Genomic landscape of CD34⁺ hematopoietic cells in myelodysplastic syndrome and gene mutation profiles as prognostic markers

Lan Xu^{a,1}, Zhao-Hui Gu^{a,1}, Yang Li^{a,1}, Jin-Li Zhang^{a,1}, Chun-Kang Chang^{b,1}, Chun-Ming Pan^a, Jing-Yi Shi^a, Yang Shen^a, Bing Chen^a, Yue-Ying Wang^a, Lu Jiang^a, Jing Lu^a, Xin Xu^a, Jue-Ling Tan^a, Yu Chen^a, Sheng-Yue Wang^c, Xiao Li^{b,2}, Zhu Chen^{a,c,2}, and Sai-Juan Chen^{a,2}

^aState Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, and Key Laboratory of Systems Biomedicine of Ministry of Education, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200025, China; ^bDepartment of Hematology, Shanghai No. 6 People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200233, China; and ^cShanghai-Ministry of Science and Technology Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China

Contributed by Zhu Chen, April 26, 2014 (sent for review February 15, 2014)

Myelodysplastic syndrome (MDS) includes a group of diseases characterized by dysplasia of bone marrow myeloid lineages with ineffective hematopoiesis and frequent evolution to acute myeloid leukemia (AML). Whole-genome sequencing was performed in CD34⁺ hematopoietic stem/progenitor cells (HSPCs) from eight cases of refractory anemia with excess blasts (RAEB), the high-risk subtype of MDS. The nucleotide substitution patterns were found similar to those reported in AML, and mutations of 96 protein-coding genes were identified. Clonal architecture analysis revealed the presence of subclones in six of eight cases, whereas mutation detection of CD34⁺ versus CD34⁻ cells revealed heterogeneity of HSPC expansion status. With 39 marker genes belonging to eight functional categories, mutations were analyzed in 196 MDS cases including mostly RAEB ($n = 89$) and refractory cytopenia with multilineage dysplasia (RCMD) ($n = 95$). At least one gene mutation was detected in 91.0% of RAEB, contrary to that in RCMD (55.8%), suggesting a higher mutational burden in the former group. Gene abnormality patterns differed between MDS and AML, with mutations of activated signaling molecules and *NPM1* being rare, whereas those of spliceosome more common, in MDS. Finally, gene mutation profiles also bore prognostic value in terms of overall survival and progression free survival.

prognostic stratification | clonal evolution | gene mutation pattern

Myelodysplastic syndrome (MDS) represents heterogeneous clonal hematopoietic stem cell disorders, which are often seen in elderly patients and characterized by abnormal cell proliferation and differentiation, peripheral blood cytopenia, and risk of progression to acute myeloid leukemia (AML) (1). Evidence has been provided by molecular cytogenetics, gene expression profiles, and xenograft animal studies to suggest that hematopoietic stem/progenitor cells (HSPCs) involved in MDS pathogenesis, like in most AML settings, are enriched in CD34⁺ cell population. Among different subtypes of MDS according to World Health Organization (WHO) nomenclature, refractory anemia with excess blasts (RAEB) is of poor prognosis because a significant percentage of the patients progress to AML; patients with refractory cytopenia with multilineage dysplasia (RCMD) show a longer lifespan with heterogeneous risks, whereas other subtypes are usually considered as intermediate or low risk groups (2, 3). Furthermore, the Revised International Prognostic Scoring System (IPSS-R) incorporates the common karyotype, the depth of cytopenias, and the percentage of blasts in bone marrow (BM) for improved prognostic prediction (4, 5).

Recently, massive parallel sequencing has successively provided an unbiased comprehensive screen to identify genetic alterations in human hematological diseases including AML and MDS (6–12). So far, studies using the whole-exome sequencing

(WES) approach on MDS were reported, and mutations of several functional gene categories including those of RNA-splicing machinery were revealed. On the other hand, these mutations were already used as biomarkers to predict the risk of disease progression and/or poor overall survival (OS) among large cohorts of MDS patients (13–15). However, no investigation has yet been reported at the level of whole genome of HSPCs in MDS, which should be important to identify genomic “scars” related to the pathogenesis, as well as the clonality of those key cells. Moreover, the correlation of different MDS phenotypes, particularly RAEB and RCMD because of their distinct risk degrees, and gene mutation patterns need to be addressed in a more systematic way.

In this work, we have characterized genomic variations in MDS by using whole-genome sequencing (WGS) in CD34⁺ cells among eight RAEB cases. This approach has allowed us to analyze the features of genomic damage, clonal architecture, and survival/growth potential of HSPCs in this disease. We also used a panel of molecular markers for mutation detection in 188 MDS

Significance

Myelodysplastic syndrome (MDS) represents a common hematopoietic disease, often in elderly patients, with heterogeneous clinical phenotypes and complex disease mechanisms. Here, we report on characteristic genome lesions, clonal architecture, and distinct tumor clone expansion patterns in a group of patients with refractory anemia with excess blasts, the MDS subtype with the highest propensity to acute myeloid leukemia. An integrative gene mutation analysis in 196 patients with different MDS subtypes allowed a regulatory network of mutually cooperative or exclusive molecules to be discovered among eight functional categories, whereas the combination of a panel of marker genes of prognostic value with the revised-International Prognostic Scoring System may provide a better stratification system for MDS.

Author contributions: X.L., Z.C., and S.-J.C. designed research; L.X., Y.L., C.-M.P., J.-Y.S., Y.-Y.W., L.J., J.L., X.X., J.-L.T., and Y.C. carried out the experiments; L.X., Z.-H.G., Y.L., J.-L.Z., C.-K.C., B.C., Y.-Y.W., S.-Y.W., Z.C., and S.-J.C. analyzed data; and L.X., Y.L., Y.S., Z.C., and S.-J.C. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The whole-genome sequencing data have been deposited in the the Sequence Read Archive, <http://www.ncbi.nlm.nih.gov/sra> (accession no. [SRP0141438](#)), and the SNP microarray data have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. [GSE57229](#)).

¹L.X., Z.-H.G., Y.L., J.-L.Z., and C.-K.C. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: zchen@stn.sh.cn, lixiao3326@163.com, or sjchen@stn.sh.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1407688111/-DCSupplemental.

cases in an attempt to improve the currently used prognostic system of MDS.

Genome Landscape of CD34⁺ Hematopoietic Cells in RAEB

General Information. High-quality whole-genome DNA sequences were obtained from CD34⁺ HSPCs and control skin samples of eight patients with RAEB (cases A1 to A8; Table 1 and *SI Appendix, Table S1*) through strict WGS and information analysis procedures (*SI Appendix, Tables S2–S4*). WGS was performed to reach a haploid median depth of 32.8× and 31.9× for genomes of the CD34⁺ BM cells and skin tissues, respectively. The number of single-nucleotide variants (SNVs) per genome ranged from 1,166 to 2,972, whereas that of short insertions and deletions (INDELs) ranged from 1,994 to 5,905. No correlation was observed between ages and the number of SNVs or nonsilent mutations (*SI Appendix, Fig. S1 A and B*). We also analyzed the genomic rearrangements (GRs) and discovered in each individual genome, on average, 2.9 DEL of large fragment (≥1,000 bp), 1.4 DEL of small fragment (50–1,000 bp), 5.4 intrachromosome translocations (ITXs), and 12.8 interchromosome translocations (CTXs) (*SI Appendix, Table S5*).

Patterns of Genomic Lesions. In terms of nucleotide substitution, the average proportion of transitions and transversions was 62.3% vs. 37.7%, showing a predominance of transitions (*SI Appendix, Fig. S1C*), which was in agreement with previously reported data in MDS and AML, because the proportion of transitions and transversions were 65.0% vs. 35.0% and 67.7% vs. 32.3%, respectively, in the two disease entities (*SI Appendix, Fig. S2A*). The most prevalent changes were G→A/C→T transitions, followed by A→G/T→C transitions and G→T/C→A transversions (Fig. 1A), which mirrored the pattern seen in AML (*SI Appendix, Fig. S2B*) (16). This situation was different from that in non-small cell lung cancer (NSCLC), where G→T/C→A transversions prevailed (17–19), followed by G→A/C→T and then A→G/T→C, or from UV light-associated skin cancers, where C→T and CC→TT transitions prevailed (20, 21). Triplet base patterns harboring SNVs have also been considered as a genome signature in cancer. Indeed, when the adjacent 5' and 3' bases were taken into account, we found ApCpG→ApTpG was the dominant triplet change in RAEB (Fig. 1C). (The underlined base was the substitution site.) These triplet profiles were similar to the previously reported patterns of the M0 through M5 subtypes of AML (Fig. 1C) (16) but different from those of solid tumors [e.g., the XpCpG→XpTpG (here X could be either G, A, C, or T) transitions prevailing in breast cancer (22) or the XpCpG→XpTpG/XpCpG→XpApG combination pattern commonly detected in NSCLC (23)]. In addition, copy number variations (CNVs) were analyzed with regard to their chromosomal positions (see details in *SI Appendix, Fig. S3*).

Mutations in Coding Genes. Through WGS and Sanger sequencing validation, we identified a total of 105 somatic mutations in 96 protein-coding genes in CD34⁺ cells of eight RAEB cases, including 76 missense, 10 nonsense, 6 splice mutations, and 13

INDELs (*SI Appendix, Fig. S1D* and *Dataset S1*). The average coding sequence mutation number of 13.1 per sample was significantly higher than the previous reports using exome sequencing (7.1 and 7.8 per sample; $P=0.042$ and 0.036 , respectively) in MDS but was significantly lower than those in solid tumors (around 100 up to several hundred). Among these 96 mutated genes, 4 represented recurrent abnormalities, namely, additional sex comb-like 1 (*ASXL1*) (Y591*, V962fs, G658*, and A716fs), stromal antigen 2 (*STAG2*) (K705*, R216*, and W485*), tet methylcytosine dioxygenase 2 (*TET2*) (Q966*, S1898F, and R1179fs), and tumor protein p53 (*TP53*) (L35fs and L330fs) (*SI Appendix, Table S7*). Of note, the two *TET2* abnormalities S1898F and R1179fs were found in the same case (A4). Moreover, GR events were checked for intra- and interchromosomal fusion genes (Fig. 1B). Among 10 such events identified in three cases, including five in-frame and five out-of-frame fusions (*SI Appendix, Table S8*), one involving EWS RNA-binding protein 1 (*EWSR1*) and *ASXL1* genes might be of significance, because it generated two fusion transcripts with the N-terminal 483 or 431 aa of *EWSR1* to the exon 3 or exon 1 of *ASXL1* in a tail-to-tail manner (*SI Appendix, Fig. S4*), which were confirmed by RNAseq in this case (*SI Appendix, Table S9*).

Three microRNAs (miRNAs) mutations were detected, namely, hsa-mir-1296 (chromosome 10: 65132732, T→A) and hsa-mir-302e (chromosome 11: 7256045, T→C) in case A6 and hsa-mir-3687 (chromosome 21: 9826247, C→T) in case A5.

Clonal Evolution of CD34⁺ HSPCs in RAEB

Architecture of Clonality. Previous studies established MDS as clonal diseases by virtue of X-linked polymorphic markers and molecular cytogenetics (24–27). The fact that about half of RAEB patients progressed to AML, whereas the other half develop progressive pancytopenia, suggests that the disease HSPC clones should be highly instable. Here, the variant allele frequency (VAF) based on genome sequences of confident SNVs was used to address the clonal architecture. VAF should be 50% if heterozygous mutations exist in all tumor cells, but it is often lower because of nontumor elements in clinical samples. With the sensitivity of this approach, the peak of greater VAF value should represent the founding clone, whereas the peak(s) with smaller VAF may represent subclone(s), although very small subclone(s) could be under the limit of detection. As shown on Fig. 2A, among eight cases, coexistence of one major clone and a minor subclone was found in six patients (A1 to A6, with VAFs varying from 37.62% to 48.86% for major peaks and those from 15.69% to 19.88% for minor peaks). Intriguingly, examination of mutations present in the founding clones and minor subclones revealed that mutations in 87 of 96 (90.6%) genes, particularly recurrently mutated genes such as *TET2*, *ASXL1*, *STAG2*, and *TP53*, were present in the founding clone, whereas nine of the mutations identified only in minor subclones in four specimens were all mutations without recurrence (*SI Appendix, Table S7*). Given that these founding mutations were detected in CD34⁺ cells, they should reflect the aberrant cellular activities at very high levels of disease involvement along with the hematopoietic cell differentiation/proliferation.

Table 1. Overview of eight RAEB patients analyzed with WGS

Patient	WHO subtype	Age, y	Sex	BM blasts, %	IPSS-R	Gene mutations,* <i>n</i>	GRs, <i>n</i>	Fusions, <i>n</i> (RNAseq)
A1	RAEB-1	43	Male	7	4.5	12	5	0
A2	RAEB-1	53	Male	9	8	11	70	7
A3	RAEB-1	62	Female	8	5	22	5	1
A4	RAEB-1	80	Male	8	5.5	6	5	NA
A5	RAEB-2	48	Male	17	5.5	12	79	0
A6	RAEB-2	61	Female	14	NA	17	30	0
A7	RAEB-2	63	Male	14	6.5	11	0	0
A8	RAEB-2	74	Female	11	5	14	0	NA

NA, not available.

*Nonsilent mutations.

Fig. 1. Genomic features of SNVs and gene fusions in RAEB. (A) Eight cases showed similar proportions of each base substitution class. (B) Circos plot of gene fusions in RAEB cases A2, A3, and A6. Green, gene fusions within the same chromosome, apart from ≤ 500 Kb; blue, gene fusions within the same chromosome, apart from >500 Kb; red, fusion genes between different chromosomes. These fusions were validated by RNAseq. Gray, fusion genes not present in transcripts detected by RNAseq. (C) Genomic heat map of the frequencies of distinct trinucleotide patterns harboring each class of nucleotide substitutions among eight RAEB cases (this work) and AML subtypes. (Original data were from the Cancer Genome Atlas Research Network. For each of the AML-M1 through -M5 subtypes, data were chosen from three cases; for M0 subtype, data were from 2 cases.) Log-transformed values of the ratios are marked in the heat map. The 5' base to each substituted base is shown on the vertical axis and the 3' base on the horizontal axis.

Aberrant Proliferation/Differentiation Models of CD34⁺ Versus CD34⁻ Cells. It is well known that HSPCs are enriched in CD34⁺ cell population, whereas CD34⁻ BM cells contain most precursors and mature hematopoietic cells (28, 29). Taking advantage of sample collection for both cell populations, we used Sanger sequencing to validate the mutated genes in CD34⁺ and CD34⁻ cells, as well as normal skin tissues from eight RAEB cases. The comparison of the heights of peaks of mutated bases and wild-type alleles was performed to generate estimated mutant allele frequencies (MAFs) (Fig. 2B). When the MAFs between CD34⁺ and CD34⁻ cell populations were scrutinized, two situations emerged. In cases A1, A2, and A7, the same high peak heights of mutated base were detected in the two populations, suggesting that abnormal HSPCs could expand to more differentiated elements. In cases A3 to A6 and A8, however, MAFs were higher in CD34⁺ cells, whereas these frequencies ranged from 5% to 30% in CD34⁻ cells (Fig. 2B and Dataset S1). Because gene mutations originating from a diseased HSPC cell should be retained during its proliferation and differentiation, the presence of very low MAFs in CD34⁻ cells suggests that the tumor clone failed to form precursors/differentiated elements. Alternatively, maturing hematopoietic precursor cells carrying mutations may undergo apoptosis, causing ineffective hematopoiesis and leaving room for mature cells derived from the residual normal hematopoietic cells to populate the marrow (Fig. 2B).

Integrative Analysis of Mutation Features Using a Panel of Marker Genes in One Hundred Ninety-Six MDS Cases

Mutation Profiles of Thirty-Nine Marker Genes in Distinct Subtypes of MDS. Analysis of WGS data in eight RAEB test cases, combined with a literature search for recurrent gene abnormalities in MDS, yielded a panel of 39 genes (Dataset S2), chosen for targeted deep sequencing of BM samples from a validation group of 188 patients

using Illumina MiSeq and/or GAIIX platforms. The results of mutation detection among all 196 cases were then pooled to form a dataset including 89 cases of RAEB, 95 cases of RCMD, 6 cases of refractory anemia with ringed sideroblasts (RARS), and 6 cases of refractory cytopenia with unilineage dysplasia (RCUD) (Table 2). In total, 287 mutations in 38 genes were discovered in 145 of 196 cases (Dataset S2), and these mutations could be classified into eight functional categories (16): members of the cohesin complex; DNA modifiers (methylation/demethylation); chromatin modifiers; spliceosome genes; transcription factors; activated signaling molecules; tumor suppressors; and AML gene nucleophosmin (*NPM1*), as well as other myeloid disease genes (Circos graph in SI Appendix, Fig. S5 and Table S10). The average number of mutations in RAEB, RCMD, RARS, and RCUD patients were, respectively, 1.99 (177/89), 0.97 (92/95), 1.67 (10/6), and 1.33 (8/6) ($P < 0.001$ between RAEB and RCMD). When the gene mutation frequencies of RAEB and RCMD were compared, 81 of 89 cases of RAEB (91.0%) had gene mutations, whereas only 53 of 95 RCMD (55.8%) exhibited mutations ($P < 0.001$), in the panel of 39 genes. This difference in mutational burdens corresponded to statistically distinct survival outcomes between RAEB and RCMD (3-y OS: $33.0 \pm 9.8\%$ vs. $82.7 \pm 6.1\%$; $P < 0.001$) (SI Appendix, Table S11). Moreover, gene mutation frequencies in three functional categories, including spliceosome genes, transcription factors, and tumor suppressors, were statistically higher in RAEB than in RCMD (SI Appendix, Table S10). As previously reported, all of the six RARS patients in this series bore *SF3B1* gene mutations (12, 30).

In 51 cases without detectable mutations of the marker gene set, 16 had cytogenetic abnormalities. Of note, cases with *TP53* mutations were found to be associated with complex chromosomal abnormalities ($P < 0.001$). When the molecular/cytogenetic data were combined, we found at least one genetic abnormality in 161 of 196 cases (82.1%) of MDS.

Cooperative or Exclusive Relationships Among Distinct Gene Categories. The landscape of somatic alterations in MDS revealed several known and unknown associations of gene mutations, and mutual exclusion of other genetic alterations. For example, each of the following four gene mutation pairs appeared in at least three cases: cohesin family gene *STAG2* with spliceosome gene serine/arginine-rich splicing factor 2 (*SRSF2*) ($P = 0.028$); *U2AF1* with runt-related transcription factor 1 (*RUNX1*) ($P = 0.005$); and chromatin modifier *ASXL1* with *STAG2* ($P = 0.001$) and *ASXL1* with *SRSF2* ($P = 0.01$); the first two pairs are previously unreported (Fig. 3). Moreover, mutual exclusivity could exist among genes of the same functional categories, because no concurrent presence of gene mutations of cohesin family was detected (Fig. 3). Mutual exclusion of gene mutations was also observed among different epigenetic modifiers, such as between *TET2* and *DNMT3A*, between *TET2* and isocitrate dehydrogenases 1/2 (*IDH1/IDH2*), or a previously unidentified mutual exclusion between *ASXL1* and *TP53* (Fig. 3).

Comparison of Gene Mutation Patterns Between MDS and AML. In view of the propensity of MDS to progress to AML, we compared gene mutation profiles of the two diseases using data from three recently reported large genotyping studies of AML as a reference (16, 31, 32). We found that cohesin family genes, epigenetic modifiers including those for DNA and histones, and tumor suppressor genes were mutated at a comparable rate in MDS and AML (SI Appendix, Table S10). However, a striking difference of involvement of activated signaling molecules emerged. In AML, *KRAS/NRAS*, *FLT3*, proto-oncogene C-Kit (*KIT*), *PTPN11*, *JAK2*, etc. anomalies accounted for 49.8% of all gene mutations, whereas they occurred in only 14.3% of MDS (SI Appendix, Fig. S6 and Table S10). Of particular note, there were no cases positive for *KIT* mutation in MDS. By contrast, the mutation rate of spliceosome genes reached 35.2% in MDS, compared with 5.5% in AML, whereas the frequency of transcription factors was lower in MDS than in AML patients (14.3% vs. 22.8%). Moreover, *NPM1*

Fig. 2. Clonality analysis and distinct models of HSPC expansion defect in RAEB. (A) VAF of SNVs in eight RAEB cases. In each plot, the density curve depicts the clustered VAF to determine the number of clusters. One major clone and one subclone were found in patients A1 to A6 (with VAFs varying from 37.62% to 48.86% for major peaks and those from 15.69% to 19.88% for minor peaks), whereas only one major clone was found in patients A7 and A8. (B, Left) HSPC clonal expansion status in RAEB patients. In cases A1, A2, and A7 (Upper), the same peak heights of mutated base were detected in the two populations. In case A3 to A6 and A8 (Lower), mutations were detected with high intensity in CD34⁺ but weakly or not detectable in CD34⁻ cells. Estimated mutant allele frequency: +, between 1% and 15%; ++, between 16% and 30%; +++, between 31% and 50%. (B, Right) It has been suggested that in addition to the abnormalities of eight gene categories detected in MDS, mutations involving activated signaling molecules (*FLT3*, *KIT*) or *IDH1/IDH2* or *NPM1* genes could contribute to the progression to AML.

was mutated at a high rate (24.2%) in AML, whereas this gene was affected in only 2.6% of cases of MDS.

Prognostic Impact of Gene Mutations in One Hundred Ninety-Six MDS Patients. Using the IPSS-R system, our patients could be classified into five subgroups; however, the survival difference between the low and intermediate subgroups was not obvious (Fig. 4A and SI Appendix, Table S11). We then stratified the patients for the presence of mutation of 21 genes with a mutation frequency $\geq 2.5\%$ among the 39 gene markers. Notably, patients with two or more mutations had a median survival of less than 2 y, patients with one mutation had a 70% survival rate at 40 mo, and patients with no mutations had a 90% survival at 40 mo (Fig. 4B and SI Appendix, Fig. S7).

We next evaluated the prognostic importance of specific gene mutations. Univariate analyses were performed to determine the significance of specific gene mutations. Mutations of *STAG2* ($P < 0.001$), cohesin gene family ($P = 0.016$), *IDH1/IDH2* ($P = 0.001$), enhancer of zeste homolog 2 (*EZH2*) ($P = 0.043$), *RUNX1* ($P = 0.026$), and *TP53* ($P = 0.016$) were found to be adverse prognostic factors for OS (SI Appendix, Table S12). Of note, contrarily to some previous reports mostly from Western population, we did not find an inferior prognostic value of the *ASXL1* mutation in univariate analysis ($P = 0.659$), and a similar situation was reported in two recent studies from China (33, 34). These observations might suggest ethnic differences in terms of mutation profiles and their clinical relevance, although larger cohort study is required to clarify the issue. In multivariate analyses, *IDH1/IDH2* [hazard ratio (HR), 6.67; 95% confidence interval (CI), 2.28–19.5; $P = 0.001$], *EZH2* (HR, 8.23; 95% CI, 2.43–27.9; $P = 0.001$), *RUNX1* (HR, 4.59; 95% CI, 1.89–11.2; $P = 0.001$), and

TP53 (HR, 4.35; 95% CI, 1.85–10.2; $P = 0.001$) were independent adverse prognostic factors for OS (SI Appendix, Table S13). We therefore defined *IDH1/IDH2*, *EZH2*, *RUNX1*, and *TP53* mutations as a panel of high-risk factors that could exert a stronger

Table 2. Clinical characteristics of 196 MDS patients

Characteristic	Value
Median age, y	56
Male sex, n (%)	109 (55.6)
Median BM blasts, %	3.75
Median ANC, $\times 10^9/L$	1.06
Median Hb, g/L	74.5
Median PLT, $\times 10^9/L$	52
WHO subtypes, n (%)	196
RAEB-1	44 (22.4)
RAEB-2	45 (23.0)
RCMD	95 (48.5)
RARS	6 (3.1)
RA	3 (1.5)
RT	3 (1.5)
IPSS-R risk group, n (%)	192*
Very low	4 (2.1)
Low	34 (17.7)
Intermediate	66 (34.4)
High	52 (27.1)
Very High	36 (18.8)

ANC, absolute neutrophil count; Hb, hemoglobin; PLT, platelets.

*Cytogenetic analyses were failed in four patients.

Fig. 3. Mutations of functional gene categories in MDS. Distribution of gene mutations in 145 MDS patients with at least one identified mutation of 38 marker genes. Cytogenetic risk categories are listed below mutation distribution according to IPSS-R.

adverse impact on survival of MDS than other molecular events. Of 287 gene mutations detected among 196 patients, 51 mutations observed in 49 cases fell into the high-risk panel (*SI Appendix, Table S16*).

Based on the above findings, we constructed a molecular marker-based system for risk stratification: low, no mutation; intermediate-1, the presence of one mutation; intermediate-2, the presence of two or more mutations; and high, the presence of any high risk gene mutations. Using this system, 196 patients could be divided into four risk groups with significantly different 3-y OS rates ($P < 0.001$) (Fig. 4A and *SI Appendix, Table S16*).

To further optimize the prognostic stratification, we integrated this molecular marker-based system and the clinical/hematological parameters from IPSS-R to form a new IPSS-R-molecular marker (M) system. In the latter model, five prognostic subgroups with reasonably distributed 3-y OS curves emerged (Fig. 4B and *SI Appendix, Tables S14 and S15*).

Discussion

In this work, we found that similar genomic signatures, such as mono- or triplet base substitution patterns, existed between RAEB and AML, in support of the view that the two diseases might share common etiologic factors and be affected by similar DNA damage. Meanwhile, although the molecular mechanisms of both diseases require joint actions of transcription/epigenetic factors, cohesin molecules, and tumor suppressors, the lack of *NPM1* and activated signaling molecules and the presence of spliceosome defect could underlie a limited expansion capacity of abnormal HSPC clones in RAEB, leading to ineffective hematopoiesis. Hence, the molecular abnormalities that distinguish RAEB and AML may mainly reside on the functional gene categories hit by mutations. Our study provides genomic evidence for MDS, RAEB in particular, to be a “preleukemia” disease status.

Because MDS has long been considered an oligoclonal hematopoietic tumor, analysis of its clonal architecture by means of WGS can shed new light on the pathogenesis. Indeed, we detected two distinct clones in six out of eight RAEB cases by virtue of VAF analysis of SNVs. The finding that all recurrent gene mutations were detected in the founding clones suggests that these are “driver” mutations and that most RAEB cases are characterized by HSPC proliferation originated from one major clone. Because HSPCs are enriched in $CD34^+$ cell populations, whereas $CD34^-$ populations contain a wide range of BM cells encompassing most precursors and mature hematopoietic cells,

we used MAFs in these two populations to address tumor clone expansion potential. Interestingly, in some cases, gene mutations were detected with high intensity in $CD34^+$ cells, but the estimated MAFs were much lower or even not detectable in $CD34^-$ cells, suggesting that the expansion defect might be situated at the level of HSPCs and early precursors. In other cases, the same MAFs were detected in both $CD34^+$ and $CD34^-$ cells, indicating the abnormal HSPCs should be able to grow and differentiate toward more mature stages and ineffective hematopoiesis might happen at a later stage of differentiation. This mode may explain the two major outcomes of RAEB, ineffective hematopoiesis leading to progressive pancytopenia or propensity to progress to AML once additional gene mutations take place. It remains under intense investigation that how mutations characteristic of RAEB lead to differentiation failure and ineffective hematopoiesis. Serial analysis of RAEB patients will be required to understand which molecular events lead to progression to AML.

In large series of MDS, the main subtypes are RAEB and RCMD (35, 36), and these diseases have many genes mutated in common. Indeed, screening using a panel of 39 recurrently mutated genes showed that the mutational burden of RAEB was heavier than RCMD, which might explain the more benign clinical course of the latter. From the prognostic point of view, what is

Fig. 4. Kaplan-Meier estimates of OS using the molecular marker (M) or IPSS-R-M-based stratification systems. (A) Kaplan-Meier estimates of OS according to M-based system; 3-y OS rates for low, intermediate-1, intermediate-2, and high-risk cases were $93.2 \pm 3.3\%$, $69.4 \pm 11.2\%$, $53.0 \pm 13.9\%$, and $49.7 \pm 8.3\%$, respectively ($P < 0.001$). n, number of cases. (B) OS according to the IPSS-R-M system; 3-y OS rates for very low, low, intermediate, high, and very high subgroups were 100%, $89.5 \pm 5.9\%$, $83.7 \pm 5.7\%$, $64.9 \pm 9.7\%$, and $41.7 \pm 8.4\%$, respectively ($P < 0.001$).

certain is that mutational burden profoundly affects the clinical behavior and prognosis of MDS. Specifically, we found that the more mutations of a panel of 21 recurrently mutated genes carried by a patient, the worse the clinical outcome. Furthermore five specific genes (*IDH1/IDH2*, *EZH2*, *RUNX1*, and *TP53*), when mutated, were independent markers of risk. We therefore established a new IPSS-R-M system, combining clinical parameters and molecular information, to better stratify prognostic groups in MDS. This system was particularly effective in distinguishing patients in the intermediate category into higher and lower risk groups. This new system may allow more refined therapeutic interventions to be developed in the future, although prospective investigation in another patient cohort is warranted.

Materials and Methods

Patients and Samples. BM samples were obtained by aspiration from eight RAEB patients with informed consent (detailed histories for these patients are provided in *SI Appendix, Table S1*). Skin biopsy was obtained as normal tissue. A cohort of 188 newly diagnosed patients with various MDS subtypes was also used for integrative gene mutation analysis (*SI Appendix, SI Methods*).

DNA Library Preparation and Massively Parallel Sequencing. Libraries were prepared using genomic DNA from CD34⁺ BM cells and skin biopsies. Paired-end sequencing was performed on the Illumina GAIIx and/or HiSeq2000 platform following the manufacturer's standard protocol. Among the 33-Mb coding regions according to the RefSeq database, 94.7% on average were covered with sufficient depth ($\geq 10\times$) (*SI Appendix, Table S2 and SI Methods*).

High-Throughput Sequencing Data Analysis. We developed a pipeline to identify somatic mutations including SNVs and short INDELs by comparing variants

identified in dataset of CD34⁺ BM cells against germ-line sequence in skin samples and dbSNP135 (*SI Appendix, Tables S3 and S4 and SI Methods*).

Targeted Gene Resequencing. Thirty-nine marker genes were screened for mutation detection in 188 MDS patients (Table 2) with the Fluidigm Access Array microfluidic platform, followed by sequencing on Illumina GAIIx and/or MiSeq platform (*SI Appendix, SI Methods*).

Somatic CNV and Uniparental Disomy Detection. To identify somatic CNVs and uniparental disomies, the genomic DNA of the seven of eight paired whole-genome sequenced RAEB patients were genotyped using Illumina high-density Genome Wide Human 660W Quad_v1 SNPs array according to the manufacturer's protocol (*SI Appendix, SI Methods*).

Clonality Analysis. Clonality analysis was performed according to a previous report (37).

Statistical Analysis. Student *t* test, Fisher's exact test, χ^2 test, Kaplan–Meier analysis, and Cox models were used to analyze the clinical and genetic data of 196 MDS cases. The statistical analysis was performed with the statistical software SPSS 19.0 (SPSS) (*SI Appendix, SI Methods*).

ACKNOWLEDGMENTS. We thank G.-B. Zhou for constructive discussions; S.-M. Xiong and X.-Q. Weng for morphology and flow cytometry analysis; and X.-D. Gao, F. Xue, S.-C. Gu, C.-M. Fei, and J. Guo for sample collection. This work was supported by Chinese National Key Basic Research Project 973 Grant 2013CB966800, National High Tech Program for Biotechnology Grant 863 2012AA02A505, Ministry of Health Grant 201202003, Mega-projects of Scientific Research for the 12th Five-Year Plan Grant 2013ZX09303302, State Key Laboratories Project of Excellence Grant 81123005, and the Samuel Waxman Cancer Research Foundation Co-Principal Investigator Program.

- Tefferi A, Vardiman JW (2009) Myelodysplastic syndromes. *N Engl J Med* 361(19):1872–1885.
- Greenberg PL, Young NS, Gattermann N (2002) Myelodysplastic syndromes. *Hematology (Am Soc Hematol Educ Program)* 136–161.
- Vardiman JW, et al. (2009) The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: Rationale and important changes. *Blood* 114(5):937–951.
- Greenberg P, et al. (1997) International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 89(6):2079–2088.
- Greenberg PL, et al. (2012) Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* 120(12):2454–2465.
- Ley TJ, et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456(7218):66–72.
- Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11(10):685–696.
- Mardis ER, et al. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 361(11):1058–1066.
- Yan X-J, et al. (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 43(4):309–315.
- Ley TJ, et al. (2010) DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 363(25):2424–2433.
- Yoshida K, et al. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478(7367):64–69.
- Papaemmanuil E, et al.; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 365(15):1384–1395.
- Graubert TA, et al. (2012) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* 44(1):53–57.
- Makishima H, et al. (2012) Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* 119(14):3203–3210.
- Bejar R, et al. (2012) Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *J Clin Oncol* 30(27):3376–3382.
- Ley T, et al.; Cancer Genome Atlas Research Network (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22):2059–2074.
- Pleasant ED, et al. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463(7278):184–190.
- Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489(7417):519–525.
- Lawrence MS, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.
- Pfeifer GP, You Y-H, Besaratinia A (2005) Mutations induced by ultraviolet light. *Mutat Res* 571(1–2):19–31.
- Krauthammer M, et al. (2012) Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet* 44(9):1006–1014.
- Nik-Zainal S, et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979–993.
- Imielinski M, et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150(6):1107–1120.
- Mongkolsritragoon W, Letendre L, Li CY (1998) Multiple lymphoid nodules in bone marrow have the same clonality as underlying myelodysplastic syndrome recognized with fluorescent in situ hybridization technique. *Am J Hematol* 59(3):252–257.
- Janssen JW, et al. (1989) Clonal analysis of myelodysplastic syndromes: Evidence of multipotent stem cell origin. *Blood* 73(1):248–254.
- Boultonwood J, Wainscoat JS (2001) Clonality in the myelodysplastic syndromes. *Int J Hematol* 73(4):411–415.
- Gerritsen WR, et al. (1992) Clonal analysis of myelodysplastic syndrome: Monosomy 7 is expressed in the myeloid lineage, but not in the lymphoid lineage as detected by fluorescent in situ hybridization. *Blood* 80(1):217–224.
- Bonnet D (2003) Biology of human bone marrow stem cells. *Clin Exp Med* 3(3):140–149.
- Krause DS, Fackler MJ, Civin CI, May WS (1996) CD34: Structure, biology, and clinical utility. *Blood* 87(1):1–13.
- Malcovati L, et al.; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium and of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative (2011) Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* 118(24):6239–6246.
- Shen Y, et al. (2011) Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood* 118(20):5593–5603.
- Patel JP, et al. (2012) Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 366(12):1079–1089.
- Wang J, et al. (2013) TET2, ASXL1 and EZH2 mutations in Chinese with myelodysplastic syndromes. *Leuk Res* 37(3):305–311.
- Shih L-Y, et al. (2013) Clonal leukemic evolution in myelodysplastic syndromes with ASXL1 and EZH2 mutations: A comparative analysis of 58 paired samples. *Blood* 122(21):2786.
- Malcovati L, et al. (2007) Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. *J Clin Oncol* 25(23):3503–3510.
- Haferlach T, et al. (2014) Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28(2):241–247.
- Welch JS, et al. (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150(2):264–278.