**Table 1. Clinical variables of MDS patients**

| | No. of patients | % of patients | Survival years, median | Dxy (95% CI; P*) | No. of patients | % of patients | AML/25% y | Dxy (95% CI; P*) |
|---|---|---|---|---|---|---|---|---|
| **Cytogenetics** | 7012 | 100 | | .25 | 6485 | 100 | | .27 |
| Very good | 255 | 4 | 5.4 | (.23-.26) | 255 | 3 | NR | (.24-.31) |
| Good | 5069 | 72 | 4.8 | | 4657 | 72 | 9.4 | |
| Intermediate | 947 | 13 | 2.7 | | 875 | 14 | 2.5 | |
| Poor | 283 | 4 | 1.5 | | 276 | 4 | 1.7 | |
| Very poor | 458 | 7 | 0.7 | | 452 | 7 | 0.7 | |
| **BM blasts** | 7012 | 100 | | .30 | 6485 | 100 | | .47 |
| 0-2% | 3279 | 47 | 5.9 | (.28-.32) | 3004 | 46 | NR | (.44-.50) |
| > 2- < 5% | 1266 | 18 | 4.2 | | 1172 | 18 | 8.5 | |
| 5-10% | 1377 | 19 | 2.3 | | 1263 | 20 | 2.2 | |
| > 10%-30% | 1090 | 16 | 1.3 | | 1046 | 16 | 1.0 | |
| > 10%-20% | (901) | (13) | (1.3) | | (860) | (13) | (0.93) | |
| > 20%-30% | (189) | (3) | (1.4) | | (186) | (3) | (1.0) | |
| **Hemoglobin, g/dL** | 7012 | 100 | | .21 | 6485 | 100 | | .16 |
| ≥ 10 | 3377 | 48 | 5.5 | (.19-.23) | 3109 | 48 | 9.5 | (.12-.19) |
| 8-< 10 | 2464 | 35 | 2.9 | | 2286 | 35 | 5.5 | |
| < 8 | 1171 | 17 | 2.0 | | 1090 | 17 | 2.4 | |
| **Platelets†** | 7012 | 100 | | .23 | 6485 | 100 | | .17 |
| ≥ 100 | 4195 | 60 | 5.1 | (.21-.25) | 3823 | 59 | 8.7 | (.14-.21) |
| 50- < 100 | 1469 | 21 | 2.8 | | 1368 | 21 | 3.1 | |
| < 50 | 1348 | 19 | 1.6 | | 1294 | 20 | 3.1 | |
| **ANC†** | 7012 | 100 | | .11 | 6485 | 100 | | .16 |
| ≥ 0.8 | 5758 | 82 | 4.4 | (.10-.13) | 5303 | 82 | 9.2 | (.13-.19) |
| < 0.8 | 1254 | 18 | 1.9 | | 1182 | 18 | 1.9 | |
| **IPSS-R** | 7012 | 100 | | .43 | 6485 | 100 | | .52 |
| Very low | 1313 | 19 | 8.8 | (.42-.45) | 1212 | 19 | NR | (.49-.55) |
| Low | 2646 | 38 | 5.3 | | 2395 | 37 | 10.8 | |
| Intermediate | 1433 | 20 | 3.0 | | 1310 | 20 | 3.2 | |
| High | 898 | 13 | 1.6 | | 857 | 13 | 1.4 | |
| Very high | 722 | 10 | 0.8 | | 711 | 11 | 0.7 | |
| **Sex** | 7012 | 100 | | .07 | 6485 | 100 | | .04 |
| Male | 4243 | 61 | 3.3 | (.05-.09) | 3962 | 61 | 5.8 | (.00-.07) |
| Female | 2769 | 39 | 4.8 | | 2523 | 39 | 8.0 | (.030) |
| **Age** | 7012 | 100 | | .05 | 6485 | | | −.02 |
| ≤ 60 y | 1582 | 23 | 5.7 | (.03-.06) | 1489 | 23 | 8.1 | (−.05-.01) |
| > 60 y | 5430 | 77 | 3.5 | | 4996 | 77 | 6.1 | (.082) |
| **ECOG Performance Status** | 2496 | 36 | | .16 | 2489 | 38 | | .09 |
| 0 | 751 | 30 | 4.3 | (.13-.18) | 748 | 30 | 8.8 | (.04-.15) |
| 1 | 1477 | 59 | 2.2 | | 1473 | 59 | 6.3 | (.005) |
| 2-4 | 268 | 11 | 1.6 | | 268 | 11 | 3.5 | |
| **Serum ferritin** | 3049 | 43 | | .16 | 2747 | 42 | | .11 |
| ≤ 350 ng/mL | 1602 | 53 | 6.3 | (.13-.20) | 1435 | 52 | NR | (.05-.17) |
| > 350 ng/mL | 1447 | 47 | 4.2 | | 1312 | 48 | 14.5 | (.004) |
| **Serum LDH** | 4257 | 61 | | .12 | 4130 | 64 | | .12 |
| Normal | 3103 | 73 | 4.1 | (.10-.14) | 3007 | 73 | 9.2 | (.08-.16) |
| High | 1154 | 27 | 2.1 | | 1123 | 27 | 3.2 | |
| **Serum $\beta_2$-microglobulin** | 1005 | 14 | | .14 (.10-.18) | 1005 | 15 | | .02 (−.08-.11) |
| ≤ 2 g/mL | 263 | 26 | 3.8 | | 263 | 26 | 6.7 | (.498) |
| > 2 g/mL | 742 | 74 | 1.7 | | 742 | 74 | 4.6 | |
| **Marrow fibrosis** | 1323 | 19 | | .04 | 1183 | 18 | | .05 |
| No | 1158 | 88 | 5.2 | (.01-.07) | 1055 | 89 | 14.5 | (−.01-.12) |
| Yes | 165 | 12 | 3.2 | (.004) | 128 | 11 | 4.8 | (.069) |
| **RBC transfusion dependence** | 2933 | 42 | | .26 | 2645 | 41 | | .27 |
| No | 2003 | 68 | 6.9 | (.23-.29) | 1808 | 68 | 14.5 | (.22-.32) |
| Yes | 930 | 32 | 2.3 | | 837 | 32 | 2.1 | |
| **IPSS** | 7008 | 100 | | .37 | 6481 | 100 | | .48 |
| Low | 2625 | 37 | 7.0 | (.35-.39) | 2394 | 37 | NR | (.45-.51) |
| Intermediate-1 | 2778 | 40 | 3.6 | | 2541 | 39 | 6.1 | |
| Intermediate-2 | 1126 | 16 | 1.5 | | 1074 | 17 | 1.2 | |
| High | 479 | 7 | 0.9 | | 472 | 7 | 0.7 | |

AML/25% indicates time for 25% of patients to develop AML.

*All univariate P values not explicitly stated are $P < .001$.

†$\times 10^9$/L.

**Table 2. MDS Cytogenetic Scoring System**

| Prognostic subgroups, % of patients | Cytogenetic abnormalities | Median survival,* y | Median AML evolution, 25%,* y | Hazard ratios OS/AML* | Hazard ratios OS/AML† |
|---|---|---|---|---|---|
| Very good (4%*/3%†) | −Y, del(11q) | 5.4 | NR | 0.7/0.4 | 0.5/0.5 |
| Good (72%*/66%†) | Normal, del(5q), del(12p), del(20q), double including del(5q) | 4.8 | 9.4 | 1/1 | 1/1 |
| Intermediate (13%*/19%†) | del(7q), +8, +19, i(17q), any other single or double independent clones | 2.7 | 2.5 | 1.5/1.8 | 1.6/2.2 |
| Poor (4%*/5%†) | −7, inv(3)/t(3q)/del(3q), double including −7/del(7q), complex: 3 abnormalities | 1.5 | 1.7 | 2.3/2.3 | 2.6/3.4 |
| Very poor (7%*/7%†) | Complex: > 3 abnormalities | 0.7 | 0.7 | 3.8/3.6 | 4.2/4.9 |

OS indicates overall survival; and NR, not reached.
*Data from patients in this IWG-PM database, multivariate analysis (n = 7012).
†Data from Schanz et al[8] (n = 2754).

than merely the number of these abnormalities; and modification of the ANC cutpoint to $0.8 \times 10^9$/L from $1.8 \times 10^9$/L in the IPSS.

Regarding the cytogenetic classification, good correlation was demonstrated regarding the proportional hazard ratios for clinical outcomes (survival and AML evolution) of the subgroups from the recently developed cytogenetic system[8] on which we based our analysis and from our data (Table 2). Because of the higher number of patients analyzed in our database, more cytogenetic subtypes were analyzable for prognosis in our study than had been possible for the IPSS (15 vs 6). A double independent review of the cytogenetic data was performed by the IWG-PM Cytogenetic Committee. Differences between this categorization and that of the IPSS included the finding of complex karyotypes with > 3 abnormalities being distinct from those with 3 abnormalities and with poorer prognosis; chromosome 7 abnormalities were similarly prognostically separable from the Very poor category (when observed in karyotypes with ≤ 3 abnormalities; supplemental Figures 1 and 2, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). The specific cytogenetic abnormalities that now were able to be placed into distinct prognostic subgroups included inv(3)/t3q)/del(3q), del(11q), del(12p), i(17q), +19, double anomalies including del(5q), double abnormalities including del(7q) or monosomy 7, and any other double changes [in addition to the previously IPSS-denoted −Y, del(5q), del(20q), all as single abnormalities] (Table 2).

The distributions of the IPSS cytogenetic categories in the present IWG-PM database was similar to those in the IMRAW database, which generated the IPSS (1): IPSS Good/Intermediate/Poor 73%/15%/12% (IMRAW 70/14/16%). This contrasted with the cytogenetic categorization in the IPSS-R: Very good/Good/Intermediate/Poor/Very poor 4%/72%/13%/4/7%. The IPSS clinical subgroups in our patient cohort were: Low 37%, Intermediate-1 40%, Intermediate-2 16%, and High 7%. These groups were also similar to the IMRAW patients: 33%/38%/22%/7% (1). As both FAB and WHO morphologic classifications were used, refractory anemia with excess blasts in transformation 6%, chronic myelomonocytic leukemia 9%, and isolated del(5q) 4% were represented in our patients.

### Model development

Analysis of the marrow blast cutpoints indicated that striking differences were evident for both survival and AML evolution for patients with blasts 0- ≤ 2% (lower risk) versus blasts > 2- < 5%: Cox univariate pairwise comparison hazard ratio 1.4 (95% CI, 1.3-1.5, $P < .001$) for survival and 2.4 (95% CI, 1.9-2.9 $P < .001$) for AML evolution (Figures 1 and 2). Multivariate results confirmed this finding. Thus, we incorporated these distinct categorical values into the scoring model. Further, the statistical analysis of clinical outcomes for blasts > 10-≤ 20% vs > 20-≤ 30% indicated that these values had similar risk: hazard ratio = 1.0 (95% CI, 0.8-1.2, $P = .996$) for survival and 0.8 (95% CI, 0.6-1.1 $P = .174$) for AML evolution (supplemental Figures 3 and 4). Thus, we combined these 2 categories in the scoring model. In addition, this finding of the statistical relevance of the specific blast cutpoints in the combined database was also present in the individual databases, including that of the IMRAW database from which the IPSS[1] had been derived. For the IMRAW patients, the hazard ratios for survival and AML evolution were substantially the same as those for the combined database for both the lower and higher blast group analyses.

Review of the data indicated that baseline depths of cytopenias were statistically and clinically important (Table 3). The relevant cutpoints were: hemoglobin values of < 8, 8-< 10, and ≥ 10 g/dL, platelets of < 50, 50-100 and ≥ $100 \times 10^9$/L, and ANC of < 0.8 versus ≥ $0.8 \times 10^9$/L.

The changes from the cutpoints used in the IPSS-R compared with those from the IPSS include (1) separating marrow blasts < 5% into 0%-2% and > 2-< 5%; and (2) providing differing depths of cytopenias; also, as patients with marrow blasts of 10%-20% had similar outcomes as those with 21%-30%; and (3)

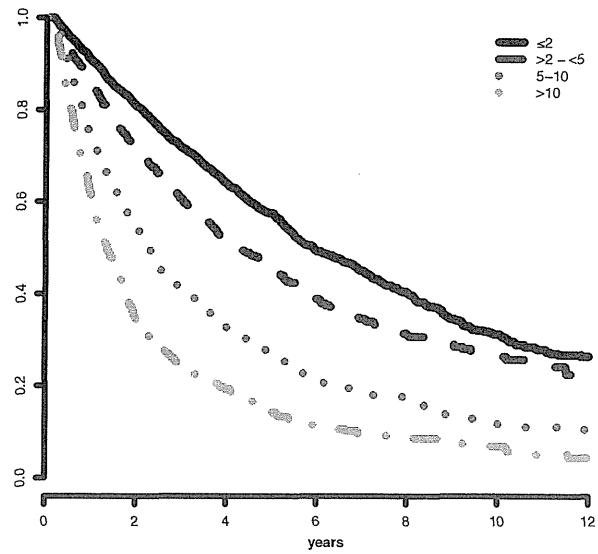



**Figure 1. IWG-PM patients marrow blast subgroups. Impact on survival.** Survival related to MDS patients' individual marrow blast percent categories (Kaplan-Meier curves, Dxy 0.3, $P < .001$). The number of patients in each category and their proportional representation are shown in Table 1.

**Table 3. IPSS-R prognostic score values**

| Prognostic variable | 0 | 0.5 | 1 | 1.5 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| Cytogenetics | Very good | — | Good | — | Intermediate | Poor | Very poor |
| BM blast, % | ≤ 2 | — | > 2%- < 5% | — | 5%-10% | > 10% | — |
| Hemoglobin | ≥ 10 | — | 8- < 10 | < 8 | — | — | — |
| Platelets | ≥ 100 | 50-< 100 | < 50 | — | — | — | — |
| ANC | ≥ 0.8 | < 0.8 | — | — | — | — | — |

— indicates not applicable.

the category of marrow blasts > 10%-30% usefully described the statistical impact of this parameter compared with having separated these groups in the IPSS.

The IPSS-R prognostic risk categories were determined by combining the scores of these main 5 features (Table 4). The model permitted the definition of 5 well-separated prognostic categories for both survival and AML evolution in the IPSS-R (Very low, Low, Intermediate, High, Very high) rather than the 4 categories that are present in the IPSS (Tables 4 and 5; Figures 3 and 4). These risk categories describe scores for 70-year-old patients.

Survival duration and time to AML evolution for patients within these 5 prognostic categories are shown in Table 5 and Figures 3 and 4. As indicated in Table 5, in which hazard ratios are shown, ~ 56% of the patients were in the lower risk (Very low and Low) and ~ 23% were in the higher (High and Very high) risk prognostic subgroups for both of these clinical outcomes. For both survival and time to AML evolution, the individual centers' Dxy's were in good agreement with that for the total patient cohort.

Ready extrapolation is available to adjust the score for patients of any age by use of the following formula: (years − 70) × [0.05 − (IPSS-R risk score × 0.005)], add the result to the sum of the 5 major variables. Patient age clearly had major impact on survival (ie, decreased survival with aging), but not for AML evolution (Figure 5; supplemental Figure 5). Table 6 provides specific survival data within each risk category for patients of differing ages. Figure 6 provides a nomogram, based on the just noted formula, which visually describes the method to determine predicted survival based on patient's age and risk status, generating age-adjusted IPSS-R categorization (IPSS-RA).

Additional significant differentiating features for predicting survival were found, although their impact on prognostic score was relatively low compared with the 5 major features and age. These were: performance status, serum ferritin, LDH, and possibly $\beta_2$-microglobulin (supplemental Table 1). For determining the contribution of each of these features to the patient's risk category, the numerical values (with the categorized values for each variable) are indicated in the table and should be added to the raw scores of the major variables. This table provides multivariate *P* values as well as indicating the incremental contribution of a feature to the already defined score. Of note, none of these variables was a statistically significant additive feature for predicting AML evolution.

As shown in Table 7, differences were noted in the proportion of our patient cohort who died with or without developing AML in relation to their initial prognostic risk category. Of the patients who died, if observed until death, the proportions dying with leukemia in the groups were 13%-33%, positively related to their higher-risk categories.

### Distinction between IPSS-R and IPSS

A summary of the refinements of the IPSS-R beyond the IPSS is shown in Table 8. The IPSS-R model showed effective separation of the IPSS patient risk categories and more effectively discriminated prognostic risk for these patients than the IPSS, as indicated by the higher Dxy values (.43 vs .37 for survival, .52 vs .48 for AML evolution; Table 1). Data indicated that 99% of the patients in the IPSS-R Very low and Low risk subgroups encompassed those who had been classified as IPSS Low and Intermediate-1; 81% of those in the IPSS-R High and Very high risk subgroups had been classified as IPSS Intermediate-2 and High (Figure 7, Kendall tau = 0.73). The IPSS-R Intermediate category (20% of the patients) was composed of 73% IPSS Intermediate-1, 19% Intermediate-2, 7% Low, 1% High (Table 9). In the IPSS lower risk group (Low/Intermediate-1), 27% of these patients were shifted into higher risk IPSS-R categories (mainly Intermediate). At the other prognostic extreme, 18% of the IPSS higher-risk (Intermediate-2/High) were downstaged into lower-risk IPSS-R categories (predominantly IPSS-R Intermediate).

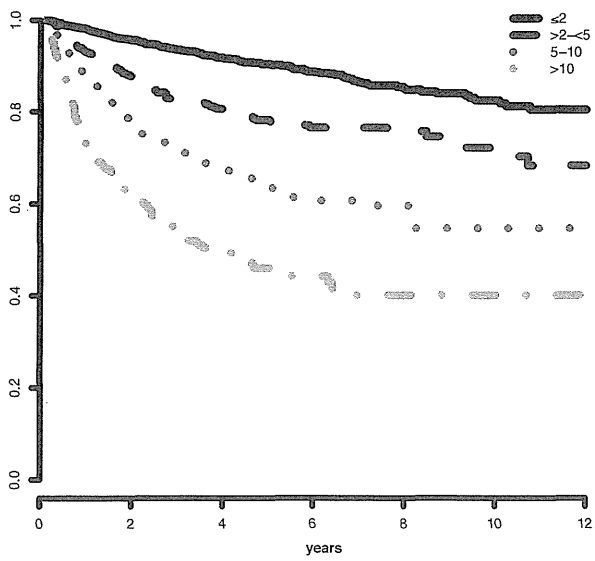



**Figure 2. IWG-PM patients marrow blast subgroups: Impact on AML evolution.** Progression to AML related to MDS patients' individual marrow blast percent categories (Kaplan-Meier curves, Dxy 0.47, $P < .001$). The number of patients in each category and their proportional representation are shown in Table 1.

**Table 4. IPSS-R prognostic risk categories/scores**

| Risk category | Risk score |
|---|---|
| Very low | ≤ 1.5 |
| Low | > 1.5-3 |
| Intermediate | > 3-4.5 |
| High | > 4.5-6 |
| Very high | > 6 |

**Table 5. IPSS-R prognostic risk category clinical outcomes**

| | No. of patients | Very low | Low | Intermediate | High | Very high |
|---|---|---|---|---|---|---|
| Patients, % | 7012 | 19 | 38 | 20 | 13 | 10 |
| Survival, all* | | 8.8 | 5.3 | 3.0 | 1.6 | 0.8 |
| | | (7.8-9.9) | (5.1-5.7) | (2.7-3.3) | (1.5-1.7) | (0.7-0.8) |
| Hazard ratio | | 0.5 | 1.0 | 2.0 | 3.2 | 8.0 |
| (95% CI) | | (0.46-0.59) | (0.93-1.1) | (1.8-2.1) | (2.9-3.5) | (7.2-8.8) |
| Patients, % | 6485 | 19 | 37 | 20 | 13 | 11 |
| AML/25%*† | | NR | 10.8 | 3.2 | 1.4 | 0.73 |
| | | (14.5-NR) | (9.2-NR) | (2.8-4.4) | (1.1-1.7) | (0.7-0.9) |
| Hazard ratio | | 0.5 | 1.0 | 3.0 | 6.2 | 12.7 |
| (95% CI) | | (0.4-0.6) | (0.9-1.2) | (2.7-3.5) | (5.4-7.2) | (10.6-15.2) |

NR indicates not reached.
*Medians, years (95% CI), $P < .001$.
†Median time to 25% AML evolution (95% CIs), $P < .001$.

### Model validation

After construction and acceptance of the IPSS-R within the IWG-PM, an external validation cohort of 200 MDS patients from the Medical University of Vienna was evaluated and demonstrated good comparability of demographic features with the global IWG-PM database (supplemental Table 2). Their median age was 71 years, 83% were > 60 years, and the male/female ratio was 1.2:1 with a median follow-up time of 4.6 years. A similar proportion of these patients composed the cytogenetic, clinical, and IPSS-R subgroups, as did the global IWG-PM cohort (supplemental Table 2). Our IPSS-R multivariate model fit these data well, exhibiting high prognostic power, as indicated by the high Dxy's and the clearly differing temporal medians and hazard ratios between prognostic risk categories for both survival and AML evolution (Table 10, supplemental Table 2), including age-related survival (supplemental Table 3). Cox model analyses also supported the improved prognostic ability compared with the IPSS.

## Discussion

We herein describe the IPSS-R, which provides useful advances and more discriminatory prognostic risk assessment beyond the IPSS for assessing clinical outcomes in MDS (Table 8). Although the IPSS has been an important standard for assessing prognosis of primary untreated adult MDS patients over the past decade, additional refinements[27] and prognostic variables have been suggested as providing meaningful differences for patient clinical outcomes.[2-5,9-20] In addition, cytogenetic subgroups have recently been defined as providing improved prognostic evaluation of clinical outcomes of primary MDS patients.[8] A number of prior prognostic systems in addition to the IPSS have demonstrated merit, although the relative value of each variable was unclear. Thus, in this collaborative IWG-PM project, with the large number of patients evaluated from multiple coalesced databases (n = 7012), we have integrated the various recently independently defined

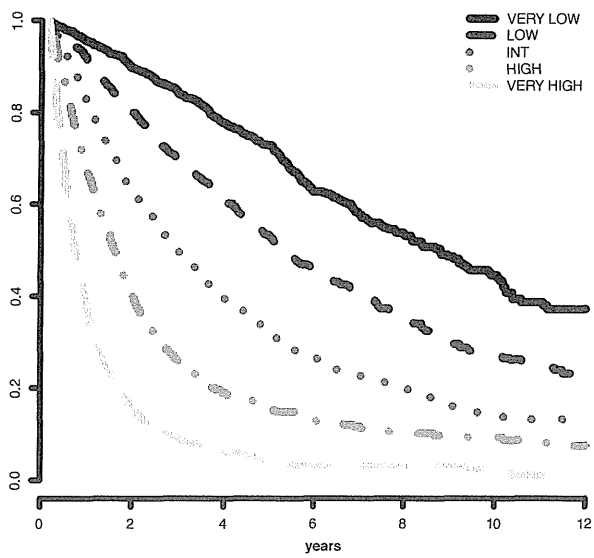



**Figure 3. Survival based on IPSS-R prognostic risk-based categories.** Survival related to MDS patients' prognostic risk categories (Kaplan-Meier curves, n = 7012; Dxy 0.43, $P < .001$). The number of patients in each category and their proportional representation are shown in Table 1.

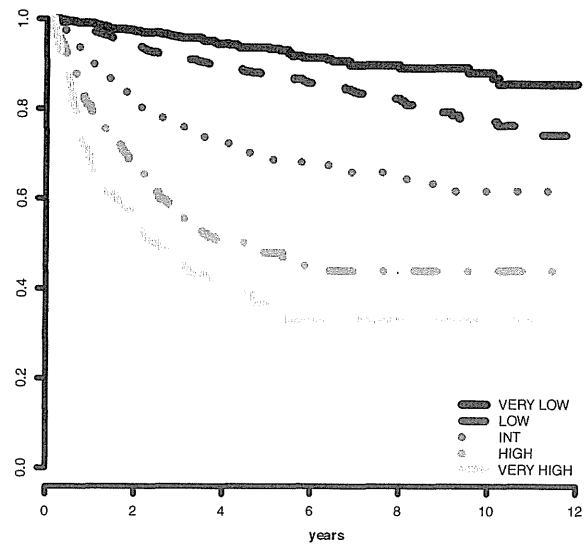



**Figure 4. AML evolution based on IPSS-R prognostic risk-based categories.** Progression to AML related to MDS patients' prognostic risk categories (Kaplan-Meier curves, n = 6485; Dxy 0.52, $P < .001$). The number of patients in each category and their proportional representation are shown in Table 1.

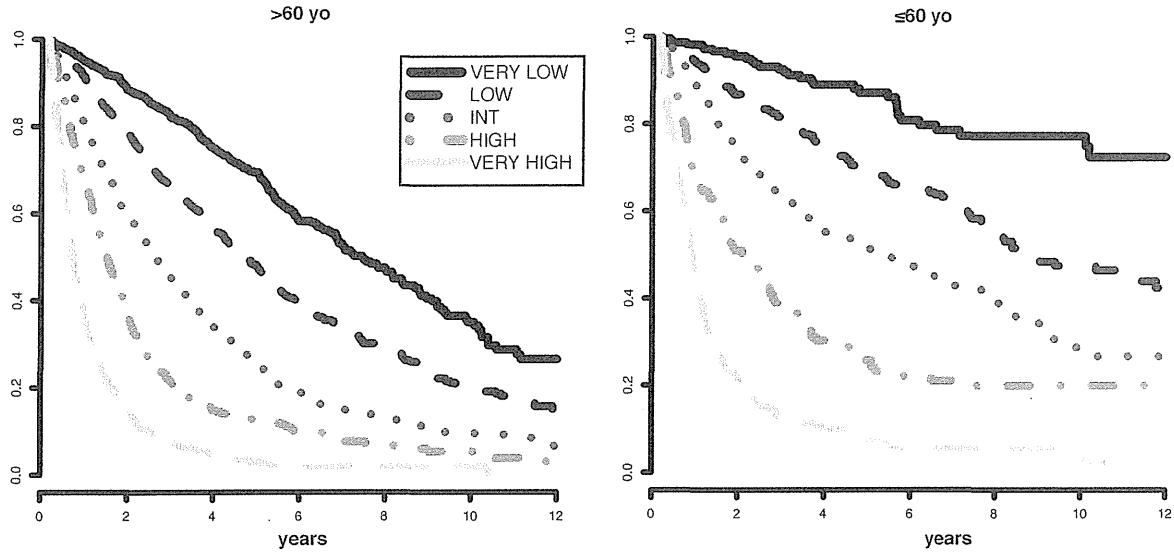



**Figure 5. Survival based on patient ages > 60 years vs ≤ 60 years related to their IPSS-R prognostic risk-based categories (Kaplan-Meier curves).** Age-related differential survivals are shown for patients in all groups, particularly for those in lower risk categories.

clinical factors for MDS in a comprehensive method and assessed their relative prognostic impact. Multivariate analyses demonstrated that the same major features present in the IPSS (cytogenetic subgroups, marrow blast percentage, and cytopenias) retained major prognostic impact in IPSS-R (in descending order of statistical weight). However, more precise prognostication of these clinical outcomes (survival and AML evolution) in the IPSS-R was demonstrated by effective refinement of these features within the IPSS-R (depth of cytopenias, splitting of marrow blasts < 5%, and more precise cytogenetic subgroups). The IPSS-R also demonstrated improved predictive prognostic power with more precise prognostic categories[5] versus 4 groups in the IPSS. In particular, a substantial proportion of those patients previously placed within the IPSS Intermediate-1 and -2 categories were more precisely separated into all 5 IPSS-R categories.

Analogous to the IPSS, based on our clinical outcome data, lower-risk patients are composed within the IPSS-R Very low and Low categories; higher-risk patients are composed within the High and Very high categories. However, as shown in Table 9, the IPSS-R has permitted improved refinement of risk categories for the IPSS Intermediate-1 and Intermediate-2 patients because a substantial portion of the patients who would have been categorized as IPSS Intermediate-1 are now in the IPSS-R Low category; a substantial portion of the patients who would have been categorized as IPSS Intermediate-2 are now in the IPSS-R High category. In other words, the "better risk" IPSS Intermediate-1 patients have been categorized into the lower-risk IPSS-R category; the "poorer risk" IPSS Intermediate-2 patients are now in the higher-risk IPSS-R category. Remaining within the IPSS-R Intermediate category are those who, indeed, have "intermediate" risk (Tables 5 and 7). On review, the clinical outcome data indicate that the IPSS-R Intermediate category appears closer to the initial IMRAW IPSS Intermediate-1 group (survival 3.0 years for IPSS-R Intermediate vs 3.5 years for IPSS Intermediate-1) than it is to IPSS Intermediate-2[1]; the proximity of AML evolution for the IPSS-R Intermediate category is also closer to Intermediate-1 than to Intermediate-2 (ie, the "lower risk" patient group). However, per Table 7, the proportion of patients dying with leukemia for the

**Table 6. IPSS-R survival related to age**

| | IPSS-R prognostic risk categories | | | | |
|---|---|---|---|---|---|
| **Ages, y** | **Very low** | **Low** | **Intermediate** | **High** | **Very high** |
| All | 8.8 | 5.3 | 3.0 | 1.6 | 0.8 |
| ≤ 60 | NR | 8.8 | 5.2 | 2.1 | 0.9 |
| | (13.0-NR) | (8.1-12.1) | (4.0-7.7) | (1.7-2.8) | (0.8-1.0) |
| > 60-70 | 10.2 | 6.1 | 3.3 | 1.6 | 0.8 |
| | (9.1-NR) | (5.3-7.4) | (2.5-4.0) | (1.5-2.0) | (0.7-1.0) |
| > 70-80 | 7.0 | 4.7 | 2.7 | 1.5 | 0.7 |
| | (5.9-9.0) | (4.3-5.3) | (2.4-3.1) | (1.3-1.7) | (0.6-0.8) |
| > 80 | 5.2 | 3.2 | 1.8 | 1.5 | 0.7 |
| | (4.2-5.9) | (2.8-3.8) | (1.6-2.6) | (1.2-1.7) | (0.5-0.8) |
| ≤ 60 (median, 52) | NR | 8.8 | 5.2 | 2.1 | 0.9 |
| > 60 (median, 74) | 7.5 | 4.7 | 2.6 | 1.5 | 0.7 |
| ≤ 70 (median, 62) | 13.3 | 7.7 | 3.9 | 1.7 | 0.9 |
| > 70 (median, 77) | 5.9 | 4.2 | 2.5 | 1.4 | 0.7 |

Data are median (95% CI).
NR indicates not reached.

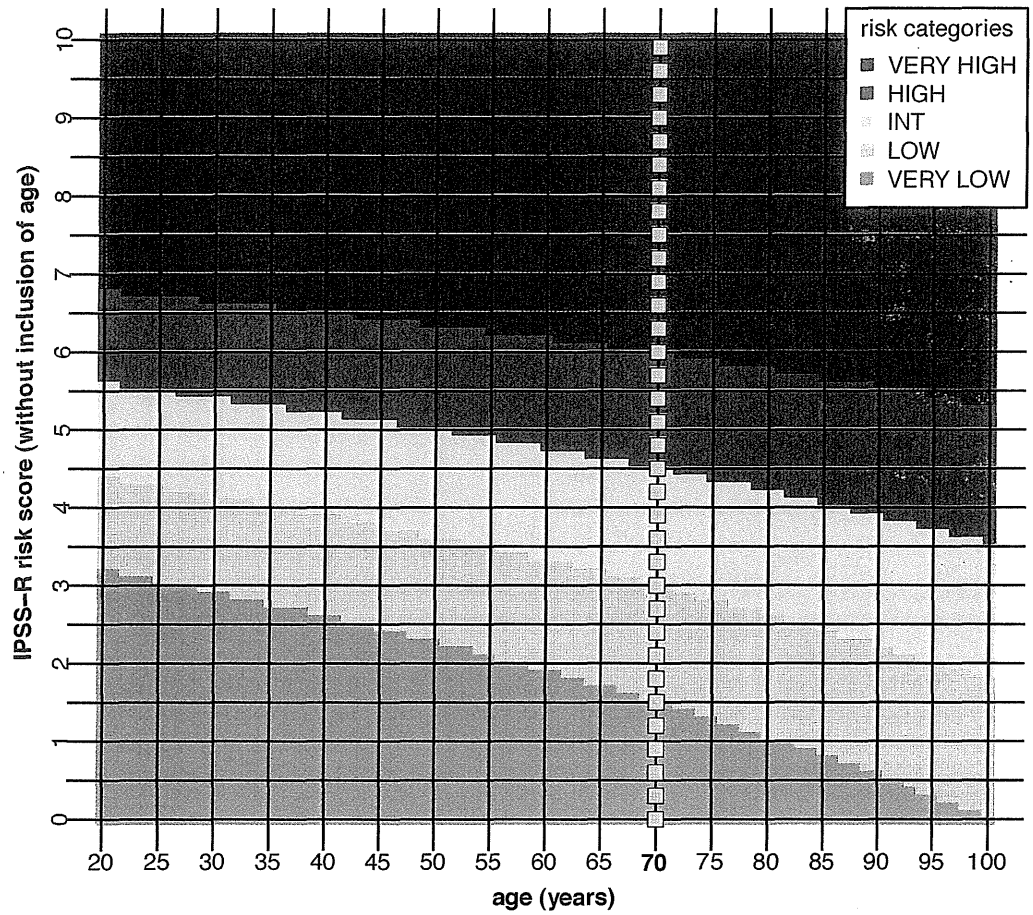



**Figure 6. Age-adjusted IPSS-R risk categories.** The nomogram describes predicted survival based on patient age and IPSS-R risk status (IPSS-RA). To determine an age-adjusted risk categorization, for example, follow the horizontal line, starting at the IPSS-R risk score 3.5 on the vertical axis (Int [Intermediate] risk category per Table 4) to the age of the patient and record the color at that point. If the patient is 45 years, the 3.5′-line and the vertical 45-year line cross in the gray field, placing the patient in the Low risk category, whereas if the patient is 95 years the 3.5′-line and the 95-year line cross in the yellow field, placing the patient in the Intermediate risk category. As indicated, for most patients in the Very high risk category there is no change of risk group, whereas for most patients in the lower risk categories there is greater possibility of category change. Note the "dotted" vertical line at 70 years, which is at the median age of the IWG-PM patient cohort from which the basic risk category scores were calculated (ie, without need for age correction for these patients). The formula to generate the age-adjusted risk score in the figure: (years − 70) × [0.05 − (IPSS-R risk score × 0.005)]. Example: For the 45-year-old patient with an IPSS-R risk score of 3.5 (Intermediate risk): (45-70) × [0.05 − (3.5 × 0.005)] = −0.81. Thus, 3.5-0.8 = 2.7 [age-adjusted IPSS-R score, IPSS-RA: Low risk].

IPSS-R Intermediate category is distinctively worse than for the lower-risk categories. As survival is the major endpoint for most MDS clinical trials, and of predominant concern to patients and caregivers, it seems reasonable to suggest placement of IPSS-R Intermediate patients into the lower-risk group regarding their potential therapeutic management. However, given the distinctiveness of this patient category, assessment of these patients within both lower- and higher-risk treatment protocols appears warranted. Clinical trial evaluation and recommendations by practice guidelines committees will be needed to substantiate this point. Use of the additional differentiating features (eg, age, performance status, ferritin, LDH; see below) could be of particular value for categorization of these patients.

A major component of this schema was the provision of 5 cytogenetic subgroups (vs 3 in the IPSS) based on an increased number of specific prognostic chromosomal categories[15] versus 6 in the IPSS. This increase in defined cytogenetic categories, with their increased prognostic weight, was the result of a larger number

**Table 7. Mortality of MDS patients with or without AML evolution**

| Risk category | No. (%) of patients | Patients who | | |
|---|---|---|---|---|
| | | Died, no. (%) | Died with AML, no. (%) | Died without AML, no. (%) |
| Very low | 1313 | 350 (27) | 46 (13) | 304 (87) |
| Low | 2646 | 1053 (40) | 174 (17) | 861 (83) |
| Intermediate | 1433 | 782 (55) | 205 (26) | 568 (74) |
| High | 898 | 633 (71) | 207 (33) | 421 (67) |
| Very high | 722 | 619 (86) | 193 (31) | 422 (69) |
| Total | 7012 | 3437 (49) | 825 (24) | 2576 (76) |

**Table 8. Refinements of the IPSS-R beyond the IPSS**

| |
|---|
| 1. New marrow blast categories |
| ≤ 2%, > 2%-< 5%, 5%-10%, > 10%-30% |
| 2. Refined cytogenetic abnormalities and risk groups |
| 16 (vs 6) specific abnormalities, 5 (vs 3) subgroups |
| 3. Evaluation of depth of cytopenias |
| Clinically and statistically relevant cutpoints used |
| 4. Inclusion of differentiating features* |
| Age, Performance Status, serum ferritin, LDH; $\beta_2$-microglobulin† |
| 5. Prognostic model with 5 (vs 4) risk categories |
| Improved predictive power |

*For survival.
†Provisional.

**Table 9. Distribution (%) of IWG-PM patients who would previously have been categorized by IPSS now categorized by IPSS-R**

| IPSS | Very low | Low | Intermediate | High | Very High |
|---|---|---|---|---|---|
| Low (37) | 44 | 52 | 4 | 0 | 0 |
| Intermediate-1 (40) | 6 | 45 | 38 | 10 | 1 |
| Intermediate-2 (16) | 0 | 1 | 24 | 45 | 30 |
| High (7) | 0 | 0 | 3 | 19 | 78 |
| Total | 19 | 38 | 20 | 13 | 10 |

% indicated within rows. Kendall tau = 0.73.

of patients available for analysis of some of the relatively rare cytogenetic categories. This increased number of cases permitted specific characterization of many of the cytogenetic subgroups that had previously been labeled in the IPSS as "other" and also separated the prior Good and Poor groups into Very good and Good and Poor and Very poor, respectively, thus improving their prognostic accuracy.

Splitting patients with marrow blasts < 5% into those with 0%-2% and > 2-< 5% provided groups with very low risk versus low risk features. The issue of splitting this "low blast group" into 2 separate subgroups may present a challenge for reporting these values in some routine clinical laboratories. However, these differences in blast enumeration were reproducible within the various databases from the different institutions in our study. The discriminatory lower blast percentages should be of particular importance in helping to ensure balanced representation of patients in clinical trials.

The presence of 10%-20% marrow blasts had similar impact on clinical outcomes as did 21%-30% blasts. Thus, these 2 categories were combined in the prognostic model. The underlying reason(s) for this finding is unclear but could relate to the stringent entry criteria for our patient cohort (eg, excluding treated patients and those with high circulating blasts or patients' innate biologic similarity). Of interest, this similarity of clinical outcomes for patients within both the low marrow blast group and the 10%-30% blast group was also demonstrated in the IMRAW database from which the initial IPSS was generated.

Scoring the depth of cytopenias by subdivision at clinically and statistically relevant cutpoints rather than solely counting their

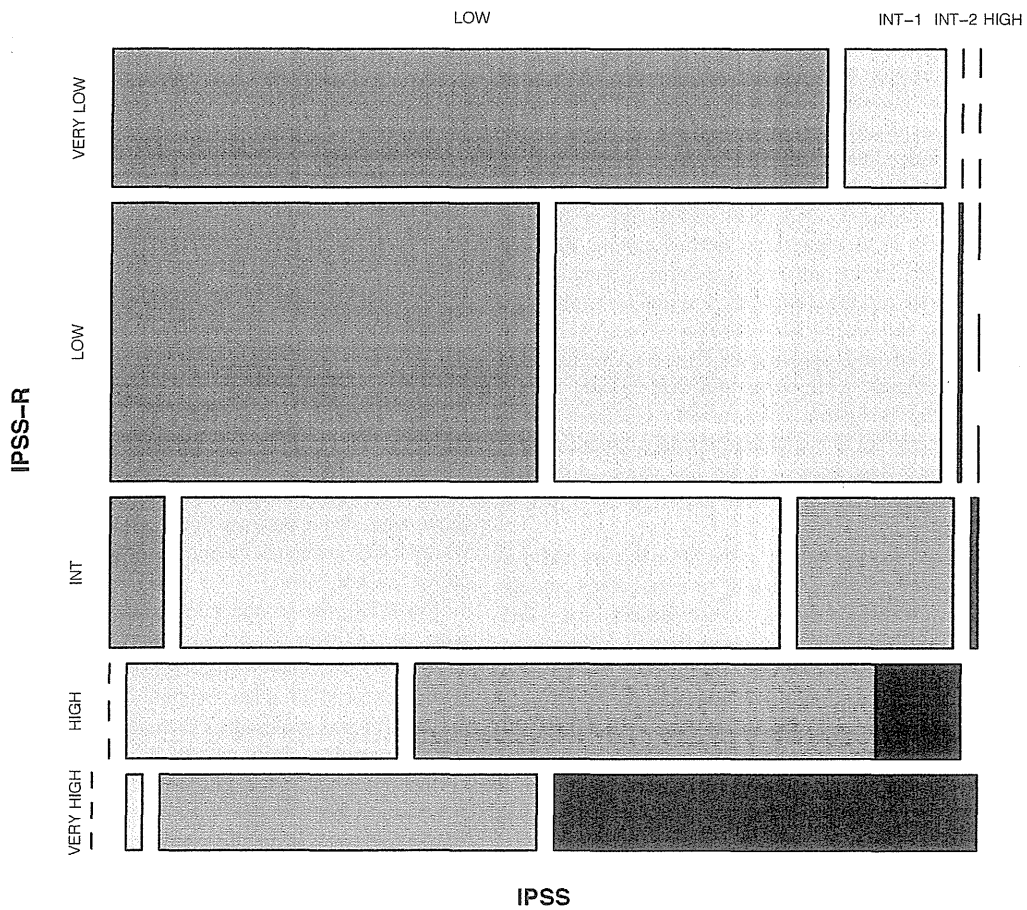



**Figure 7. Comparison of IPSS-R and IPSS subgroups within the IWG-PM database patient cohort.** Vertical axis represents IPSS-R categories' and horizontal axis, IPSS categories. The proportion of patients in each category is shown in Table 9. Kendall $\tau$ = 0.73.

**Table 10. IPSS-R prognostic risk category: clinical outcomes of Medical University of Vienna patients (n = 200)**

| | Very low | Low | Intermediate | High | Very high |
|---|---|---|---|---|---|
| **Patients, %** | 21 | 38 | 18 | 14 | 8 |
| **Survival** | | | | | |
| All* | 9.3 | 6.3 | 3.4 | 1.2 | 0.6 |
| Hazard ratio | 0.8 | 1 | 2.1 | 4.3 | 9.4 |
| 95% CI | 0.4-1.5 | 0.7-1.5 | 1.3-3.5 | 2.4-7.7 | 4.3-20.8 |
| **AML transformation** | | | | | |
| AML/25%*† | NR | NR | 2.4 | 0.8 | 0.6 |
| Hazard ratio | 0 | 1 | 8.0 | 18.7 | 52.2 |
| 95% CI | 0-∞ | 0.3-3.9 | 3.1-20.5 | 7.0-49.7 | 13.8-198.2 |

NR indicates not reached.
*Medians, years (95% CI; $P < .001$).
†Median time to 25% AML evolution (95% CI; $P < .001$).

presence was demonstrated as being useful. The degree of anemia is an important correlate of poor clinical outcomes in MDS[27] and appears to be a good surrogate for RBC transfusion dependence.[28] In this regard, low hemoglobin levels have recently replaced RBC transfusion dependence as a prognostic parameter of the WPSS.[28] Underlying this finding, chronic anemia may contribute to the high nonleukemic mortality related to cardiac disease in MDS patients.[2,3,28]

The other cytopenia cutpoints also have clinical relevance. The ANC of $0.8 \times 10^9$/L (in IPSS-R) is associated with higher potential infectious risk rather than that of $1.8 \times 10^9$/L (in the IPSS).[29,30] Severe thrombocytopenia has been associated with increased morbidity and poor survival in MDS patients.[29,31-33]

The impact of age was a major prognostic parameter for overall survival, although not for AML evolution. This effect has previously been shown with the IPSS and in other studies.[1,29,34-36] In the IPSS-R prognostic model, the data are shown for age 70 years (the near median age of our patient cohort). However, to incorporate the model for different patient ages (IPSS-RA), we provide a formula (in "Results"), which permits statistical adjustment of survival prognosis for patients of all ages. This formula and the Figure 6 nomogram for calculating the impact of age for modifying the risk score/category provide resources for clinicians and for trial design and analysis. These age-related survival data are also shown in Table 6, Figures 5 and 6, and supplemental Figure 5. Our approach herein for providing age as an important, although optional, feature to assess predicted survival permits this variable to be used if total mortality risk is the aim but to not be used if solely disease-related risk is the objective. Age as a variable has some prognostic influence in all risk groups, but with more impact in lower than in higher-risk patients.

Additional differentiating features in the IPSS-R were additive to the 5 major parameters for predicting survival, albeit not for AML evolution: performance status, serum ferritin, and LDH levels. Serum LDH,[9-11] ferritin,[12] and $\beta_2$-microglobulin[13,14] have previously been shown to have prognostic importance for survival in MDS. Thus, our analyses have helped determine that these clinical features, of the many previously reported, were also reproducible in our large patient cohort after multivariate analysis. Relevant is that, although these features had some additive impact on survival (often moving patients either into a higher or lower risk category based on dichotomized values), this effect was relatively minor for determining prognostic risk categories compared with that of the 5 major features (see gains in Dxy's and score points shown in supplemental Table 1, which would be added to the basic IPSS-R prognostic score values seen in Table 3). None of these features had additive prognostic impact on the potential for AML transformation.

Our data indicated the importance of performance status as contributing to prognosis for survival in MDS. Other studies have demonstrated the impact of comorbidities on survival in MDS,[5,18-20,37-39] which may in part be reflected by performance status.

The negative impact of elevated serum ferritin levels for survival in our patients may relate to prior RBC transfusions contributing to iron overload and its complications or may reflect the severity of the anemia and degree of ineffective erythropoiesis because of the patients' poorer innate marrow function.[2,12,40] In addition, as serum ferritin is an inflammatory marker and this value was obtained early in our patients' disease courses (ie, before high RBC transfusion burden), this abnormality may reflect the effects of inflammatory cytokines in MDS.[41,42]

Although high serum $\beta_2$-microglobulin levels had significant negative impact on survival in our patients, as this feature was essentially reported from only one institution in our cohort, we have included this as a provisional predictive parameter. In addition, renal dysfunction alters these levels and could confound these results.

Marrow fibrosis did not show incremental prognostic value for clinical outcomes in our study despite previous reports demonstrating poor prognosis of this morphologic feature.[15-17] The absence of this variable as an additive factor could relate to the low number of patients assessed for this feature (19%) as well as the variable ways the degree of fibrosis was reported from the different institutions in our study.

Comparison of the IPSS-R with the IPSS categorizations of our IWG-PM patients indicated that a substantial proportion of patients within the IPSS lower-risk group (27%) would be "upstaged" with the IPSS-R categorization, and 18% in the higher risk group would be "downstaged" by IPSS-R. Such findings have implications for more precisely evaluating patient prognosis and their potential management.

Our data showed that the risk of dying related to leukemic evolution, for our patients observed until death, was increased in those with more advanced prognostic risk categories, Table 7). Thus, mortality from complications of bone marrow failure (without leukemic evolution) and patient comorbidities plays a major role in the clinical outcome of the lower-risk patients in contrast to the more prominent role of leukemic evolution in the higher-risk patients. The proportion of patients dying is lower in our patent cohort compared with the patient group within the original IMRAW patient sample,[1] predominantly because of a lower proportion of our patients dying after AML transformation. This was probably related to the more stringent entry criteria used in our current patient cohort.

Regarding the stringent inclusion criteria in our study, to be more precise with the diagnostic entity of MDS, and as recommended by National Comprehensive Cancer Network practice guidelines for MDS, relative stability of peripheral blood counts for 1-2 months was required to exclude other possible etiologies for their cytopenias, such as drugs, other diseases, or incipient evolution to AML.[43] Exclusion of these patients had minimal influence on the estimates of survival and time to AML evolution (data not shown). In addition, we excluded patients with secondary MDS as their clinical and biologic features (higher degree of AML progression, decreased survival and differing distribution, incidence and types of aberrant and poor risk cytogenetics) distinctively differ from those of primary MDS patients.[44]

An external validation cohort of untreated primary MDS patients from the Medical University of Vienna was evaluated and demonstrated that the IPSS-R model also fit these data well. In particular, the validity of the model for this cohort was indicated by the high prognostic power (ie, Dxy's) and clearly differing temporal medians and hazard ratios between prognostic risk categories for both survival and AML evolution, including age-related survival. Further validation of the IPSS-R in other patient cohorts is warranted.

Some of the patients in the IWG-PM project were also assessed by the WPSS parameters.[2,3] However, because of the relatively low proportion of our patients having several of this system's parameters reported (cellular dysplasia, RBC transfusions), these clinical variables were not included in our analysis. Modification and refinement of the WPSS (WPSS-R) based on the additional features present in the IWG-PM database will be the subject of a separate publication.

In conclusion, the IPSS-R retained continuity with the IPSS and was shown to possess improved prognostic ability for survival and AML evolution compared with the IPSS along with determining additional predictive features, particularly age, having significant impact on survival in primary untreated MDS patients. As such, the IPSS-R should prove beneficial for determination of prognostic status of untreated patients with this disease and aid design and analysis of clinical trials for this disease. Given recent molecular[45-47] and flow cytometric studies[48,49] showing prognostic value in MDS, further investigations to determine the impact of these technologies on the IPSS-R are warranted and ongoing.

## Acknowledgments

The authors thank the staff of the MDS Foundation Inc, for logistical support; Ms Tracey Iraca for helpful logistical assistance for the IWG-PM project; Sherry Pierce, Canan Alhan, Norene Keenan, Ann Hyslop, Michael Groves, Rosa Sapena, Fatiha Chermat, Friedrich Wimazal, and Susanne Herndlhofer for aid in preparing institutional databases; and Barbara Hildebrandt for aid in performing cytogenetic analyses.

Investigators and institutions providing data from the Spanish, French, Piemonte (Italy), and Brazilian MDS Registries are listed in supplemental Table 4.

This work was supported by Celgene Inc and Amgen Inc (unrestricted grants), the Tayside Leukemia Research Endowment Fund for enabling the Tayside Registry (S.T.), Associazione Italiana per la Ricerca sul Cancro (grant 1005) "Special Program Molecular Clinical Oncology 5 × 1000," and Fondazione Cariplo, Milan, Italy (L.M. and M.C.).

The URLs for a Web-based calculator tool to the IPSS-R are located at http://www.ipss-r.com and at http://www.mds-foundation.org/calculator/index.php.

## Authorship

Contribution: P.L.G. designed, performed, and coordinated the research, collected, contributed, analyzed and interpreted the data, and wrote the manuscript; H.T. designed and performed the research, performed the statistical analyses, produced the figures, and edited the manuscript; J.S. collected and contributed data, performed the research, and analyzed and interpreted the data; G.S., G.G.-M., F.S., D.B., P.F., A.L., J.C., O.K., M.L., J.M., S.M.M.M., Y.M., M.P., M.S., W.R.S., R.S., S.T., P.V., T.V., A.A.v.d.L., and U.G. collected and contributed data, analyzed the results and critically revised the paper; J.M.B. collected, contributed, analyzed, and interpreted the data and critically revised the paper; C.F., M.M.L.B., and M.L.S. analyzed and interpreted the data and critically revised the paper; F.D., H.K., A.K., L.M., and M.C. collected and contributed data and reviewed the manuscript; the Cytogenetics Committee members reviewed the cytogenetics data and formulations; and D.H. collected, contributed, analyzed, and interpreted data, designed and performed the research, and edited the manuscript.

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

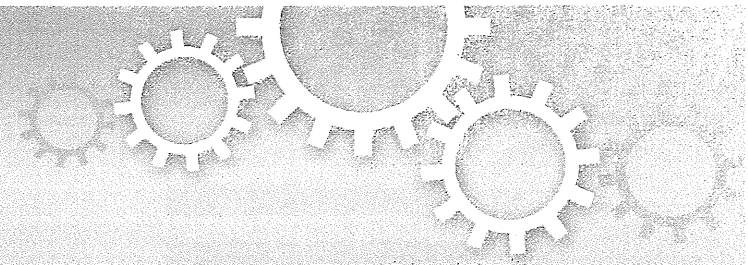
Correspondence: Peter L. Greenberg, Hematology Division, Stanford University Cancer Center, 875 Blake Wilbur Dr, Rm 2335, Stanford, CA 94305-5821; e-mail: peterg@stanford.edu.

## References

1. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood.* 1997;89(6):2079-2088.
2. Malcovati L, Porta M, Pascutto C, et al. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria, a basis for clinical decision making. *J Clin Oncol.* 2005;23(30):7594-7603.
3. Malcovati L, Germing U, Kuendgen A, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. *J Clin Oncol.* 2007;25(23):3503-3510.
4. Kantarjian H, O'Brien S, Ravandi F, et al. Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. *Cancer.* 2008;113(6):1351-1361.
5. Della Porta MG, Luca Malcovati L, Strupp C, et al. Risk stratification based on both disease status and extra-hematologic comorbidities in patients with myelodysplastic syndrome. *Haematologica.* 2011;96(3):441-449.
6. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the WHO classification of myeloid neoplasms and acute leukemia, rationale and important changes. *Blood.* 2009;114(5):937-951.
7. Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol.* 1982;51(2):189-199.
8. Schanz J, Tüchler H, Solé F, et al. New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes and oligoblastic AML following MDS derived from an international database merge. *J Clin Oncol.* 2012;30(8):820-829.
9. Wimazal F, Sperr WR, Kundi M, et al. Prognostic value of lactate dehydrogenase activity in myelodysplastic syndromes. *Leuk Res.* 2001;25(4):287-294.
10. Germing U, Hildebrandt B, Pfeilstocker M, et al. Refinement of the international prognostic scoring system (IPSS) by including LDH as an additional prognostic variable to improve risk assessment in patients with primary myelodysplastic syndromes. *Leukemia.* 2005;19(12):2223-2231.
11. Wimazal F, Sperr WR, Kundi M, et al. Prognostic significance of serial determinations of lactate dehydrogenase (LDH) in the follow-up of patients with myelodysplastic syndromes. *Ann Oncol.* 2008;19(5):970-976.
12. Sanz G, Nomdedeu B, Such E, et al. Independent impact of iron overload and transfusion dependency on survival and leukemic evolution in patients with myelodysplastic syndrome [abstract]. *Blood.* 2008;112(11). Abstract 640.
13. Gatto S, Ball G, Onida F, et al. Contribution of beta-2 microglobulin levels to the prognostic stratification of survival in patients with myelodysplastic syndrome (MDS). *Blood.* 2003;102(5):1622-1625.
14. Neumann F, Gattermann N, Barthelmes HU, et al. Levels of beta 2 microglobulin have a prognostic relevance for patients with myelodysplastic syndrome with regard to survival and the risk of transformation into acute myelogenous leukemia. *Leuk Res.* 2009;33(2):232-236.
15. Verburgh E, Achten R, Maes B, et al. Additional prognostic value of bone marrow histology in patients subclassified according to the International Prognostic Scoring System for myelodysplastic syndromes. *J Clin Oncol.* 2003;21(2):273-282.
16. Buesche G, Teoman H, Wilczak W, et al. Marrow fibrosis predicts early fatal marrow failure in patients with myelodysplastic syndromes. *Leukemia.* 2008;22(2):313-322.

17. Della Porta MG, Malcovati L, Boveri E, et al. Clinical relevance of bone marrow fibrosis and CD34-positive cell clusters in primary myelodysplastic syndromes. *J Clin Oncol.* 2009;27(5):754-762.
18. Wang R, Gross CP, Halene S, Ma X. Comorbidities and survival in a large cohort of patients with newly diagnosed myelodysplastic syndromes. *Leuk Res.* 2009;33(12):1594-1598.
19. Naqvi K, Garcia-Manero G, Sardesai S, et al. Association of comorbidities with overall survival in myelodysplastic syndrome: development of a prognostic model. *J Clin Oncol.* 2011;29(16): 2240-2246.
20. Pfeilstöcker M, Tüchler H, Schönmetzler A, et al. Time changes in predictive power of established and recently proposed clinical, cytogenetic and comorbidity scores for myelodysplastic syndromes. *Leuk Res.* 2012;36(2):132-139.
21. Shaffer LG, Slovak ML, Campbell LJ, eds. *An International System for Human Cytogenetic Nomenclature: Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature.* Basel, Switzerland: Karger; 2009.
22. Grambsch PM, Therneau TM, Fleming TR. Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics.* 1995;51(4):1469-1482.
23. Minder CE, Bednarski T. A robust method for proportional hazards regression. *Stat Med.* 1996; 15(10):1033-1047.
24. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4): 361-387.
25. R Development Core Team, R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing [Internet]. 2009. http://www.r-project.org. Accessed March 1, 2012.
26. Therneau TM, Lumley T. Survival analysis, including penalized likelihood. http://www.r-project.org. Accessed March 1, 2012.
27. Kao JM, McMillan A, Greenberg PL, et al. Impact of cytopenias on clinical outcomes in myelodysplastic syndrome. *Am J Hematol.* 2008;83(10): 765-770.
28. Malcovati L, Della Porta MG, Strupp C, et al. Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS). *Haematologica.* 2011;96(10):1433-1440.
29. Sanz GF, Sanz MA, Vallespi T, et al. Two regression models and a scoring system for predicting survival and planning treatment in myelodysplastic syndromes: a multivariate analysis of prognostic factors in 370 patients. *Blood.* 1989;74(1):395-408.
30. Cordoba I, Gonzalez-Porras JR, Such E, et al. The degree of neutropenia has a prognostic impact in low risk myelodysplastic syndrome. *Leuk Res.* 2012;36(3):287-292.
31. Kantarjian H, Giles F, List A, et al. The incidence and impact of thrombocytopenia in myelodysplastic syndromes. *Cancer.* 2007;109(9):1705-1714.
32. Aul C, Gattermann N, Heyll A, et al. Primary myelodysplastic syndromes: analysis of prognostic factors in 235 patients and proposals for an improved scoring system. *Leukemia.* 1992;6(1):52-59.
33. Gonzalez-Porras JR, Cordoba I, Such E, et al. Spanish Myelodysplastic Syndrome Registry: prognostic impact of severe thrombocytopenia in low-risk myelodysplastic syndrome. *Cancer.* 2011;117(24):5529-5537.
34. Morel P, Declercq C, Hebbar M, et al. Prognostic factors in myelodysplastic syndromes: critical analysis of the impact of age and gender and failure to identify a very-low-risk group using standard mortality ratio techniques. *Br J Haematol.* 1996;94(1):116-119.
35. Kuendgen A, Strupp C, Aivado M, et al. Myelodysplastic syndromes in patients younger than age 50. *J Clin Oncol.* 2006;24(34):5358-5365.
36. Nösslinger T, Tüchler H, Germing U, et al. Prognostic impact of age and gender in 897 untreated patients with primary myelodysplastic syndromes. *Ann Oncol.* 2010;21(1):120-125.
37. Stauder R, Noesslinger T, Pfeilstöcker M, et al. Impact of age and comorbidity in myelodysplastic syndromes. *J Natl Compr Canc Netw.* 2008;6(9): 927-934.
38. Zipperer E, Pelz D, Nachtkamp K, et al. The hematopoietic stem cell transplantation comorbidity index is of prognostic relevance for patients with myelodysplastic syndrome. *Haematologica.* 2009;94(5):729-732.
39. Sperr WR, Wimazal F, Kundi M, et al. Comorbidity as prognostic variable in MDS: comparative evaluation of the HCT-CI and CCI in a core data set of 582 patients of the Austrian MDS Platform. *Ann Oncol.* 2010;21(1):114-119.
40. Park S, Sapena R, Kelaidi C, Ferritin level at diagnosis is not correlated with poorer survival in non RBC transfusion dependent lower risk de novo MDS. *Leuk Res.* 2011;35(11):1530-1533.
41. Deeg HJ, Beckham C, Loken MR, et al. Negative regulators of hemopoiesis and stroma function in patients with myelodysplastic syndrome. *Leuk Lymphoma.* 2000;37(3):405-414.
42. Verma A, List AF. Cytokine targets in the treatment of myelodysplastic syndromes. *Curr Hematol Rep.* 2005;4(6):429-435.
43. Greenberg PL, Attar E, Bennett JM, et al. NCCN Practice Guidelines for in Oncology: myelodysplastic syndromes. *J Natl Compr Canc Netw.* 2011;9(1):30-56.
44. Smith SM, Le Beau MM, Huo D, et al. Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series. *Blood.* 2003;102(1):43-52.
45. Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med.* 2011;364(26):2496-2506.
46. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011;365(15): 1384-1395.
47. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med.* 2012;366(12):1090-1098.
48. van de Loosdrecht AA, Westers TM, Westra AH, et al. Identification of distinct prognostic subgroups in low- and intermediate-1-risk myelodysplastic syndromes by flow cytometry. *Blood.* 2008;111(3):1067-1077.
49. Westers TM, Ireland R, Wolfgang Kern W, et al. Standardization of flow cytometry in myelodysplastic syndromes: a report from an International Consortium and the European LeukemiaNet Working Group. *Leukemia.* 2012;26(7):1730-1741.
50. Greenberg P, Tuechler H, Schanz J, et al. Revised International Prognostic Scoring System (IPSS-R), developed by the International Prognostic Working Group for Prognosis in MDS (IWG-PM) [abstract]. *Leuk Res.* 2011;35(Suppl 1):S6. Abstract 14.

SCIENTIFIC REPORTS



SUBJECT AREAS:
CANCER GENOMICS
HAEMATOLOGICAL CANCER
DNA DAMAGE AND REPAIR
ONCOGENES

Received
2 August 2012

Accepted
10 October 2012

Published
13 November 2012

Correspondence and requests for materials should be addressed to K.S. (shind009@kuhp.kyoto-u.ac.jp)

# APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells

Masanobu Shinohara, Katsuhiro Io, Keisuke Shindo, Masashi Matsui, Takashi Sakamoto, Kohei Tada, Masayuki Kobayashi, Norimitsu Kadowaki & Akifumi Takaori-Kondo

Department of Hematology and Oncology, Graduate school of medicine, Kyoto University, Kyoto 606-8507, Japan.

**Human APOBEC3 proteins play pivotal roles in intracellular defense against viral infection by catalyzing deamination of cytidine residues, leading to base substitutions in viral DNA. Activation-induced cytidine deaminase (AID), another member of the APOBEC family, is capable of editing immunoglobulin (Ig) and non-Ig genes, and aberrant expression of AID leads to tumorigenesis. However, it remains unclear whether APOBEC3 (A3) proteins affect stability of human genome. Here we demonstrate that both A3A and A3B can induce base substitutions into human genome as AID can. A3B is highly expressed in several lymphoma cells and somatic mutations occur in some oncogenes of the cells highly expressing A3B. Furthermore, transfection of *A3B* gene into lymphoma cells induces base substitutions in *cMYC* gene. These data suggest that aberrant expression of A3B can evoke genomic instability by inducing base substitutions into human genome, which might lead to tumorigenesis in human cells.**

It is widely recognized that the accumulation of genetic changes in tumor-related genes is essential for cancer development[1]. With the innovation of high-throughput sequencing technology, genome-wide analyses on various types of cancer cells have revealed numerous somatic mutations in tumor-related genes[2]. Some of these mutations are caused by defects in DNA repair systems (e.g., DNA mismatch repair deficiencies give rise to hereditary non-polyposis colon cancer[3]), whereas mechanisms that account for the majority of genetic changes in cancer cells are poorly understood. Referring to somatic base substitution spectra in cancer cells, C/G to T/A transitions are most prevalent, especially in gastric cancer, colorectal cancer, glioma, and melanoma[2,4,5]. This strong bias in somatic mutations suggests the existence of active mechanisms that induce C/G to T/A transitions into genomic DNA. It is obviously attributable to ultraviolet irradiation and following repair process against pyrimidine dimer in case of melanoma, but not in others.

The human APOBEC family proteins can induce C to T (G to A, in complementary sequences) transitions into target DNA through cytidine deamination. The APOBEC family is comprised of a series of molecules with conserved cytidine deaminase domains (CDAs), including AID, APOBEC1, APOBEC2, APOBEC3A to H, and APOBEC4[6,7]. Among them, AID plays a crucial role in somatic hypermutation and class switch recombination of Ig genes, which enables diversification of immune system[8]. AID has been considered the only molecule that can induce C/G to T/A transitions into genomic DNA. The expression of AID is highly regulated and restricted in germinal center B-cells under physiological conditions, but with inflammatory stimulations, AID can be overexpressed in not only B-cells but also other types of cells (e.g., epithelial cells) via activation of NF-κB[9]. Aberrant expression of AID results in the accumulation of mutations in non-Ig genes[10], which leads to development of various cancers such as gastric and hepatic cancers as well as lymphomas[9,11–13].

A series of seven A3 genes are tandemly arrayed on human chromosome 22, and the main function of the resulting gene products is to protect the cells from retroviruses and endogenous mobile retroelements[14,15]. A3B, A3D, A3F, and A3G contain two CDAs, instead of one in A3A, A3C, and A3H. A3G is a powerful anti-retroviral molecule that induces cytidine deamination in viral genome and acts as a host defensive factor against viruses such as HIV-1[16]. A3A and A3B have been reported as potent inhibitors of retrotransposons[17]. Thus, A3 proteins act as sentinels in innate immunity against mobile DNA/RNA including viruses, while little is known about the effect of these proteins on nuclear DNA, in other words, host human genome. Recent studies have demonstrated that A3A impairs nuclear DNA under the condition of suppressing uracil DNA-glycosylase (UNG) which prevents base alterations by eliminating uracil from DNA and initiating the base-excision repair pathway[18,19].

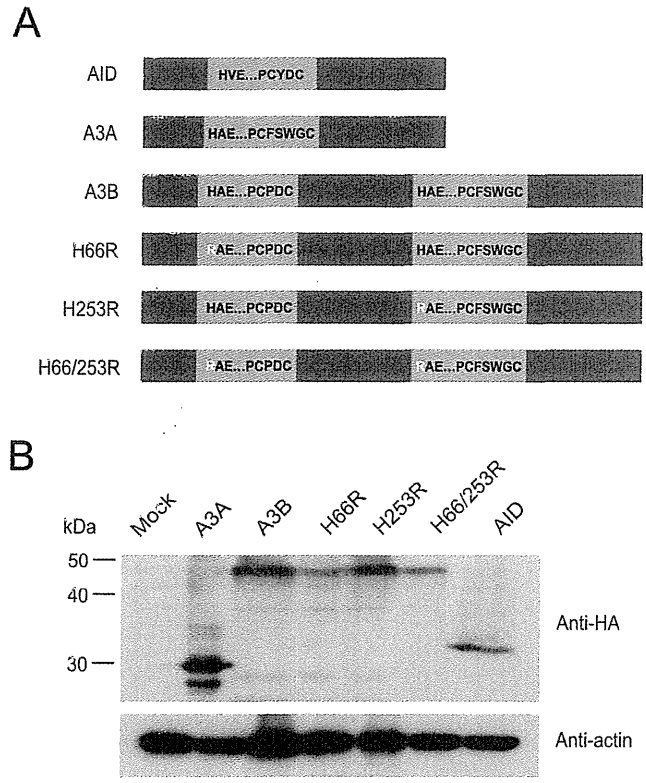



**Figure 1 | Expression of A3A, A3B wild-type and mutants, and AID.** (A) Schematic of expression vectors. The consensus amino acid residues for zinc-coordinating motifs are shown. Substituted residues are shown in white. (B) Expression of HA-tagged proteins. Expression vectors were transfected into HEK293 cells, and cell lysates were analyzed by immunoblotting with anti-HA antibody (top panel) and anti-ß-actin antibody (bottom panel) for loading control.

However, it is still unclear whether A3 proteins can induce somatic mutations into human genome with intact DNA repair systems. Here we first demonstrate that expression of A3B and A3A as well as AID can induce somatic mutations in genomic DNA in human cells even in the presence of UNG. We also find that high expression of A3B leads to somatic mutations in tumor-related genes. These data suggest that aberrant expression of A3B might be one of the active mechanisms that induce somatic mutations in cancer cells.

## Results

**A3 and AID induce hypermutations into foreign DNA.** Besides A3A, we focused on A3B because it is localized predominantly in the nucleus[20,21] and highly expressed in many types of cancer cells[14] referring to microarray database (e.g., NextBio: http://www.nextbio.com). Previous studies have shown that A3B contains two enzymatically active CDAs in restricting HIV-1[22], whereas only carboxyl-terminal CDA is responsible for inhibiting HBV replication[23,24] and editing bacterial DNA[22]. A3B is also shown to restrict foreign DNA in mammalian cells[25], but it has not been tested which CDA is active in this context. First, to examine whether A3 and AID induce mutations in foreign DNA in human cells and which CDA is responsible for this DNA editing, we constructed amino- and/or carboxyl-terminal CDA mutants (H66R, H253R, and H66/253R) by site directed mutagenesis (Fig. 1a) and confirmed their expression in HEK293 cells by immunoblotting (Fig 1b). We transfected expression vectors for these together with EGFP expression vector into HEK293 cells and examined base substitutions in *EGFP* sequences. The expression vector for UNG inhibitor (UGI) was also co-transfected to avoid UNG-triggered degradation of uracil-containing foreign DNA as described previously[25]. We recovered total DNA from the cells 2 days after transfection and performed differential DNA denaturation PCR (3D-PCR) to efficiently recover edited DNA sequences[26]. 3D-PCR is based on the principle that DNA sequences with fewer interstrand hydrogen bonds dissociates easier. If cytidine deamination takes place frequently, resulting AT-rich *EGFP* gene can be amplified at lower denaturation temperatures. Although PCR products were obtained from all samples at 92°C of denaturation temperature (Td), we obtained robust PCR products at 83.8°C of Td only from A3A-, A3B wild-type (WT)-, and AID-expressing cells (Fig 2a). Amplification of *EGFP* at the lowest Td was impaired in H66R-expressing cells compared to A3B WT-expressing cells and undetectable in H253R- or H66/253R-expressing cells (Fig 2a, bottom). To ascertain whether *EGFP* gene was actually hyperedited, we cloned and sequenced the amplicons at 83.8°C of Td. As can be seen from the mutation matrices, high levels of C/G to T/A transitions were introduced into *EGFP* sequences (Fig 2b). To compare the extent of baseline mutations and that of A3B-induced mutations, we also cloned and sequenced the amplicons at 94.0°C of Td. Mutation frequency of A3B-expressing cells were about 6 times higher than that of mock-transfected cells (Supplementary Fig. S1 online). The mutation frequency in H66R-expressing cells was approximately a half compared to that in A3B WT-expressing cells in the amplicons at the lowest Td (Fig 2c). These data suggest that carboxyl-terminal CDA of A3B is mainly responsible for foreign DNA editing, but both domains are requisite for full editing activity. It is worth noting that AID is also capable of inducing cytidine deamination into foreign DNA.

Human A3 proteins have preferred target dinucleotide sequences in the substrate DNA; A3A and A3B prefer to deaminate cytosine residues flanked by 5′ thymine residue, 5′-TC, whereas A3G prefers to deaminate cytosine residues flanked by 5′ cytosine residue, 5′-CC[25,27–29]. We analyzed the context of C/G to T/A transitions in hyperedited *EGFP* sequences. We observed a strong bias toward deamination at 5′-TC dinucleotides in A3A-, A3B WT-, and H66R-expressing cells, but not in AID-expressing cells (Fig 2d). 5′-TC dinucleotide preference of A3B was also confirmed by sequencing amplicon at 94.0°C of Td which is supposed to be unbiased (Supplementary Fig. S1 online). These data suggest that the preference of editing sites in foreign DNA by A3s coincides with that seen in viral DNA.

**A3A and A3B can edit genomic DNA in human cells.** We next investigated whether A3 proteins induce C/G to T/A transitions into not only foreign DNA but also nuclear DNA in human cells. We first established a HEK293 cell line stably expressing EGFP (HEK293/EGFP) using retrovirus vector that carries *EGFP*. We transfected HEK293/EGFP cells with expression vectors for A3A, A3B WT or mutant (H66R, H253R, or H66/253R), or AID by lipofection, and then recovered total DNA from these cells after 7-day culture. We performed 3D-PCR of *EGFP* gene and obtained amplicons from A3A-, A3B WT-, H66R-, and AID-expressing cells at lower Td (Fig 3a). *EGFP* gene was recovered at Td as low as 86.3°C from A3B WT-expressing cells, while as low as 86.5°C from A3A-, H66R-, and AID-expressing cells. By contrast, *EGFP* gene was not amplified below Td of 87°C from cells transfected with mock, H253R or H66/253R. We repeated this procedure consisting of transfection, DNA extraction, and 3D-PCR three times and obtained similar results (Fig 3b). To unambiguously confirm the presence of C/G to T/A transitions, we cloned and sequenced amplicons obtained at the lowest Td in A3A-, A3B WT-, H66R-, and AID-expressing cells (Fig 3c). These analyses revealed approximately 2 to 5 C/G to T/A transitions per *EGFP* sequence from each sample (Fig 3d). The transitions were detected most frequently in A3A-expressing cells, and deaminase activity of H66R mutant was approximately a half

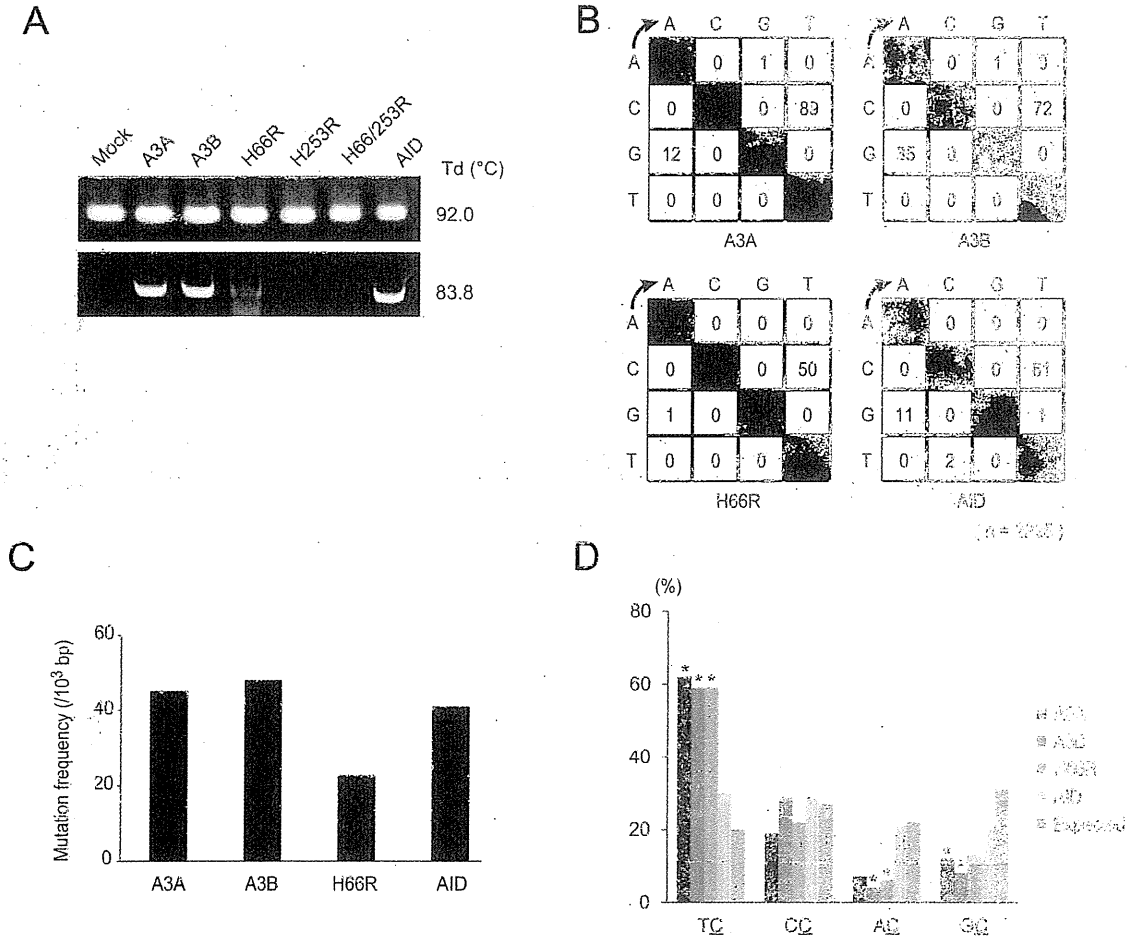



**Figure 2 | Foreign DNA editing by A3A, A3B, and AID.** (A) Agarose gel analyses of 3D-PCR products from HEK293 cells. Cells were transfected with expression vector for A3A, A3B wild-type or mutant, or AID together with pEGFP-N3 and pEF-UGI. Total DNA was recovered 2 days after transfection, and *EGFP* gene was amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Mutation matrices of hyperedited *EGFP* sequences derived from cloned amplicons at 83.8°C of Td. "n" indicates the number of bases sequenced. We sequenced 5 clones (2,225 base pairs in total) in each group. (C) Frequencies of C/G to T/A transitions in hyperedited *EGFP* genes. C/G to T/A transitions per 1,000 sequenced base pairs are shown. (D) Dinucleotide contexts in foreign DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Asterisks indicate statistical significance in a $\chi^2$ test ($p < 0.01$).

compared to that of A3B WT as seen in foreign DNA assays. The contexts of C/G to T/A transitions detected from the lowest Td amplicons in genomic DNA editing in A3-expressing cells were distinct from those in foreign DNA editing (Fig 3e). A preference for 5′-TC dinucleotide was not apparently observed, alternatively, 5′-GC dinucleotides were preferred in all samples. However, this bias fails to reach statistical significance ($p < 0.01$) in a $\chi^2$ test. The preferred target sequences of AID editing were 5′-GC and 5′-AC dinucleotides as described by many prior studies[27,30]. Mutation frequencies and preferred target sequence of A3B was also analyzed by using amplicons at 94.0°C of Td. Mutation frequency of A3B-expressing cells were about 3 times higher than that of mock-transfected cells (Supplementary Fig. S2 online). A preference for 5′-TC dinucleotide was impaired, compared to that in foreign DNA editing assays (Supplementary Fig. S2 online). Our results reveal that in addition to AID, A3A and A3B can induce C/G to T/A transitions into human nuclear DNA without repressing proofreading enzymes (e.g., UNG). Mutation frequencies were 6 to 9 per 1000 base pairs in A3A-, A3B WT-, and AID-expressing cells, and much less frequent compared to those in foreign DNA editing. As seen with foreign DNA editing, carboxyl-terminal CDA is mainly responsible for catalytic activity but not sufficient for full editing activity. The preference context of genomic DNA editing by A3A and A3B is different from that of viral or foreign DNA editing.

**Deep sequencing reveals hyperediting of human genomic DNA by A3 proteins.** Amplicon sequencing by next-generation sequencer has enabled to detect extremely low levels of mutations of targeted regions in genomic DNA. To verify more certainly that A3 proteins edit human nuclear DNA, we performed deep sequencing of A3-expressing cells. HEK293/EGFP cells were transfected with an empty vector or expression vectors for A3A, A3B WT, H66/253R, or AID by lipofection, and total DNA were extracted after 7-day culture. We amplified a portion of *EGFP* gene with 443 base pair length (from thymine 56 to cytosine 498) by conventional PCR protocol, not by 3D-PCR, and performed amplicon sequencing with the coverage of 1337 to 2654 reads per sample. This analysis revealed that extremely large numbers of nucleotides were substituted over the full length of amplicons in A3A-, A3B WT-, and AID-expressing cells, whereas very few mutations were detected in mock and H66/253R-expressing cells (Supplementary Table 1 online). C/G to T/A transitions were observed most frequently in A3A-expressing cells as variation rates reach approximately 7% at the maximum, while below 3% at most in A3B- and AID-expressing cells (Fig 4a

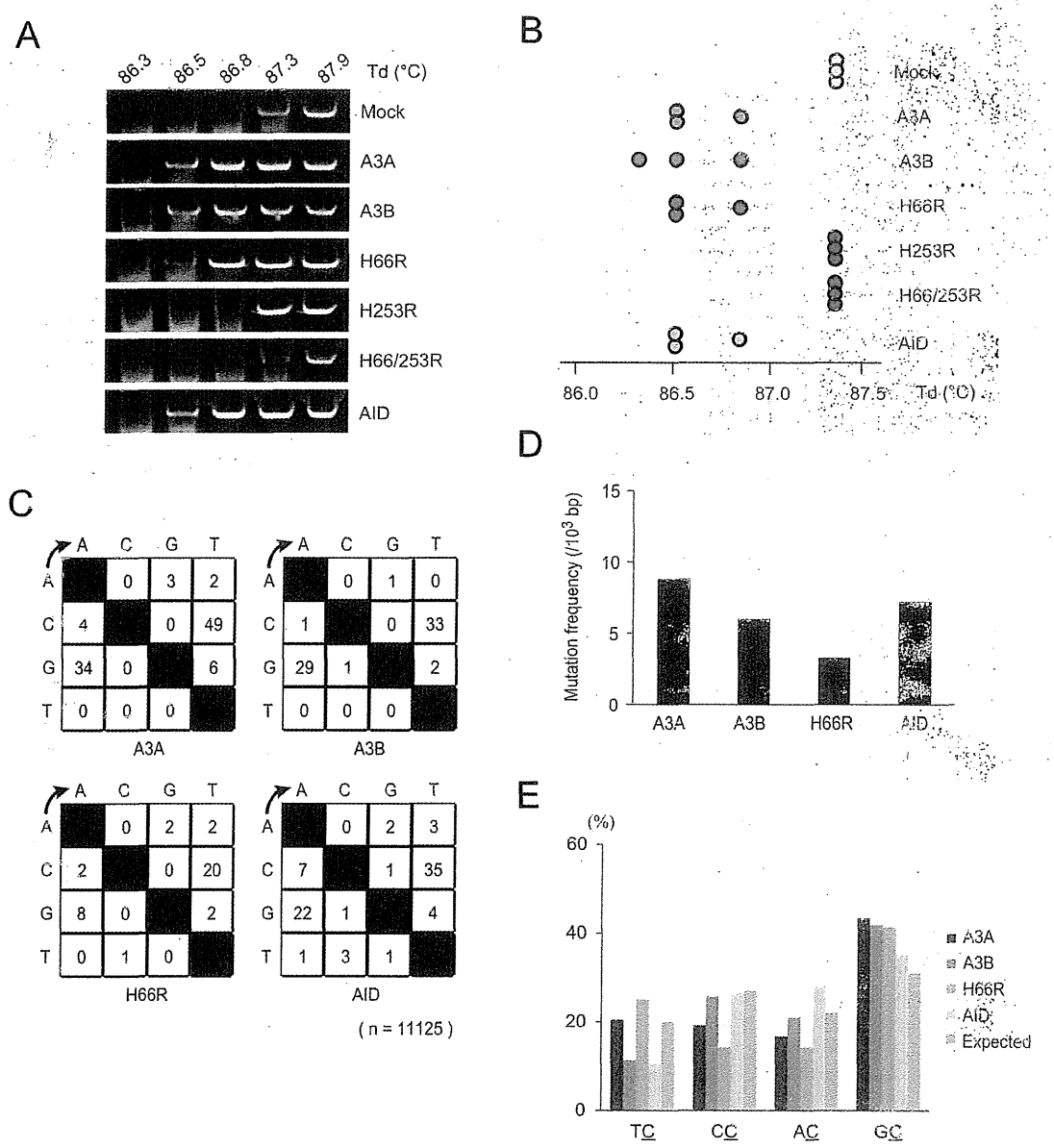



**Figure 3 | Hypermutations in *EGFP* genes integrated in genomic DNA of HEK293 cells.** (A) Agarose gel analyses of 3D-PCR products of *EGFP* genes extracted from HEK293/EGFP cells. Cells were transfected with expression vector for A3A, A3B wild-type or mutant, or AID. Total DNA was recovered 7 days after transfection and *EGFP* genes were amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Distributions of the lowest denaturation temperatures for positive PCR amplification in each sample. Each circle represents independent experiment consisting of transfection, DNA extraction, and 3D-PCR. (C) Mutation matrices of hyperedited *EGFP* sequences derived from cloned PCR products at Td lower than 87°C. "n" indicates the number of bases sequenced. We sequenced 25 clones (11,125 base pairs in total) in each group. (D) Frequencies of C/G to T/A transitions in hyperedited *EGFP* genes. C/G to T/A transitions per 1,000 sequenced base pairs are shown. (E) Dinucleotide contexts in genomic DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Deviations in the editing contexts do not reach statistical significance ($p < 0.01$) in a $\chi^2$ test.

and Supplementary Table 1 online). The mutation frequency analysis revealed that large numbers of C/G to T/A substitutions were induced in A3A-, A3B-, and AID-expressing cells, whereas other types of base substitutions were very few (Fig 4b). These results are similar to the data obtained by 3D-PCR and clonal sequencing of A3- and AID-expressing cells, and further demonstrated that A3A and A3B as well as AID can induce C/G to T/A transitions into genomic DNA in human cells with intact DNA repair systems. Dinucleotide preference of target sequence for deamination by A3A, A3B and AID was also analyzed, however, we did not found any preference in this experiment (Fig. 4C), suggesting the difference between foreign DNA editing and genomic DNA editing.

**Expression of A3B and somatic mutations in lymphoma cells.** Although AID has been reported to play important roles in lymphomagenesis by inducing mutations in both Ig and non-Ig genes[11,12,31–34], AID-independent mechanisms are also suggested, because AID is not expressed in all types of B-cell lymphomas[31,35]. We hypothesized that A3 may contribute to somatic mutations in some lymphoma cells. To examine this hypothesis, we first determined expression levels of *A3A*, *A3B*, and *AID* by quantitative RT-PCR in several B-cell lymphoma cell lines using peripheral blood lymphocytes (PBL) as control (Fig 5a). Our analysis revealed that *A3B* was highly expressed in 3 of 4 cell lines, particularly, markedly high in KIS1 cells, whereas expression of *A3A* transcripts was not detected in any

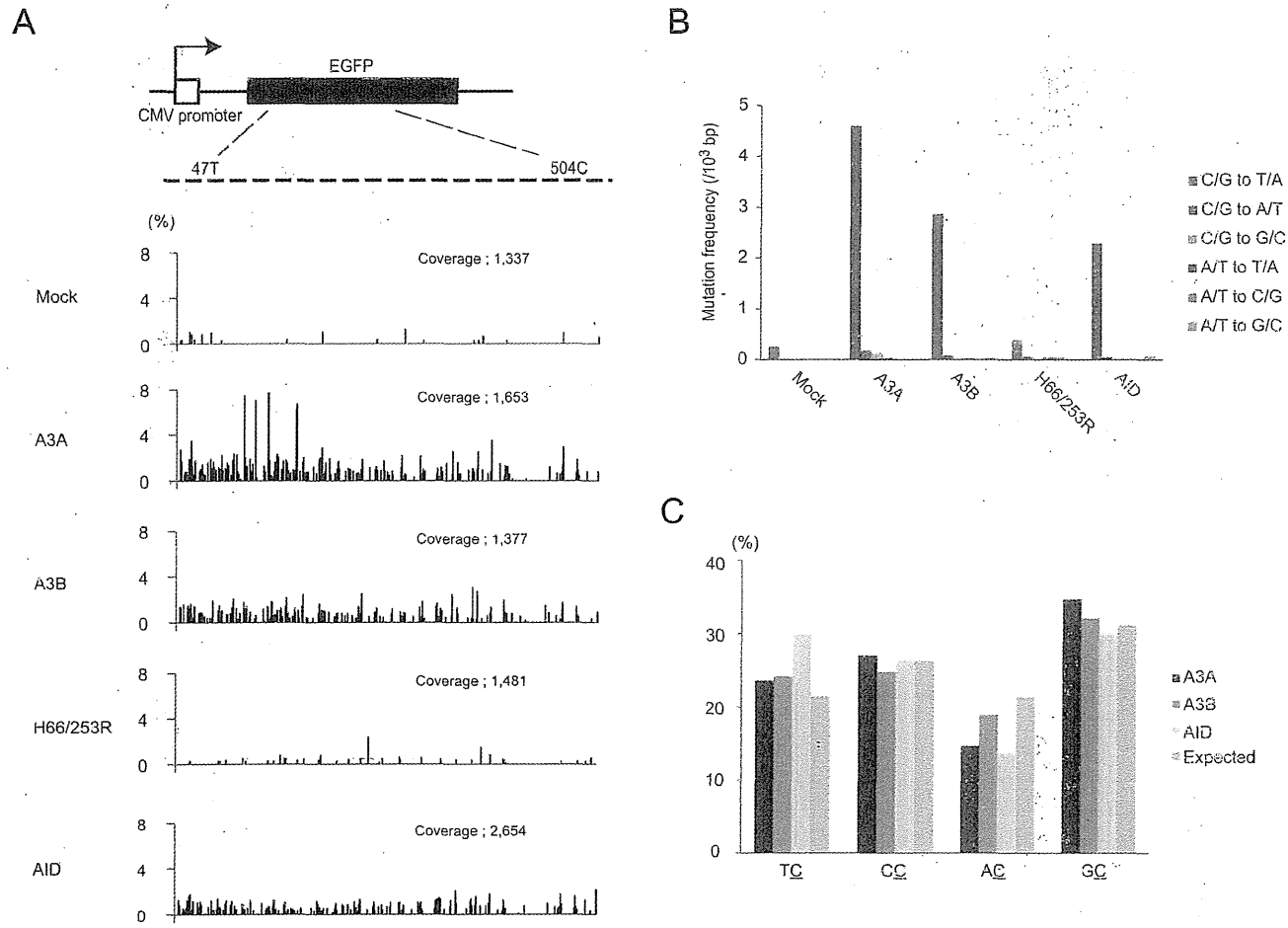



**Figure 4 | Deep sequencing of *EGFP* genes in genomic DNA.** (A) The distributions of C/G to T/A substitutions in the *EGFP* sequences. Total DNA was recovered form HEK293/EGFP cells 7 days after transfection with expression vector for A3A, A3B wild type or H66/253R or AID. We amplified a portion of *EGFP* sequence from thymine 47 to cytidine 504 (top schematic) by PCR with high-fidelity polymerase and sequenced the amplicons by GS-junior bench top system (Roche). Sequence data were analyzed with equipped software. "Coverage" indicates the total numbers of sequenced reads. (B) Frequencies of base substitutions in hyperedited *EGFP* genes. Base substitutions were classified to 6 groups and substituted base number of each group per 1,000 sequenced base pairs are show. (C) Dinucleotide contexts in genomic DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Deviations in the editing contexts do not reach statistical significance ($p < 0.01$) in a $\chi^2$ test.

lymphoma cell lines consistent with prior work suggesting myeloid specificity[25,36]. *AID* transcripts were detected in 2 of 4 cell lines, which is consistent with previous studies[31,32,37]. We also examined expression of *A3B* in two lymph node samples of diffuse large B-cell lymphoma, and found that A3B is actually expressed (supplementary Fig. 3 online).

To investigate the correlation between *A3B* expression and frequency of somatic mutations, we next performed direct sequencing of *cMYC*, *PAX5*, and *A20* genes which are exemplary genes mutated frequently in B-cell lymphoma[33,38]. We compared mutation frequencies of these genes in SUDHL-6 and KIS-1, because the expression of *A3B* was the lowest in the former and the highest in the latter, while AID was not expressed in either cell line. DNA sequences between exon 1 and intron 1 of these three genes were analyzed (899 base pairs of *cMYC*, 1550 base pairs of *Pax5*, and 1088 base pairs of *A20*), since it have been reported that somatic mutations induced by cytidine deaminases were concentrated within 2 kb downstream from transcription initiation sites[33,34]. We found nine mutations within investigated sequences of *cMYC* and *PAX5* in KIS-1, but not in SUDHL-6, in which five of nine mutations detected were C/G to T/A transitions. On the other hand, no mutation was detected within sequenced region of *A20* in either cells (Fig 5b). To analyze ongoing mutations in the genome in individual cells, we next sequenced the same region of *cMYC* sub-cloned from KIS-1 and SUDHL-6, and found several more C to T mutations in KIS-1 cells, but not in SUDHL-6 cells (Supplementary Fig. S4 online). We next determined expression of these tumor-related genes by quantitative RT-PCR and found that the transcripts of *cMYC* and *PAX5* were highly expressed in both SUDHL6 and KIS1 cells as compared to PBL, whereas *A20* was less transcribed in these lymphoma cells. These results suggest that high expression of *A3B* resulted in accumulation of base alterations, especially C/G to T/A transitions, in actively transcribed tumor-related genes in lymphoma cells.

To ascertain more definitely that A3B can edit tumor-related genes in lymphoma cells, we introduced *A3B* into a lymphoma cell line and analyzed somatic mutations in *cMYC*. SUDHL-6 cells were transfected with expression vector for A3B WT, H66/253R, or mock by electroporation and total DNA was extracted after 7-day culture. With 3D-PCR analysis of *cMYC*, we obtained the amplicon from only A3B WT-expressing cells at the lower Td (Fig 6a). Clonal sequencing of amplicons at 85.9°C revealed 2 to 7 nucleotide substitutions per strand, and more than 80% of these mutations were C/G to T/A transitions, with a preference for 5′-GC dinucleotide sites (Fig 6b and c). We also sequenced the amplicons at 94.0°C of Td and

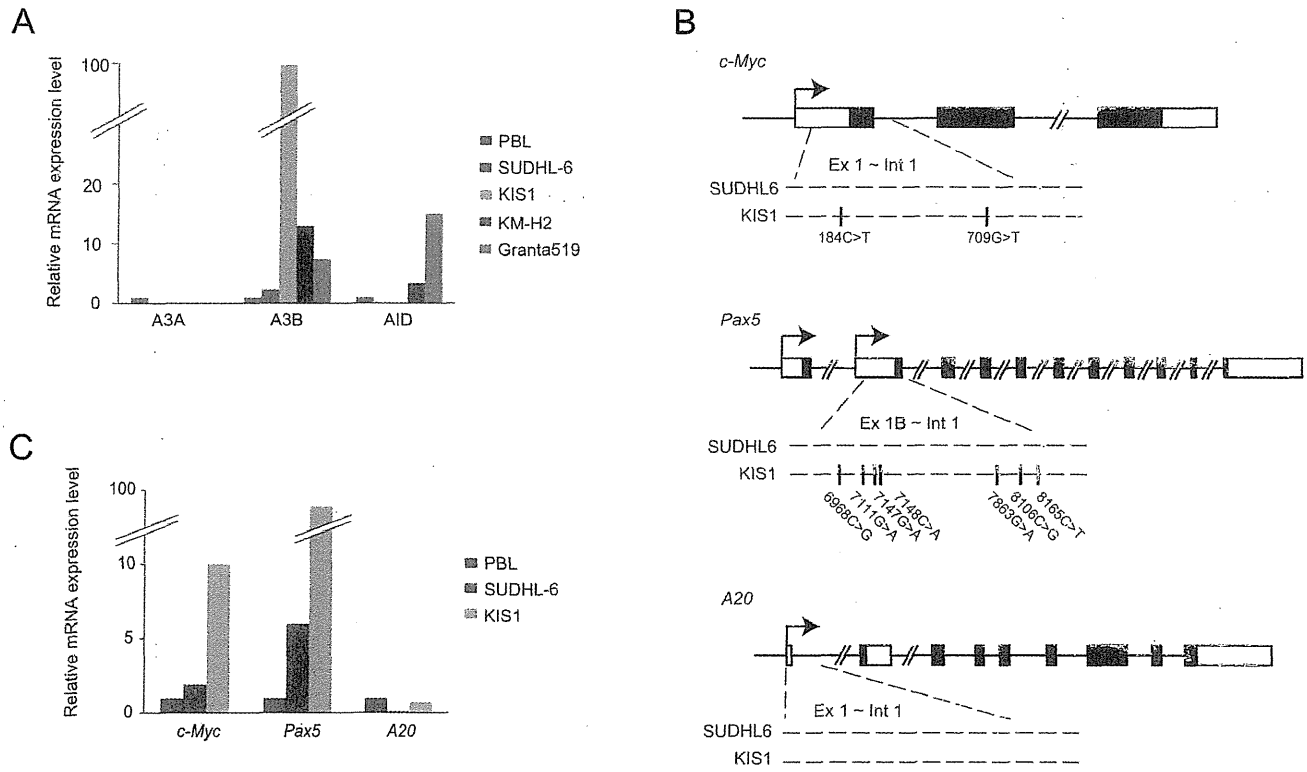



**Figure 5 | Expression of *A3B* and somatic mutations in oncogenes in human lymphoma cell lines.** (A) Quantitative RT-PCR for *A3A*, *A3B*, and *AID* in lymphoma cell lines. The levels of target cDNA were normalized to the endogenous hypoxanthine phosphoribosyl transferase 1 (*HPRT1*) and then compared to those in peripheral blood lymphocytes. (B) Mutational analyses of *C-myc*, *Pax5*, and *A20* in SUDH6 and KIS1 cells. We recovered total DNA from the cells and amplified the sequence between exon1 and intron1 of *C-myc*, *Pax5* and *A20* by PCR and performed direct sequencing of the amplicons. Locations of somatic mutations are shown below the loci with their positions. (C) The expression levels of transcripts of *C-myc*, *Pax5*, and *A20* in KIS1 and SUDHL6 cells. Quantitative RT-PCR was similarly performed with (a).

found A3B-induced C/G to T/A transitions without 5′-TC dinucleotide preference (supplementary Fig. 5 online). These data demonstrate that expression of A3B can induce somatic mutations into actively transcribed tumor-related genes in lymphoma cells.

## Discussion

To date, most studies on A3 proteins have focused on their abilities as antiviral or antitransposon factors, whereas the capability of A3 proteins to induce mutations into genomic DNA in host cells has been scarcely verified. In contrast, many studies have elucidated that AID induces somatic mutations into not only Ig genes, but also tumor-related genes in human cells and that ubiquitous expression of AID in mice leads to cancers of various organs as well as lymphomas, with the accumulation of nucleotides alterations[9–13]. Thus, AID has been considered as the only DNA cytosine deaminase that can induce somatic mutations into human genome and has potential to cause cancers or hematologic malignancies.

Suspène *et al.* have recently reported that hyperediting of both mitochondrial and nuclear DNA was detected in human cells defective for UNG derived from hyper IgM syndrome patients and demonstrated that A3A-induced mutations in nuclear DNA are detectable under the condition of suppressing UNG in human cells[19]. In their report, deamination of nuclear DNA was not observed in cells expressing other A3 proteins or in cells expressing A3A without UNG suppression. Furthermore, Landry *et al.* have reported that expression of A3A together with UGI in mammalian cell lines resulted in breaking of DNA and activation of DNA damage response in a deaminase-dependent manner[18]. In these two reports, the effect on genomic integrity by A3A was dependent on the presence of UGI. Thus, there has been no direct evidence that A3 proteins induce mutations in genomic DNA in the cells with intact DNA repair systems.

In this study, we demonstrate that A3A and A3B as well as AID can induce C/G to T/A transitions into nuclear DNA without suppressing UNG by two different assays, 3D-PCR and deep sequencing. We assume that increased number of cytidine deamination catalyzed by highly expressed A3A or A3B exceeded the processivity of DNA repair enzymes such as UNG and resulted in leaving C/G to T/A transitions in nuclear DNA. Mutation frequencies were considerably lower compared to Suspène's report. However, Yoshikawa *et al.* reported that AID induced hypermutations into an actively transcribed gene in fibroblasts and that the mutation frequency was approximately 4 to 6 per 1000 base pairs[10], almost to the same extent as our results. Hence the frequency of mutations induced by cytidine deaminases into nuclear DNA is probably this extent or less in the cells with intact DNA repair systems.

We also find that *A3B* is highly expressed in several lymphoma cell lines and that the cells expressing high levels of *A3B* actually possesses somatic mutations, especially C/G to T/A transitions, in actively transcribed tumor-related genes. Furthermore, we reveal that introduction of *A3B* into lymphoma cells induces the accumulation of C/G to T/A transitions in *cMYC* gene. This is the first report that suggests the involvement of A3B in inducing somatic mutations of oncogenes in tumor cells. Together with the microarray database of A3B expression in miscellaneous cancer cell lines[14] (NextBio: http://www.nextbio.com), it is possible that A3B may induce somatic mutations into tumor related genes in various types of cancers.

Several questions remain open. First, it remains unclear what is preferred target sequences of A3 proteins in genomic DNA. Because

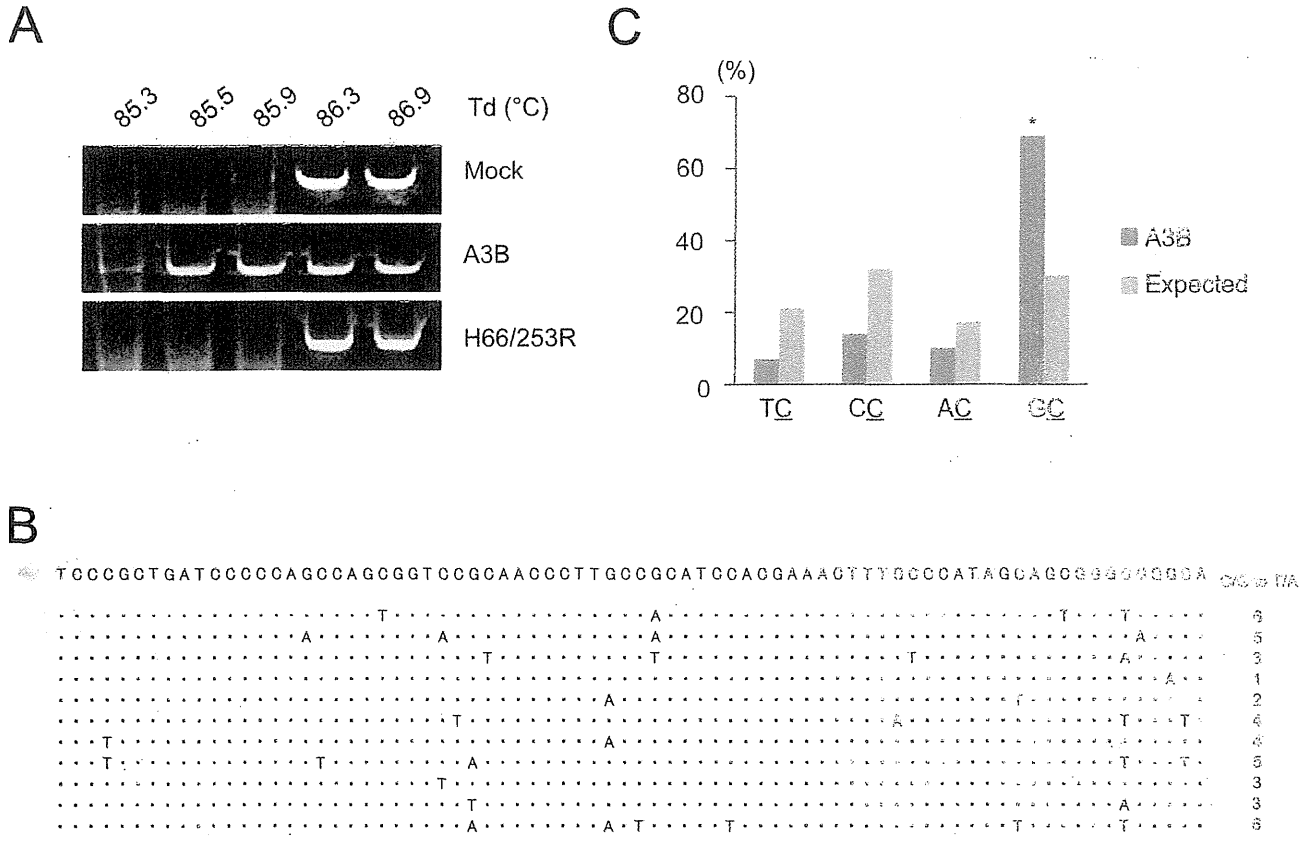



**Figure 6 | A3B induced somatic mutations into *c-myc* gene in human lymphoma cells.** (A) Agarose gel analyses of 3D-PCR products of *c-Myc* genes in SUDHL6. We transfected expression vector for A3B wild-type or H66/253R or empty vector and recovered total DNA 7 days after transfection. *C-myc* genes were amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Clonal sequencing of amplicons from A3B-WT expressing SUDHL6 cells. We sequenced 11 clones (5104 base pairs in total). Seventy six bases from thymine 310 to adenine 385 in which mutations are concentrated among sequenced 464 base pairs are shown. The numbers of C/G to T/A substitutions in sequenced 464 base pair length are shown at the right end. (C) Dinucleotide contexts of somatic mutations in *c-Myc* gene by A3B. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Asterisks indicate statistical significance in a $\chi^2$ test ($p < 0.01$).

in several cancers such as breast cancer and melanoma, 5′-TC is the most prevalent target in C to T base substitutions, A3 is the most potential candidate to induce these mutations[25,28]. However, in our results, neither A3A nor A3B had a definite preference of editing site in nuclear DNA editing, whereas a preference for 5′-TC dinucleotide was observed in foreign DNA editing as previously reported. It is possible that A3A and A3B have no distinct favorite context in nuclear DNA editing, unlike viral and bacterial DNA editing, because human genomic DNA is more profoundly protected in transcription than viral or bacterial DNA and is under survey of DNA repair systems. However, Suspene *et al.* reported that target contexts of cytidine deamination in A3A+UGI-expressing cells were 5′-TC and 5′-CC dinucleotides, which were identical to the contexts of viral or bacterial DNA editing[19]. We assume that this discrepancy might be attributable to cell types or expression levels in cells. Hence, further analyses should be required to clarify the favorite target contexts of A3 proteins in nuclear DNA editing. The second question is how transcriptional control and post-translational modification of A3 proteins regulate A3 activity. Because the molecules that possess a capability of editing nuclear DNA threaten cell homeostasis, expression and activity of A3A and A3B must be strictly controlled. AID is known to be regulated at multiple steps[39], for example, transcriptional regulation[40–42], post-transcriptional regulation by micro-RNA[43,44], regulation of intracellular localization[45,46], and phosphorylation by PKA[47,48]. In contrast to AID, little is known about how A3 proteins are regulated. It has been reported that A3A is abundantly expressed in CD14+ monocytes and upregulated by interferon-α stimulation[25,49,50]. Meanwhile, it is not clear where A3B is expressed normally[25,42,49,50] and how it is regulated. As for post-translational modification, we previously reported that PKA-mediated phosphorylation of A3G regulates the interaction between A3G and HIV Vif[51]. To better understand the physiological roles of A3 proteins, it is important to elucidate how their expression and activity are regulated. The last question is whether A3 proteins can serve as an "initiator" of tumorigenesis. Our results suggest A3B indeed induces somatic mutations into genomic DNA in various human tumor cells, however, it is unclear whether A3B proteins impair genomic DNA from the early stage of oncogenesis. To address this question, hereafter, histopathological and genetic analyses of transgenic mouse constitutively expressing A3 proteins are necessary.

In conclusion, our findings provides the first evidence that A3A and A3B can induce C/G to T/G transitions into genomic DNA without suppressing DNA repair system. Our data also show that high expression of *A3B* is related to mutation frequencies of oncogenes in lymphoma cells. Our results suggest that *A3B* is an oncogene, like AID, which may have the capacity to evoke genomic instability through base substitutions in human cells. Further studies will be required to test whether endogenous A3B is capable of impairing genomic integrity as a DNA mutator and contributing to the development of human cancers and hematologic malignancies.

## Methods

**DNA constructs and cell lines.** Plasmids containing coding sequence of human *A3A* and *A3B* were kindly provided by Dr. Kenzo Tokunaga[21]. Expression vectors for HA-tagged A3A, A3B and AID were generated by sub-cloning of coding sequences into pCAG-GS vector. A3B catalytic domain mutants (H66R, H253R, and H66/253R) were generated by KOD-plus mutagenesis Kit (Toyobo). Expression vector for Uracil-DNA glycosylase inhibitor, pEF-UGI was kindly provided by Dr. Ruben S Harris[25]. HEK293 and HEK293T cells were maintained with Dulbecco's modified Eagle's medium containing 10% of fetal bovine serum (FBS) and penicillin, streptomycin, and glutamine (PSG). All B-cell lymphoma cell lines were maintained with RPMI1640 containing 10% FBS and PSG. Retrovirus containing *EGFP* sequence was produced by co-transfection of pMLV gag-pol, VSV-G, and pDON-EGFP into HEK293T cells. HEK293/EGFP cells were generated by retroviral transduction of *EGFP* and selection of 1 mg/ml G418 for two months.

**Immunoblotting.** HEK293 cells were transfected with expression vector for A3A, A3B wild-type or mutant (H66R, H253R or H66/253R) or AID, and lysed with RIPA buffer (50 mM Tris-HCl pH7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton-X, 0.1% SDS, 0.1% DOC) after 2-day culture. After centrifugation at 20,000 x g for 15 min, supernatant was mixed with sample buffer (Biorad), boiled for 5 minutes, resolved on 12% (w/v) polyacrylamide gel, transferred to PVDF membrane (Immobilon, Millipore), and analyzed by standard immunoblotting procedure with anti-HA monoclonal antibody (12CA5, Roche) or anti-β-actin monoclonal antibody (AC-15, Sigma).

**3D-PCR and clonal sequencing.** For foreign DNA editing assay, HEK293 cells were transfected with pEGFP-N3, pEF-UGI, and expression vector for A3A, A3B WT or mutant, or AID by using Fugene HD (Roche). After two-day culture, total DNA was extracted by using Quick Gene DNA whole blood kit (Fuji Film). First round PCR was performed with primers listed in Supplementary Table S2 using rTaq DNA polymerase (Takara), with the following reaction profile; 30 s at 94°C, 25cycles of 30 s at 94°C, 40 s at 62°C, and 90 s at 72°C followed by 10 min at 72°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, and extracted from the gel using Qiaquick Gel Extraction kit (Qiagen). We used 25 ng of first-round PCR products as template for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen), with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 83–92°C, 60 s at 62°C, 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 83.8°C were cloned into pT7-blue vector (Novagen). For nuclear DNA editing assay, HEK293/EGFP cells were transfected with expression vector for A3A, A3B WT or mutants, or AID using Fugene HD (Roche). Seven days after transfection, we extracted total DNA from these cells with the same method of foreign DNA editing assay. First round PCR was performed using Advantage HF2 polymerase kit (Clontech), with the following reaction profile;1 min at 94°C, 30 cycles of 30 s at 94°C followed by 2 min at 68°C, followed by 3 min at 68°C. We used 25 ng of first-round PCR products for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen) with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 86–89°C, 60 s at 62°C, and 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 86.5°C and 83.8°C were cloned into pT7-blue vector (Novagen). For *c-myc* gene editing assay in lymphoma cells, we transfected SUDHL6 cells with expression vectors for A3B WT or H66/253R by electroporation using Nucleofector (Amaxa) and extracted total DNA from the cells 7 days after transfection. First round PCR and gel extractions of amplicons were performed with the same methods of nuclear DNA editing assay. We used 25 ng of first-round PCR products for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen), with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 85–88°C, 60 s at 62°C, 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 85.3°C were cloned into pT7-blue vector (Novagen) and sequenced using 3130xl Genetic Analyzer (Applied Biosystems).

**Deep sequencing.** Total DNA was extracted from HEK293/EGFP transfected with expression vectors for A3A, A3B WT or H66/253R or AID 7 days after transfection. A portion of *EGFP* gene with 443 base pair length, from thymine 56 to cytosine 498, was amplified with the primers listed in Supplementary Table 2 using Advantage HF2 polymerase kit (Clontech), with the following reaction profile; 1 min at 94°C, 30 cycles of 30 s at 94°C followed by 2 min at 68°C, and followed by 3 min at 68°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, and extracted from the gel using Quiaquick Gel Extraction Kit (Qiagen). Purified amplicons were sequenced using GS junior bench top system (Roche) according to the manufacturer's protocol and analyzed with equipped software, GS Amplicon Variant Analyzer.

**Lymphoma cell lines and patient samples.** Four B-cell lymphoma cell lines (SUDHL6, KIS1, KM-H2, and Granta519) were cultured in RPMI1640 containing 10% of fetal bovine serum (FBS) and penicillin, streptomycin, and glutamine (PSG). We extracted total DNA from these cells by using Quick Gene DNA whole blood kit (Fuji Film) and total RNA by using mir Vana miRNA isolation kit (Ambion). Tumor biopsy specimens prior to treatment were obtained from two patients with diffuse large B-cell lymphoma. The study was approved by the Kyoto University Institutional Review Board and written informed consent was obtained from each patient. Total RNA were extracted similarly to lymphoma cell lines. Naïve B-cells were isolated from healthy donor's peripheral blood by using MACS® naive B cell isolation kit (Miltenyi Biotec).

**Quantitative RT-PCR.** Complementary DNA was synthesized from 200 ng of total RNA using Revertra Ace qPCR RT Master Mix (Toyobo). Real-time PCR were performed with Thunderbird SYBR qPCR Mix (Toyobo) according to manufacturer's protocol. Target cDNAs were normalized to the endogenous expression level of the house keeping reference gene for hypoxanthine-guanine phosphoribosyl transferase 1 (*HPRT1*) or glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). All primers for real-time PCR are listed in Supplementary Table 2.

**Sequencing of oncogenes from lymphoma cell lines.** We amplified portions of *C-myc*, *Pax5*, and *A20* from with the primers listed in Supplementary Table 2 using Advantage HF2 polymerase kit (Clontech), with the following reaction profile; 1 min at 94°C, 30 cycles of 30 s at 94°C followed by 4 min at 68°C, and followed by 3 min at 68°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, extracted from the gel using Qiaquick Gel Extraction kit (Qiagen), and sequenced using 3130xl Genetic Analyzer (Applied Biosystems). In *c-Myc* clonal sequencing, the amplicon was subcloned into pTA2-vector (TOYOBO) and subsequently sequenced.


1. Hahn, W. C. & Weinberg, R. A. Rules for making human tumor cells. *N Engl J Med* **347**, 1593–603 (2002).
2. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–6 (2010).
3. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–61 (1994).
4. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–8 (2007).
5. Prickett, T. D. *et al.* Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nat Genet* **41**, 1127–32 (2009).
6. Macduff, D. A. & Harris, R. S. Directed DNA deamination by AID/APOBEC3 in immunity. *Curr Biol* **16**, R186–9 (2006).
7. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229 (2008).
8. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–63 (2000).
9. Matsumoto, Y. *et al.* Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* **13**, 470–6 (2007).
10. Yoshikawa, K. *et al.* AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**, 2033–6 (2002).
11. Okazaki, I. M. *et al.* Constitutive expression of AID leads to tumorigenesis. *J Exp Med* **197**, 1173–81 (2003).
12. Pasqualucci, L. *et al.* AID is required for germinal center-derived lymphomagenesis. *Nat Genet* **40**, 108–12 (2008).
13. Endo, Y. *et al.* Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* **135**, 889–98, 898 e1-3 (2008).
14. Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–96 (2002).
15. Goila-Gaur, R. & Strebel, K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* **5**, 51 (2008).
16. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
17. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 8780–5 (2006).
18. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep* **12**, 444–50 (2011).
19. Suspène, R. *et al.* Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci U S A* **108**, 4858–63 (2011).
20. Lackey, L. *et al.* APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol* **419**, 301–14 (2012).
21. Kinomoto, M. *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* **35**, 2955–64 (2007).
22. Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486–93 (2007).
23. Suspène, R. *et al.* Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci U S A* **102**, 8321–6 (2005).
24. Bonvin, M. & Greeve, J. Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *J Gen Virol* **88**, 3270–4 (2007).
25. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**, 222–9 (2010).
26. Suspène, R., Henry, M., Guillot, S., Wain-Hobson, S. & Vartanian, J. P. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol* **86**, 125–9 (2005).

27. Vartanian, J. P. *et al.* Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog* **6**, e1000928 (2010).
28. Bishop, K. N. *et al.* Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**, 1392–6 (2004).
29. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–9 (2003).
30. Beale, R. C. *et al.* Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* **337**, 585–96 (2004).
31. Greeve, J. *et al.* Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood* **101**, 3574–80 (2003).
32. Deutsch, A. J. *et al.* MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation. *Blood* **109**, 3500–4 (2007).
33. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–6 (2001).
34. Kotani, A. *et al.* A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc Natl Acad Sci U S A* **102**, 4506–11 (2005).
35. Smit, L. A. *et al.* Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype. *Cancer Res* **63**, 3894–8 (2003).
36. Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol* **16**, 480–5 (2006).
37. Mottok, A., Hansmann, M. L. & Bräuninger, A. Activation induced cytidine deaminase expression in lymphocyte predominant Hodgkin lymphoma. *J Clin Pathol* **58**, 1002–4 (2005).
38. Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712–6 (2009).
39. Delker, R. K., Fugmann, S. D. & Papavasiliou, F. N. A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat Immunol* **10**, 1147–53 (2009).
40. Crouch, E. E. *et al.* Regulation of AID expression in the immune response. *J Exp Med* **204**, 1145–56 (2007).
41. Gonda, H. *et al.* The balance between Pax5 and Id2 activities is the key to AID gene expression. *J Exp Med* **198**, 1427–37 (2003).
42. Pauklin, S., Sernández, I. V., Bachmann, G., Ramiro, A. R. & Petersen-Mahrt, S. K. Estrogen directly activates AID transcription and function. *J Exp Med* **206**, 99–111 (2009).
43. Teng, G. *et al.* MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* **28**, 621–9 (2008).
44. de Yébenes, V. G. *et al.* miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* **205**, 2199–206 (2008).
45. Ito, S. *et al.* Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* **101**, 1975–80 (2004).
46. Patenaude, A. M. *et al.* Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* **16**, 517–27 (2009).
47. Basu, U. *et al.* The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–11 (2005).
48. McBride, K. M. *et al.* Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J Exp Med* **205**, 2585–94 (2008).
49. Koning, F. A. *et al.* Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474–85 (2009).
50. Berger, G. *et al.* APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. *PLoS Pathog* **7**, e1002221 (2011).
51. Shirakawa, K. *et al.* Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat Struct Mol Biol* **15**, 1184–91 (2008).


## Acknowledgments

We thank R. Harris for thoughtful feedback and for UGI vector, and K. Tokunaga for A3A and A3B vectors.

## Author contributions

M.S. performed most of the experiments, analyzed the data and wrote the manuscript; I.K., M.M., T.S. and K.T. performed sequencing of lymphoma cell lines; K.S. and N.K. wrote the manuscript; M.K. established q-PCR; A.T.K. designed the experiments and wrote the manuscript.

## Additional information

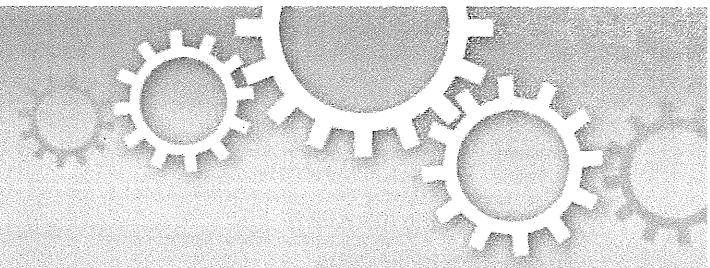
**Supplementary information** accompanies this paper at http://www.nature.com/scientificreports

**Competing interest statements:** The authors declare no competing financial interests.



**How to cite this article:** Shinohara, M. *et al.* APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci. Rep.* 2, 806; DOI:10.1038/srep00806 (2012).

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
STEM CELLS
DIFFERENTIATION
EPIGENETICS
MEDICAL RESEARCH

Received
6 September 2011

Accepted
16 January 2012

Published
17 February 2012

Correspondence and requests for materials should be addressed to H.N. (hnakajim@sc.itc.keio.ac.jp)

# *Tet2* disruption leads to enhanced self-renewal and altered differentiation of fetal liver hematopoietic stem cells

Hiroyoshi Kunimoto[1], Yumi Fukuchi[1], Masatoshi Sakurai[1], Ken Sadahira[1], Yasuo Ikeda[1,2], Shinichiro Okamoto[1] & Hideaki Nakajima[1]

[1]Division of Hematology, Department of Internal Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan, [2]Department of Life Science and Medical Bioscience, Faculty of Science and Engineering, Waseda University, Tokyo 162-8480, Japan.

**Somatic mutation of ten-eleven translocation 2 (*TET2*) gene is frequently found in human myeloid malignancies. Recent reports showed that loss of *Tet2* led to pleiotropic hematopoietic abnormalities including increased competitive repopulating capacity of bone marrow (BM) HSCs and myeloid transformation. However, precise impact of *Tet2* loss on the function of fetal liver (FL) HSCs has not been examined. Here we show that disruption of *Tet2* results in the expansion of Lin$^-$Sca-1$^+$c-Kit$^+$ (LSK) cells in FL. Furthermore, *Tet2* loss led to enhanced self-renewal and long-term repopulating capacity of FL-HSCs in *in vivo* serial transplantation assay. Disruption of *Tet2* in FL also led to altered differentiation of mature blood cells, expansion of common myeloid progenitors and increased resistance for hematopoietic progenitor cells (HPCs) to differentiation stimuli *in vitro*. These results demonstrate that *Tet2* plays a critical role in homeostasis of HSCs and HPCs not only in the BM, but also in FL.**

Myelodysplastic syndrome (MDS) and myeloproliferative neoplasm (MPN) are two major hematopoietic malignancies, thought to arise from hematopoietic stem or progenitor cells. MDS is characterized by dysregulated hematopoietic differentiation with propensity to develop acute myeloid leukemia (AML)[1]. MPN is a spectrum of disease characterized by the expansion of maturing hematopoietic elements with minimal dysplasia. Some of the patients with MPN also progress to myeloid leukemia by acquiring additional mutations altering differentiation. Recent advances in high throughput genome-wide sequencing have made it possible to identify a number of recurrent somatic mutations in human MDS and MPN. One of the most frequently found mutations is a loss-of-function mutation of ten-eleven translocation 2 (*TET2*), which is located at chromosome 4q24 where uniparental disomy was frequently observed in human myeloid malignancies. Recurrent *TET2* mutation was identified in 10–20% of MDS and MPN[2,3], and subsequent studies reported high incidence of *TET2* mutations in secondary AML (20–40%) and chronic myelomonocytic leukemia (CMML) (40–50%)[4,5]. These results underscore the importance of *TET2* in maintaining homeostasis and malignant transformation of hematopoietic system.

DNA methylation is one of the most important epigenetic modifications, and aberrant DNA methylations are hallmark of cancers including AML[6]. It was recently reported that *TET* family proteins could convert 5-methylcytosine (5 mC) to 5-hydroxymethylcytosine (5 hmC)[7–9] and 5 hmC production by *Tet1* is critical for ES cell self-renewal and inner cell mass specification[8,10]. Moreover, *TET2* mutations associated with myeloid malignancies disrupt its catalytic activity, and BM cells from patients with *TET2* mutations contained lower levels of 5 hmC compared to normal controls[9]. On the other hand, mutations of isocitrate dehydrogenase (*IDH*) 1 and *IDH2*, enzymes involved in citrate metabolism, are seen in AML and brain tumors in mutually exclusive manner with *TET2*[11–14]. These mutations inhibit catalytic activity of *TET2* by 2-hydroxyglutarate (2-HG), an oncometabolite generated by mutant *IDH1/2*[14]. TET2 is dependent on alpha-keto-glutarate (α-KG) for its catalytic activity, and 2-HG has been shown to inhibit TET2 by competing with α-KG[15]. These results strongly suggest that impaired 5 hmC generation by mutant TET2 or by mutant IDH1/2 is one of the critical steps in myeloid transformation.

FL-HSCs and adult HSCs differ in several aspects of their phenotypes and functions. For example, FL-HSCs are different from adult HSCs in the expression of CD11b/ Mac-1 and CD34, while some markers such as CD150 and endomucin are commonly expressed in both populations[16,17]. FL-HSCs have higher, more robust capacity to reconstitute hematopoietic compartment in irradiated recipients as compared to adult HSCs[18,19]. Moreover,