Published in final edited form as:

Nat Genet. 2017 February ; 49(2): 204–212. doi:10.1038/ng.3742.

Dynamics of clonal evolution in myelodysplastic syndromes

Hideki Makishima[1,2,12], **Tetsuichi Yoshizato**[2,12], **Kenichi Yoshida**[2,12], **Mikkael A Sekeres**[1,3], **Tomas Radivoyevitch**[1], **Hiromichi Suzuki**[2], **Bartlomiej Przychodzen**[1], **Yasunobu Nagata**[2], **Manja Meggendorfer**[4], **Masashi Sanada**[5], **Yusuke Okuno**[2], **Cassandra Hirsch**[1], **Teodora Kuzmanovic**[1], **Yusuke Sato**[2], **Aiko Sato-Otsubo**[2], **Thomas LaFramboise**[6], **Naoko Hosono**[1], **Yuichi Shiraishi**[7], **Kenichi Chiba**[7], **Claudia Haferlach**[5], **Wolfgang Kern**[5], **Hiroko Tanaka**[7], **Yusuke Shiozawa**[2], **Inés Gómez-Seguí**[1], **Holleh D Husseinzadeh**[1], **Swapna Thota**[1], **Kathryn M Guinta**[1], **Brittney Dienes**[1], **Tsuyoshi Nakamaki**[8], **Shuichi Miyawaki**[9], **Yogen Saunthararajah**[1], **Shigeru Chiba**[10], **Satoru Miyano**[7], **Lee-Yung Shih**[11], **Torsten Haferlach**[5], **Seishi Ogawa**[2,13], **Jaroslaw P Maciejewski**[1,13]

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA.

[2]Department of Pathology and Tumor Biology, Kyoto University, Kyoto, and Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[3]Leukemia Program, Department of Hematology and Medical Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA.

[4]MLL Munich Leukemia Laboratory, Munich, Germany.

[5]Clinical Research Center, Nagoya Medical Center, Nagoya, Japan.

Correspondence should be addressed to J.P.M. (maciejj@ccf.org) or S.O. (sogawa-tky@umin.ac.jp).

AUTHOR CONTRIBUTIONS

H.M.: project leader, analysis coordination, variant validation production and manuscript preparation. T.Y.: project leader, sequence analysis, microarray data analysis and manuscript preparation. K.Y.: data analysis team and variant validation production. M.A.S.: clinical data and specimen acquisition and manuscript preparation. T.R.: data analysis, manuscript preparation. H.S.: data analysis team and variant validation production. B.P.: supervisor data analysis team and variant validation production. Y.N.: hybridization-based capture next-generation platform development. M.M.: clinical data and specimen acquisition, data analysis. M.S.: sequence analysis. Y.O.: auto-analysis and manuscript preparation. C. Hirsch: clinical data and specimen acquisition, data analysis. T.K.: clinical data and specimen acquisition, data analysis. Y. Sato: microarray data analysis. A.S.-O.: data management, data analysis. T.L.: data analysis and manuscript preparation. N.H.: variant validation production. Y. Shiraishi: data management, data analysis. K.C.: data management, data analysis. C. Haferlach: clinical data and specimen acquisition, data analysis. W.K.: clinical data and specimen acquisition, data analysis. H.T.: data management, data analysis. Y. Shiozawa: data analysis. I.G.-S.: clinical data management. H.D.H.: clinical data management. S.T.: clinical data management. K.M.G.: specimen acquisition. B.D.: specimen acquisition. T.N.: clinical data management and specimen acquisition. S. Miyawaki: clinical data management and specimen acquisition. Y. Saunthararajah: clinical data management and specimen acquisition. S.C.: clinical data management and specimen acquisition. S. Miyano: data management, data analysis automation leader. L.-Y.S.: clinical data and specimen acquisition, study design. T.H.: clinical data and specimen acquisition, study design. S.O. and J.P.M.: project leaders, study design, execution and analysis, manuscript preparation.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Accession codes. The mutation calls from the WES are available from NCBI under accession PRJNA203580.

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

[6]Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, Ohio, USA.

[7]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[8]Division of Hematology, Department of Medicine, Showa University School of Medicine, Tokyo, Japan.

[9]Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, Tokyo, Japan.

[10]Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan.

[11]Division of Hematology-Oncology, Chang Gung Memorial Hospital–Linkou, Chang Gung University, Taoyuan, Taiwan.

[12]These authors contributed equally to this work.

[13]These authors jointly directed this work.

Abstract

To elucidate differential roles of mutations in myelodysplastic syndromes (MDS), we investigated clonal dynamics using whole-exome and/or targeted sequencing of 699 patients, of whom 122 were analyzed longitudinally. Including the results from previous reports, we assessed a total of 2,250 patients for mutational enrichment patterns. During progression, the number of mutations, their diversity and clone sizes increased, with alterations frequently present in dominant clones with or without their sweeping previous clones. Enriched in secondary acute myeloid leukemia (sAML; in comparison to high-risk MDS), *FLT3, PTPN11, WT1, IDH1, NPM1, IDH2* and *NRAS* mutations (type 1) tended to be newly acquired, and were associated with faster sAML progression and a shorter overall survival time. Significantly enriched in high-risk MDS (in comparison to low-risk MDS), *TP53, GATA2, KRAS, RUNX1, STAG2, ASXL1, ZRSR2* and *TET2* mutations (type 2) had a weaker impact on sAML progression and overall survival than type-1 mutations. The distinct roles of type-1 and type-2 mutations suggest their potential utility in disease monitoring.

MDS are a heterogeneous group of chronic myeloid neoplasms in which progression to sAML is common[1-3]. In the past decade, advanced high-throughput sequencing technologies have led to the identification of a broad range of genes recurrently mutated in MDS and sAML[4-17]. Using deep sequencing of exemplary sAML samples, accurate determination of allelic composition enabled early insights into intratumor heterogeneity during progression from MDS to sAML[18-20]. The clinical impact of selected genetic lesions has been studied using large sets of clinical samples[21-24]. However, clonal architecture was largely inferred from the allelic burden of a limited number of driver mutations at a single time point[22-25], and to date whole-genome sequencing or whole-exome sequencing (WES) of serially collected samples has been performed for only a small number of patients[18,19]. To better understand disease progression in terms of gene mutations and their clonal architecture and dynamics, a more comprehensive study involving a large number of serially collected

samples is needed. Such a study could identify MDS biomarkers that improve prognostication, differential diagnosis and therapy selection.

In the present study, we performed WES and/or targeted deep sequencing of a large set of MDS samples, including serially collected ones, to elucidate the dynamics of clonal structure (Supplementary Fig. 1). Leveraging previously published genotyping data to comprise a total of 2,250 MDS and sAML patients[18,20,23,24], we also analyzed differential impact of driver mutations on disease progression and clinical outcomes.

RESULTS

Clonal architecture in MDS and myelodysplastic/myeloproliferative neoplasms (MDS/MPN)

We analyzed 204 samples from 193 patients, using WES with a mean coverage of 107.3× (Supplementary Fig. 2): 45 with refractory anemia with excess blasts (RAEB) (high-risk MDS) and 79 with other MDS subtypes (low-risk MDS), 45 with myelodysplastic/myeloproliferative neoplasms (MDS/MPN) and 24 with sAML (Supplementary Fig. 1 and Supplementary Table 1). We analyzed 11 patients at two separate time points (Supplementary Fig. 1). After removing polymorphisms and sequencing errors, we identified a total of 2,322 mutations in 1,383 genes (Supplementary Figs. 3 and 4), of which 49 genes were significantly mutated compared to the background mutation rate ($q < 0.01$) (Supplementary Table 2). These driver gene mutations were present in 77% of the patients analyzed by WES. Mutations were dominated by C to T transitions at CpG dinucleotides, suggesting age-related deamination of methylated cytosines as the major source of mutations (Supplementary Fig. 5). Although most of these drivers have been previously reported, we identified new possible drivers, including *C1QTNF3, IRF2, NEURL1, GNL2, PCDHA1* and *PDGFRB* (Supplementary Table 3). All patients had at least one genetic defect, including copy-number alterations (Supplementary Fig. 4). The mean number of nonsynonymous mutations was 11.4 per patient (or 0.34/Mb) for all samples, with a significantly higher mutation rate in sAML (0.38/Mb) compared to low-risk MDS (0.19/Mb) ($P = 0.00195$; Fig. 1a); the number of mutations also increased significantly over time in serially analyzed cases ($P = 0.032$; Fig. 1b).

Unbiased detection of mutations enabled comparison of clonal architecture across different disease subtypes. Estimated from the highest variant allele frequency (VAF) among mutations in individual samples, the average tumor burden was higher in sAML than in low- or high-risk MDS (Fig. 1c). Clustering of mutations according to their VAFs revealed the subclonal structure of MDS and MDS/MPN in patients harboring multiple mutations (Supplementary Fig. 6)[26]. On the basis of WES, we observed intratumor heterogeneity in more than 80% of cases, which was substantially higher than previously reported (48%) using targeted sequencing of AML/MDS-related genes[23]. Intratumor heterogeneity was more frequent in sAML than low-risk MDS ($P = 0.03$) (Supplementary Fig. 7). As measured by Shannon index, diversity of tumor population in terms of mutations was lowest in low-risk MDS, followed by MDS/MPN, high-risk MDS and sAML (Fig. 1d).

Analysis of serially collected samples from MDS patients

To illustrate the dynamics of clonal architecture during progression from MDS to sAML, we analyzed WES data of longitudinally collected specimens in 11 cases (Fig. 2 and Supplementary Fig. 8). We observed two distinct patterns of clonal evolution. The first pattern we observed in UPN18 (Fig. 2a), UPN01, UPN19, UPN29 and UPN33 (Supplementary Fig. 8), where several subclones recursively evolved from within the dominant population, taking over the latter population (linear evolution) (Fig. 2c). In the second pattern, exemplified by UPN83 (Fig. 2b), UPN05, UPN08, UPN10, UNP11 and UPN14 (Supplementary Fig. 8), a new or preexisting subclone(s) 'swept out' one or more preexisting subclones and eventually populated the entire tumor fraction (Fig. 2c). In both scenarios, emergence of new subclones at the second sampling was quite common and found in the majority (9/11) of cases (Fig. 2d). We identified one or more subclones in 10/11 cases at disease presentation, where a newly emerged ($n = 4$) and/or preexisting ($n = 5$) subclone outcompeted other subclones and populated the entire compartment (clone sweeping) (Fig. 2d). Although new driver mutations emerged in 6/11 cases, we also saw disappearance of preexisting subclones (mostly with uncommon mutations) in 6/11 cases and this was accompanied by clone sweeping (Fig. 2d). As indicated by their larger variance of VAFs, mutations in the initial sampling tended to be less uniformly clustered compared to later sampling, in which mutations were more discretely clustered with a smaller variance of VAFs ($P = 0.03$), indicating that disease progression is typically accompanied by clone sweeping (Fig. 2a,b and Supplementary Fig. 9).

We also investigated clonal dynamics in a larger set of MDS patients ($n = 122$) (Supplementary Table 4), focusing on the temporal profiles of common driver mutations and their relationship to disease progression, using targeted sequencing of serially collected samples (Fig. 3a). In total, we detected 401 mutations in known driver genes in 109 (82.0%) patients at one or more time points, of whom 97 (79.5%) had multiple driver mutations. In the majority of cases (70/122), the number of mutations was higher at the second time point (2.7 on average) than at diagnosis (1.9 on average) (Supplementary Fig. 10). VAFs of mutations were significantly higher at the second sampling than at the first sampling (Fig. 3b). Of 401 mutations, 13 showed the equal VAFs at the both samplings. Among these, 281 mutations were newly acquired (138 mutations in 80 cases) or accompanied by an increased allelic burden (143 mutations in 75 cases), whereas 107 mutations in 66 cases had decreased VAFs, of which 41 in 30 cases totally disappeared at the second sampling (Fig. 3c and Supplementary Figs. 11 and 12). Stem cell transplantation and chemotherapy often led to a temporal eradication of one or more subclones (Fig. 3a and Supplementary Fig. 13). We less frequently observed clone sweeping, or a replacement of one clone by another, in targeted sequencing than in WES, and confirmed clone sweeping in only 10 out of 44 patients who progressed from MDS to sAML (Fig. 3a and Supplementary Fig. 13). This could be explained by the strong limitation of detecting subclones defined by passenger or less common driver mutations in targeted sequencing (Fig. 2a,b and Supplementary Fig. 13). In fact, none of the subclones that were subjected to clone sweeping in the six cases with WES harbored mutations in common drivers (except for *PRPF8* in UPN11) included in our targeted gene panel (Supplementary Table 5).

Driver mutation enrichment in different subgroups of MDS/sAML

To clarify the relationship between driver mutations and disease progression, we first looked at the enrichment of major driver mutations in different disease subtypes by combining previously published data sets[18,20,23,24] with the current one. In total, genotyping data from 2,250 samples with low-risk (n = 1,207) and high-risk (n = 683) MDS as well as sAML (n = 360) were available for 25 driver genes commonly mutated in myeloid malignancies. We anticipated that a homogeneous cohort consisting of only cases with present or past history of MDS would eliminate experimental noise owing to the presence of a variety of distinct nosologic entities. In univariate comparison between low-risk and high-risk MDS cases, the majority of differentially mutated genes were enriched in high-risk MDS, except for *SF3B1*, which was more frequently mutated in low-risk MDS (Supplementary Fig. 14). To exclude the effects of correlations between different mutations that might confound the results, we performed multivariate analysis, in which we selected variables using a least absolute shrinkage and selection operator model (LASSO). Mutations in seven genes, including *FLT3, PTPN11, WT1, IDH1, NPM1, IDH2* and *NRAS* (designated as 'type-1' mutations), were significantly enriched in sAML compared to high-risk MDS, whereas *ASXL1* mutations were inversely enriched in high-risk MDS (Fig. 4a and Table 1). When we compared high-risk and low-risk MDS samples, mutations in 10 genes were enriched in high-risk MDS, including *GATA2, NRAS, KRAS, IDH2, TP53, RUNX1, STAG2, ASXL1, ZRSR2* and *TET2* (Fig. 4a and Table 1). We designated mutations in this second gene set as 'type-2' mutations, excluding *NRAS* and *IDH2* mutations, which we already assigned as type-1 mutations (Fig. 4a). *SF3B1* mutations were strongly enriched in low-risk MDS, compared to high-risk MDS (Fig. 4a). These results indicated that the two novel sets of gene mutations might be simply associated with sAML evolution from MDS, because the examined cohort included only MDS cases even though sAML is sometimes derived from other types of myeloid neoplasms, including MDS/MPN.

Driver mutation dynamics and clinical course

In light of the above results, we next evaluated the role of type-1, type-2 and other mutations by determining their clinical behavior in serially collected samples from 122 MDS cases, of which 90 progressed to sAML. This cohort consisted of patients of purely MDS history to validate the effects of type-1 and type-2 mutations on sAML evolution strictly from MDS. Overall, driver mutations tended to increase their clone sizes and were more likely to be newly acquired than lost between two time points (Fig. 3b,c). In accordance with their significant enrichment in sAML compared to high-risk MDS, type-1 mutations were more frequently acquired at the second time points and upon progression to sAML, compared to type-2 and other mutations (P = 0.0001) (Fig. 4b). We confirmed this trend in analyses confined to mutations associated with sAML progression from high-risk MDS (P = 0.002) (Supplementary Fig. 15a) or from high or low-risk MDS (P = 0.0001) (Supplementary Fig. 15b). In patients who had high-risk MDS at the second time point, type-2 mutations tended to dominate type-1 mutations, and most of the type-2 mutations (88%) occurred in clones with increased size (Supplementary Fig. 16a). Similarly, type-2 mutations in high-risk MDS and sAML patients evolving from low-risk MDS increased in clone size (30 of 38 mutations; 79%) more frequently, compared to those found in patients stably in low-risk MDS group (4 of 11; 36%) (P = 0.02) (Supplementary Fig. 16b). As expected, *SF3B1*

mutations were rarely newly acquired at the second sampling (1 of 17 cases; 5.9%) or upon sAML progression (1 of 13 cases; 7.7%).

Although commonly seen during disease progression, clone sweeping may not necessarily be caused by type-1 mutations, as type-1 mutations tended to have a smaller clonal burden compared to other mutations ($P = 5.40 \times 10^{-12}$) (Supplementary Fig. 17). We found a total of 74 type-1 mutations in sAML samples, of which only 25 (34%) were involved in clone sweeping, whereas the remaining 49 (66%) were found in one or more subclones. One could speculate that the diagnosis of sAML (increased blast count of ≥20%) is closely associated with the evolution of a new subclone, rather than the dominant clones in the proceeding MDS phase, but may not necessarily require sweeping of the entire tumor population by the newly evolved clone.

Effects of driver mutations on sAML progression and survival

As described above, type-1 mutations in sAML were more likely to be newly acquired than present before progression, whereas type-2 mutations were more frequently carried over than newly acquired (Fig. 4b). This finding suggests that the time from the acquisition of type-1 mutations to sAML progression was, on average, shorter than the time from the acquisition of other mutations to sAML and/or that more non-leukemic deaths occurred in MDS patients with type-2 gene mutations compared to patients with type-1 mutations. To address this, we evaluated the effects of the different classes of mutations on progression to sAML among 429 patients with MDS. In accordance with these observations in serially collected samples, patients with type-1 mutations (group I) had a significantly shorter time to progression to sAML compared to patients who had type-2 mutations but lacked type-1 mutations (group II) (hazard ratio (HR) = 1.82, 95% confidence interval (CI): 1.08–3.05; $P = 0.025$) and those who had no type-1, type-2 or *SF3B1* mutations (group IV) (Supplementary Table 6) (HR = 4.48, 95% CI: 2.53–7.94; $P = 0.0001$). This was in stark contrast to a very low leukemia progression seen in patients carrying *SF3B1* mutation with no type-1 or type-2 mutations (group III) (Fig. 4c).

Although group-II patients were less likely to progress to sAML, time to progression to sAML (or progression free survival; PFS) in group-II cases was still significantly shorter than that in group-IV cases (HR = 2.46, 95% CI: 1.43–4.23; $P = 0.001$) and other patients (Fig. 4c). Of interest, in the analysis of serially collected samples, 35 patients having type-2, but not type-1, mutations progressed to sAML, of whom 15 (43%) newly acquired type-1 mutations, suggesting that a higher risk of sAML progression in group-II patients could be partly explained by subsequently acquired type-1 mutations (Fig. 4d and Supplementary Fig. 18). Because there was no significant difference between group-I and group-II patients with regard to treatment, the effects of the category of mutations were not likely to be influenced by treatment (Supplementary Table 7). In multivariate analysis, the mutational category (i.e., group I, II or III) was shown to be an independent significant predictor of PFS, together with the International Prognostic Scoring System (IPSS) score (Table 2).

We also assessed the effects of mutations on overall survival (OS) in a larger cohort of MDS patients ($n = 1,347$). Group-I cases showed a significantly shorter OS than group-II cases (HR = 1.50, 95% CI: 1.20–1.86; $P = 0.001$) (Fig. 4e). OS in group-II cases was also

significantly shorter than in group-IV cases (HR = 1.40, 95% CI: 1.14–1.71; $P = 0.0011$), but that in group-III cases was significantly longer compared to any other categories (Fig. 4e), suggesting that the faster progression to sAML was translated into a shorter OS. In multivariate analysis, the group-I category remained an independent negative predictor of OS, together with other factors, including complex karyotype, IPSS score, −7/del(7q), age and del(20q) (Table 2).

Effects of clone size of driver mutations on prognosis

To further clarify the roles of type-1 and type-2 mutations on disease progression, we examined the effects of clone size of these mutations on PFS and OS, where clones were dichotomized into larger and smaller ones by the median VAF of mutations and survivals were compared between patients with larger and smaller clones having type-1 and type-2 mutations. Among patients with type-1 mutations, those with larger clones had a significantly shorter PFS than those with smaller type-1 clones (HR = 2.22, 95% CI: 1.01–4.88; $P = 0.047$) (Fig. 5a), although clone size did not significantly affect OS (Supplementary Fig. 19a). In contrast, even though clone size of type-2 mutations did not affect PFS (Fig. 5b), larger clones were associated with a significantly shorter OS (HR = 1.46, 95% CI: 1.18–1.80; $P = 0.0005$) (Supplementary Fig. 19b).

Effects of combinations of driver mutations on sAML progression

Given the distinct roles of type-1, type-2 and *SF3B1* mutations, it would be of particular interest to see the interactions between these different sets of mutations in terms of co-occurrence and mutual exclusiveness. Although gene interactions between mutations have been investigated previously in the two large cohorts of MDS patients[23,24], they were more comprehensively interrogated in a combined, larger cohort of 2,250 patients. Thus, the present analysis not only confirmed 34 positive and 17 negative associations previously reported, but also newly identified 75 positive and 20 negative associations (Fig. 6a). Most conspicuous among these interactions were those found between *SF3B1* and other mutations. *SF3B1* mutations were mutually exclusive not only with other splicing factor mutations[9], but also with most of common mutations, except for *DNMT3A* and *JAK2* mutations, which significantly co-occurred with *SF3B1* mutations. Particularly, *SF3B1* mutations were mutually exclusive with type-1 and type-2 mutations, which is in line with the very rare progression to sAML in *SF3B1*-mutated patients. Also of interest was a global trend of co-occurring relationships within type-1 and type-2 mutations, with prominent exceptions seen in *TP53, TET2* and *NPM1*, which were largely mutually exclusive with other type-1 and type-2 mutations. Thus, the interactions seem to depend on gene context: some type-1 mutations may cooperate with type-2 mutations during progression to sAML, whereas others may not. Despite the global trend of co-occurrence between type-1 and type-2 mutations, type-1 mutations had significantly lower VAFs compared to type-2 mutations in the patients with both mutations (Fig. 6b), suggesting that type-1 mutations were likely acquired during disease progression after previous type-2 hits.

DISCUSSION

Representing an unbiased analysis of somatic mutations in the largest cohort of MDS ever described, to our knowledge, our results not only confirmed the spectrum of major driver genes involved in MDS patients, but also disclosed the complexity of clonal structure and its dynamics during MDS disease progression. Disease progression in MDS patients is not merely shaped by simple rounds of linear evolutions as previously described[18,19], but is also accompanied by frequent clone sweeping of existing subclones, in which driver mutations are thought to play critical roles. Moreover, our analysis, which combined previously published large genotyping data sets[18,20,23,24] and newly generated WES and targeted sequencing data, identified distinct effects of different driver mutations on disease progression.

Frequently involving genes in signal transduction, type-1 mutations tend to be more newly acquired during progression from MDS to sAML and when present in MDS, are associated with higher risk of progression to sAML and shorter overall survival, compared to other mutations[21,23,27-30]. Thus, close monitoring of the emergence of type-1 mutations might allow for early diagnosis of progression to sAML. A role of preemptive use of RAS pathway, IDH1, IDH2 and FLT3 inhibitors in this context should also be investigated. In contrast, type-2 genes are largely characterized by transcription and epigenetic regulations. Significantly enriched in high-risk vs. low-risk MDS, type-2 mutations might typically drive disease progression to high-risk MDS, which suggests that mutational burden of type-2 genes might be smaller in lower-risk MDS and increase during progression to high-risk MDS. Such analysis of clonal size in different genes highlighted the distinct characteristics of type-1 and type-2 mutations differentially contributing to sAML evolution. In contrast to the cases for type-1 and type-2 mutations, which were associated with progression to higher-risk diseases, as previously reported, patients with *SF3B1* mutations are less likely to progress to sAML[8,31,32]. In accordance with this, *SF3B1* mutations are highly mutually exclusive with most of the type-1 and type-2 mutations. These findings suggest that *SF3B1* mutations might have a distinct biological role in MDS pathogenesis that may define a unique entity of myeloid neoplasms[33]. Mutations in other genes, including *DNMT3A* and *U2AF1*, exhibited no significant enrichment in specific disease subtypes and thus are likely to represent founder or ancestral mutations that initiate the early stage of MDS, rather than secondary mutations involved in disease progression. Of interest in this regard are the frequent *DNMT3A* mutations reported in age-related clonal hematopoiesis (ARCH) or clonal hematopoiesis of indeterminate potential (CHIP)[34-37], which is associated with an increased risk of blood cancer.

Although a common finding during the entire clinical course in MDS, clone sweeping may not necessarily be obligatory for leukemic progression, but a more frequent scenario would be the emergence of clones with type-1 mutations that remained in a subclone without clone sweeping. This might have been anticipated, because ≥20% blasts in the diagnostic criteria for sAML might be a rather arbitrary, and patients with increasing blast counts in MDS showed a strong trend to progress to sAML earlier or later. Accordingly, the majority of the remaining cells are thought to represent pre-leukemic MDS clones that should be discriminated from the sAML clone. These pre-leukemic MDS cells may be comparable to

pre-leukemic cells recently proposed to exist in the pathogenesis of *de novo* AML[38], in that sAML clones derived from this cell population by newly acquiring additional mutations, typically of type-1 genes. Thus, it would be an interesting question whether or not the emergence of type-1 mutations better predicts survival or guide therapy than blast counts, which should be addressed.

Our gene categories were based on the statistically significant enrichment in different disease stages, and supported by distinct effects on disease progression and co-occurring patterns; as such, these are clinically relevant to predict disease progression. However, except for the distinct *SF3B1* -mutated tumors, it is still an open question whether these categories provide biologically relevant classification. For example, genes in type-1 and type-2 categories showed some segregation into distinct gene function pathways: type-1 genes in RAS pathway and signal transduction and type-2 genes in transcription and epigenetic regulation. In this regard, type-1 and type-2 mutations in MDS might be comparable to class-I and class-II mutations previously proposed to have distinct roles in the pathogenesis of primary acute myeloid leukemia[39-41], respectively. However, there were some overlaps between these functional pathways in type-1 and type-2 categories. This suggests that the biological impact of mutations on disease progression may not be simply anticipated based on the pathway they belong to, but that different mutations in a similar functional category could have distinct roles at different stages during clonal selection and disease progression in MDS. Further evaluation of the biological basis for these gene categories should be warranted. Also, the impact of a driver mutation may differ between MDS and other myeloid neoplasms. For example, CMML is characterized by a high frequency of RAS pathway mutations, which are common type-1 mutations in MDS and are typically acquired upon progression to sAML but have been shown to be present at initial diagnosis of CMML without excess of blasts[42].

In conclusion, we elucidated how clone diversity and dynamics associate with MDS outcomes and modes of progression. Our molecular analysis parallels risk classification of MDS, showing that progression steps previously defined by pathologic criteria are accompanied or mediated by distinct molecular changes. Our results indicate that the driver genes can be classified into molecular subtypes differentially associated with low-risk MDS, high-risk MDS or sAML. This new categorization provides insights into clonal dynamics and allows for the use of subclonal events as MDS progression biomarkers.

ONLINE METHODS

Patients.

For the initial screening, paired tumor and normal germline DNA was obtained from cases with myeloid neoplasms (699 MDS cases and 45 MDS/MPN cases) for WES (n = 193) and targeted deep sequencing (n = 699), including low-risk MDS (refractory cytopenia with unilineage dysplasia (RCUD), refractory cytopenia with multilineage dysplasia (RCMD), refractory anemia with ring sideroblasts (RARS), isolated del(5q), and MDS unclassifiable (MDS-u)), high-risk MDS (RAEB1 and RAEB2), chronic myelomonocytic leukemia (CMML), other MDS/MPN, and sAML (Supplementary Table 1); 122 cases were serially studied. All samples were obtained after written informed consent, according to protocols

approved by the ethics boards of participating institutions. Publicly available information on genotyping and diagnosis from the other cases (n = 1,551) were collected from recent publications[18,20,23,24]. In total, 2,250 cases with MDS and sAML were enrolled in this study (Supplementary Fig. 1).

Whole exome sequencing.

WES was performed as previously described[9,15,43,44]. Briefly, tumor DNAs were extracted from patients' bone marrow or peripheral blood mononuclear cells. For germline control, DNA was obtained from either buccal mucosa or paired CD3-positive T cells with or without prior culture in the presence of phytohemagglutinin and IL-2. Whole-exome capture was accomplished based on liquid phase hybridization of sonicated genomic DNA having 150–200 bp mean length to the bait cRNA library synthesized on magnetic beads (SureSelect, ver. 3 or 4 (Agilent Technology)), according to the manufacturer's protocol. The captured targets were subjected to massive sequencing using the HiSeq 2000 with the paired-end 75–108 bp read option, according to the manufacturer's instructions. Subsequent validation and confirmatory sequencing are described below. Briefly, sequencing reads were aligned to human genome (hg19) by Burrows-Wheeler aligner (http://bio-bwa.sourceforge.net/). Using a GATK pipeline algorithm, comparison of sequencing reads between tumor and germline DNA (Fisher's exact test) was performed to extract candidate variations, and polymorphisms and to remove sequencing errors. Confirmatory validation by Sanger sequencing or PCR amplicon sequencing was performed as described below (Supplementary Fig. 3).

Targeted deep sequencing.

Multi-amplicon deep sequencing (TruSeq; Illumina) for 61 gene targets was performed according to the manufacturer's instructions (Supplementary Table 5). For amplicon sequencing, TruSeq custom amplicon generation protocol was applied to customized probe sets to amplify target exons of target genes. The sequencing libraries were generated according to an Illumina paired-end library protocol and subjected to deep sequencing on MiSeq (Illumina) sequencers according to the standard protocol. Targeted-capture sequencing was performed as previously described[23]. Subsequent validation and confirmatory sequencing are described as below (Supplementary Fig. 3).

Confirmatory Sanger sequencing and PCR amplicon sequencing.

Exons of selected genes were amplified and underwent direct genomic sequencing by standard techniques on the ABI 3730xl DNA analyzer (Applied Biosystems), as previously described[45-47]. All mutations were detected by bidirectional sequencing and scored as pathogenic if not present in nonclonal paired germline DNA. When the marginal volume of mutant clone size was not confirmed by Sanger sequencing, cloning and sequencing individual colonies (TOPO TA cloning, Invitrogen) was performed for validation. For detecting allelic frequency of mutations or SNPs, we applied PCR amplicon sequencing to targeted exons as previously described[9]. Briefly, we analyzed for possible or validated mutations in amplicons of around 250 bp, targeting the locus with each specific primer pair. The sequencing libraries were generated according to an Illumina paired-end library

protocol and subjected to deep sequencing on HiSeq2000 or MiSeq sequencers according to the standard protocol (Illumina)[9,32].

Single-nucleotide polymorphism array analysis.

Single-nucleotide polymorphism (SNP) array karyotyping for confirming metaphase cytogenetics and detecting copy-number normal loss of heterozygosity was performed as previously described[47,48]. Briefly, Affymetrix 250K and 6.0 SNP arrays were used to evaluate copy number and loss of heterozygosity. Using our internal and a publicly available database (http://dgv.tcag.ca/dgv/app/home), the screening algorithm validated each lesion as somatic[49,50]. Non-somatic lesions were excluded from further analysis. Affected genomic positions in each lesion were visualized and extracted by CNAG (v3.0)[51] or the Genotyping Console (Affymetrix) software.

Significantly mutated gene analysis.

To detect significantly mutated genes, the global background mutation rate was calculated on the basis of data from the current WES study. The average of mutation rate (per megabase) was calculated from the total number of somatic mutations and the length of whole coding region (50 Mb) as previously described[52]. The number of mutations in each gene was tested based on the Poisson distribution, where the mean lambda is a product of the background mutation rate and coding length of each gene. The obtained *P* value was adjusted by Benjamini-Hochberg correction. A false detection rate (FDR) *q* value of 0.01 was used as the cutoff.

PyClone.

Clustering analysis of mutations was performed according to the Beta Binomial emission model implemented in PyClone[53] as described previously[54].

Diversity of mutational spectrum.

Shannon-Weaver index for estimating the diversity of mutated genes in each disease subtype was calculated using R package ‘vegan’. Because the number of cases in each category was different, we performed multiple testing toward the randomly selected cases with a common number of 37 samples drawn, which was the minimum number of cases among the categories. We performed 1,000 random draws of 37 cases from each category in our calculations of the Shannon-Weaver index.

Comparison of mutational spectra among disease subtypes.

For the comparison of mutational spectrum in a large data set, we combined our data set with data for four other cohorts[18,20,23,24]. Among disease subtypes, we compared the mutational spectra of the genes commonly tested in all the enrolled cases. In univariate analysis, mutational frequencies were compared by odds ratio. Multivariate analysis was executed using logistic regression model and the best model was selected with least absolute shrinkage and selection operator (LASSO)[55] with R package ‘glmnet’; lambda determined by leave-one-out cross-validation was minimized to select the best combination of

explanatory variables. When there was any zero cell in the table, we corrected the number by adding 0.5.

Statistical analysis.

Pairwise comparisons were performed by Wilcoxon test for continuous variables and by two-sided Fisher's exact test for categorical variables. Paired data were analyzed by Wilcoxon signed-rank test. Whisker plot boxes denote median and 25th and 75th percentiles, and ends of the whiskers display minimum and maximum values. Trends of categorical variables were assessed by the Cochran-Armitage test. The Kaplan-Meier method was used to analyze survival outcomes (progression-free or overall survival) by the log-rank test. For multivariate analyses, a Cox proportional hazards model was conducted for progression free or overall survival. Variables considered for model inclusion were IPSS risk group, age, sex, cytogenetic abnormality and gene mutation status. Statistical analyses were performed with R (https://www.r-project.org) or JMP9 software (SAS). Significance was determined at a two-sided α level of 0.05, except for *P* values in multiple comparisons, in which multiple testing was adjusted according to the method described by Benjamini and Hochberg[56]. Methods of detailed statistical analyses are described in each section above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health (NIH) grants RO1HL-082983 (J.P.M.), U54 RR019391 (J.P.M.) and K24 HL-077522 (J.P.M.), a grant from the AA & MDS International Foundation (J.P.M., M.A.S. and H.M.), the Robert Duggan Charitable Fund (J.P.M.), a grant from Edward P. Evans Foundation (J.P.M. and M.A.S.), Scott Hamilton CARES grant (H.M.), Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan, the Japanese Agency for Medical Research and Development (Health and Labour Sciences Research Expenses for Commission and Applied Research for Innovative Treatment of Cancer) and KAKENHI (26221308, 23249052, 22134006, 15H05909, hp160219 and 21790907; S.O.), (15km0305018h0101, 16H05338; H.M.), (26890016; K.Y.), project for development of innovative research on cancer therapies (FIRST, p-direct; S.O.), the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology', initiated by the Council for Science and Technology Policy (CSTP) (S.O.), NHRI-EX100-10003NI Taiwan (L.-Y.S.), National Aeronautics and Space Administration grant NNJ13ZSA001N (T.R.), and a grant from the American Cancer Society (Research Scholar Grant 123436-RSG-12-159-01-DMC, to T.L.). Supercomputing resources were also provided by the Human Genome Center, the Institute of Medical Science, The University of Tokyo.

References

1. Harris NL et al. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting–Airlie House, Virginia, November 1997. J. Clin. Oncol 17, 3835–3849 (1999). [PubMed: 10577857]
2. Vardiman JW et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 114, 937–951 (2009). [PubMed: 19357394]
3. Malcovati L et al. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: a basis for clinical decision making. J. Clin. Oncol 23, 7594–7603 (2005). [PubMed: 16186598]
4. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med 368, 2059–2074 (2013). [PubMed: 23634996]

5. Mardis ER et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N. Engl. J. Med 361, 1058–1066 (2009). [PubMed: 19657110]
6. Ley TJ et al. *DNMT3A* mutations in acute myeloid leukemia. N. Engl. J. Med 363, 2424–2433 (2010). [PubMed: 21067377]
7. Delhommeau F et al. Mutation in *TET2* in myeloid cancers. N. Engl. J. Med 360, 2289–2301 (2009). [PubMed: 19474426]
8. Papaemmanuil E et al. Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med 365, 1384–1395 (2011). [PubMed: 21995386]
9. Yoshida K et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64–69 (2011). [PubMed: 21909114]
10. Gelsi-Boyer V et al. Mutations of Polycomb-associated gene *ASXL1* in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br. J. Haematol 145, 788–800 (2009). [PubMed: 19388938]
11. Grossmann V et al. Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype. Blood 118, 6153–6163 (2011). [PubMed: 22012066]
12. Sanada M et al. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature 460, 904–908 (2009). [PubMed: 19620960]
13. Ernst T et al. Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. Nat. Genet 42, 722–726 (2010). [PubMed: 20601953]
14. Maxson JE et al. Oncogenic *CSF3R* mutations in chronic neutrophilic leukemia and atypical CML. N. Engl. J. Med 368, 1781–1790 (2013). [PubMed: 23656643]
15. Makishima H et al. Somatic *SETBP1* mutations in myeloid malignancies. Nat. Genet 45, 942–946 (2013). [PubMed: 23832012]
16. Kurtovic-Kozaric A et al. *PRPF8* defects cause missplicing in myeloid malignancies. Leukemia 29, 126–136 (2015). [PubMed: 24781015]
17. Huang D et al. *BRCC3* mutations in myeloid neoplasms. Haematologica 100, 1051–1057 (2015). [PubMed: 26001790]
18. Walter MJ et al. Clonal architecture of secondary acute myeloid leukemia. N. Engl. J. Med 366, 1090–1098 (2012). [PubMed: 22417201]
19. Walter MJ et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia 27, 1275–1282 (2013). [PubMed: 23443460]
20. Lindsley RC et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood 125, 1367–1376 (2015). [PubMed: 25550361]
21. Bejar R et al. Clinical effect of point mutations in myelodysplastic syndromes. N. Engl. J. Med 364, 2496–2506 (2011). [PubMed: 21714648]
22. Bejar R et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J. Clin. Oncol 30, 3376–3382 (2012). [PubMed: 22869879]
23. Haferlach T et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 28, 241–247 (2014). [PubMed: 24220272]
24. Papaemmanuil E et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616–3627, quiz 3699 (2013). [PubMed: 24030381]
25. Rocquain J et al. Combined mutations of *ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2* and *WT1* genes in myelodysplastic syndromes and acute myeloid leukemias. BMC Cancer 10, 401 (2010). [PubMed: 20678218]
26. Roth A et al. PyClone: statistical inference of clonal population structure in cancer. Nat. Methods 11, 396–398 (2014). [PubMed: 24633410]
27. Badar T et al. Detectable *FLT3-ITD* or *RAS* mutation at the time of transformation from MDS to AML predicts for very poor outcomes. Leuk. Res 39, 1367–1374 (2015). [PubMed: 26547258]
28. Schnittger S et al. Characterization of *NPM1*-mutated AML with a history of myelodysplastic syndromes or myeloproliferative neoplasms. Leukemia 25, 615–621 (2011). [PubMed: 21233837]
29. Shih LY et al. Acquisition of *FLT3* or N-*ras* mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia. Leukemia 18, 466–475 (2004). [PubMed: 14737077]

30. Dicker F et al. Mutation analysis for *RUNX1, MLL-PTD, FLT3-ITD, NPM1* and *NRAS* in 269 patients with MDS or secondary AML. Leukemia 24, 1528–1532 (2010). [PubMed: 20520634]
31. Malcovati L et al. Clinical significance of *SF3B1* mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood 118, 6239–6246 (2011). [PubMed: 21998214]
32. Makishima H et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 119, 3203–3210 (2012). [PubMed: 22323480]
33. Malcovati L et al. *SF3B1* mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. Blood 126, 233–241 (2015). [PubMed: 25957392]
34. Busque L et al. Recurrent somatic *TET2* mutations in normal elderly individuals with clonal hematopoiesis. Nat. Genet 44, 1179–1181 (2012). [PubMed: 23001125]
35. Genovese G et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med 371, 2477–2487 (2014). [PubMed: 25426838]
36. Xie M et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med 20, 1472–1478 (2014). [PubMed: 25326804]
37. Jaiswal S et al. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med 371, 2488–2498 (2014). [PubMed: 25426837]
38. Shlush LI et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature 506, 328–333 (2014). [PubMed: 24522528]
39. Gilliland DG & Griffin JD The roles of FLT3 in hematopoiesis and leukemia. Blood 100, 1532–1542 (2002). [PubMed: 12176867]
40. Fröhling S, Scholl C, Gilliland DG & Levine RL Genetics of myeloid malignancies: pathogenetic and clinical implications. J. Clin. Oncol 23, 6285–6295 (2005). [PubMed: 16155011]
41. Kihara R et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia 28, 1586–1595 (2014). [PubMed: 24487413]
42. Merlevede J et al. Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents. Nat. Commun 7, 10767 (2016). [PubMed: 26908133]
43. Yoshida K et al. The landscape of somatic mutations in Down syndrome–related myeloid disorders. Nat. Genet 45, 1293–1299 (2013). [PubMed: 24056718]
44. Kon A et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat. Genet 45, 1232–1237 (2013). [PubMed: 23955599]
45. Dunbar AJ et al. 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. Cancer Res. 68, 10349–10357 (2008). [PubMed: 19074904]
46. Jankowska AM et al. Loss of heterozygosity 4q24 and *TET2* mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood 113, 6403–6410 (2009). [PubMed: 19372255]
47. Makishima H et al. *CBL, CBLB, TET2, ASXL1,* and *IDH1/2* mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. Blood 117, e198–e206 (2011). [PubMed: 21346257]
48. Makishima H et al. Mutations of E3 ubiquitin ligase Cbl family members constitute a novel common pathogenic lesion in myeloid malignancies. J. Clin. Oncol 27, 6109–6116 (2009). [PubMed: 19901108]
49. Huh J et al. Characterization of chromosome arm 20q abnormalities in myeloid malignancies using genome-wide single nucleotide polymorphism array analysis. Genes Chromosom. Cancer 49, 390–399 (2010). [PubMed: 20095039]
50. Tiu RV et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. J. Clin. Oncol 27, 5219–5226 (2009). [PubMed: 19770377]
51. Nannya Y et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res. 65, 6071–6079 (2005). [PubMed: 16024607]

52. Sato Y et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat. Genet 45, 860–867 (2013). [PubMed: 23797736]

53. Shah SP et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486, 395–399 (2012). [PubMed: 22495314]

54. Suzuki H et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nat. Genet 47, 458–468 (2015). [PubMed: 25848751]

55. Tibshirani R Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Series B Stat. Methodol 58, 267–288 (1996).

56. Benjamini Y & Hochberg Y Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995).

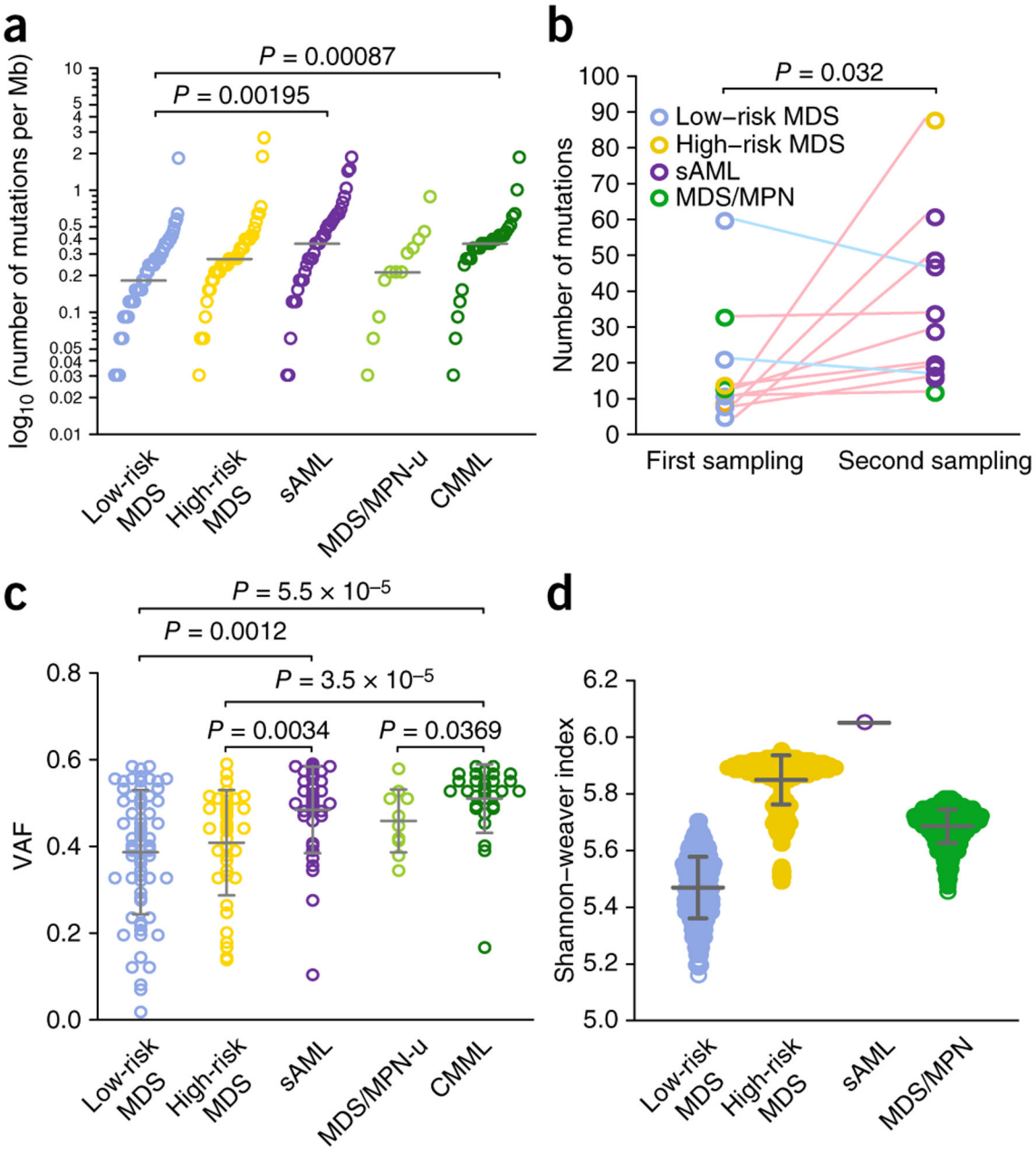

Figure 1.

Numbers, allele frequency and diversity of somatic nonsynonymous mutations. (**a**) Numbers and means of nonsynonymous mutations (per megabase) in individual samples for different disease subtypes. MDS/MPN-u, MDS/MPN unclassifiable. (**b**) Numbers of nonsynonymous mutations in 11 paired serial samples. Red and blue lines indicate cases with increasing and decreasing numbers of nonsynonymous mutations during disease progression, respectively. (**c**) Maximum values of VAF of mutations in patients for each disease subtype (25th, 50th and 75th percentiles are shown). (**d**) Shannon indices of mutation diversity calculated by randomly resampling 37 patients (25th, 50th and 75th percentiles are shown).

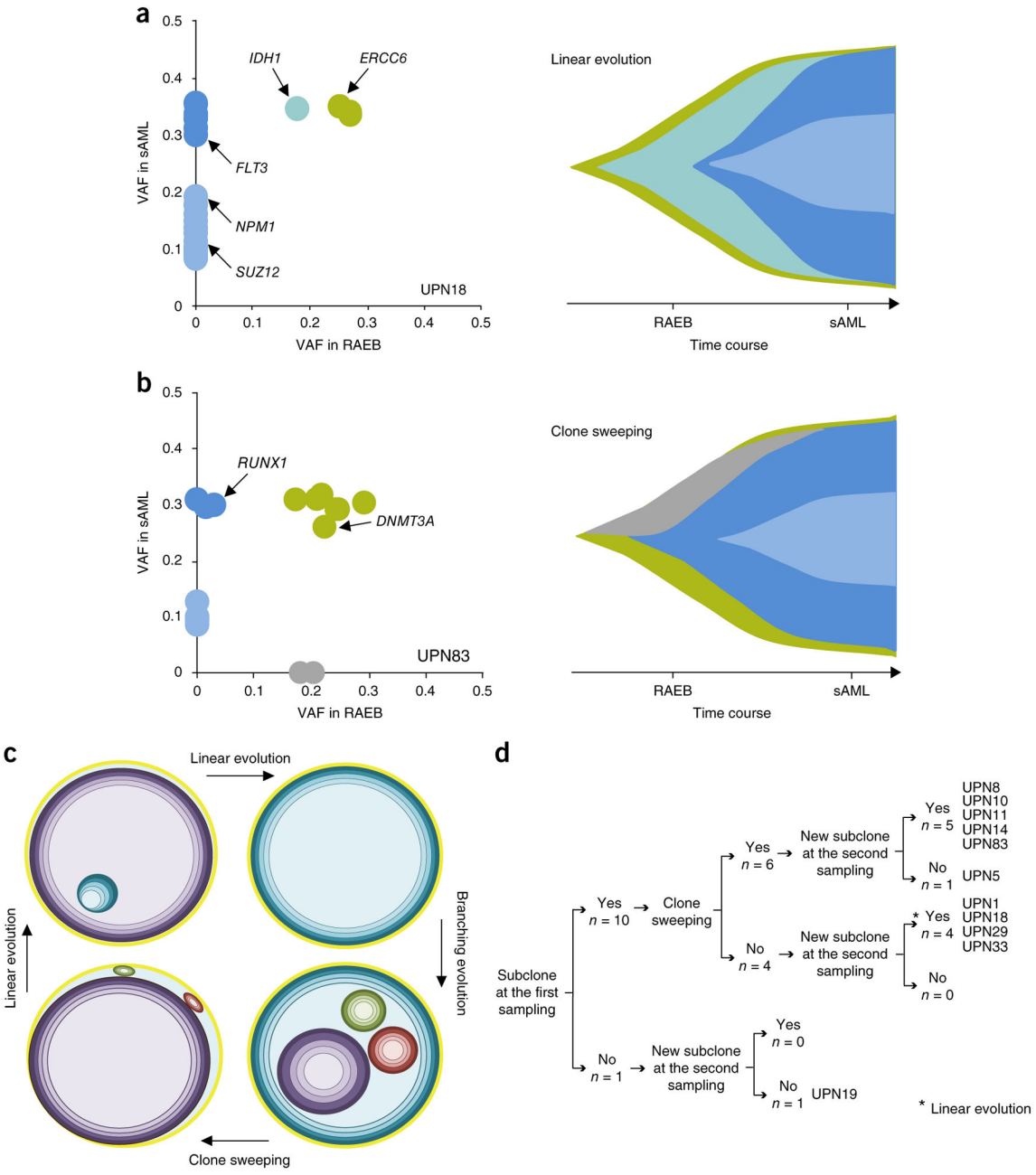

Figure 2.

Clonal evolution from MDS to sAML analyzed by WES. (**a–c**) Depictions of clonal evolution during transformation from RAEB to sAML found in two representative patients, UPN18 (**a**) and UPN83 (**b**) as examples of linear evolution and clone sweeping, respectively, which can sequentially arise and recur in a single patient (**c**). Clonal populations in diagnostic (RAEB) and sAML samples were inferred from mutations detected by WES. VAFs of mutations in each sample are shown in diagonal plots. Known driver mutations are shown on the left; during leukemic evolution, the clonal structure evolved dynamically as shown on the right (**a,b**). Each concentric circle with a color gradient indicates a population

of clones undergoing evolution, where each circle represents clones sharing the same set of mutations. Newly emerging clones can evolve with or without clone sweeping of other subpopulations. (**d**) Summary of 11 cases in which two longitudinally collected samples were analyzed by WES, in terms of presence or absence of subclones at presentation, clone sweeping and emergence of new subclones during disease progression.

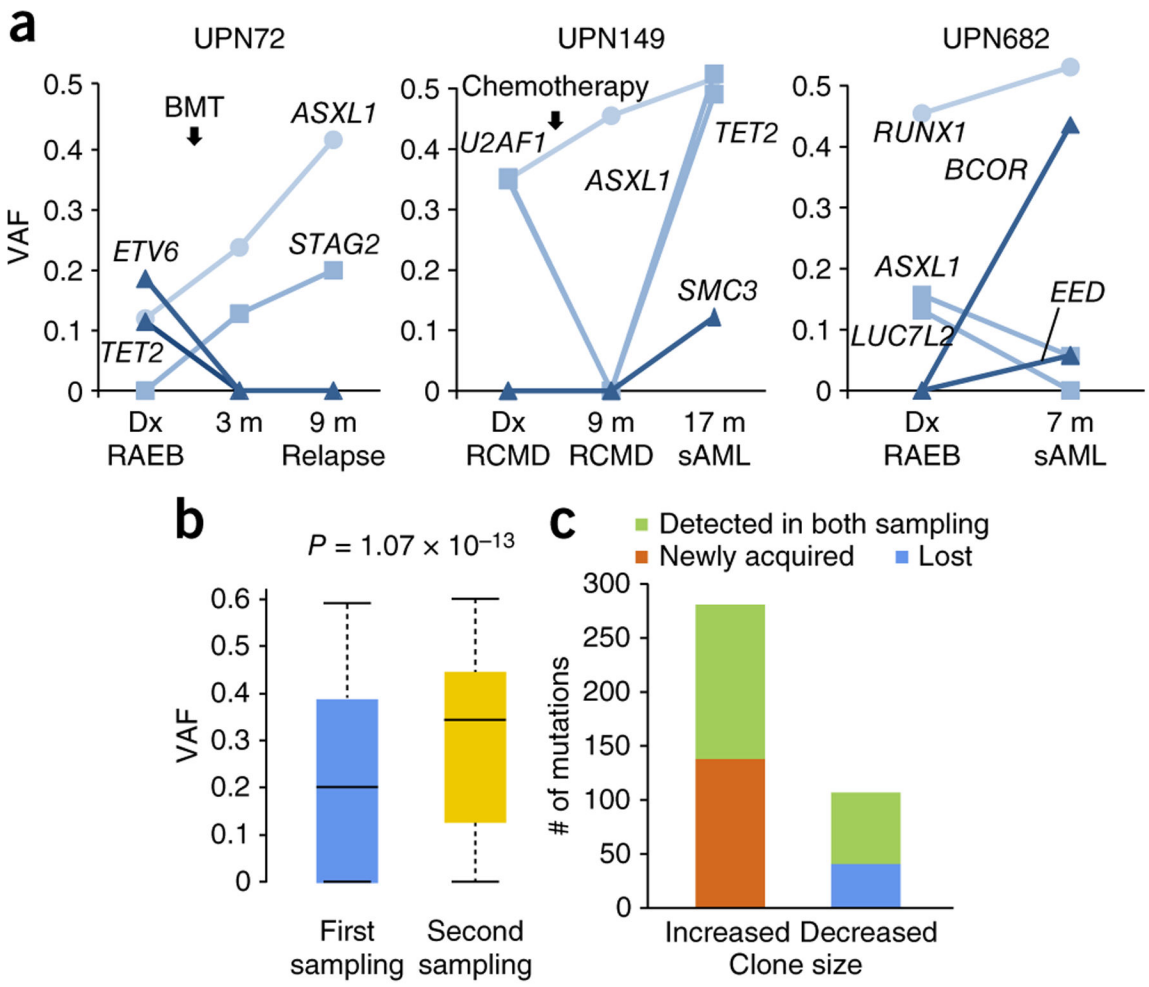

Figure 3.

Dynamics of major driver mutations revealed by targeted sequencing. (**a**) Time charts of driver mutations showing dynamic changes in their VAFs during disease progression (UPN149 and UPN682) or relapse (UPN72) in three representative cases out of 122 patients in whom driver mutations were analyzed in serially collected samples using targeted sequencing of a panel of 61 major driver mutations reported in MDS and AML. Dx, diagnosis; m, months. BMT, bone marrow transplantation; RCMD, refractory cytopenia with multilineage dysplasia. (**b**) Difference in VAFs of driver mutations (401 mutations) between the first and the second sampling. In box-and-whisker plots, boxes denote the median and 25th and 75th percentiles, and the ends of the whiskers correspond to minimum and maximum values. (**c**) The number of mutations showing an increased or a decreased clone size or VAF between two consecutive samples, in which newly acquired or lost mutations in the second samples are indicated in orange or blue, respectively, and the number of persistent mutations are shown in green.

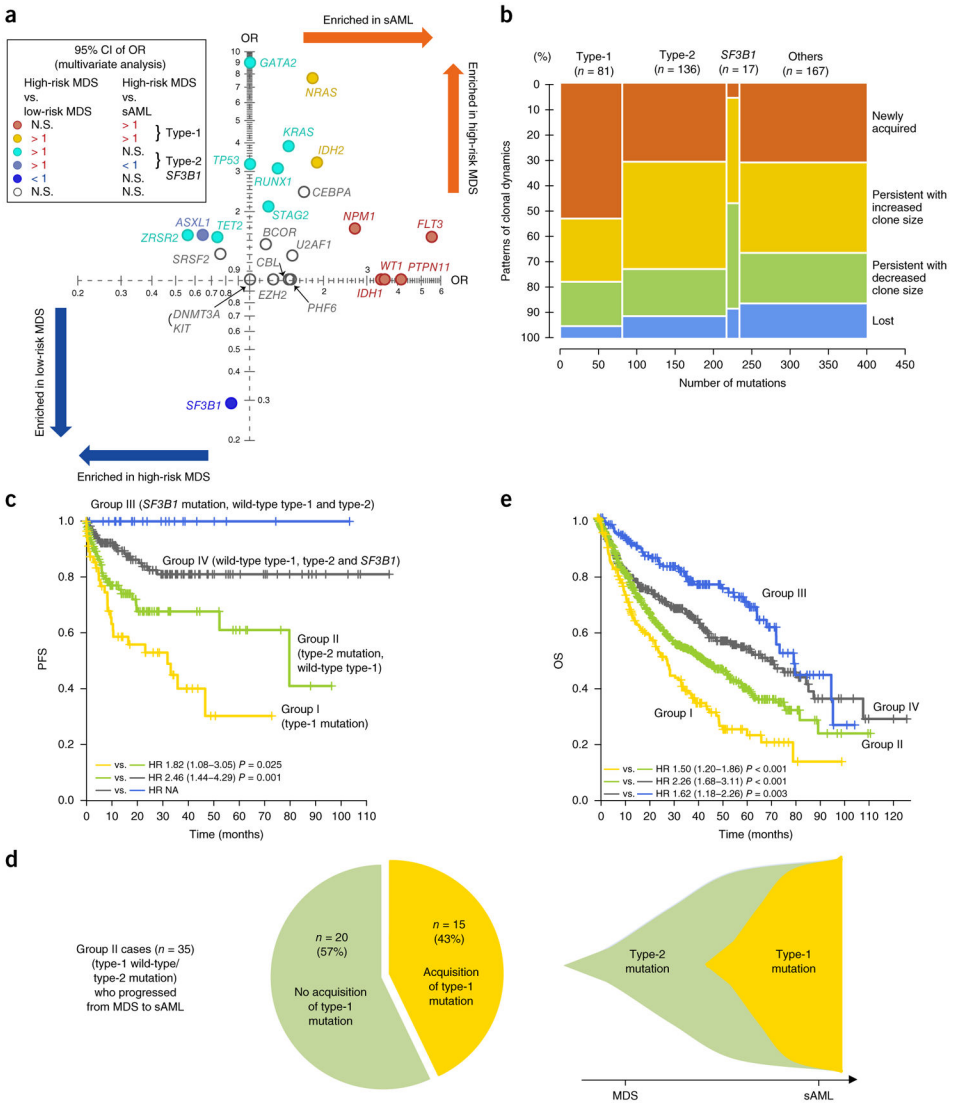

Figure 4.

Distinct sets of driver mutations in MDS and their impact on clinical outcomes. (**a**) Enrichment of driver mutations in sAML and high-risk MDS relative to high-risk and low-risk MDS, respectively. Enrichment is expressed as an odds ratio (OR) of mutation rates in sAML (n = 360) vs. high-risk MDS (n = 683) on the x axis and high-risk (n = 683) vs. low-risk (n = 1,207) on the y axis. Logistic regression analysis was applied to 25 driver genes measured in whole cohorts in 2,250 MDS and sAML patients, and the best model was selected by the least absolute shrinkage and selection operator. Mutations showing significant enrichment in either comparison are indicated by colors according to OR 95% CI limits being above (if OR >1) or below (if OR <1). According to their distinct enrichment patterns, mutations are classified into type 1 or type 2, as indicated. (**b**) Compositions of type-1, type-2 and other mutations are shown for each set of mutations that were newly acquired, that persisted with increased or decreased clone size, and that were lost in the second sampling. The Cochran-Armitage trend test was applied. (**c,e**) Kaplan-Meier curves

for PFS (n = 429) (**c**) and OS (n = 1,347) (**e**) of patients with type-1 mutations (group I), with type-2 but not type-1 mutations (group II), with *SF3B1* but no type-1 or type-2 mutations (group III), and other patients with no type-1, type-2 or *SF3B1* mutations (group IV). (**d**) Clonal dynamics in group-II cases (n = 35) during sAML progression. Type-1 mutations were acquired in 15 (43%) cases.

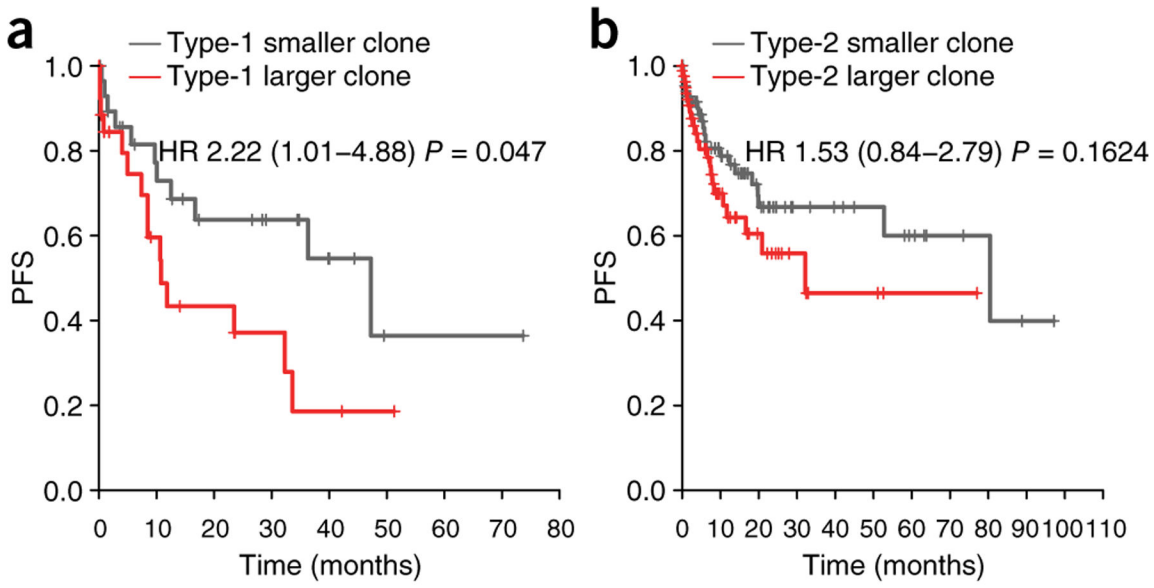

Figure 5.

Effects of clone size of driver mutations on prognosis. (**a,b**) Effects of clone size of type-1 (**a**) and type-2 (**b**) mutations on progression to sAML. Clones are dichotomized into larger and smaller ones by the median VAF of mutations and survivals are compared between patients with larger and smaller clones. In cases where multiple mutations in the corresponding category are observed, the maximum value of VAFs was used.

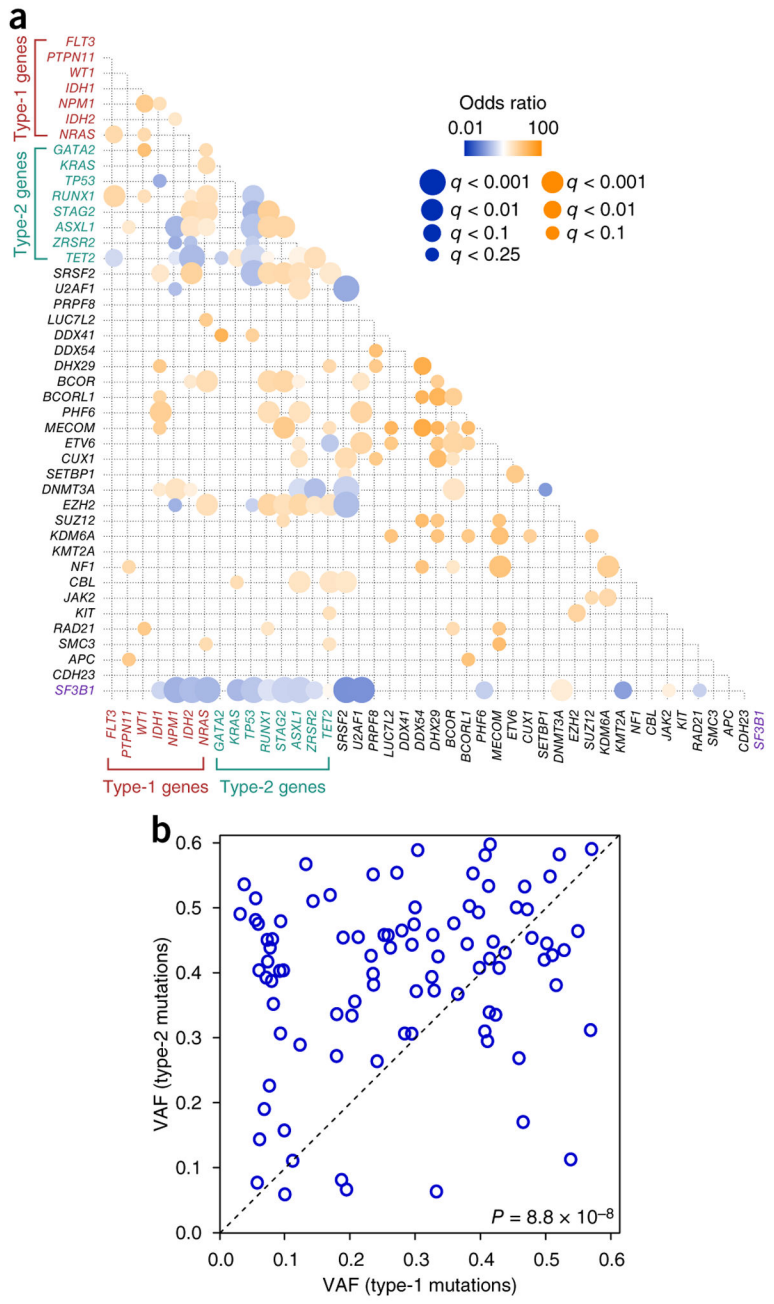

Figure 6.

Combination of type-1 and type-2 mutations in MDS. (**a**) Correlation between mutations found in 43 genes associated with MDS pathogenesis. Correlation coefficients and associated *q* values are indicated by the size of circles and color gradient as indicated. (**b**) Clone size of type-1 and type-2 mutations which were concomitantly identified in a case. Cases with both type-1 and type-2 mutations ($n = 93$) were represented in two-dimensional plots according to the VAFs of the mutations.

Table 1

Multivariate analysis of enrichment of mutations in disease subtypes

High-risk MDS vs. sAML (*n* = 683 vs. 360)			
Gene	**Odds ratio (95% CI)**		***P***
FLT3[a]	5.40	(2.71–10.8)	1.74×10^{-6}
PTPN11[a]	4.05	(1.89–8.71)	0.000333
WT1[a]	3.48	(1.23–9.86)	0.0191
IDH1[a]	3.36	(1.78–6.33)	0.000176
NPM1[a]	2.63	(1.31–5.27)	0.00646
IDH2[a]	1.84	(1.11–3.05)	0.0177
NRAS[a]	1.77	(1.12–2.79)	0.0141
ASXL1	0.63	(0.44–0.90)	0.0121
Low-risk vs. high-risk MDS (*n* = 1,207 vs. 683)			
Gene	**Odds ratio (95% CI)**		***P***
GATA2[b]	8.86	(2.90–27.1)	0.000131
NRAS	7.58	(3.26–17.6)	2.44×10^{-6}
KRAS[b]	3.82	(1.56–9.39)	0.00345
IDH2	3.24	(1.77–5.95)	0.000144
TP53[b]	3.19	(2.18–4.68)	2.73×10^{-9}
RUNX1[b]	3.06	(2.03–4.60)	8.24×10^{-8}
STAG2[b]	2.08	(1.33–3.26)	0.00133
ASXL1[b]	1.56	(1.17–2.09)	0.00234
ZRSR2[b]	1.56	(1.02–2.39)	0.0395
TET2[b]	1.53	(1.19–1.97)	0.000882
SF3B1	0.29	(0.21–0.39)	1.33×10^{-16}

[a]Type-1 genes.

[b]Type-2 genes.

For graphical depiction, see Figure 4a.

Table 2

Multivariate analysis of PFS and OS in clinical variables

Progression-free survival[a]			
Variables	**Hazard ratio (95% CI)**		***P***
IPSS (int-2 vs. low)	5.54	(2.26–13.6)	0.00018
Group I (vs. group IV)	3.12	(1.43–6.80)	0.00427
IPSS (int-1 vs. low)	2.80	(1.20–6.56)	0.0177
Group II (vs. group IV)	2.31	(1.17–4.57)	0.0161
Group III (vs. group IV)	NA	No events	NA
Overall survival[b]			
Variables	**Hazard ratio (95% CI)**		***P***
Complex karyotype	3.24	(2.42–4.33)	2.55×10^{-15}
IPSS (high vs. low)	2.30	(1.52–3.47)	7.93×10^{-5}
Group I (vs. group IV)	2.09	(1.55–2.81)	1.17×10^{-6}
−7/del(7q)	2.08	(1.44–3.03)	0.000113
Age (>60 year old)	2.03	(1.51–2.73)	3.30×10^{-6}
IPSS (int-2 vs. low)	1.96	(1.50–2.57)	9.98×10^{-7}
Del(20q)	1.64	(1.17–2.30)	0.00434
IPSS (int-1 vs. low)	1.56	(1.24–1.96)	0.00014

Variables include IPSS grades, cytogenetics, age, gender and groups of cases. The groups of cases are defined in Supplementary Table 6. NA, not applicable; int, intermediate.

[a] Analyzed in the cohort shown in Figure 4c.

[b] Analyzed in the cohort shown in Figure 4e.