$), copy number deletions in *CDKN2A* at 9p21 ($n = 17$; including 4 homodeletions) and gains in *MYC* at 8q24 ($n = 24$), a total of 42 cases (40%) had genetic alterations involved in p53 signaling and/or cell cycle checkpoints (**Fig. 5b**). The pathways involved in mRNA processing were not previously implicated in ccRCC. Frequent mutations in splicing machinery were reported in myelodysplastic syndromes (MDS) and in other hematopoietic neoplasms, with mutations affecting spliceosome components involved in recognizing the 3′ splice site[37]. In contrast, in ccRCC, mutated genes that were involved in mRNA processing were related to later steps in splicing, including the release of introns, 3′ end processing and mRNA export to the cytoplasm (**Supplementary Fig. 13c**).

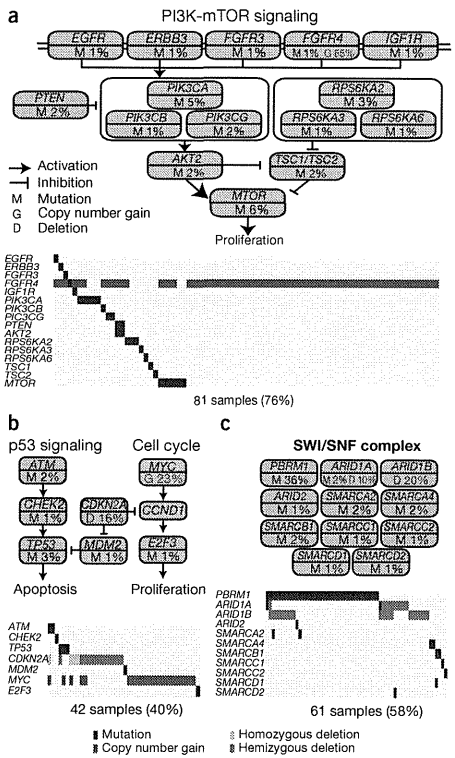

Figure 5 Significantly mutated pathways for 106 ccRCC specimens. (**a**–**c**) PI3K-mTOR signaling (**a**), p53 signaling (**b**) and the SWI/SNF complex (**c**) are commonly altered in ccRCC. Alterations are defined as somatic mutations and DNA copy number changes identified by GISTIC 2.0 analysis. Genes considered to be oncogenes or tumor suppressor genes are colored pink and blue, respectively.

PBRM1 was the second most commonly mutated gene in ccRCC and encodes a component of the SWI/SNF complex that regulates chromatin structure through ATP-dependent nucleosome remodeling. Other components of the SWI/SNF complex, including *ARID1A* and *ARID1B*, were within significantly deleted regions at 1p36 and 6q25, respectively. Taking into consideration these cases and those with mutations in *ARID1A* ($n = 2$), *SMARCA2* ($n = 2$), *SMARCA4* ($n = 2$), *SMARCB1* ($n = 2$) and other genes, a total of 61 cases (58%) had mutations and/or deletions in 11 different components of the SWI/SNF complex (**Fig. 5c**). These changes were not mutually exclusive, and 5 cases had ≥2 mutations. Multiple components of the SWI/SNF complex might be mutated in single cases in various types of cancer[38]. Notably, these five ccRCC cases had significantly worse prognosis compared to individuals with less than two mutations in the SWI/SNF complex (HR = 5.40, 95% CI = 1.56–14.5; $P = 0.0113$; Cox proportional hazards model), indicating that multiple mutations in the SWI/SNF complex could lead to aggressive phenotype in ccRCC.

Integrated molecular analysis of ccRCC

RNA sequencing was also performed in 100 ccRCC cases (**Supplementary Table 10**). In total, 44 fusion transcripts were identified in 25 specimens (**Supplementary Table 11**), of which about one-third ($n = 14$) were in frame. The lack of recurrent lesions largely obscured the significance of these fusions, except for *SETD2-QRICH1* and *NONO-TFE3* fusions. *TFE3* on Xp11 participates in promiscuous gene fusions with several partners, including *NONO*[39]. *TFE3*-containing fusions have been implicated in RCCs that are histologically distinct from typical ccRCCs (RCCs with Xp11 translocation)[39,40]. The current case with a *NONO-TFE3* fusion was positive for TFE3 expression in immunohistochemistry but showed histology otherwise indistinguishable from that typical for ccRCC cases[41] (**Supplementary Fig. 17**).

As previously reported, gene expression profiling of 101 ccRCC cases identified 2 major clusters—ccA and ccB—which were characterized by upregulated angiogenic factors and enhanced expression of genes involved in cell cycle progression, respectively (**Supplementary Fig. 18a–c**)[42]. In accordance with recent reports[10,20], Gene Set Enrichment Analysis (GSEA) showed discrete expression profiles in *BAP1*- and *PBRM1*-mutated tumors. *PBRM1*-mutated tumors were enriched for upregulated expression of the gene set with a hypoxia signature, whereas *BAP1*-mutated cases were associated with downregulated expression of target genes of polycomb repressive complex 2 (PRC2) (**Supplementary Fig. 19**).

DNA methylation profiling based on hierarchical clustering identified 3 distinct clusters that were characterized by different DNA methylation levels, including clusters with high, intermediate and low mean methylation levels, as determined using a total of 1,288 differentially methylated genes (**Fig. 6a, Online Methods** and **Supplementary Fig. 20**). Functional annotation analysis using The Database for Annotation, Visualization and Integrated Discovery (DAVID)[43] showed over-representation of homeobox genes in the 1,228 differentially methylated genes (**Supplementary Table 12**), and GSEA of the differentially methylated genes in the high and low-intermediate subgroups showed a marked enrichment in the cluster with high methylation of the gene set regulated by PRC2 and of the genes undergoing methylation at histone H3 lysine 27 (H3K27) (**Fig. 6b**). The cluster with high methylation was also characterized by a higher *BAP1* mutation rate ($P = 5.22 \times 10^{-3}$) and lower *BAP1* ($P = 2.11 \times 10^{-2}$) and higher *EZH2* ($P = 1.57 \times 10^{-5}$) expression levels (**Fig. 6c,d**). Notably, *BAP1* mutation ($P = 4.86 \times 10^{-3}$), decreased *BAP1* expression ($P = 7.46 \times 10^{-3}$) and increased *EZH2* expression ($P = 8.10 \times 10^{-3}$) were shown to be significantly associated with increased methylation of PRC2 target genes (**Supplementary Fig. 21**). These

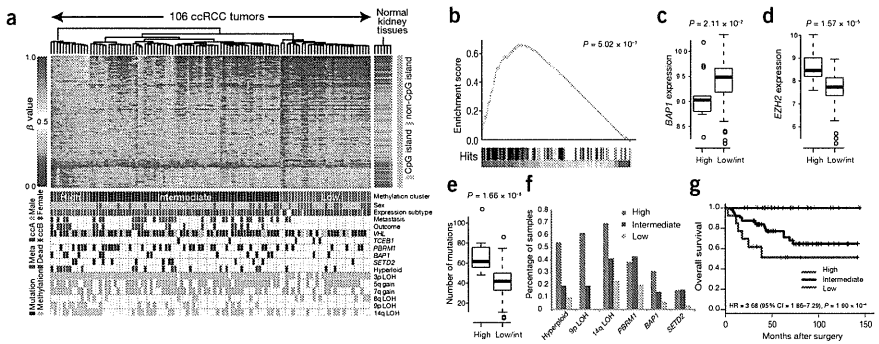

Figure 6 Correlations between DNA methylation and other genetic lesions. (**a**) Integrated view of DNA methylation clustering combined with mutation status of common driver genes, gene expression profiles and genome copy numbers. β values are for DNA methylation. Probes corresponding to CpG islands and non-CpG islands are shown on the right. Clinical outcomes and the presence-absence of metastasis are also shown. Meta, cases with distant metastasis. (**b**) GSEA showing significant enrichment of PRC2-regulated genes in differentially methylated genes in the subgroup with high DNA methylation. Hits displayed below the graph show where the members of the gene set appear in the ranked list of genes. (**c**–**g**) Integrated genetic and epigenetic analyses showed close correlation between DNA methylation status and *BAP1* expression (**c**), *EZH2* expression (**d**), total number of somatic mutations (**e**), hyperploidy, LOH at 9p, LOH at 14q and *PBRM1*, *BAP1* and *SETD2* mutations (**f**) and clinical outcome (**g**). *P* values were calculated using the *t* test in **c**–**e**, the Cochran-Armitage trend test in **f** and the log-rank test in **g**. Box plots show median, 25% and 75% quartile ranges.

findings strongly indicate that the cluster with high methylation was closely related to deregulated PRC2 activity.

Other prominent features of the cluster with high methylation included a higher number of somatic mutations ($P = 1.66 \times 10^{-6}$), more hyperploid cases ($P = 1.57 \times 10^{-3}$), cases with LOH at 14q ($P = 2.0 \times 10^{-4}$) and 9p ($P = 1.57 \times 10^{-3}$), enrichment of the ccB gene expression profile ($P = 2.46 \times 10^{-10}$) (**Fig. 6a,e,f**) and strong association with higher probability of metastasis (HR = 1.96, 95% CI = 1.17–3.30; $P = 1.11 \times 10^{-2}$) and poor overall survival (HR = 3.68, 95% CI = 1.86–7.29; $P = 1.90 \times 10^{-4}$) (**Fig. 6g**), which combined with the deregulated PRC2 profile, indicate that deregulation of polycomb-mediated gene silencing may contribute to aggressive tumor phenotype and poor clinical outcome.

DISCUSSION

Our comprehensive molecular study involving 106 ccRCC cases provided new insights into ccRCC genetics and biology and identified potential therapeutic targets. Similar to what has been observed in other cancers, the development of ccRCC seems to be shaped by the acquisition of a number of somatic gene mutations and chromosomal lesions that predominantly affect a handful of genes (*VHL*, *PBRM1*, *SETD2* and *BAP1*) and chromosome regions (3p, 5q, 9p and 14q), with many less frequent mutations additionally occurring in key targets. Pathway analysis of gene mutations disclosed several functional gene pathways, including PI3K-AKT-mTOR signaling, the KEAP1-NRF2 apparatus and mRNA processing, that were commonly affected by multiple, less frequent mutations.

A cardinal feature of ccRCC is a very high frequency of *VHL* inactivation caused by gene deletion, mutation and/or silencing via promoter methylation, leading to HIF accumulation. From this perspective, the discovery of *TCEB1* mutations and the obligatory loss of chromosome 8 in 36–42% of ccRCC cases with intact *VHL* was among the most notable findings of this study, demonstrating a new mechanism for inactivation of the VHL complex during ccRCC pathogenesis. These changes not only abolished the recruitment of VHL to the CUL2–RING ubiquitin ligase complex (CRL2), resulting in HIF accumulation, but could also compromise the recruitment of Elongin A, which is an essential component of the RNA polymerase II Elongin complex[21], and of other BC-box proteins, such as SOCS3, FEM1B and LRR1, to the CRL2 complex via Elongin C (**Supplementary Fig. 11b–e**). Loss of chromosome 8 was obligatory and resulted in the total loss of wild-type *TCEB1* alleles. Nevertheless, no nonsense or truncating *TCEB1* mutations were observed, suggesting that mutated Elongin C could still retain its interactions with unknown molecules, which would be critical for tumor cell viability.

By integrating multiple layers of different comprehensive analyses, we unmasked unique correlations between somatic mutations, DNA methylation, gene expression and copy number alterations, which were also closely linked to the clinical behaviors of tumors. In particular, correlation of the cluster with hypermethylation with hyperploid status, LOH at 9p and poor prognosis was conspicuous and was in stark contrast to the correlation of the cluster with low methylation with less frequent hyperploidy or LOH at 9p and an excellent prognosis. Clinically, the discovery of the cluster of cases with hypermethylation was a notable finding, and special attention should be paid to these cases in their management and in early detection or prevention of recurrent and/or metastatic disease. Except for frequent hyperploidy and a higher *BAP1* mutation rate, the genetic basis of this cluster of cases with hypermethylation remains unclear and will require further investigation to clarify its pathophysiology.

URLs. dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; 1000 Genomes Project, http://www.1000genomes.org/; RepeatMasker, http://www.repeatmasker.org/; Genomonfusion (in Japanese), http://genomon.hgc.jp/rna/; CNAG/AsCNAR, http://www.genome.umin.jp/; GISTIC 2.0, http://www.broadinstitute.org/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=216&p=t; Gene Set Enrichment Analysis (GSEA), http://www.broadinstitute.org/gsea/; MSigDB, http://www.broadinstitute.org/gsea/msigdb/index.jsp; the European Genome-phenome Archive, https://www.ebi.ac.uk/ega/.

METHODS

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

Accession code. Sequencing and genotype data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000509.

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

ACKNOWLEDGMENTS

We thank Y. Mori, M. Nakamura, N. Mizota and S. Ichimura for their technical assistance. We also thank M. Nangaku and N. Takeda for fruitful discussion and comments. We thank T. Kitamura (University of Tokyo) for providing pMXs-puro, M. Onodera (National Center for Child Health and Development, Japan) for providing pGCDNsamIRESEGFP and R.C. Mulligan (Boston Children's Hospital) for providing 293gp cells. This work was supported by KAKENHI (22134006), the Industrial Technology Research Grant Program from the New Energy and Industrial Technology Development Organization (NEDO) (08C46598a) 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.

AUTHOR CONTRIBUTIONS

Y. Sato, S. Maekawa, Y.N., H.S., Y. Suzuki, S.S., K.Y. and A.K. performed DNA sequencing. Y. Shiraishi, Y.O., K.C., H.T., A.F., T.T. and S. Miyano performed bioinformatics analyses of the sequencing data. T.Y., M.S. and T.K. performed the functional analyses of Elongin C mutants. Y. Sato, A.S.-O., A.N. and M.S. performed SNP array and expression array analyses. T.S., G.N. and H.A. performed methylation analysis. H.K. and Y.H. provided specimens and were also involved in planning the project. T.M., D.M. and M.F. confirmed histological diagnosis and performed immunostaining for HIF proteins. Y. Sato, T.Y., Y.O., A.S.-O. and S.O. generated figures and tables and wrote the manuscript. S.O. led the entire project. All authors participated in the discussion and interpretation of data and results.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
2. Rini, B.I., Campbell, S.C. & Escudier, B. Renal cell carcinoma. *Lancet* **373**, 1119–1132 (2009).
3. Ljungberg, B. *et al.* EAU guidelines on renal cell carcinoma: the 2010 update. *Eur. Urol.* **58**, 398–406 (2010).
4. Gnarra, J.R. *et al.* Mutations of the *VHL* tumour suppressor gene in renal carcinoma. *Nat. Genet.* **7**, 85–90 (1994).
5. Gallou, C. *et al.* Mutations of the *VHL* gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. *Hum. Mutat.* **13**, 464–475 (1999).
6. Schraml, P. *et al.* VHL mutations and their correlation with tumour cell proliferation, microvessel density, and patient prognosis in clear cell renal cell carcinoma. *J. Pathol.* **196**, 186–193 (2002).
7. Herman, J.G. *et al.* Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA* **91**, 9700–9704 (1994).
8. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
9. Dalgliesh, G.L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
10. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
11. Guo, G. *et al.* Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat. Genet.* **44**, 17–19 (2012).
12. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
13. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
14. Huang, J. *et al.* Exome sequencing of hepatitis B virus–associated hepatocellular carcinoma. *Nat. Genet.* **44**, 1117–1121 (2012).
15. Sung, W.K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
16. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
17. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
18. Hakimi, A.A. *et al.* Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators *BAP1* and *SETD2*: a report by MSKCC and the KIRC TCGA Research Network. *Clin. Cancer Res.* **19**, 3259–3267 (2013).
19. Hakimi, A.A. *et al.* Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur. Urol.* **63**, 848–854 (2013).
20. Kapur, P. *et al.* Effects on survival of *BAP1* and *PBRM1* mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol.* **14**, 159–167 (2013).
21. Aso, T., Lane, W.S., Conaway, J.W. & Conaway, R.C. Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science* **269**, 1439–1443 (1995).
22. Kamura, T. *et al.* Activation of HIF1α ubiquitination by a reconstituted von Hippel-Lindau (VHL) tumor suppressor complex. *Proc. Natl. Acad. Sci. USA* **97**, 10430–10435 (2000).
23. Stebbins, C.E., Kaelin, W.G. Jr. & Pavletich, N.P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
24. Takagi, Y., Pause, A., Conaway, R.C. & Conaway, J.W. Identification of elongin C sequences required for interaction with the von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **272**, 27444–27449 (1997).
25. Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
26. Langemeijer, S.M. *et al.* Acquired mutations in *TET2* are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
27. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561–564 (2013).
28. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
29. Zimmerman, E.S., Schulman, B.A. & Zheng, N. Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20**, 714–721 (2010).
30. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol. Cell* **21**, 689–700 (2006).
31. Shibata, T. *et al.* Cancer related mutations in *NRF2* impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl. Acad. Sci. USA* **105**, 13568–13573 (2008).
32. Kim, Y.R. *et al.* Oncogenic *NRF2* mutations in squamous cell carcinomas of oesophagus and skin. *J. Pathol.* **220**, 446–451 (2010).
33. Adam, J. *et al.* Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20**, 524–537 (2011).
34. Kinch, L., Grishin, N.V. & Brugarolas, J. Succination of Keap1 and activation of Nrf2-dependent antioxidant pathways in FH-deficient papillary renal cell carcinoma type 2. *Cancer Cell* **20**, 418–420 (2011).
35. Ooi, A. *et al. CUL3* and *NRF2* mutations confer an NRF2 activation phenotype in a sporadic form of papillary renal cell carcinoma. *Cancer Res.* **73**, 2044–2051 (2013).
36. Kucejova, B. *et al.* Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma. *Mol. Cancer Res.* **9**, 1255–1265 (2011).
37. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
38. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
39. Clark, J. *et al.* Fusion of splicing factor genes *PSF* and *NonO*(*p54^{nrb}*) to the *TFE3* gene in papillary renal cell carcinoma. *Oncogene* **15**, 2233–2239 (1997).
40. Ross, H. & Argani, P. Xp11 translocation renal cell carcinoma. *Pathology* **42**, 369–373 (2010).
41. Kuroda, N. *et al.* Review of renal carcinoma associated with Xp11.2 translocations/*TFE3* gene fusions with focus on pathobiological aspect. *Histol. Histopathol.* **27**, 133–140 (2012).
42. Brannon, A.R. *et al.* Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns. *Genes Cancer* **1**, 152–163 (2010).
43. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

ONLINE METHODS

Subjects and materials. Paired tumor-normal DNA was isolated from 240 ccRCC specimens and subjected to comprehensive molecular analyses after written informed consent was obtained. Matched normal specimens for germline controls were obtained from adjacent normal kidney tissue or from peripheral blood specimens (**Supplementary Note**). No subjects received preoperative treatments, including immunotherapies or molecular targeted therapies. Histopathological specimens were reviewed to confirm that the tumor specimens were histologically consistent with ccRCC (**Supplementary Figs. 14** and **22**). This study was approved by the ethics committee of the Graduate School of Medicine at the University of Tokyo.

Whole-genome sequencing. For whole-genome sequencing, genomic DNA was sonicated to generate approximately 400-bp fragments and was analyzed by HiSeq 2000 with the 100-bp paired-end read option according to the manufacturer's protocol.

FASTQ sequences generated by CASAVA 1.8 were aligned to the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA)[44] version 0.5.8. We attempted to realign unmapped or poorly mapped reads using BLAT[45]. For both SNVs and indels, the variant bases found in ≥7 reads in tumor samples and in ≤1 (SNVs) or 0 (indels) reads in germline samples were designated as variants. Within noncoding regions, variants found on average in 1% of the total reads in 13 unrelated germline samples were excluded from further analysis owing to the high probability that they represented false positives. Synonymous variants and variants found in either an in-house SNP database constructed from exome sequence data from 98 germline samples, dbSNP131 or the 1000 Genomes Project database were excluded. We also excluded SNVs that were in a tandem repeat region identified using tandem repeats finder[46] and indels that were in RepeatMasker.

To detect structural variations, soft-clipped sequences that could be mapped to unique genomic positions were collected. Structural variation candidates were called if they had >4 supporting read pairs in total and at least 1 read pair from each shore of the breakpoint. Structural variations having a contig sequence that could be aligned to an alternate assembly of the hg19 genome with >93% identity were excluded as false positives. Structural variations with a read depth of >150 at either breakpoint were considered to be in a repeat element and were excluded.

Whole-exome and RNA sequencing. Whole-exome sequencing was performed using target capture with SureSelect v.4 followed by sequencing as previously described[37]. Mutation calling used the EBcall algorithm[47]. For targeted sequencing, 3 mg of whole-genome DNA was amplified using the REPLI-g kit (Qiagen) and sonicated to generate a peak target size of 200 bp. Captured targets were sequenced using the Illumina HiSeq2000 platform with the 100-bp paired-end read option. A custom bait (SureSelect) library for a panel of eight genes was used for target capture, and high-throughput sequencing was performed using the 100-bp paired-end read option. Libraries for RNA sequencing were generated using the TruSeq RNA Sample Preparation kit (Illumina) and were analyzed using the Illumina HiSeq 2000 platform with 100-bp paired-end reads according to the manufacturer's protocol. Gene fusion detection was performed using Genomon-fusion.

To detect significantly mutated genes, the background mutation rate was calculated for each gene on the basis of data from the current whole-exome sequencing study, taking into consideration the effect of the replication timing of the gene during DNA replication on the background mutation rate as previously described[48,49]. Mutations showing significantly higher mutation frequencies compared to the corresponding background mutation rate were adopted as significantly mutated genes, with Benjamini-Hochberg correction applied.

Analysis of significantly mutated pathways. To detect significantly mutated pathways, each pathway registered in the Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, Reactome, Sigma-Aldrich and Signaling Transduction KE was tested with the PathScan package[50] on the basis of the background mutation rate observed in whole-exome sequencing data. We excluded somatic mutations that were not observed in RNA sequencing because we expected that they would make only a small contribution to pathogenesis. Benjamini-Hochberg correction was applied.

PCR-based deep sequencing. To validate somatic mutations and estimate mutant allele frequencies, we conducted deep sequencing using the Illumina MiSeq platform. We randomly selected SNVs and indels called through our pipeline, finding that TPRs were 99% (476/477) for coding SNVs, 99% (93/94) for noncoding SNVs, 96% (29/30) for coding indels and 97% (32/33) for noncoding indels in whole-genome sequencing. In whole-exome sequencing, TPRs were 96% (504/525) for SNVs and 96% (55/57) for indels. Primer sequences are shown in **Supplementary Tables 13** and **14**.

Microarray analyses. Global DNA methylation profiles were analyzed using the Infinium HumanMethylation450 BeadChip (Illumina) according to the manufacturer's instructions. To determine DNA methylation profiles, the following filtering steps were adopted to select probes for unsupervised clustering analysis. We first removed probes that were designed for sequences on the X and Y chromosomes. Second, we selected probes with variance ranked in the top 1% of the remaining probes. We then performed unsupervised hierarchical clustering with 3,562 probes, identifying 3 distinct clusters. The β values of 1,672 probes, containing 1,228 genes, that were differentially methylated in the high and low-intermediate groups were represented graphically using a heatmap. Genome-wide analysis of DNA copy number was conducted for 240 tumor-normal specimens using the Affymetrix GeneChip Human Mapping 250K NspI Array according to the manufacturer's protocol. Microarray data were analyzed to determine total and allele-specific copy numbers using CNAG/AsCNAR[51,52]. Significant focal copy number alterations were identified using GISTIC 2.0. Gene expression analysis used 500 ng of total RNA extracted from each tumor (n = 101) using the Human Gene Expression 4x44K v2 Microarray (Agilent) according to the manufacturer's instructions. GSEA was performed for both gene expression and DNA methylation using the curated gene sets (c2) acquired from MSigDB.

Multivariate analysis. To evaluate the impact of the mutation status of *PBRM1*, *BAP1* and *SETD2* on overall survival and disease-free survival, we performed multivariate analysis of the three genes for an extended cohort of 240 ccRCC cases.

Immunostaining for HIF-1α, HIF-2α and TFE3. Immunohistochemistry used mouse monoclonal antibody against HIF-1α (1:300 dilution; clone H1alpha67, NB100-105, Novus Biomedicals) and rabbit polyclonal antibody against HIF-2α (1:1,000 dilution; NB100-122, Novus Biomedicals). Cases were considered positive for expression when >5% of tumor cells showed nuclear immunoreactivity. Expression-positive cases were further classified on the basis of the intensity of nuclear immunoreactivity: 1+, mild; 2+, moderate; 3+, strong. Immunohistochemistry for TFE3 was performed as previously described[53] (sc-5958, Santa Cruz Biotechnology).

doi:10.1038/ng.2699

Plasmid construction. Full-length *TCEB1* cDNA was obtained by PCR amplification of DNA extracted in normal human colon and was cloned adjacent to a sequence encoding three copies of a hemagglutinin (HA) tag at the N terminus of the protein. This construct was subcloned into pMXs-puro[54], pGCDNsamIRESEGFP[55] and pCI-neo (Promega). cDNA encoding the Tyr79Cys and Ala100Pro mutants was obtained from *in vitro* mutagenesis using wild-type *TCEB1* cDNA as the template and the QuikChange Site-Directed Mutagenesis kit (Stratagene). In the pGCDNsamIRESEGFP-*TCEB1* constructs, four synonymous nucleotide substitutions, c.210A>C (p.Leu70Leu) and c.211_213TCG>AGC (p.Ser71Ser), were introduced to protect the transcripts from targeting by siRNA designed for endogenous *TCEB1* transcripts using PrimeSTAR HS DNA Polymerase (Takara Bio) according to the manufacturer's protocol. Human *TCEB2* cDNA was obtained with the same method as for *TCEB1* and was cloned into pCI-neo. For *VHL*, *FEM1B*, *TCEB3*, *LRR1* and *SOCS3*, each cDNA was tagged with three copies of sequence encoding a Flag epitope (Flag tag) at its N terminus and was cloned into pcDNA3/Puro[56].

Cell culture and gene transfer. HeLa, HEK 293T and 293gp cells were maintained in DMEM (Gibco) supplemented with 10% FCS and 1% penicillin and streptomycin in a humidified atmosphere with 5% CO_2 or 10% CO_2 at 37 °C. HeLa cells were engineered to stably express wild-type or mutant *TCEB1* or indicated empty vectors (mock) using retrovirus-mediated gene transfer (pMXs-puro-*TCEB1* or pGCDNsamIRESEGFP-*TCEB1*)[37]. HeLa cells transfected with pMXs-puro were subjected to puromycin selection. Cells transduced with pGCDNsamIRESEGFP were sorted for green fluorescent protein (GFP) signal using a flow cytometer (BD FACSAria III, Becton Dickinson). HEK 293T cells were transiently cotransfected with the indicated genes by the calcium phosphate transfection method. Cells were used for experiments after 48 h in culture.

RNA interference. Synthetic siRNAs were obtained from Takara Bio (**Supplementary Table 7b**). HeLa cells were transfected with each siRNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were used for experiments after incubation for 48 h.

Quantitative RT PCR. RNA (500 ng) extracted from the indicated cells using the RNA RNeasy Mini kit (Qiagen) was subjected to reverse transcription using the ReverTra Ace qPCR RT kit (Toyobo) according to the manufacturer's protocol. Quantitative expression levels of mRNA were measured as described previously[37]. Primers used for quantitative RT-PCR are listed in **Supplementary Table 7c**.

Antibodies. Antibodies used for immunoblot analysis are described in **Supplementary Table 15**. Antibodies were from commercial companies except for antibody against Elongin B[57].

Protein extraction and immunoblot analysis. Cells were lysed in RIPA buffer (Santa Cruz Biotechnology). Lysates were subjected to SDS-PAGE. Proteins were separated and electrophoretically transferred to polyvinylidene difluoride (PVDF) membranes. Membranes were incubated with the indicated antibodies, and proteins were detected using Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Immunoprecipitation. Cells transfected with the indicated vectors were lysed. For endogenous VHL stabilization, HeLa cells transfected with pMXs-puro-*TCEB1* were treated with 10 mM MG132 for 6 h before harvesting. HEK 293T cells cotransfected with constructs encoding HA-TCEB1, untagged TCEB2 and Flag-tagged binding proteins were subjected to immunoprecipitation using an antibody to HA or Flag, and immunoblotting was performed as described.

44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
45. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
46. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
47. Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41**, e89 (2013).
48. Hellmann, I. *et al.* Why do human diversity levels vary at a megabase scale? *Genome Res.* **15**, 1222–1231 (2005).
49. Stamatoyannopoulos, J.A. *et al.* Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
50. Wendl, M.C. *et al.* PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27**, 1595–1602 (2011).
51. 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).
52. 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).
53. Tsuji, K., Ishikawa, Y. & Imamura, T. Technique for differentiating alveolar soft part sarcoma from other tumors in paraffin-embedded tissue: comparison of immunohistochemistry for TFE3 and CD147 and of reverse transcription polymerase chain reaction for *ASPSCR1-TFE3* fusion transcript. *Hum. Pathol.* **43**, 356–363 (2012).
54. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**, 1063–1066 (2000).
55. 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).
56. Kamura, T. *et al.* VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev.* **18**, 3055–3065 (2004).
57. Garrett, K.P. *et al.* Positive regulation of general transcription factor SIII by a tailed ubiquitin homolog. *Proc. Natl. Acad. Sci. USA* **92**, 7172–7176 (1995).

doi:10.1038/ng.2699

Published November 11, 2013

Naive and memory human B cells have distinct requirements for STAT3 activation to differentiate into antibody-secreting plasma cells

Elissa K. Deenick,[1,2] Danielle T. Avery,[1] Anna Chan,[1] Lucinda J. Berglund,[1,2,3] Megan L. Ives,[1,2] Leen Moens,[1] Jennifer L. Stoddard,[4] Jacinta Bustamante,[6,7,8] Stephanie Boisson-Dupuis,[6,9] Miyuki Tsumura,[10] Masao Kobayashi,[10] Peter D. Arkwright,[11] Diana Averbuch,[12] Dan Engelhard,[12] Joachim Roesler,[13] Jane Peake,[14] Melanie Wong,[15] Stephen Adelstein,[16] Sharon Choo,[17] Joanne M. Smart,[17] Martyn A. French,[18,19] David A. Fulcher,[3] Matthew C. Cook,[20,21,22] Capucine Picard,[6,7,8] Anne Durandy,[7,8,23] Christoph Klein,[24] Steven M. Holland,[5] Gulbu Uzel,[5] Jean-Laurent Casanova,[6,8,9] Cindy S. Ma,[1,2] and Stuart G. Tangye[1,2]

[1]Immunology and Immunodeficiency Group, Immunology Research Program, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[2]St. Vincent's Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[3]Department of Immunology, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, NSW 2145, Australia
[4]Clinical Center; and [5]Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases; National Institutes of Health, Bethesda, MD 20892
[6]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, French Institute of Health and Medical Research (INSERM) U980, Necker Medical School, University Paris Descartes, Paris, 75993 Paris, France
[7]Study Center for Primary Immunodeficiencies, AP-HP, Necker Hospital, 75015 Paris, France
[8]Université Paris Descartes, Institut Imagine, 75015 Paris, France
[9]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065
[10]Department of Pediatrics, Hiroshima University Graduate School of Biomedical and Health Sciences, Hiroshima 739-0511, Japan
[11]University of Manchester, Royal Manchester Children's Hospital, Manchester M13 9WL, England, UK
[12]Department of Pediatrics and Pediatric Infectious Diseases, Hadassah-Hebrew University Medical Centre, Jerusalem 91120, Israel
[13]Department of Pediatrics, University Clinic Carl Gustav Carus, 01307 Dresden, Germany
[14]Department of Paediatrics and Child Health, Royal Children's Hospital Brisbane, Brisbane, QLD 4029, Australia
[15]Department of Allergy and Immunology, Children's Hospital at Westmead, Westmead, NSW 2145, Australia
[16]Department of Clinical Immunology, Royal Prince Alfred Hospital, Sydney, NSW 2050, Australia
[17]Department of Allergy and Immunology, Royal Children's Hospital Melbourne, Parkville, VIC 3052, Australia
[18]Department of Clinical Immunology, Royal Perth Hospital, Perth, WA 6000, Australia
[19]School of Pathology and Laboratory Medicine, University of Western Australia, Crawley, WA 6009, Australia
[20]Australian National University Medical School and [21]John Curtin School of Medical Research, Australian National University, Canberra, ACT 0200, Australia
[22]Department of Immunology, The Canberra Hospital, Canberra, ACT 2601, Australia
[23]INSERM, U768, Hôpital Necker Enfants-Malades, 75743 Paris, France
[24]Department of Pediatrics, Dr. von Hauner Children's Hospital, Ludwig-Maximilians University Munich, 80337 Munich, Germany

Long-lived antibody memory is mediated by the combined effects of long-lived plasma cells (PCs) and memory B cells generated in response to T cell–dependent antigens (Ags). IL-10 and IL-21 can activate multiple signaling pathways, including STAT1, STAT3, and STAT5; ERK; PI3K/Akt, and potently promote human B cell differentiation. We previously showed that loss-of-function mutations in *STAT3*, but not *STAT1*, abrogate IL-10– and IL-21–mediated differentiation of human naive B cells into plasmablasts. We report here that, in contrast to naive B cells, STAT3-deficient memory B cells responded to these STAT3-activating cytokines, differentiating into plasmablasts and secreting high levels of IgM, IgG, and IgA, as well as Ag-specific IgG. This was associated with the induction of the molecular machinery necessary for PC formation. Mutations in *IL21R*, however, abolished IL-21–induced responses of both naive and memory human B cells and compromised memory B cell formation in vivo. These findings reveal a key role for IL-21R/STAT3 signaling in regulating human B cell function. Furthermore, our results indicate that the threshold of STAT3 activation required for differentiation is lower in memory compared with naive B cells, thereby identifying an intrinsic difference in the mechanism underlying differentiation of naive versus memory B cells.

CORRESPONDENCE
Stuart G. Tangye:
s.tangye@garvan.org.au

Abbreviations used: Ab, antibody; AD-HIES, autosomal-dominant hyper-IgE syndrome; Ag, antigen; PC, plasma cell; Tfh cell, T follicular helper cell.

E.K. Deenick and D.T. Avery contributed equally to this paper.

The Rockefeller University Press $30.00
J. Exp. Med. 2013 Vol. 210 No. 12 2739-2753
www.jem.org/cgi/doi/10.1084/jem.20130323

Supplemental Material can be found at:
http://jem.rupress.org/content/suppl/2013/11/06/jem.20130323.DC1.html

Published November 11, 2013

Long-lived immunological memory is mediated by the combined effects of long-lived plasma cells (PCs) and memory B cells generated in response to T-dependent antigens (Ags) and underlies the success of most currently available vaccines (Ahmed and Gray, 1996; Rajewsky, 1996; Tangye and Tarlinton, 2009; Goodnow et al., 2010). PCs reside in survival niches in bone marrow and secondary lymphoid tissues and constantly produce high titers of neutralizing antibodies (Abs; Tangye and Tarlinton, 2009; Tangye, 2011). In contrast, memory B cells recirculate throughout peripheral blood, secondary lymphoid tissues, and bone marrow. Upon reexposure to Ag, they can proliferate and differentiate into Ab-secreting plasmablasts more rapidly than naive cells, thereby replenishing the PC pool and simultaneously expanding the memory cell population (Ahmed and Gray, 1996; Rajewsky, 1996; Tangye and Tarlinton, 2009).

Analysis of gene-targeted mice and humans with monogenic primary immunodeficiencies has identified some of the molecular requirements for memory B cell generation. Thus, mutations in B cell–intrinsic genes (*CD19/CD81*, *CD40*, *IKBKG*, *DOCK8*, and *IL2RG*) or genes expressed by $CD4^+$ T helper cells (*CD40LG*, *ICOS*, and *SH2D1A* [SAP]) all result in reductions in the frequencies of memory B cells and associated deficiencies in total serum Ig levels or Ag-specific Ab (Tangye and Tarlinton, 2009; Recher et al., 2011; Jabara et al., 2012; Tangye et al., 2012). We also have some understanding of the mechanisms that enable memory B cells to respond more rapidly and vigorously than naive cells to cognate Ag. First, memory B cells are recruited into division significantly earlier and undergo more rounds of division than naive cells (Bernasconi et al., 2002; Tangye et al., 2003a,b; Macallan et al., 2005). Second, memory B cells have higher expression of cell surface receptors, TLRs (TLR7/9/10), CD21, CD27, and TACI, that could enable them to respond more efficiently to co-stimulatory signals (Tangye et al., 1998; Bernasconi et al., 2002, 2003; Darce et al., 2007; Good et al., 2009). Third, memory B cells express heightened levels of CD80 and CD86 (Liu et al., 1995; Tangye et al., 1998; Ellyard et al., 2004; Good et al., 2009), which facilitate soliciting help from T helper cells. Fourth, memory B cells express lower levels of genes that restrict the entry of naive B cells into division, limiting their activation (Good and Tangye, 2007; Horikawa et al., 2007). Lastly, distinct signaling pathways downstream of the B cell receptor expressed by naive (i.e., IgM) or memory (IgG) cells have been identified that preferentially promote responsiveness of memory cells (Martin and Goodnow, 2002; Engels et al., 2009; Davey and Pierce, 2012). However, the requirements for cytokine-mediated regulation of naive and memory B cells remain to be determined.

Human B cell differentiation is regulated by the actions of numerous cytokines, with IL-10 and IL-21, produced by T follicular helper cells (Tfh cells), being key factors in promoting proliferation, isotype switching, PC differentiation, and secretion of most Ig isotypes by not only naive B cells, but also memory B cells, including both IgM^+ and isotype-switched subsets (Banchereau et al., 1994; Arpin et al., 1997; Pène et al., 2004; Ettinger et al., 2005; Bryant et al., 2007; Avery et al., 2008a,b). Although the functions of IL-10 and IL-21 on human B cells are similar, the effects of IL-21 exceed those of IL-10 by 10–100-fold (Bryant et al., 2007). The importance of IL-21 to immune regulation has been validated by the recent identification of IL-21R–deficient humans, who exhibit infectious susceptibility to several pathogens (Kotlarz et al., 2013). The predominance of IL-21 in regulating human B cell function over IL-10 is also indicated by the fact that *IL21R* mutations result in poor Ab responses after vaccination (Kotlarz et al., 2013), whereas specific Abs are produced at normal levels in individuals with mutations in *IL10/IL10R* (Kotlarz et al., 2012). IL-10 and IL-21 activate STAT1, STAT3, STAT5, as well as MAPK/ERK and PI3K/Akt pathways (Asao et al., 2001; Zeng et al., 2007; Avery et al., 2008b, 2010; Diehl et al., 2008). Autosomal-dominant hyper-IgE syndrome (AD-HIES) is caused by heterozygous mutations in *STAT3* (Holland et al., 2007; Minegishi et al., 2007; Casanova et al., 2012). These mutations operate in a dominant-negative manner, effectively reducing the level of functional STAT3 by 75%. Loss-of-function mutations in *STAT1* also underlie several immunodeficiency states, such as those characterized by selective susceptibility to infection with environmental mycobacteria and, depending on the nature of the mutation (i.e., dominant/recessive), some viruses (Boisson-Dupuis et al., 2012; Casanova et al., 2012). By examining these patients, we previously found that functional STAT3 deficiency not only severely compromised the generation of memory (i.e., $CD27^+$) B cells in vivo, but prevented IL-10– and IL-21–mediated induction of *PRDM1* (Blimp-1 [B lymphocyte induced maturation protein-1]) and *XBP1* (*X-box binding protein 1*) in naive B cells and their subsequent differentiation to the PC lineage in vitro. However, STAT3 mutant ($STAT3_{MUT}$) naive B cells could still acquire expression of *AICDA* (*activation-induced cytidine deaminase*) and undergo IL-21–induced isotype switching in vitro. In contrast, STAT1 was dispensable for human B cell differentiation in vivo and in vitro (Avery et al., 2010).

These findings led us to investigate further the role of STATs in governing human B cell differentiation. We have now discovered that the small number of memory B cells generated in STAT3-deficient patients are unaffected by these mutations; thus, they are capable of differentiating into Ab-secreting cells in response to STAT3-actvating cytokines as efficiently as normal memory cells. These findings demonstrate that the threshold of STAT3 activation required for B cell differentiation is significantly lower in memory compared with naive cells. Consequently, limiting amounts of functional STAT3 are sufficient to mediate memory, but not naive, B cell differentiation, thereby revealing an intrinsic difference in the requirements for activating naive versus memory B cells. The memory B cell deficiency in AD-HIES patients likely contributes to impaired Ag-specific Ab responses characteristic of these individuals. Thus, by targeting the residual population of STAT3-deficient memory B cells to respond to IL-21, it may be possible to improve humoral immunity in AD-HIES.

Downloaded from jem.rupress.org on February 16, 2014

Table 1. Characteristics of CD27⁻ and CD27⁺ B cells in STAT3-deficient individuals

Parameter	CD27⁻ B cells		CD27⁺ B cells	
	Normal	$STAT3_{MUT}$	Normal	$STAT3_{MUT}$
% Cells	75.2 ± 2.7	94.9 ± 0.77	24.8 ± 2.7	5.1 ± 0.77
% IgM⁺	87.0 ± 2.2	90.0 ± 2.0	47 ± 3.8	48 ± 3.1
% IgD⁺	88.0 ± 1.9	93.0 ± 1.8	41 ± 2.2	45 ± 2.0
% IgG⁺	3.2 ± 0.45	1.6 ± 0.6[a]	25 ± 2.1	32 ± 2.8[a]
% IgG1⁺ (% total IgG⁺ cells)	2.3 ± 1.0 (62.1)	0.63 ± 0.24 (40.1)	11 ± 1.7 (51.4)	22 ± 2.8 (72.7)[b]
% IgG2⁺ (% total IgG⁺ cells)	0.72 ± 0.14 (19.4)	0.54 ± 0.1 (34.4)	6.7 ± 2.1 (31.3)	3.9 ± 0.9 (12.9)
% IgG3⁺ (% total IgG⁺ cells)	0.62 ± 0.15 (16.7)	0.36 ± 0.12 (22.9)	3.6 ± 0.9 (16.8)	4.1 ± 1.1 (13.5)
% IgG4⁺ (% total IgG⁺ cells)	0.065 ± 0.02 (1.8)	0.04 ± 0.014 (2.5)	0.11 ± 0.04 (0.5)	0.27 ± 0.08 (0.9)
% IgA⁺	1.2 ± 0.2	0.64 ± 0.15[a]	17 ± 1.2	8.6 ± 1.1[c]
% CD23⁺	62.0 ± 3.9	86.0 ± 3.7[c]	18 ± 3.9	46 ± 4.7[c]

Values represent percentage (or absolute number for CD4⁺CXCR5⁺ T cells) of cells expressing the indicated surface molecule; each value represents the mean ± SEM; normal donors: n = 8–24; STAT3 patients: n = 9–27. P-values were determined by Student's t test, comparing normal with $STAT3_{MUT}$ B cells.
[a]$P < 0.05$.
[b]$P < 0.01$.
[c]$P < 0.001$.

RESULTS

$STAT3_{MUT}$ CD27⁺ B cells phenotypically resemble normal memory B cells

The population of circulating CD27⁺ memory B cells is significantly reduced in STAT3-deficient patients (n = 27) compared with normal donors (Table 1; Avery et al., 2010). In contrast, the frequency of memory B cells in STAT1-deficient individuals is comparable with normal donors (i.e., 24.4 ± 6.1%; n = 9). Although it is generally accepted that CD27 is expressed on human memory B cells (Tangye and Tarlinton, 2009), recent studies have suggested that B1 cells (Griffin et al., 2011) and some bone marrow progenitor B cells (Nilsson et al., 2005) are also CD27⁺. Conversely, a small proportion of memory B cells lack CD27 (Tangye and Tarlinton, 2009). Thus, it was important to establish the nature of the residual population of CD27⁺ B cells in $STAT3_{MUT}$ patients.

The size and granularity of CD27⁻ and CD27⁺ B cells were determined by flow cytometry. This demonstrated that CD27⁺ B cells from both normal donors and $STAT3_{MUT}$ patients were significantly larger and more granular than corresponding CD27⁻ B cells. However, these morphological features were not significantly different between cells from normal donors and $STAT3_{MUT}$ patients (Fig. 1, A–C).

We next examined a series of surface receptors that are differentially expressed by human naive and memory B cells (Liu et al., 1995; Tangye et al., 1998; Ellyard et al., 2004; Good et al., 2009). CD24, CD80, CD95, and TACI were significantly higher on CD27⁺ B cells from normal donors than on corresponding CD27⁻ B cells (Fig. 1, A and C). CD86 also tended to be higher on normal CD27⁺ versus CD27⁻ B cells (Fig. 1, A and C). The same pattern was seen for samples from $STAT3_{MUT}$ patients, with CD80, CD86, CD95, and TACI being significantly higher on CD27⁺ than on CD27⁻ B cells (Fig. 1, B and C). In contrast to these molecules, CD23 is present on normal naive B cells but is significantly down-regulated on normal memory B cells (Fig. 1, A and C; and Table 1). Interestingly, CD23 expression was dysregulated on $STAT3_{MUT}$ B cells inasmuch as its level exceeded that on normal naive B cells by >10-fold, whereas it was detected on a substantial proportion of $STAT3_{MUT}$ memory B cells (Fig. 1, B and C; and Table 1).

We also determined expression of Ig isotypes by CD27⁻ and CD27⁺ B cells from normal and STAT3-deficient individuals. Approximately ~90% of CD27⁻ B cells and ~40–50% of CD27⁺ B cells from normal and STAT3-deficient individuals expressed IgM and IgD, with the remaining memory B cells expressing predominantly IgG or IgA (Table 1). Although the proportion of $STAT3_{MUT}$ CD27⁺ B cells that expressed IgG was greater than that observed for normal CD27⁺ B cells (Table 1), the distribution of IgG subclasses within the CD27⁻ and CD27⁺ B cell subsets was comparable, with IgG1 being preferentially expressed by both CD27⁻ and CD27⁺ normal and $STAT3_{MUT}$ B cells (Table 1). Furthermore, we did not detect an enrichment of B1 cells, based on cells with a CD20⁺CD43⁺CD27⁺ phenotype, in $STAT3_{MUT}$ individuals compared with healthy donors (not depicted). Collectively, these findings confirm there is a significant contraction of the memory B cell compartment in AD-HIES and provide evidence that the small population of CD27⁺ B cells in $STAT3_{MUT}$ individuals are indeed memory B cells (Table 1).

IL-21 activates STAT1, STAT3, and STAT5 in naive and memory B cells

IL-21 has been reported to activate multiple signaling pathways in different cell types (Asao et al., 2001; Zeng et al., 2007; Diehl et al., 2008). To determine whether different subsets of human B cells used similar signaling pathways downstream of the IL-21 receptor, we examined phosphorylation of STAT proteins in normal naive, IgM memory, and isotype-switched memory B cells in response to IL-21. Phosphorylation of STAT1, STAT3, and STAT5 was greater in naive B cells than in memory B cells; however, STAT1 and STAT3 were both clearly activated by IL-21 in both IgM-expressing and Ig-switched memory B cell subsets (Fig. 2, A and B). In contrast to these STATs, IL-21 had minimal, if any, effect on phosphorylation of STAT4 or STAT6 (Fig. 2 A). Although IL-21 has been reported to activate ERK and AKT (Zeng et al., 2007), we observed no ERK or AKT phosphorylation in response to IL-21 in naive or memory B cells (Fig. 2, C and D). These results demonstrate that IL-21R signaling activates similar pathways in human naive and memory B cells, predominantly STAT1, STAT3, and STAT5.

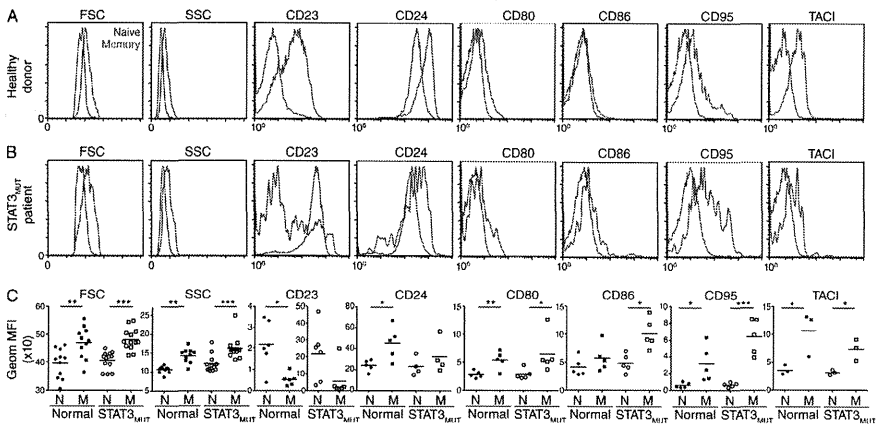

Figure 1. Morphology and phenotype of CD27⁺ B cells in $STAT3_{MUT}$ individuals resemble normal memory B cells. (A–C) PBMCs from normal donors and patients with AD-HIES caused by mutations in *STAT3* were stained with mAbs specific for CD20, CD27, and CD23, CD24, CD80, CD86, CD95, or TACI. The forward scatter (FSC) and 90° light/side scatter (SSC) and surface expression of the indicated molecules on CD27⁻ (naive) and CD27⁺ (memory) B cells were determined. The histograms in A and B are from a representative normal donor and patient, respectively, whereas the graphs in C depict the geometric mean fluorescence intensity (MFI) of each of the indicated cellular features for CD27⁻ naive (N) and CD27⁺ memory (M) B cells from 3–12 normal donors and $STAT3_{MUT}$ patients. Each value represents an individual donor or patient; the horizontal lines correspond to the mean. Because of the large difference in the level of expression of CD23 on normal versus $STAT3_{MUT}$ naive and memory B cells, individual graphs are depicted for normal and patient B cell subsets. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Mutations in *STAT3* impair the response of naive, but not memory, B cells to the stimulatory effects of IL-10 and IL-21

STAT3 is required for some, but not all, aspects of naive B cell differentiation. For instance, isotype switching was intact in $STAT3_{MUT}$ B cells, as indicated by the detection of circulating IgG⁺ and IgA⁺ B cells in AD-HIES patients ex vivo (Table 1) and the ability of their naive B cells to up-regulate *AICDA* and switch to IgG in response to IL-21 in vitro (Avery et al., 2010). Consistent with intact *AICDA* expression in $STAT3_{MUT}$ naive B cells in vitro, somatic hypermutation was comparable in normal and $STAT3_{MUT}$ memory B cells (Avery et al., 2010). These observations lead us to question whether there were also differences between naive and memory B cells in their requirement for STAT3 function to respond to cytokines such as IL-10 and IL-21, which are well known for their abilities to induce human B cell differentiation (Banchereau et al., 1994; Arpin et al., 1997; Pène et al., 2004; Ettinger et al., 2005; Bryant et al., 2007; Avery et al., 2008a,b).

Naive and memory B cells isolated from normal donors or $STAT3_{MUT}$ or $STAT1_{MUT}$ patients were cultured with CD40L alone or together with IL-21, and Ig secretion was determined after 10–12 d. Because of the limited numbers of memory cells recovered from $STAT3_{MUT}$ patients, we could only culture ~5,000 sorted B cells/well. Under these conditions, IL-21 potently promoted secretion of IgM and induced production of IgG and IgA by normal and $STAT1_{MUT}$ naive B cells (Fig. 3 A and Table 2). IL-21 substantially increased production of IgM, IgG, and IgA by CD40L-stimulated memory B cells from normal donors and STAT1-deficient individuals; however, the response of $STAT1_{MUT}$ memory B cells was significantly less (approximately fourfold) than that of normal memory cells (Fig. 3 B and Table 2). Ig secretion by $STAT3_{MUT}$ naive B cells in response to IL-21 was ~30-fold less than normal naive cells (Fig. 3 A and Table 2; $P < 0.001$). In distinct contrast, Ig secretion by IL-21–stimulated $STAT3_{MUT}$ memory B cells (either total memory cells [Fig. 3 B and Table 2] or IgM⁺ and switched subsets [not depicted]) was largely comparable with normal memory B cells,

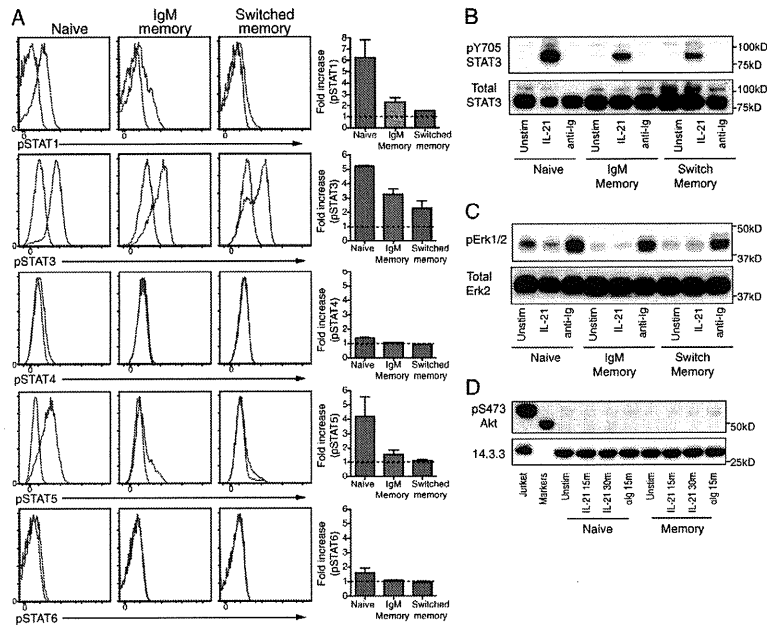

Figure 2. IL-21 induces activation of STAT1, STAT3, and STAT5 in human naive and memory B cells. Human naive, IgM memory and isotype-switched memory, or total memory, B cells were sort-purified from normal donor spleens. (A) These B cell subsets were cultured for ~18 h with anti-Ig, rested, and then cultured in the absence (red histograms) or presence (blue histograms) of IL-21 for 30 min. Phosphorylation of STAT1, STAT3, STAT4, STAT5, and STAT6 was determined by intracellular staining. Histograms on the left show representative staining in naive and memory B cells. Right panels plot increase in mean fluorescence intensity of pSTATs in naive, IgM memory, and isotype-switched memory B cells cultured with IL-21; response of unstimulated cells were normalized to a value of 1.0. These values represent the mean ± SEM of two independent experiments using B cells from different donor spleens. Identical results were obtained when the B cell subsets were prestimulated with CD40L/anti-Ig. (B–D) Human B cell subsets were cultured for ~18 h with anti-Ig, rested, and then left unstimulated or stimulated with IL-21 or anti-Ig for 15–30 min. Cells lysates were prepared and subjected to SDS-PAGE and Western blotting to detect phosphorylated or total STAT3 (B), phosphorylated or total ERK (C), or phosphorylated AKT or 14.3.3 as a loading control (D). B–D are representative of three to four similar experiments.

with the only significant difference (less than twofold) being noted for the levels of IgM secreted by normal versus $STAT\text{-}3_{MUT}$ memory B cells (Table 2). Furthermore, both IgM memory (i.e., IgM^+CD27^+) and switched (IgM/D^-CD27^+) memory cells from STAT3-deficient patients could respond to the stimulatory effects of IL-10 (not depicted).

The inability of $STAT3_{MUT}$ naive B cells to respond to IL-21 did not reflect a general impairment in differentiation because these cells produced IgM after stimulation with CD40L plus CpG (Fig. 3 C) and up-regulated expression of *AICDA* (Fig. 3 E) and Ig ε germline transcripts, a precursor to producing mature IgE (Geha et al., 2003), in response to CD40L/IL-4 (Fig. 3 F) as efficiently as normal naive B cells. Consistent with the intact response of $STAT3_{MUT}$ memory B cells to CD40L together with IL-10 or IL-21 (Fig. 3 B, Table 2, and not depicted), these cells also exhibited normal responses to stimulation with CD40L/CpG (Fig. 3 D) or CD40L/IL-4 (Fig. 3, E and F). Because many facets of lymphocyte differentiation are linked to cell division (Hodgkin et al., 1996; Deenick et al., 1999; Tangye et al., 2003a,b; Avery et al., 2005, 2008a), we also assessed the proliferative potential of $STAT3_{MUT}$ naive and memory B cells. IL-21 enhanced proliferation of both normal and $STAT3_{MUT}$ naive B cells over that induced by CD40L alone, yet the response of $STAT3_{MUT}$ naive B cells was approximately threefold less than that of normal B cells (Fig. 3 G). Consistent with the differential dependency on STAT3 function for IL-21–induced Ig secretion by naive versus memory B cells, $STAT3_{MUT}$ memory

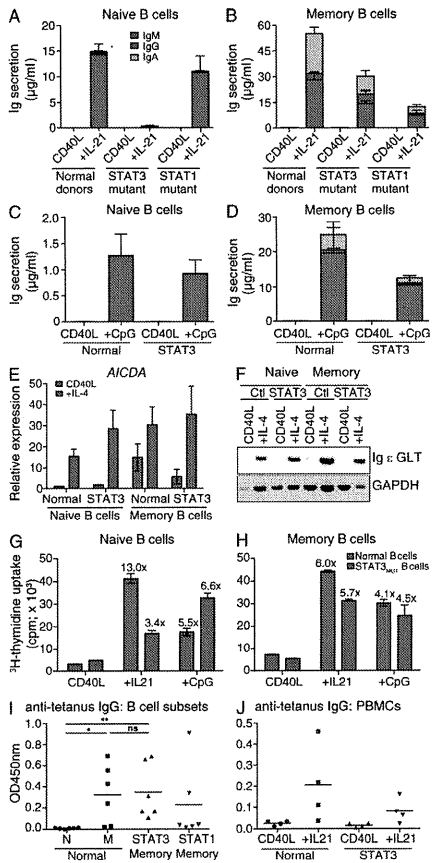

Figure 3. STAT3-deficient memory B cells differentiate into Ab-secreting cells in response to IL-21. (A–D) Naive ($CD20^+CD10^-CD27^-IgG^-$) and memory ($CD20^+CD10^-CD27^+$) B cells were sort-purified from normal donors (A and B, $n = 16$; C and D, $n = 7$), $STAT3_{MUT}$ patients (A and B, $n = 8$; C and D, $n = 7$), or $STAT1_{MUT}$ patients ($n = 6$), and then cultured with CD40L alone or together with IL-21 (A and B) or CpG (C and D). The levels of secreted IgM, IgG, and IgA were determined by ELISA after 10–12 d. The columns represent the mean ± SEM of experiments performed using naive B cells from 7–16 normal donors, 7–8 $STAT3_{MUT}$ patients, or 6 $STAT1_{MUT}$ patients. (E and F) Naive and memory B cells were sort-purified from normal donors or $STAT3_{MUT}$ patients and then cultured with CD40L alone or together with IL-4. Expression of *AICDA* (E) and Ig ε germline transcript (GLT; F) was determined by qPCR and PCR, respectively. The graphs in E represent the mean ± SEM of three experiments using B cells from different donors or patients. The gel depicted in F is representative of experiments performed using B cells from two to three different donors or patients. (G and H) Naive and memory B cells were sort-purified from a single normal donor or $STAT3_{MUT}$ patient and then cultured with CD40L alone or together with IL-21 or CpG. Proliferation was assessed after 5 d by determining incorporation of $[^3H]$thymidine during the last 18 h of culture. The graphs are the mean ± SEM of replicate cultures of naive or memory B cells from one normal donor or one $STAT3_{MUT}$ patient. The annotated values indicate the fold increase in proliferation of normal or $STAT3_{MUT}$ naive or memory B cells cultured with CD40L/IL-21 or CD40L/CpG over that induced by CD40L alone. (I and J) Normal naive (N) or normal, $STAT3_{MUT}$, or $STAT1_{MUT}$ memory (M) B cells ($n = 6$; I) or total PBMCs from normal donors or $STAT3_{MUT}$ patients ($n = 4$; J) were cultured with CD40L and IL-21 for 10–12 d. The levels of antitetanus IgG in culture supernatants were determined by ELISA using immobilized tetanus toxoid as solid phase Ag. Each symbol represents the response of B cells from an individual control or patient; the horizontal bars represent means. ns, no significant; *, $P < 0.05$; **, $P < 0.01$.

B cells proliferated to a similar extent as normal memory B cells in response to IL-21 (Fig. 3 H). Not surprisingly, CD40L/CpG induced comparable proliferation in STAT3-sufficient and -deficient naive and memory B cells (Fig. 3, G and H). Thus, impaired proliferation of naive $STAT3_{MUT}$ B cells to IL-21 correlates with poor differentiation of these cells to plasmablasts under these culture conditions. However, naive $STAT3_{MUT}$ B cells do undergo some proliferation to IL-21 (Fig. 3 G; Avery et al., 2010), indicating that the block in differentiation is not simply caused by *STAT3* mutations abrogating cell division.

The memory cell pool in STAT3-deficient individuals contains Ag-specific B cells

Although the total levels of Ig produced by $STAT3_{MUT}$ memory B cells in response to STAT3 cytokines were normal, it was unknown whether these B cells could contribute to an Ag-specific Ab response. To address this, we quantified the relative amounts of antitetanus IgG produced by B cells from normal donors or $STAT3_{MUT}$ or $STAT1_{MUT}$ patients after in vitro culture with CD40L/IL-21. As expected, the levels of antitetanus IgG produced by normal naive B cells were very low/undetectable, whereas memory cells from most normal donors produced significantly higher amounts of tetanus-specific IgG (Fig. 3 I). Importantly, $STAT3_{MUT}$ memory B cells from all patients tested produced significantly higher amounts of tetanus-specific IgG than normal naive B cells (Fig. 3 I). Memory B cells from some $STAT1_{MUT}$ patients exhibited a lower response than others, but in general this exceeded that of normal naive B cells and, on average, approximated that of normal and $STAT3_{MUT}$ memory B cells (Fig. 3 I). We also assessed production of antitetanus IgG in cultures of total PBMCs from normal donors and $STAT3_{MUT}$ individuals that had been stimulated with CD40L and IL-21. On average, normal PBMCs produced approximately threefold higher levels of antitetanus IgG than did $STAT3_{MUT}$ PBMCs (Fig. 3 J). However, there are several caveats to screening PBMCs, rather than purified B cells, for the production of Ag-specific Ab.

Table 2. Ig secretion by IL-21–stimulated normal, $STAT3_{MUT}$, and $STAT1_{MUT}$ naive and memory B cells

Cell type and culture	Ig secretion								
	IgM			IgG			IgA		
	Normal	$STAT3_{MUT}$	$STAT1_{MUT}$	Normal	$STAT3_{MUT}$	$STAT1_{MUT}$	Normal	$STAT3_{MUT}$	$STAT1_{MUT}$
	ng/ml	ng/ml	ng/ml	ng/ml	ng/ml	ng/ml	ng/ml	ng/ml	ng/ml
Naive B cells									
CD40L	2.0 ± 1.0	5.0 ± 2.4[a]	<0. 1[a]	<1	<1[a]	<1[a]	<1	<1[a]	<1[a]
+IL-21	14,640 ± 2,021	444 ± 177[d]	10,801 ± 3,183[a]	267 ± 80	7.3 ± 2.7[b]	86 ± 40[a]	365 ± 106	4.6 ± 2.6[b]	174 ± 40[a]
Memory B cells									
CD40L	166 ± 51	100 ± 25[a]	2.5 ± 2.5[a]	6.1 ± 1.4	11.8 ± 4.7[a]	0.2 ± 0.14[a]	21.3 ± 6.1	26.2 ± 10.6[a]	<1[a]
+IL-21	29,389 ± 4,107	15,214 ± 1,905[b]	7,396 ± 3,097[d]	4,202 ± 1,122	5,651 ± 1,743[a]	953 ± 195[a]	25,208 ± 4,134	13,330 ± 4,027[a]	4,265 ± 989[c]

Naive and memory B cells were sort-purified from normal healthy donors ($n = 17$), $STAT3_{MUT}$ patients ($n = 8$), or $STAT1_{MUT}$ patients ($n = 6$) and then cultured with CD40L alone or together with IL-21. The levels of secreted IgM, IgG, and IgA were determined by ELISA after 10–12 d. The values represent the mean ± SEM from the indicated number of donors/patients and correspond to the data depicted in Fig. 3 (A and B). Statistical analyses were performed using one-way ANOVA; differences are indicated for normal donors compared with $STAT3_{MUT}$ or $STAT1_{MUT}$ B cells.

[a]Not significant.
[b]$P < 0.01$.
[c]$P < 0.001$.
[d]$P < 0.0001$.

First, there is substantial variability in the frequencies of B cells within the population of all PBMCs, as well as in the proportion of B cells that are memory cells. Indeed, there is a strong correlation between the frequency of memory B cells and production of antitetanus IgG in vitro (not depicted). Second, the addition of CD40L to cultures of PBMCs will activate myeloid cells (monocytes, macrophages, and DCs) to secrete molecules such as BAFF and APRIL (Litinskiy et al., 2002; Craxton et al., 2003), which can preferentially promote differentiation and Ig secretion by human memory B cells presumably in a STAT3-independent manner (Avery et al., 2003). Even taking these into account, it is clear that STAT3-deficient individuals are capable of generating Ag-specific Ab responses and that these Ag-specific cells reside within the residual subset of $CD27^+$ B cells. This further substantiates that these cells are indeed memory B cells and are likely to contribute to Ag-specific Ab responses in vivo.

STAT1 deficiency does not affect the early differentiation of memory B cells into Ig-secreting plasmablasts

The accumulated levels of Ig secreted by $STAT1_{MUT}$ memory B cells were 2.5–5-fold less than those by normal memory cells (Fig. 3 B and Table 2). To determine whether this reflected a quantitative defect in generating Ab-secreting cells from $STAT1_{MUT}$ memory B cells, we performed kinetic analyses of plasmablast formation and Ig secretion by normal and $STAT1_{MUT}$ naive and memory B cells that had been stimulated with CD40L alone or together with IL-21. CD40L alone resulted in <0.5% of normal and $STAT1_{MUT}$ naive B cells and ~1–3% of memory B cells acquiring a $CD38^{hi}CD27^{hi}$ phenotype, which corresponds to plasmablasts (Fig. 4, A, B, D, and E; Avery et al., 2005). Addition of IL-21 had minimal effect on naive B cell differentiation, in terms of the frequency of plasmablasts and Ig secretion, after 4 d of culture (Fig. 4, A–C); however, a substantial proportion of memory B cells had differentiated to become Ig-secreting plasmablasts at this time (Fig. 4, D–F). The rate of plasmablast formation from naive and memory B cells increased after 5.5 d and tended to plateau or decline at later times (7 d; Fig. 4, B and E). Co-incident with this was a dramatic increase in Ig secretion by both naive and memory B cells between 4 and 5.5 d of in vitro culture (Fig. 4, C and F). The rate of formation of plasmablasts and Ig secretion by IL-21–stimulated naive and memory B cells was not affected by *STAT1* mutations (Fig. 4). Collectively, these results suggested that $STAT1_{MUT}$ memory B cells could initially generate normal numbers of functional plasmablasts. However, in contrast to this normal rate of differentiation of $STAT1_{MUT}$ memory B cells between days 4 and 7 of culture (Fig. 4, D–F), Ig secretion by these cells after 11 d of culture was consistently less than that by normal memory B cells (Fig. 3 B and Table 2). Thus, STAT1 may play a role in sustaining Ig secretion by differentiated memory B cells.

Commitment of memory B cells to the PC lineage is unaffected by mutations in *STAT1* or *STAT3*

The differentiation of human and mouse B cells into PCs is regulated by the coordinated actions of several transcription factors. PAX5 is down-regulated in activated B cells, thereby relieving PAX5-mediated repression of Blimp-1, resulting in Blimp-1 expression. Although Blimp-1 is not required for initial commitment to the PC lineage, it is indispensable for the generation of terminally differentiated PCs. Other transcription factors, XBP-1 and IRF4 (interferon-induced regulatory

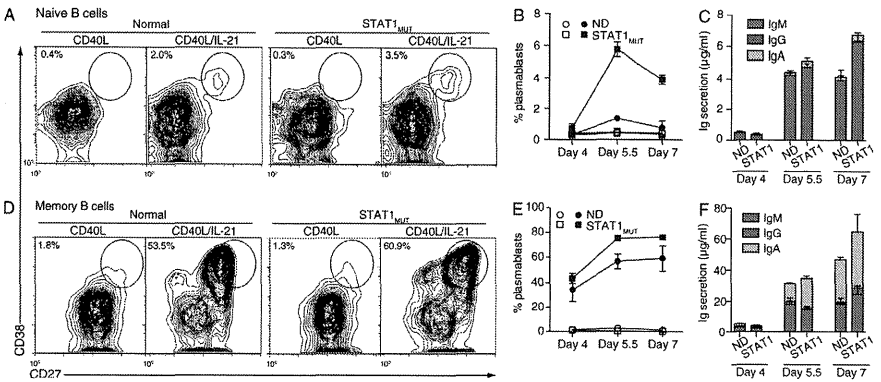

Figure 4. *STAT1* mutations do not affect the generation of plasmablasts but impair sustained Ig secretion. (A–F) Naive (A–C) and memory (D–F) B cells were sort-purified from normal donors or $STAT1_{MUT}$ patients and then cultured (25×10^3/well/100 µl) with CD40L alone (open symbols in B and E) or together with IL-21 (closed symbols in B and E). The generation of plasmablasts, defined as cells acquiring a $CD38^{hi}CD27^{hi}$ phenotype (A, B, D, and E), as well as secretion of IgM, IgG, and IgA (C and F), was determined after 4, 5.5, and 7 d. The contour plots (A and D) are representative of plasmablasts detected after 5 d of culture. The graphs depicting Ig secretion are from cultures of CD40L/IL-21–stimulated B cells. The values represent the mean (±SEM for B, C, E, and F) of experiments using cells from two normal donors and two $STAT1_{MUT}$ patients. Similar results were obtained in a second independent experiment.

factor-4), are also involved in PC differentiation (Nutt et al., 2011). A key mechanism by which IL-21 mediates differentiation of naive B cells into PCs is by modulating expression of these transcription factors. Thus, naive B cells lose PAX5 and acquire Blimp-1 and XBP-1 in response to IL-21 in vitro (Ettinger et al., 2005; Bryant et al., 2007).

Our finding that $STAT3_{MUT}$ memory, but not naive, B cells were capable of secreting near-normal levels of Ig in response to IL-21 (Fig. 3 and Table 2) led us to investigate transcriptional changes in normal and $STAT3_{MUT}$ naive and memory B cells after stimulation with IL-21. We also examined $STAT1_{MUT}$ naive and memory B cells as the latter had some defects in secreting normal levels of Ig (Fig. 3 and Table 2). Addition of IL-21 to cultures of B cells from normal donors resulted in the down-regulation of *PAX5* and up-regulation of *PRDM1* and *XBP1* in naive and memory B cells (Fig. 5, A and B). *PRDM1* and *XBP1* were substantially higher, and PAX5 much lower, in memory versus naive cells (Fig. 5, A and B). Furthermore, although IL-21 had no detectable effect on *IRF4* expression in normal naive B cells (Fig. 5 A), it induced an approximately threefold increase in *IRF4* in normal memory B cells (Fig. 5 B). These differences likely contribute to memory B cells secreting 10–20-fold more Ig than naive cells (Figs. 3 and 4 and Table 2).

$STAT1_{MUT}$ naive and memory B cells modulated expression of *PAX5*, *PRDM1*, *XBP1*, and *IRF4* in a manner indistinguishable from normal B cells (Fig. 5, A and B), consistent with normal Ig secretion during short-term cultures (Fig. 4, C and F). However, $STAT3_{MUT}$ B cells revealed marked differences in the behavior of naive and memory cells. Although naive $STAT3_{MUT}$ B cells down-regulated *PAX5* in response to IL-21, they failed to up-regulate *PRDM1* and *XBP1* (Fig. 5 A). In stark contrast, IL-21–mediated induction of *PRDM1*, *XBP1*, and *IRF4* in $STAT3_{MUT}$ memory B cells was intact (Fig. 5 B), mirroring the ability of these cells to secrete large amounts of Ig in response to IL-21 (Fig. 3 B). Induction of *AICDA* in $STAT3_{MUT}$ naive B cells by IL-21 was comparable with normal and $STAT1_{MUT}$ naive B cells (Fig. 5 A), further demonstrating that $STAT3_{MUT}$ naive B cells can respond to IL-21 under the culture conditions used here. In contrast to naive B cells, IL-21 reduced *AICDA* expression in memory B cells from all individuals compared with stimulation with CD40L alone (Fig. 5 B). This is probably a result of memory B cells expressing much higher levels of *PRDM1* (Fig. 5 B), which directly represses *AICDA* (Nutt et al., 2011). Thus, differentiation of naive and memory B cells into Ab-secreting cells, as determined both at the cellular and molecular level, exhibit distinct sensitivity to mutations in *STAT3*.

Memory cells exhibit greater sensitivity to the stimulatory effects of STAT3-activating cytokines IL-21 and IL-10

One possible explanation for this differential susceptibility to mutations in *STAT3* would be that in memory B cells IL-21 activates an alternate STAT3-independent signaling pathway. However, we observed little activation of pathways other than STAT1 and STAT3 in memory B cells (Fig. 2). An alternative possibility was that memory B cells expressed higher levels of STAT3 or were enriched for expression of the wild-type

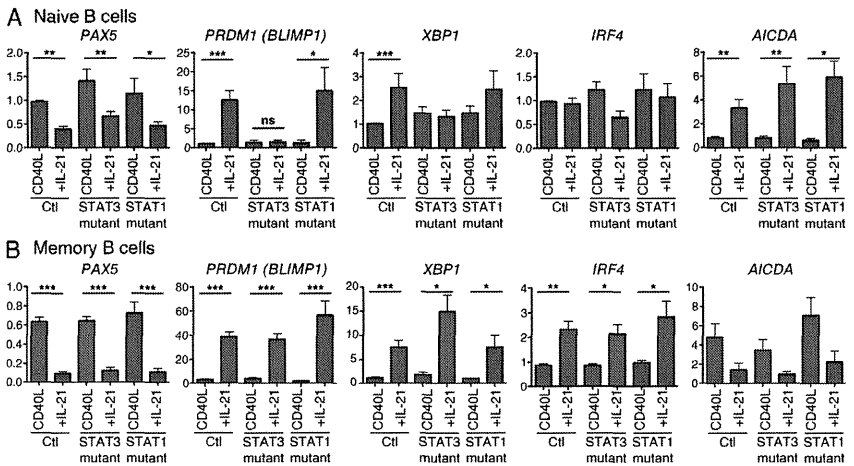

Figure 5. Induction of the PC transcriptional program is intact in IL-21–stimulated $STAT3_{MUT}$ memory B cells. (A and B) Naive ($CD20^+CD10^-CD27^-IgG^-$; A) and memory ($CD20^+CD10^-CD27^+$; B) B cells were sort-purified from normal donor controls (Ctl; $n = 10$ [or 7 for *AICDA*]), $STAT3_{MUT}$ patients ($n = 5$), or $STAT1_{MUT}$ patients ($n = 4$) and then cultured with CD40L alone or together with IL-21 (+IL-21) for 5 d. Expression of *PAX5*, *PRDM1*, *XBP1*, *IRF4*, and *AICDA* was determined by qPCR. The columns represent the mean ± SEM of experiments performed using naive B cells from 7–10 normal donors, 5 $STAT3_{MUT}$ patients, or 4 $STAT1_{MUT}$ patients. Levels of expression are relative to the amount of *GAPDH*. ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

STAT3 allele compared with naive B cells. Yet we found that naive and memory B cells from normal donors or AD-HIES patients expressed comparable levels of STAT3 both ex vivo and after in vitro culture (not depicted) and that the mutant allele still accounted for ∼50% of total STAT3 that was expressed by memory B cells (not depicted). A final explanation for the differential effects of STAT3 deficiency on the function of naive versus memory B cells derives from the heterozygous nature of the *STAT3* mutations and the fact that STATs exert their effect as dimers. This predicts that the mutant allele acts in a dominant-negative manner, thereby inhibiting the function of up to 75% of STAT3 dimers and leaving only 25% intact (Holland et al., 2007; Minegishi et al., 2007). Thus, the differential sensitivity of naive and memory B cells to *STAT3* mutations may reflect an increased responsiveness of memory cells to STAT3 action such that the residual wild-type STAT3 dimers in $STAT3_{MUT}$ memory, but not naive, B cells are sufficient to render these cells responsive to the effects of STAT3-activating cytokines. To test whether there are differences in the threshold of activation of naive and memory B cells, these B cell subsets were purified from normal donor spleens and cultured with CD40L and increasing concentrations of IL-21 or IL-10, and induction of expression of *PRDM1* and *XBP1* was determined after 4 d. The concentrations of IL-21 (Fig. 6 A, left) and IL-10 (Fig. 6 B, left) required to induce *PRDM1* in naive B cells (i.e., 10 ng/ml IL-21 and 100 U/ml IL-10) were at least five times higher than those required for induction in memory B cells (i.e., 2 ng/ml IL-21 and 20 U/ml IL-10). Furthermore, these cytokines induced *PRDM1* in memory B cells at levels that exceeded those in naive B cells by two- to five-fold (Fig. 6, A and B, left). Induction of *XBP1* followed a similar pattern, with expression being detected in memory B cells at much reduced cytokine concentrations than in naive B cells and memory B cells expressing substantially more *XBP1* than naive B cells (Fig. 6, A and B, right). Collectively, these results demonstrate that memory B cells have greater sensitivity to the stimulatory effects of these STAT3-activating cytokines, especially when present at limiting concentrations. Thus, it is likely that the small percentage of wild-type STAT3 dimers that can form in $STAT3_{MUT}$ memory B cells are sufficient to integrate signals provided by IL-10 and IL-21 to facilitate the differentiation of memory B cells into Ig-secreting plasmablasts.

Loss-of-function mutations in *IL21R* abolishes responses of naive and memory B cells to IL-21

To establish that the differences in responses of naive and memory $STAT3_{MUT}$ B cells to IL-21 involved direct signaling through the IL-21R, rather than interactions between IL-21 and a putative alternate receptor that may be expressed only on memory B cells and functions independently of

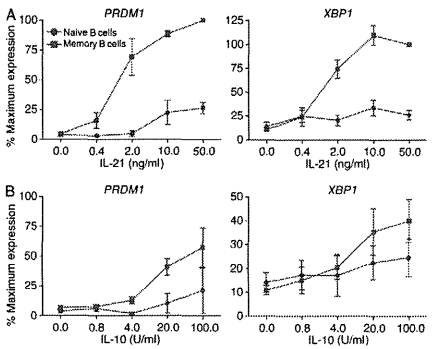

Figure 6. Memory B cells exhibit greater sensitivity to the differentiation-inducing effects of STAT3-activating cytokines IL-21 and IL-10. (A and B) Naive and memory B cells were sort-purified from normal donor spleens and then cultured with CD40L alone or with increasing concentrations of IL-21 (A) or IL-10 (B). After 4 d, expression of *PRDM1* and *XBP1* was determined by qPCR. The values are presented as the percentage of the maximum response, defined as the levels of expression induced in memory B cells by the highest dose of IL-21 tested (50 ng/ml). The data represent the mean ± SEM of experiments using naive and memory B cells from three different normal donor spleens.

STAT3, we examined the B cell compartment in recently identified individuals with loss-of-function mutations in *IL21R* (Kotlarz et al., 2013). Phenotypic analysis of three individuals revealed a marked deficiency in memory B cells that was comparable with that observed for $STAT3_{MUT}$ patients (Fig. 7 A and Table 1). Furthermore, in contrast to healthy donors, ∼95% of $IL\text{-}21R_{MUT}$ B cells were IgM^+IgD^+, revealing a deficiency in Ig isotype–switched cells (Fig. 7 B). We also quantified the proportion of memory B cells that expressed specific Ig isotypes and found that although ∼50% of memory B cells from normal donors had lost expression of IgD and IgM, >90% of $IL\text{-}21R_{MUT}$ memory B cells remained IgD^+ (Fig. 7 C and not depicted). Consistent with this, <5% of $IL\text{-}21R_{MUT}$ memory B cells had undergone switching to express IgG or IgA, whereas the memory B cell pool of normal donors is comprised of ∼25% IgG^+ and ∼20% IgA^+ cells (Fig. 7 C). Thus, IL-21 signaling is indispensable for the generation of not only a normal pool of memory B cells but also the generation of isotype-switched effector B cells within the memory cell subset.

When naive B cells were isolated from normal donors and $IL\text{-}21R_{MUT}$ individuals and cultured in vitro, only B cells from normal donors responded to IL-21, as revealed by the generation of $CD38^{hi}CD27^{hi}$ plasmablasts (Fig. 7 D), secretion of high levels of IgM, IgG, and IgA (Fig. 7 E), and down-regulation of *PAX5* while concomitantly acquiring *AICDA*, *PRDM1*, and *XBP1* expression (Fig. 7 F). All of these readouts of naive B cell differentiation were abolished by loss-of-function mutations in *IL21R* (Fig. 7, D–F). Not surprisingly, CD40L-stimulated $IL\text{-}21R_{MUT}$ memory B cells also failed to respond to the stimulatory effects of IL-21 (Fig. 7 E). $IL\text{-}21R_{MUT}$ naive B cells, though, are intrinsically functional, as indicated by intact responses to IL-4 and IL-10 with respect to induction of *AICDA* (IL-4 and IL-10) and *XBP1* (IL-4) and reduction in *PAX5* (IL-10; Fig. 7 F). These findings demonstrate the specificity of the IL-21 used in our experiments and reveal that both naive and memory B cells absolutely require a functional IL-21R for their response to IL-21.

DISCUSSION

Naive and memory B cells play distinct roles during humoral immune responses. Thus, naive B cells activated after primary encounter with foreign Ag initially produce Ag-specific IgM and eventually yield B cells that produce IgG or IgA. In contrast, memory cells respond much more efficiently upon subsequent exposure to such Ags, rapidly differentiating into Ab-secreting cells to produce substantially higher levels of protective Ig than naive cells (Ahmed and Gray, 1996; Rajewsky, 1996; Tangye and Tarlinton, 2009). This increased efficacy of memory B cell activation is one mechanism underlying long-term protective immunological memory.

B cell differentiation into plasmablasts is regulated by the integration of signals provided by Ag, T cell help (CD40L), and cytokines. Signal transduction pathways activated by these ligands converge to activate key transcriptional regulators, such as Blimp-1, that mediate the commitment of activated B cells to a PC fate (Nutt et al., 2011). Cytokines important for human B cell differentiation include IL-10 and IL-21, which induce isotype switching, PC generation, and Ab secretion from activated naive and memory B cells (Banchereau et al., 1994; Arpin et al., 1997; Pène et al., 2004; Ettinger et al., 2005; Bryant et al., 2007; Avery et al., 2008a). A common feature of these cytokines is their ability to activate similar signaling intermediates, such as STAT1 and STAT3. Remarkably, the effects of IL-10 or IL-21 on human naive B cells are abolished by heterozygous *STAT3* mutations, yet are unaffected by mutations in *STAT1* (Figs. 3–5; Avery et al., 2010). These defects likely explain defective Ab responses and reduced numbers of memory B cells in AD-HIES (Leung et al., 1988; Sheerin and Buckley, 1991; Avery et al., 2010) and, conversely, intact humoral responses in STAT1-deficient patients (Boisson-Dupuis et al., 2012). Interestingly, mutations in *IL21R* recapitulated the impaired Ab responses to vaccines (Kotlarz et al., 2013) and memory B cell deficit (Fig. 7) observed in AD-HIES, despite the ability of IL-21R–deficient B cells to respond to other growth and differentiation-inducing cytokines such as IL-4 or IL-10. In contrast, B cell responses in vivo appear intact in IL-10/IL-10R–deficient individuals (Kotlarz et al., 2012). Thus, STAT3, downstream of IL-21R, clearly plays a central role in establishing long-lived Ab-mediated immunity.

Our study revealed that *STAT3* mutations do not affect memory B cell function, as $STAT3_{MUT}$ memory cells underwent the molecular and cellular changes required for plasmablast

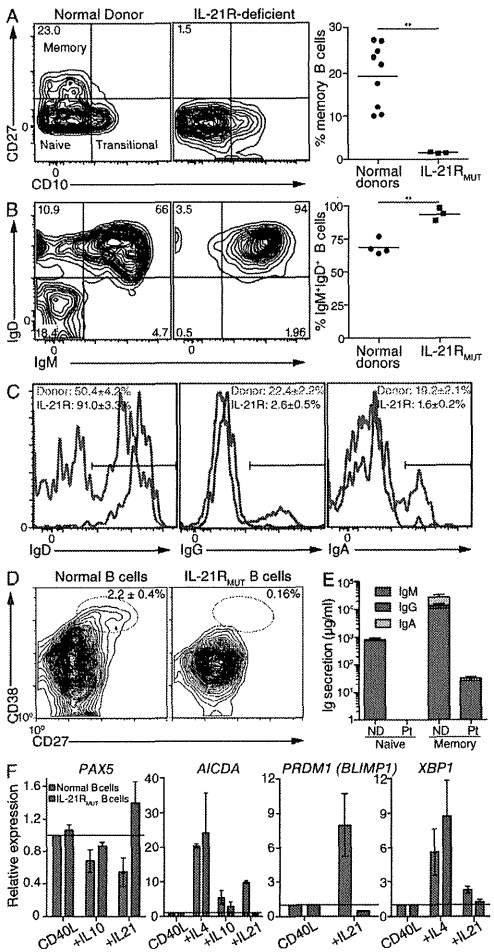

Figure 7. Loss-of-function mutation in *IL21R* abolishes B cell responses to IL-21. (A and B) PBMCs from age-matched normal donors and three patients with loss-of-function mutations in *IL21R* were labeled with mAb against CD20, CD10, CD27, IgM, IgD, IgG, and IgA. (A) Memory B cells were quantified based on the frequency of $CD20^+$ B cells that were $CD10^-CD27^+$. (B and C) The percentages of total B cells in normal donors or IL-21R–deficient patients that coexpressed IgM and IgD (B) and of memory (i.e., $CD27^+$) B cells from normal donors and IL-21R$_{MUT}$ patients that expressed IgD, IgG, or IgA (C) were determined. **, $P < 0.01$. Each symbol in A and B represents an individual normal donor or IL-21R$_{MUT}$ patient; the horizontal bars represent means. The values in C represent the mean percentage ± SEM of memory B cells from four normal donors or three IL-21R$_{MUT}$ patients that express IgD, IgG, or IgA. (D–F) Naive or memory B cells sort-purified from normal donors or IL-21R1$_{MUT}$ patients were cultured with CD40L alone or CD40L/IL-21 (D and E) or CD40L/IL-4, CD40L/IL-10, or CD40L/IL-21 (F). After 5 d, the percentage of plasmablasts (i.e., $CD38^{hi}CD27^{hi}$) generated (D) and expression of *PAX5*, *AICDA*, *PRDM1*, and *XBP1* by cultured naive B cells (F) were determined by flow cytometry or qPCR. The values in D represent the mean (±SEM) percentage of naive B cells that acquired a plasmablast phenotype in response to CD40L/IL-21 in experiments using naive B cells from three different normal donors or one IL-21R$_{MUT}$ patient. In the absence of IL-21 <0.5% plasmablasts were detected in these cultures. (E) The levels of secreted IgM, IgG, and IgA by naive and memory B cells from normal donors (ND) or IL-21R$_{MUT}$ patients (Pt) in response to stimulation with CD40L/IL-21 were determined by ELISA after 10 d. The amounts of Ig secreted by IL21R$_{MUT}$ memory B cells in response to CD40L/IL-21 did not differ from those induced by CD40L alone (not depicted). These data are from one of three experiments using naive B cells from five different healthy controls or two unrelated IL-21R$_{MUT}$ patients. The values in F represent the mean ± SEM from experiments using two to five different healthy controls or of three experiments using cells from two unrelated IL-21R$_{MUT}$ patients. The horizontal black lines on the graphs in F indicate a value of 1.0, which corresponds to the relative level of expression of the indicated gene in CD40L-stimulated normal naive B cells.

differentiation in response to IL-10 or IL-21. Importantly, although memory cells are numerically deficient in AD-HIES patients, they produced normal levels of Ag-specific IgG in vitro on a per cell basis, inferring that STAT3$_{MUT}$ memory B cells would be functional in vivo. This explains the variability in impairment in humoral immunity in AD-HIES patients (Sheerin and Buckley, 1991; Avery et al., 2010). There are several explanations for the normal response of STAT3$_{MUT}$ memory B cells to STAT3 cytokines. First, these cytokines may activate signaling pathways in memory B cells distinct from naive cells, thereby allowing memory cells to respond independently of STAT3, unlike naive B cells. This is unlikely as we detected comparable activation of STATs, yet little activation of Erk and Akt in IL-21-stimulated naive and memory B cells. Second, STAT3 may be differentially expressed by naive and memory B cells; however, this was also found to not be the case. Third, because STAT1$_{MUT}$ B cells secreted less Ig in vitro over time, STAT1 may contribute to the function of activated memory B cells. Even if correct, the reduction in Ig secretion by STAT1$_{MUT}$ memory B cells is probably not physiologically significant as these patients have intact humoral immunity (Boisson-Dupuis et al., 2012). This may reflect the intact early differentiation of STAT1$_{MUT}$ naive and memory B cells in response to IL-10 and IL-21 in vitro, the generation of normal numbers of memory B cells in vivo, and the ability of STAT1$_{MUT}$ memory B cells to produce sufficient quantities of specific Abs after reexposure to immunizing Ags or infectious pathogens. These findings suggest that STAT1 plays only a minor, if any, role in inducing and maintaining humoral immunity. A final possibility is that memory B cells require less activated STAT3 to respond to specific cytokines than do naive B cells. Thus, the residual amounts of functional STAT3 in STAT3$_{MUT}$ B cells are sufficient to mediate plasmablast differentiation induced by IL-10 and IL-21 in memory, but not naive, B cells. This is supported by our finding that IL-10 and IL-21 induced expression of the key PC transcription factors Blimp-1 and XBP-1 in memory B cells at concentrations that had no effect on gene expression in corresponding naive B cells (Fig. 6). This, therefore, is our favored model, which is also consistent with memory B cells having a lower threshold for activation than naive B cells (Yefenof et al., 1986; Poudrier and Owens, 1994), which underlies their rapid response on subsequent encounters with specific Ag (Ahmed and Gray, 1996; Rajewsky, 1996; Tangye and Tarlinton, 2009; Goodnow et al., 2010).

These findings also provide important insights into the hierarchy by which cytokines operate to induce B cell differentiation. Thus, although IL-21R–deficient naive B cells could respond to IL-4 and IL-10 in vitro to induce key events required for Ig isotype switching (*AICDA* expression) and PC generation (*PRDM1* and *XBP1*), responses to these and other cytokines such as BAFF and APRIL (Banchereau et al., 1994; Litinskiy et al., 2002; Avery et al., 2003; Craxton et al., 2003) are insufficient in vivo to compensate for a complete absence of IL-21/IL-21R signaling. This is indicated by B cells in IL-21R–deficient individuals expressing only IgM and IgD, with essentially no isotype-switched cells being detected (Fig. 7). Thus, despite IL-4 and IL-10 inducing *AICDA* expression in IL-21R–deficient B cells, a primary signal via the IL-21R appears to be the critical and rate-limiting step for B cells to undergo isotype switching, after which cytokines such as IL-4, IL-10, BAFF, and APRIL can cooperate to enhance IL-21–induced switching and B cell differentiation (Litinskiy et al., 2002; Avery et al., 2003, 2008a; Craxton et al., 2003). This is reminiscent of the immunological phenotype of X-linked or JAK3-deficient SCID patients who have undergone stem cell transplant but retain autologous (i.e., *IL2RG* or *JAK3* mutant) B cells, inasmuch that these patients have significant reductions in memory B cells, isotype-switched B cells, and serum IgM and lack serum IgG and IgA (Recher et al., 2011), thereby highlighting the requirement for intact signaling through γc/JAK3 downstream of IL-4R and IL-21R for B cell differentiation and effector function. Interestingly though, populations of IgG^+ and IgA^+ cells were detectable within the memory B cell subset of STAT3-deficient individuals, despite the reduction in total memory B cells in these patients. Because IL-21 could induce CD40L-activated STAT3$_{MUT}$ naive B cells to express *AICDA*, but not *PRDM1*, it is likely that the level of STAT3 required to mediate class switching in naive B cells is significantly less that that required for plasmablast formation. Thus, although IL-21R is indispensable for class switching in vivo, the residual amount of functional STAT3 in STAT3$_{MUT}$ naive B cells is sufficient to mediate IL-21–induced class switching in vivo. These findings demonstrate that within the same cell type (i.e., naive B cells) the thresholds of activation of STAT3 required for different biological processes (i.e., class switching versus plasmablast generation) are distinct, thereby providing a rational explanation for (a) intact class switch recombination but defective plasmablast formation by STAT3-deficient naive B cells to IL-21 and (b) phenotypic differences between memory B cells in patients with mutations in *STAT3* or *IL21R*.

As a key attribute of memory B cells is their ability to respond more rapidly than naive B cells, a question that arises is why STAT3$_{MUT}$ memory B cells do not increase in frequency over time to improve humoral immunity in AD-HIES. Because STAT3$_{MUT}$ memory B cells exhibit normal responses to IL-21 in vitro, this would suggest that availability of, or access to, stimulatory cytokines in vivo is limiting. We have reported that the proportions (Ma et al., 2012; Mazerolles et al., 2013) and absolute numbers (101 ± 7 cells/ml in normal donors vs. 51.6 ± 10 cells/ml peripheral blood in STAT3 deficiency) of circulating $CD4^+CXCR5^+$ T cells, which like Tfh cells present in secondary lymphoid tissues are enriched for IL-21–producing cells (Chevalier et al., 2011), are reduced in AD-HIES patients. Furthermore, STAT3$_{MUT}$ $CD4^+$ T cells are impaired in their ability to generate Tfh-like cells in vitro, thereby compromising IL-21–mediated help for B cell differentiation (Ma et al., 2012). These observations are consistent with a scenario whereby STAT3$_{MUT}$ memory B cells, despite their intact ability to respond to IL-21, are constrained in doing so in vivo because of diminished production of IL-21 by STAT3-deficient $CD4^+$ T cells.

The deficit in memory B cells observed in STAT3$_{MUT}$ patients is comparable with that in other immune-deficient individuals, such as patients with mutations in *SH2D1A* (XLP; Ma et al., 2005, 2006), *CD40LG* (hyper-IgM syndrome; Notarangelo et al., 2006) or *ICOS* (common variable immunodeficiency; Warnatz et al., 2006), or transplanted X-linked/JAK3-deficient SCID patients who retain autologous B cells (Recher et al., 2011). Although all of these latter conditions are characterized by reductions in serum Ig levels (Ma et al., 2005; Notarangelo et al., 2006; Warnatz et al., 2006; Recher et al., 2011), serum levels of IgM, IgG, and IgA are normal in STAT3-deficient patients despite a generalized impairment in the ability to elicit sustained Ag-specific Ab

Downloaded from jem.rupress.org on February 16, 2014

responses (Leung et al., 1988; Sheerin and Buckley, 1991; Avery et al., 2010). Our results provide a potential explanation for the apparent discrepancy between reduced memory B cells but normal serum Ig levels in STAT3 deficiency, inasmuch that the residual memory B cells can respond to B cell differentiating cytokines and thus contribute to the pool of serum Ig. The corollary of this is that the small population of memory B cells in XLP, hyper-IgM syndrome, ICOS deficiency, and posttransplant SCID are unable to access appropriate $CD4^+$ T cell–derived signals (i.e., SAP-dependent interactions; CD40L; ICOS; IL-2/IL-4/IL-21) and thus are limited in their ability to contribute to humoral immunity. An extension of our findings is that because we could detect functional Ag-specific $STAT\text{-}3_{MUT}$ memory B cells, directed targeting of these cells with IL-21–mediated signals may improve humoral immunity in AD-HIES patients.

MATERIALS AND METHODS

Human blood and tissue samples. Buffy coats and spleens from healthy donors were provided by the Australian Red Cross Blood Service. Peripheral blood was also collected from patients with loss-of-function mutations in *STAT1* ($STAT1_{MUT}$; Dupuis et al., 2001; Chapgier et al., 2006, 2009; Sampaio et al., 2012; Hirata et al., 2013; Ives et al., 2013), *STAT3* ($STAT\text{-}3_{MUT}$; Ma et al., 2008, 2012; Avery et al., 2010) or *IL21R* ($IL\text{-}21R_{MUT}$; Kotlarz et al., 2013; Ives et al., 2013; Table S1). Approval for this study was obtained from the human research ethics committees of the St. Vincent's Hospital and Sydney South West Area Health Service (Australia), the Rockefeller University Institutional Review Board (New York), and the National Institute of Allergy and Infectious Diseases Intramural Institutional Review Board (Bethesda, MD).

Lymphocyte phenotyping and isolation. PBMCs were incubated with mAb to CD20 and CD27 and an isotype control or mAb specific for CD23, CD24, CD80, CD86, CD95, TACI, IgM, IgD, IgG, IgG1, IgG2, IgG3, IgG4, and IgA, and expression of these molecules on $CD20^+CD27^-$ (naive) and $CD20^+CD27^+$ (memory) B cells was determined by flow cytometry (Ma et al., 2006). Naive and memory B cells were purified by labeling either PBMCs or total B cells with mAb against CD10, CD20, and CD27 and sorting $CD10^-CD20^+CD27^-$ (naive) and $CD10^-CD20^+CD27^+$ (memory) cells (FACSAria; BD; (Avery et al., 2008b, 2010). Splenic B cells were labeled with mAbs against CD20, CD27, IgG, and IgA, and subsets of either naive ($CD20^+CD27^-$) and total memory B cells ($CD20^+CD27^+$) or IgM memory ($CD20^+CD27^-IgG^-IgA^-$) and isotype-switched ($CD20^+CD27^-IgG^+IgA^+$) memory B cells were collected (FACSAria; Tangye et al., 2003a; Bryant et al., 2007; Good et al., 2009). The purity of the recovered populations was typically >98%. To enumerate circulating Tfh-like cells, PBMCs were labeled with mAbs against CD3, CD4, and CXCR5, and the absolute number of $CD4^+CXCR5^+$ T cells was then determined.

In vitro activation of naive and memory B cells. Naive and memory B cells isolated from normal donors or $STAT3_{MUT}$, $STAT1_{MUT}$, or $IL\text{-}21R_{MUT}$ patients were cultured (~$5\text{–}10 \times 10^3$/200 µl/well for proliferation, Ig secretion, and qPCR and ~4×10^4/400 µl/well for phenotyping; BD) with CD40L alone or together with 100 U/ml IL-4, 100 U/ml IL-10, 50 ng/ml IL-21 (PeproTech), or 1 µg/ml CpG 2006 (Sigma-Aldrich). Expression of *STAT3*, *PAX5*, *PRDM1*, *XBP1*, *IRF4*, and *AICDA* was determined after 5 d by real-time PCR and standardized to *GAPDH* (Avery et al., 2010). Expression of Ig ε germline transcripts were determined by PCR as described previously (Avery et al., 2008b). Differentiation of B cells to plasmablasts was assessed by determining the frequency of cells acquiring a $CD\text{-}38^{hi}CD27^{hi}$ phenotype during in vitro culture (Avery et al., 2005). B cell proliferation was determined by assessing the incorporation of [^{3}H]thymidine (1 µCi/ml per well; ICN Biomedicals) during the final 18 h of a 5-d culture (Good et al., 2006). Ig secretion was determined by ELISA after 4–12 d of culture (Bryant et al., 2007). Relative levels of antitetanus IgG in culture supernatants were determined by ELISA using plates precoated with tetanus toxoid (Sigma-Aldrich) and then detecting bound IgG (Avery et al., 2010).

Analysis of intracellular signaling. Naive and memory splenic B cells were cultured with $F(ab')_2$ fragments of goat anti–human Ig (Jackson ImmunoResearch Laboratories, Inc.) for ~18 h, washed, and then recultured with media alone, 100 ng/ml IL-21, or $F(ab')_2$ anti-Ig for 15–30 min. Cells were then fixed, permeabilized, labeled with anti–phospho-STAT1, STAT3, STAT4, STAT5, and STAT6 mAb (Avery et al., 2008b, 2010) and analyzed by flow cytometry. Alternatively, cells were lysed and Western blotting was performed using rabbit polyclonal anti-STAT3 (C-20), anti-ERK2 (C-14), and anti-14.3.3 (K-19; Santa Cruz Biotechnology, Inc.); and anti-STAT3 pY705 (3E2), anti-AKT pS473 (587F11), and anti-ERK1/2 pT202/Y204 (Cell Signaling Technology).

Statistical analysis. Significant differences between datasets were determined using either the unpaired Student's *t* test when comparing two variables or ANOVA for more than two variables (Prism; GraphPad Software).

Online supplemental material. Table S1 provides details of the patients analyzed in this study. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20130323/DC1.

We thank the Garvan Flow Cytometry facility for cell sorting, Rob Brink, Tri Phan, and Tony Basten for critical review of this manuscript, members of the James laboratory (Garvan Institute) for advice on Western blotting, and the patients and their families for participating in this project. We also acknowledge the late Dr. Klaus Madgorf for his contributions to our research.

This work was funded by project and program grants from the National Health and Medical Research Council (NHMRC) of Australia (to E.K. Deenick, C.S. Ma, D.A. Fulcher, M.C. Cook, and S.G. Tangye) and the Rockefeller University Center for 541 Clinical and Translational science (5UL1RR024143 to J.L. Casanova). C.S. Ma is a recipient of a Career Development Fellowship, L.J. Berglund is a recipient of a Medical Postgraduate Scholarship, and S.G. Tangye is a recipient of a Principal Research Fellowship from the NHMRC of Australia. L. Moens is the recipient of a Postdoctoral Fellowship from the Research Foundation-Flanders (FWO), Belgium.

The authors declare no competing financial interests.

Submitted: 12 February 2013
Accepted: 22 October 2013

REFERENCES

Ahmed, R., and D. Gray. 1996. Immunological memory and protective immunity: understanding their relation. *Science.* 272:54–60. http://dx.doi.org/10.1126/science.272.5258.54

Arpin, C., J. Banchereau, and Y.J. Liu. 1997. Memory B cells are biased towards terminal differentiation: a strategy that may prevent repertoire freezing. *J. Exp. Med.* 186:931–940. http://dx.doi.org/10.1084/jem.186.6.931

Asao, H., C. Okuyama, S. Kumaki, N. Ishii, S. Tsuchiya, D. Foster, and K. Sugamura. 2001. Cutting edge: the common gamma-chain is an indispensable subunit of the IL-21 receptor complex. *J. Immunol.* 167:1–5.

Avery, D.T., S.L. Kalled, J.I. Ellyard, C. Ambrose, S.A. Bixler, M. Thien, R. Brink, F. Mackay, P.D. Hodgkin, and S.G. Tangye. 2003. BAFF selectively enhances the survival of plasmablasts generated from human memory B cells. *J. Clin. Invest.* 112:286–297.

Avery, D.T., J.I. Ellyard, F. Mackay, L.M. Corcoran, P.D. Hodgkin, and S.G. Tangye. 2005. Increased expression of CD27 on activated human memory B cells correlates with their commitment to the plasma cell lineage. *J. Immunol.* 174:4034–4042.

Avery, D.T., V.L. Bryant, C.S. Ma, R. de Waal Malefyt, and S.G. Tangye. 2008a. IL-21-induced isotype switching to IgG and IgA by human naive B cells is differentially regulated by IL-4. *J. Immunol.* 181:1767–1779.

Avery, D.T., C.S. Ma, V.L. Bryant, B. Santner-Nanan, R. Nanan, M. Wong, D.A. Fulcher, M.C. Cook, and S.G. Tangye. 2008b. STAT3 is required for IL-21-induced secretion of IgE from human naive B cells. *Blood.* 112:1784–1793. http://dx.doi.org/10.1182/blood-2008-02-142745

Avery, D.T., E.K. Deenick, C.S. Ma, S. Suryani, N. Simpson, G.Y. Chew, T.D. Chan, U. Palendira, J. Bustamante, S. Boisson-Dupuis, et al. 2010. B cell–intrinsic signaling through IL-21 receptor and STAT3 is required for establishing long-lived antibody responses in humans. *J. Exp. Med.* 207:155–171. http://dx.doi.org/10.1084/jem.20091706

Banchereau, J., F. Brière, Y.J. Liu, and F. Rousset. 1994. Molecular control of B lymphocyte growth and differentiation. *Stem Cells.* 12:278–288. http://dx.doi.org/10.1002/stem.5530120304

Bernasconi, N.L., E. Traggiai, and A. Lanzavecchia. 2002. Maintenance of serological memory by polyclonal activation of human memory B cells. *Science.* 298:2199–2202. http://dx.doi.org/10.1126/science.1076071

Bernasconi, N.L., N. Onai, and A. Lanzavecchia. 2003. A role for Toll-like receptors in acquired immunity: up-regulation of TLR9 by BCR triggering in naive B cells and constitutive expression in memory B cells. *Blood.* 101:4500–4504. http://dx.doi.org/10.1182/blood-2002-11-3569

Boisson-Dupuis, S., X.F. Kong, S. Okada, S. Cypowyj, A. Puel, L. Abel, and J.L. Casanova. 2012. Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. *Curr. Opin. Immunol.* 24:364–378. http://dx.doi.org/10.1016/j.coi.2012.04.011

Bryant, V.L., C.S. Ma, D.T. Avery, Y. Li, K.L. Good, L.M. Corcoran, R. de Waal Malefyt, and S.G. Tangye. 2007. Cytokine-mediated regulation of human B cell differentiation into Ig-secreting cells: predominant role of IL-21 produced by CXCR5+ T follicular helper cells. *J. Immunol.* 179:8180–8190.

Casanova, J.L., S.M. Holland, and L.D. Notarangelo. 2012. Inborn errors of human JAKs and STATs. *Immunity.* 36:515–528. http://dx.doi.org/10.1016/j.immuni.2012.03.016

Chapgier, A., S. Boisson-Dupuis, E. Jouanguy, G. Vogt, J. Feinberg, A. Prochnicka-Chalufour, A. Casrouge, K. Yang, C. Soudais, C. Fieschi, et al. 2006. Novel STAT1 alleles in otherwise healthy patients with mycobacterial disease. *PLoS Genet.* 2:e131. http://dx.doi.org/10.1371/journal.pgen.0020131

Chapgier, A., X.F. Kong, S. Boisson-Dupuis, E. Jouanguy, D. Averbuch, J. Feinberg, S.Y. Zhang, J. Bustamante, G. Vogt, J. Lejeune, et al. 2009. A partial form of recessive STAT1 deficiency in humans. *J. Clin. Invest.* 119:1502–1514. http://dx.doi.org/10.1172/JCI37083

Chevalier, N., D. Jarrossay, E. Ho, D.T. Avery, C.S. Ma, D. Yu, F. Sallusto, S.G. Tangye, and C.R. Mackay. 2011. CXCR5 expressing human central memory CD4 T cells and their relevance for humoral immune responses. *J. Immunol.* 186:5556–5568. http://dx.doi.org/10.4049/jimmunol.1002828

Craxton, A., D. Magaletti, E.J. Ryan, and E.A. Clark. 2003. Macrophage- and dendritic cell—dependent regulation of human B-cell proliferation requires the TNF family ligand BAFF. *Blood.* 101:4464–4471. http://dx.doi.org/10.1182/blood-2002-10-3123

Darce, J.R., B.K. Arendt, X. Wu, and D.F. Jelinek. 2007. Regulated expression of BAFF-binding receptors during human B cell differentiation. *J. Immunol.* 179:7276–7286.

Davey, A.M., and S.K. Pierce. 2012. Intrinsic differences in the initiation of B cell receptor signaling favor responses of human IgG(+) memory B cells over IgM(+) naive B cells. *J. Immunol.* 188:3332–3341. http://dx.doi.org/10.4049/jimmunol.1102322

Deenick, E.K., J. Hasbold, and P.D. Hodgkin. 1999. Switching to IgG3, IgG2b, and IgA is division linked and independent, revealing a stochastic framework for describing differentiation. *J. Immunol.* 163:4707–4714.

Diehl, S.A., H. Schmidlin, M. Nagasawa, S.D. van Haren, M.J. Kwakkenbos, E. Yasuda, T. Beaumont, F.A. Scheeren, and H. Spits. 2008. STAT3-mediated up-regulation of BLIMP1 Is coordinated with BCL6 down-regulation to control human plasma cell differentiation. *J. Immunol.* 180:4805–4815.

Dupuis, S., C. Dargemont, C. Fieschi, N. Thomassin, S. Rosenzweig, J. Harris, S.M. Holland, R.D. Schreiber, and J.L. Casanova. 2001. Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. *Science.* 293:300–303. http://dx.doi.org/10.1126/science.1061154

Ellyard, J.I., D.T. Avery, T.G. Phan, N.J. Hare, P.D. Hodgkin, and S.G. Tangye. 2004. Antigen-selected, immunoglobulin-secreting cells persist in human spleen and bone marrow. *Blood.* 103:3805–3812. http://dx.doi.org/10.1182/blood-2003-09-3109

Engels, N., L.M. König, C. Heemann, J. Lutz, T. Tsubata, S. Griep, V. Schrader, and J. Wienands. 2009. Recruitment of the cytoplasmic adaptor Grb2 to surface IgG and IgE provides antigen receptor-intrinsic costimulation to class-switched B cells. *Nat. Immunol.* 10:1018–1025. http://dx.doi.org/10.1038/ni.1764

Ettinger, R., G.P. Sims, A.M. Fairhurst, R. Robbins, Y.S. da Silva, R. Spolski, W.J. Leonard, and P.E. Lipsky. 2005. IL-21 induces differentiation of human naive and memory B cells into antibody-secreting plasma cells. *J. Immunol.* 175:7867–7879.

Geha, R.S., H.H. Jabara, and S.R. Brodeur. 2003. The regulation of immunoglobulin E class-switch recombination. *Nat. Rev. Immunol.* 3:721–732. http://dx.doi.org/10.1038/nri1181

Good, K.L., and S.G. Tangye. 2007. Decreased expression of Kruppel-like factors in memory B cells induces the rapid response typical of secondary antibody responses. *Proc. Natl. Acad. Sci. USA.* 104:13420–13425. http://dx.doi.org/10.1073/pnas.0703872104

Good, K.L., V.L. Bryant, and S.G. Tangye. 2006. Kinetics of human B cell behavior and amplification of proliferative responses following stimulation with IL-21. *J. Immunol.* 177:5236–5247.

Good, K.L., D.T. Avery, and S.G. Tangye. 2009. Resting human memory B cells are intrinsically programmed for enhanced survival and responsiveness to diverse stimuli compared to naive B cells. *J. Immunol.* 182:890–901.

Goodnow, C.C., C.G. Vinuesa, K.L. Randall, F. Mackay, and R. Brink. 2010. Control systems and decision making for antibody production. *Nat. Immunol.* 11:681–688. http://dx.doi.org/10.1038/ni.1900

Griffin, D.O., N.E. Holodick, and T.L. Rothstein. 2011. Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70-. *J. Exp. Med.* 208:67–80. http://dx.doi.org/10.1084/jem.20101499

Hirata, O., S. Okada, M. Tsumura, R. Kagawa, M. Miki, H. Kawaguchi, K. Nakamura, S. Boisson-Dupuis, J.L. Casanova, Y. Takihara, and M. Kobayashi. 2013. Heterozygosity for the Y701C STAT1 mutation in a multiplex kindred with multifocal osteomyelitis. *Haematologica.* 98:1641–1649. http://dx.doi.org/10.3324/haematol.2013.083741

Hodgkin, P.D., J.H. Lee, and A.B. Lyons. 1996. B cell differentiation and isotype switching is related to division cycle number. *J. Exp. Med.* 184:277–281. http://dx.doi.org/10.1084/jem.184.1.277

Holland, S.M., F.R. DeLeo, H.Z. Elloumi, A.P. Hsu, G. Uzel, N. Brodsky, A.F. Freeman, A. Demidowich, J. Davis, M.L. Turner, et al. 2007. STAT3 mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* 357:1608–1619. http://dx.doi.org/10.1056/NEJMoa073687

Horikawa, K., S.W. Martin, S.L. Pogue, K. Silver, K. Peng, K. Takatsu, and C.C. Goodnow. 2007. Enhancement and suppression of signaling by the conserved tail of IgG memory-type B cell antigen receptors. *J. Exp. Med.* 204:759–769. http://dx.doi.org/10.1084/jem.20061923

Ives, M.L., C.S. Ma, U. Palendira, A. Chan, J. Bustamante, S. Boisson-Dupuis, P.D. Arkwright, D. Engelhard, D. Averbuch, K. Magdorf, et al. 2013. Signal transducer and activator of transcription 3 (STAT3) mutations underlying autosomal dominant hyper-IgE syndrome impair human CD8(+) T-cell memory formation and function. *J. Allergy Clin. Immunol.* 132:400–411: e9. http://dx.doi.org/10.1016/j.jaci.2013.05.029

Jabara, H.H., D.R. McDonald, E. Janssen, M.J. Massaad, N. Ramesh, A. Borzutzky, I. Rauter, H. Benson, L. Schneider, S. Baxi, et al. 2012. DOCK8 functions as an adaptor that links TLR-MyD88 signaling to B cell activation. *Nat. Immunol.* 13:612–620. http://dx.doi.org/10.1038/ni.2305

Kotlarz, D., R. Beier, D. Murugan, J. Diestelhorst, O. Jensen, K. Boztug, D. Pfeifer, H. Kreipe, E.D. Pfister, U. Baumann, et al. 2012. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. *Gastroenterology.* 143:347–355. http://dx.doi.org/10.1053/j.gastro.2012.04.045

Kotlarz, D., N. Ziętara, G. Uzel, T. Weidemann, C.J. Braun, J. Diestelhorst, P.M. Krawitz, P.N. Robinson, J. Hecht, J. Puchałka, et al. 2013. Loss-of-function mutations in the IL-21 receptor gene cause a primary immunodeficiency

Published November 11, 2013

JEM

Downloaded from jem.rupress.org on February 16, 2014

syndrome. *J. Exp. Med.* 210:433–443. http://dx.doi.org/10.1084/jem.20111229

Leung, D.Y., D.M. Ambrosino, R.D. Arbeit, J.L. Newton, and R.S. Geha. 1988. Impaired antibody responses in the hyperimmunoglobulin E syndrome. *J. Allergy Clin. Immunol.* 81:1082–1087. http://dx.doi.org/10.1016/0091-6749(88)90873-1

Litinskiy, M.B., B. Nardelli, D.M. Hilbert, B. He, A. Schaffer, P. Casali, and A. Cerutti. 2002. DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL. *Nat. Immunol.* 3:822–829. http://dx.doi.org/10.1038/ni829

Liu, Y.J., C. Barthélémy, O. de Bouteiller, C. Arpin, I. Durand, and J. Banchereau. 1995. Memory B cells from human tonsils colonize mucosal epithelium and directly present antigen to T cells by rapid up-regulation ofB7-1andB7-2. *Immunity.*2:239–248.http://dx.doi.org/10.1016/1074-7613(95)90048-9

Ma, C.S., N.J. Hare, K.E. Nichols, L. Dupré, G. Andolfi, M.G. Roncarolo, S. Adelstein, P.D. Hodgkin, and S.G. Tangye. 2005. Impaired humoral immunity in X-linked lymphoproliferative disease is associated with defective IL-10 production by CD4+ T cells. *J. Clin. Invest.* 115:1049–1059.

Ma, C.S., S. Pittaluga, D.T. Avery, N.J. Hare, I. Maric, A.D. Klion, K.E. Nichols, and S.G. Tangye. 2006. Selective generation of functional somatically mutated IgM+CD27+, but not Ig isotype-switched, memory B cells in X-linked lymphoproliferative disease. *J. Clin. Invest.* 116:322–333. http://dx.doi.org/10.1172/JCI25720

Ma, C.S., G.Y. Chew, N. Simpson, A. Priyadarshi, M. Wong, B. Grimbacher, D.A. Fulcher, S.G. Tangye, and M.C. Cook. 2008. Deficiency of Th17 cells in hyper IgE syndrome due to mutations in *STAT3*. *J. Exp. Med.* 205:1551–1557. http://dx.doi.org/10.1084/jem.20080218

Ma, C.S., D.T. Avery, A. Chan, M. Batten, J. Bustamante, S. Boisson-Dupuis, P.D. Arkwright, A.Y. Kreins, D. Averbuch, D. Engelhard, et al. 2012. Functional STAT3 deficiency compromises the generation of human T follicular helper cells. *Blood.* 119:3997–4008. http://dx.doi.org/10.1182/blood-2011-11-392985

Macallan, D.C., D.L. Wallace, Y. Zhang, H. Ghattas, B. Asquith, C. de Lara, A. Worth, G. Panayiotakopoulos, G.E. Griffin, D.F. Tough, and P.C. Beverley. 2005. B-cell kinetics in humans: rapid turnover of peripheral blood memory cells. *Blood.* 105:3633–3640. http://dx.doi.org/10.1182/blood-2004-09-3740

Martin, S.W., and C.C. Goodnow. 2002. Burst-enhancing role of the IgG membrane tail as a molecular determinant of memory. *Nat. Immunol.* 3:182–188. http://dx.doi.org/10.1038/ni752

Mazerolles, F., C. Picard, S. Kracker, A. Fischer, and A. Durandy. 2013. Blood CD4+CD45RO+CXCR5+ T cells are decreased but partially functional in signal transducer and activator of transcription 3 deficiency. *J. Allergy Clin. Immunol.* 131:1146–1156: e1–e5. http://dx.doi.org/10.1016/j.jaci.2012.12.1519

Minegishi, Y., M. Saito, S. Tsuchiya, I. Tsuge, H. Takada, T. Hara, N. Kawamura, T. Ariga, S. Pasic, O. Stojkovic, et al. 2007. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature.* 448:1058–1062. http://dx.doi.org/10.1038/nature06096

Nilsson, A., A. de Milito, F. Mowafi, G. Winberg, O. Björk, E.Z. Wolpert, and F. Chiodi. 2005. Expression of CD27-CD70 on early B cell progenitors in the bone marrow: implication for diagnosis and therapy of childhood ALL. *Exp. Hematol.* 33:1500–1507. http://dx.doi.org/10.1016/j.exphem.2005.10.005

Notarangelo, L.D., G. Lanzi, S. Peron, and A. Durandy. 2006. Defects of class-switch recombination. *J. Allergy Clin. Immunol.* 117:855–864. http://dx.doi.org/10.1016/j.jaci.2006.01.043

Nutt, S.L., N. Taubenheim, J. Hasbold, L.M. Corcoran, and P.D. Hodgkin. 2011. The genetic network controlling plasma cell differentiation. *Semin. Immunol.* 23:341–349. http://dx.doi.org/10.1016/j.smim.2011.08.010

Pène, J., J.F. Gauchat, S. Lécart, E. Drouet, P. Guglielmi, V. Boulay, A. Delwail, D. Foster, J.C. Lecron, and H. Yssel. 2004. Cutting edge: IL-21 is a switch factor for the production of IgG1 and IgG3 by human B cells. *J. Immunol.* 172:5154–5157.

Poudrier, J., and T. Owens. 1994. Co-stimulation by anti-immunoglobulin is required for B cell activation by CD40Llow T cells. *Eur. J. Immunol.* 24:2993–2999. http://dx.doi.org/10.1002/eji.1830241211

Rajewsky, K. 1996. Clonal selection and learning in the antibody system. *Nature.* 381:751–758. http://dx.doi.org/10.1038/381751a0

Recher, M., L.J. Berglund, D.T. Avery, M.J. Cowan, A.R. Gennery, J. Smart, J. Peake, M. Wong, S.Y. Pai, S. Baxi, et al. 2011. IL-21 is the primary common γ chain-binding cytokine required for human B-cell differentiation in vivo. *Blood.* 118:6824–6835. http://dx.doi.org/10.1182/blood-2011-06-362533

Sampaio, E.P., H.I. Bax, A.P. Hsu, E. Kristosturyan, J. Pechacek, P. Chandrasekaran, M.L. Paulson, D.L. Dias, C. Spalding, G. Uzel, et al. 2012. A novel STAT1 mutation associated with disseminated mycobacterial disease. *J. Clin. Immunol.* 32:681–689. http://dx.doi.org/10.1007/s10875-012-9659-2

Sheerin, K.A., and R.H. Buckley. 1991. Antibody responses to protein, polysaccharide, and phi X174 antigens in the hyperimmunoglobulinemia E (hyper-IgE) syndrome. *J. Allergy Clin. Immunol.* 87:803–811. http://dx.doi.org/10.1016/0091-6749(91)90126-9

Tangye, S.G. 2011. Staying alive: regulation of plasma cell survival. *Trends Immunol.* 32:595–602. http://dx.doi.org/10.1016/j.it.2011.09.001

Tangye, S.G., and D.M. Tarlinton. 2009. Memory B cells: effectors of long-lived immune responses. *Eur. J. Immunol.* 39:2065–2075. http://dx.doi.org/10.1002/eji.200939531

Tangye, S.G., Y.J. Liu, G. Aversa, J.H. Phillips, and J.E. de Vries. 1998. Identification of functional human splenic memory B cells by expression of CD148 and CD27. *J. Exp. Med.* 188:1691–1703. http://dx.doi.org/10.1084/jem.188.9.1691

Tangye, S.G., D.T. Avery, E.K. Deenick, and P.D. Hodgkin. 2003a. Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. *J. Immunol.* 170:686–694.

Tangye, S.G., D.T. Avery, and P.D. Hodgkin. 2003b. A division-linked mechanism for the rapid generation of Ig-secreting cells from human memory B cells. *J. Immunol.* 170:261–269.

Tangye, S.G., E.K. Deenick, U. Palendira, and C.S. Ma. 2012. T cell-B cell interactions in primary immunodeficiencies. *Ann. N. Y. Acad. Sci.* 1250:1–13. http://dx.doi.org/10.1111/j.1749-6632.2011.06361.x

Warnatz, K., L. Bossaller, U. Salzer, A. Skrabl-Baumgartner, W. Schwinger, M. van der Burg, J.J. van Dongen, M. Orlowska-Volk, R. Knoth, A. Durandy, et al. 2006. Human ICOS deficiency abrogates the germinal center reaction and provides a monogenic model for common variable immunodeficiency. *Blood.* 107:3045–3052. http://dx.doi.org/10.1182/blood-2005-07-2955

Yefenof, E., V.M. Sanders, J.W. Uhr, and E.S. Vitetta. 1986. In vitro activation of murine antigen-specific memory B cells by a T-dependent antigen. *J. Immunol.* 137:85–90.

Zeng, R., R. Spolski, E. Casas, W. Zhu, D.E. Levy, and W.J. Leonard. 2007. The molecular basis of IL-21-mediated proliferation. *Blood.* 109:4135–4142. http://dx.doi.org/10.1182/blood-2006-10-054973

Identification of a novel erythroid-specific enhancer for the *ALAS2* gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia

Kiriko Kaneko,[1,2] Kazumichi Furuyama,[1,6] Tohru Fujiwara,[3] Ryoji Kobayashi,[4] Hiroyuki Ishida,[5] Hideo Harigae,[3] and Shigeki Shibahara[1]

[1]Department of Molecular Biology and Applied Physiology, [2]Department of Endocrinology and Applied Medical Science, [3]Department of Hematology and Rheumatology, Tohoku University Graduate School of Medicine, Sendai, Miyagi; [4]Department of Pediatrics, Sapporo Hokuyu Hospital, Sapporo; [5]Department of Pediatrics, Kyoto Prefectural University of Medicine, Graduate School of Medical Science, Kyoto; and [6]Laboratory of Molecular Biochemistry, Iwate Medical University, Yahaba, Iwate, Japan

ABSTRACT

Erythroid-specific 5-aminolevulinate synthase (ALAS2) is the rate-limiting enzyme for heme biosynthesis in erythroid cells, and a missense mutation of the *ALAS2* gene is associated with congenital sideroblastic anemia. However, the gene responsible for this form of anemia remains unclear in about 40% of patients. Here, we identify a novel erythroid-specific enhancer of 130 base pairs in the first intron of the *ALAS2* gene. The newly identified enhancer contains a *cis*-acting element that is bound by the erythroid-specific transcription factor GATA1, as confirmed by chromatin immunoprecipitation analysis *in vivo* and by electrophoretic mobility shift assay *in vitro*. A promoter activity assay in K562 human erythroleukemia cells revealed that the presence of this 130-base pair region increased the promoter activity of the *ALAS2* gene by 10-15-fold. Importantly, two mutations, each of which disrupts the GATA-binding site in the enhancer, were identified in unrelated male patients with congenital sideroblastic anemia, and the lower expression level of ALAS2 mRNA in bone marrow erythroblasts was confirmed in one of these patients. Moreover, GATA1 failed to bind to each mutant sequence at the GATA-binding site, and each mutation abolished the enhancer function on ALAS2 promoter activity in K562 cells. Thus, a mutation at the GATA-binding site in this enhancer may cause congenital sideroblastic anemia. These results suggest that the newly identified intronic enhancer is essential for the expression of the *ALAS2* gene in erythroid cells. We propose that the 130-base pair enhancer region located in the first intron of the *ALAS2* gene should be examined in patients with congenital sideroblastic anemia in whom the gene responsible is unknown.

Introduction

The *ALAS2* gene encodes for erythroid-specific 5-aminolevulinate synthase (ALAS-E, EC 2.3.1.37), which is the rate-limiting enzyme of the heme biosynthetic pathway in erythroid cells.[1] It has been reported that the human *ALAS2* gene is mapped on the X chromosome,[2] and that a loss-of-function mutation of this gene causes X-linked sideroblastic anemia (XLSA),[3,4] which is the most common genetic form of congenital sideroblastic anemia (CSA). Moreover, a missense mutation of *ALAS2* was identified in a patient with non-familial CSA (nfCSA),[5] in which no family history of sideroblastic anemia was identified. In addition to *ALAS2*, several other genes were recently identified as causative genes for CSA, including *SLC25A38*,[6] *GLRX5*,[7] *ABCB7*,[8] *PUS1*,[9] and *SLC19A2*,[10] but the cause of sideroblastic anemia still remains undefined in more than 40% of patients with CSA.[11]

GATA1 transcription factor regulates the expression of several erythroid–specific genes, such as erythropoietin receptor gene,[12,13] α- and β-globin genes,[14,15] *ALAS2*[16] and the *GATA1* gene itself,[17] during erythroid differentiation.[18,19] Ablation of the *Gata1* gene in mice resulted in embryonic death because of anemia,[20] suggesting that GATA1 is essential for erythroid differentiation *in vivo*. It has been reported that GATA1 regulates transcription of human *ALAS2* through the proximal promoter region[16] and the erythroid-specific enhancer located in the eighth intron of *ALAS2*.[21] However, Fujiwara *et al.* demonstrated that the GATA1 protein binds to the *ALAS2* gene only in the middle of its first intron, where no regulatory region had so far been identified, by genome-wide analysis of K562 human erythroleukemia cells using chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq).[22]

In the present study, we have identified a novel erythroid-specific enhancer region in the first intron of the *ALAS2* gene. Moreover, we describe two mutations in the newly identified enhancer of *ALAS2*: a T-to-C transition, which changes GATA to GGTA at the GATA element in the antisense strand, in a pedigree with XLSA and one proband with nfCSA, and a 35-base pair (bp) deletion including the above-mentioned GATA element in a proband with nfCSA.

Methods

Polymerase chain reaction

DNA polymerases used for polymerase chain reaction (PCR) analysis were purchased from TAKARA BIO Inc. (Shiga, Japan). The sequence of primers and probes used in this study are listed in the *Online Supplementary Tables*.

Polymerase chain reaction-based quantitative chromatin immunoprecipitation

Real-time PCR-based quantitative chromatin immunoprecipitation (ChIP-qPCR) analysis was conducted essentially as previously described.[22]

Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) was performed using "DIG Gel Shift Kit, 2nd Generation" (Roche Diagnostics GmbH, Mannheim, Germany), according to the manufacturer's protocol. Sequences of oligonucleotides for probes are indicated by the horizontal bar in the relevant figures. Nuclear extracts were prepared, as described previously,[23] from K562 cells or HEK293 human embryonic kidney cells that were transfected with a GATA1-FLAG fusion protein expression vector or its backbone vector.

Promoter/enhancer activity assays

Each target DNA fragment was prepared from genomic DNA from normal volunteers (WT) or patients with CSA (referred to as "GGTA" or "delGATA" in each reporter construct) and was cloned into pGL3basic plasmid (Promega Corporation, Madison, WI, USA). The human ALAS2 proximal promoter region (g.4820_5115, between -267 and +29 from the transcription start site)[23,24] was cloned into the multiple-cloning site of pGL3basic [referred to as pGL3-AEpro(-267)]. A single DNA fragment (5.2 kbp), carrying the ALAS2 proximal promoter, first exon, first intron and the untranslated region of the second exon, was subcloned into the multiple cloning site of pGL3basic [referred to as pGL3-AEpro(-267)+intron1]. A DNA fragment containing the GATA1-binding region in the first intron of the *ALAS2* gene (corresponding to g.7488_7960), which was defined by ChIP-seq analysis,[22] is referred to as the ChIP-peak. The length of the WT ChIP-peak is 473 bp. In addition, a 130-bp fragment containing ALAS2int1GATA, the consensus sequence for the GATA1-binding site in the ChIP-peak, is referred to as ChIPmini. Several deletion mutants of ChIPmini were prepared using pGL3-AEpro(-267)+ChIPmini(WT) as a template. The pGL3-TKpro plasmid was constructed by cloning herpes simplex virus thymidine kinase promoter into the multiple cloning site of pGL3basic plasmid. Each reporter vector and pEF-RL25 were introduced into K562 cells or HEK293 cells. Luciferase activity was determined using a dual-luciferase reporter system (Promega).

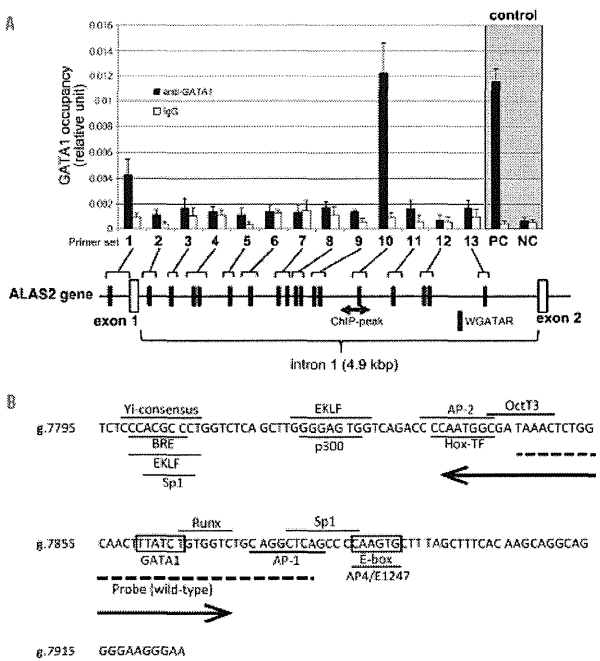

Figure 1. **Identification of a functional GATA1 element in the first intron of the *ALAS2* gene.** (A) Chromatin immunoprecipitation assay. Fragmented genomic DNA segments were immunoprecipitated with anti-GATA1 antibody or control IgG, and then precipitated fragments were quantified using real-time PCR as described in the *Online Supplementary Methods*. PC or NC indicates positive control or negative control, respectively, for the ChIP assay using anti-GATA1 in K562 cells.[23] One GATA element is present in the proximal promoter region and 17 GATA elements in the first intron (black symbols). The shaded double arrow indicates the region corresponding to ChIP-peak. (B) Nucleotide sequence of ChIPmini. The GATA binding site, ALAS2int1GATA, is located in the center of ChIPmini (boxed). A box also indicates the consensus for E-box that is bound by Scl/TAL1.[25] The sequence of ChIPmini was further analyzed for putative transcription factor binding sites using GeneQuest software (DNASTAR Inc., Madison, WI, USA), and the results are indicated by the horizontal bar. Yi-consensus, Yi transcription factor consensus site;[26] BRE, transcription factor IIB binding site;[24] EKLF, erythroid/Kruppel-like factor consensus site;[25] Sp1, stimulatory protein 1 binding site;[24] P300, P300 transcriptional coactivator consensus site;[27] AP-2, AP-2 beta consensus site;[28] Hox-TF, C1 element binding factor binding site;[29] OctT3, OctT3 binding site;[40] Runx, Runx proteins binding site;[41] AP-1, activator protein 1 binding site;[42] and AP4/E1247, AP4/E1247 binding site.[43] The sequence for the wild-type probe used in the EMSA is indicated by a dashed line. A double arrow indicates the deleted region of the delGATA mutation.

©2013 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2013.085449
The online version of this article has a Supplementary Appendix.
Manuscript received on February 4, 2013. Manuscript accepted on August 1, 2013.
Correspondence: furuyama@iwate-med.ac.jp

Identification of mutations of the ALAS2 gene

All exons including exon-intron boundaries, the proximal promoter region, and intron 1 and intron 8 of the *ALAS2* gene (GeneBank: NG_8983.1) were directly sequenced according to previously reported methods.[25]

Measurement of ALAS2 mRNA in purified erythroblasts

Total RNA was extracted from glycophorin A-positive bone marrow mononuclear cells, and was used for cDNA synthesis. ALAS2 expression was measured by real-time PCR, and was normalized to that of GAPDH mRNA.

Statistical analysis

Multiple comparisons between groups were made using the Tukey-Kramer test.

Patients

Eleven probands (eight pedigrees) with CSA of unknown cause were selected to determine the nucleotide sequence of the first intron of *ALAS2* gene. In these patients no disease-causative mutation was identified in the coding regions or reported regulatory regions in *ALAS2*, *SLC25A38*, *GLRX5*, *ABCB7*, *PUS1* and *SLC19A2*, which have been reported to be genes causing CSA[11] (see the *Online Supplementary Methods* for full details of the methods).

The genetic analyses performed in this project were approved by the ethical committee of Tohoku University School of Medicine. Blood samples were withdrawn from the probands and the family members after informed consent.

Results

Polymerase chain reaction-based quantitative chromatin immunoprecipitation analysis of the first intron of the ALAS2 gene

To identify the novel regulatory region for ALAS2 transcription, we first performed ChIP-qPCR analysis in K562 cells to localize the GATA1-binding region of the *ALAS2* gene *in vivo*, which was determined by genome-wide ChIP-seq analysis.[22] In fact, ChIP-qPCR enabled us to examine the GATA1-binding activity of an individual GATA element or two adjacent GATA elements in the first intron of the *ALAS2* gene. Based on a search of NCBI Reference Sequence (NG_8983.1) using SeqBuilder software (DNASTAR Inc., Madison, WI, USA), we identified 17 GATA elements (16 out of 17 GATA elements are present in the antisense orientation) in the first intron of human *ALAS2* (Figure 1A), which is compatible with the previous report.[21] We also included the proximal promoter region that contains a functional GATA-binding site (g.4961_4966).[16] Overall 13 primer sets were designed to amplify the GATA elements located in the proximal promoter region and the first intron of *ALAS2* (Figure 1A and *Online Supplementary Table S1*). Among the 12 primer sets targeting the first intron, using primer set 10, we could amplify genomic DNA that was precipitated with anti-GATA1 antibody at a similar level to that of the positive control, but not with other primer sets. We refer to this region amplified with primer set 10 as ChIPmini (g.7795_7924), the sequence of which is shown in Figure 1B. *In silico* analysis identified only one GATA element (g.7860_7865, boxed in Figure 1B) in ChIPmini, termed ALAS2int1GATA. In addition, primer set 1 which targets the proximal promoter region yielded notable amounts of amplified genome DNA. These results indicate that GATA1 protein bound to the regions amplified with primer sets 1 and 10 in K562 cells; that is, GATA1 protein could bind to the proximal promoter region as well as to ALAS2int1GATA in the first intron of the *ALAS2* gene *in vivo*. Since the GATA element located in the proximal promoter has been well examined *in vitro*,[15] we further determined the functional features of ALAS2int1GATA.

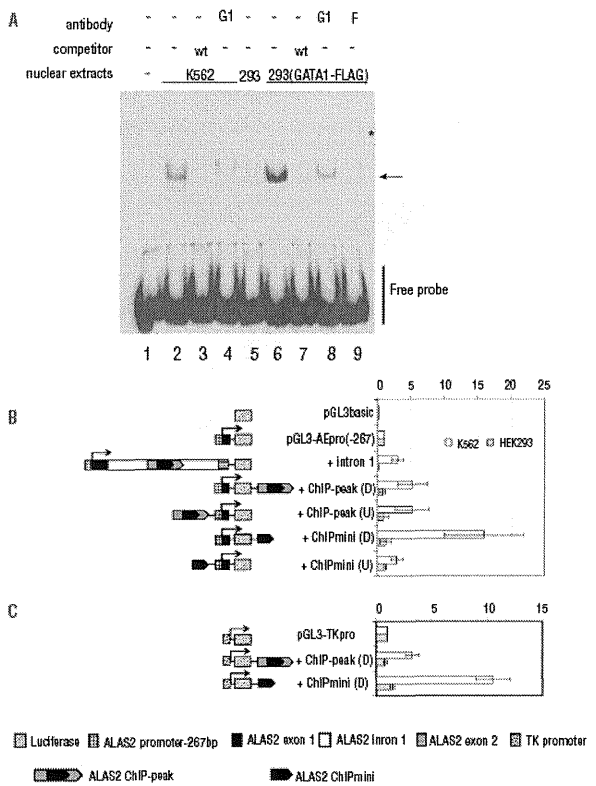

Figure 2. **Functional analyses of ChIPmini present in the first intron of the *ALAS2* gene.** (A) Electrophoretic mobility shift assay (EMSA). Wild-type (wt) probe was incubated with nuclear extracts prepared from K562 cells (lanes 2-4) or HEK293 cells expressing GATA1-FLAG (lanes 6-9). HEK293 cells were transfected with mock vector (lane 5) or FLAG-fused GATA1 expression vector before preparation of nuclear extracts. The protein-probe complex was detected as a retarded band (arrow). An excess amount of unlabeled probe (lanes 3, 7), anti-GATA1 antibody (G1) (lanes 4, 8) or anti-FLAG antibody (F) (lane 9) was included in the reaction mixture. Lane 1 shows the control without nuclear extracts. The asterisk indicates the super-shifted band (lane 9). (B) Functional analysis of ChIPmini as an enhancer for the *ALAS2* gene. Details of the fragments for each plasmid, such as intron1, ChIP-peak and ChIPmini, are described in the *Methods* section. Each DNA fragment was inserted upstream of the *ALAS2* proximal promoter or downstream of luciferase cDNA, indicated as (U) or (D), respectively. Results are expressed as a relative activity compared to that of pGL3-AEpro(-267), and are presented as the mean ± standard deviation (SD) of three independent experiments. (C) Functional analysis of ChIPmini as an enhancer for non-erythroid gene promoter. The enhancer activity of the first intron was examined using the herpes simplex virus TK promoter as a non-erythroid promoter. ChIP-peak or ChIPmini was inserted downstream of the luciferase gene of pGL3-TKpro, yielding pGL3-TKpro+ChIP-peak(D) or pGL3-TKpro+ChIPmini(D). Each of these reporter vectors was introduced into K562 cells or HEK293 cells to measure enhancer activity. Results are expressed as a relative activity compared to that of pGL3-TKpro, and are presented as the mean ± SD of three independent experiments.

GATA1 protein binds to ALAS2int1GATA located in ChIPmini

We then examined whether GATA1 protein binds to ALAS2int1GATA present in the center of ChIPmini using EMSA (Figure 2A). The WT probe contains ALAS2int1GATA (Figure 1B). The incubation of labeled WT probe with nuclear extracts of K562 cells yielded the retarded band that represents the protein-probe complex (lane 2), whereas this retarded band was undetectable with an excess amount of non-labeled WT probe (lane 3). Moreover, the addition of anti-GATA1 antibody reduced the intensity of the retarded band (lane 4), suggesting that GATA1 protein may bind to the WT probe. In fact, the retarded band was not detected when the labeled probe was incubated with nuclear extracts of mock-transfected HEK293 cells (lane 5). In contrast, the retarded band was observed when the labeled probe was incubated with the nuclear extracts of HEK293 cells expressing FLAG-fused GATA1 (lane 6). Importantly, the retarded band observed in lane 6 was not detectable in the presence of an excess amount of non-labeled probe (lane 7). The formation of the retarded band was partially inhibited by anti-GATA1 antibody (lane 8). Likewise, the inclusion of anti-FLAG antibody (lane 9) resulted in the disappearance of the retarded band and instead generated the super-shifted band (indicated by an asterisk). These results suggest that GATA1 protein binds to the WT probe containing ALAS2int1GATA.

Enhancement of ALAS2 promoter activity by the DNA segment containing ALAS2int1GATA

To examine the functional importance of ALAS2int1GATA in the promoter activity of the *ALAS2* gene (Figure 2B), we constructed the pGL3-AEpro(-267) vector, in which the expression of firefly luciferase gene is controlled under the proximal promoter of the *ALAS2* gene (g.4820_5115). The presence of the first intron of *ALAS2* (pGL3-AEpro(-267)+intron1) increased luciferase activity about 3-fold in K562 cells, whereas luciferase activity was decreased to 10% of pGL3-AEpro(-267) in HEK293 cells. When the ChIP-peak, the region determined by ChIP-seq analysis (g.7488_7960),[22] was present downstream [+ChIP-peak(D)] or upstream [+ChIP-peak(U)] of the ALAS2 proximal promoter, luciferase activity was increased about 5-fold, irrespective of the location, compared to that of pGL3-AEpro(-267) in K562 cells. Moreover, the presence of the ChIPmini fragment downstream of the luciferase gene [+ChIPmini (D)] resulted in a 16-fold increase of luciferase activity. However, when the same fragment was inserted upstream of the ALAS2 promoter [+ChIPmini(U)], luciferase activity increased only 3-fold. Thus, the enhancer activity of the ChIPmini fragment varies, depending on its location. Moreover, among the constructs examined, the ChIPmini fragment showed maximum enhancer activity downstream of the luciferase gene. The ChIP-peak or ChIPmini fragment downstream of the ALAS2 promoter influenced luciferase activity marginally (0.73- or 1.25-fold, respectively) in HEK293 cells (Figure 2B). These results suggest that the enhancer activity of each fragment containing ALAS2int1GATA is specific to erythroid cells.

To examine whether the erythroid-specific enhancer activity depends on the ALAS2 promoter, we replaced the ALAS2 promoter with the herpes simplex virus TK promoter (Figure 2C). The ChIP-peak and ChIPmini enhanced TK promoter activity 3.4- and 9.8-fold in K562 cells, respectively, whereas they did not enhance TK promoter activity in HEK293 cells. These results indicate that the erythroid-specific enhancer is present in the ChIP-peak and ChIPmini fragments. In addition, the erythroid-specific enhancer is functional in the non-erythroid gene promoter.

Identification of mutations in the first intron of the ALAS2 gene in patients with congenital sideroblastic anemia

Considering the newly identified enhancer in the first intron of the *ALAS2* gene, we examined whether some CSA patients carry the mutation in ChIP-peak or ChIPmini of *ALAS2*. We determined the nucleotide sequence of the first intron of *ALAS2* in 11 probands (eight pedigrees), and found two distinct mutations in the newly identified enhancer region in five Japanese patients (three pedigrees). The clinical features and hematologic status of the probands at diagnosis of the disease are summarized in Table 1.

Proband 1 in a pedigree with XLSA

The first male Japanese proband was referred to hospital at the age of 3 months to investigate the cause of his pale face. No problems were reported during the birth. Investigations showed microcytic/hypochromic anemia, an increased concentration of serum iron and raised serum ferritin level. Bone marrow aspiration revealed the presence of ring sideroblasts. Two maternal relatives – male cousins of the proband's mother – have sideroblastic ane-