ORIGINAL ARTICLE

Somatic *SF3B1* Mutation in Myelodysplasia with Ring Sideroblasts

E. Papaemmanuil, M. Cazzola, J. Boultwood, L. Malcovati, P. Vyas, D. Bowen, A. Pellagatti, J.S. Wainscoat, E. Hellstrom-Lindberg, C. Gambacorti-Passerini, A.L. Godfrey, I. Rapado, A. Cvejic, R. Rance, C. McGee, P. Ellis, L.J. Mudie, P.J. Stephens, S. McLaren, C.E. Massie, P.S. Tarpey, I. Varela, S. Nik-Zainal, H.R. Davies, A. Shlien, D. Jones, K. Raine, J. Hinton, A.P. Butler, J.W. Teague, E.J. Baxter, J. Score, A. Galli, M.G. Della Porta, E. Travaglino, M. Groves, S. Tauro, N.C. Munshi, K.C. Anderson, A. El-Naggar, A. Fischer, V. Mustonen, A.J. Warren, N.C.P. Cross, A.R. Green, P.A. Futreal, M.R. Stratton, and P.J. Campbell for the Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium

ABSTRACT

BACKGROUND

Myelodysplastic syndromes are a diverse and common group of chronic hematologic cancers. The identification of new genetic lesions could facilitate new diagnostic and therapeutic strategies.

METHODS

We used massively parallel sequencing technology to identify somatically acquired point mutations across all protein-coding exons in the genome in 9 patients with low-grade myelodysplasia. Targeted resequencing of the gene encoding RNA splicing factor 3B, subunit 1 (*SF3B1*), was also performed in a cohort of 2087 patients with myeloid or other cancers.

RESULTS

We identified 64 point mutations in the 9 patients. Recurrent somatically acquired mutations were identified in *SF3B1*. Follow-up revealed *SF3B1* mutations in 72 of 354 patients (20%) with myelodysplastic syndromes, with particularly high frequency among patients whose disease was characterized by ring sideroblasts (53 of 82 [65%]). The gene was also mutated in 1 to 5% of patients with a variety of other tumor types. The observed mutations were less deleterious than was expected on the basis of chance, suggesting that the mutated protein retains structural integrity with altered function. *SF3B1* mutations were associated with down-regulation of key gene networks, including core mitochondrial pathways. Clinically, patients with *SF3B1* mutations had fewer cytopenias and longer event-free survival than patients without *SF3B1* mutations.

CONCLUSIONS

Mutations in *SF3B1* implicate abnormalities of messenger RNA splicing in the pathogenesis of myelodysplastic syndromes. (Funded by the Wellcome Trust and others.)

The authors' full names, degrees, and affiliations are listed in the Appendix.

Address reprint requests to Dr. Campbell at Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom, or at pc8@sanger.ac.uk.

This article (10.1056/NEJMoa1103283) was published on September 26, 2011, at NEJM.org.

N Engl J Med 2011;365:1384-95.

Copyright © 2011 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

The myelodysplastic syndromes are a heterogeneous group of hematologic cancers characterized by low blood counts, most commonly anemia, and a risk of progression to acute myeloid leukemia.[1] These disorders have increased in prevalence and are expected to continue to do so. Blood films and bone marrow–biopsy specimens from patients with myelodysplastic syndromes show dysplastic changes in myeloid cells, with abnormal proliferation and differentiation of one or more lineages. Target genes of recurrent chromosomal aberrations have been mapped,[2,3] and several genes have been identified as recurrently mutated in these disorders, including *NRAS* (encoding neuroblastoma RAS viral oncogene homologue), *TP53* (encoding tumor protein p53), *RUNX1* (encoding runt-related transcription factor 1), *CBL* (encoding Cas-Br-M ecotropic retroviral transforming sequence),[4,5] *TET2* (encoding tet oncogene family member 2),[6,7] *ASXL1* (encoding additional sex combs–like protein 1),[8,9] and *EZH2* (encoding enhancer of zeste homologue 2).[10] With the exception of *TET2*, most of these genes are mutated in no more than 5 to 15% of cases, and generally the mutation rates are lower in the more benign subtypes of the disease.

The myelodysplastic syndromes can be divided into several categories on the basis of bone marrow and peripheral-blood morphologic characteristics and cytogenetic changes.[11] In low-risk disease, such as refractory anemia, cytopenias are the major clinical challenge, whereas high-risk disease, such as refractory anemia with excess blasts, is characterized by both cytopenias and a high rate of transformation to acute myeloid leukemia. More than a quarter of patients with myelodysplastic syndromes have large numbers of ring sideroblasts in the bone marrow,[12] a sufficiently distinctive morphologic abnormality to warrant a separate designation. Ring sideroblasts are characteristically seen on iron staining of bone marrow aspirates as differentiating erythroid cells with a complete or partial ring of iron-laden mitochondria surrounding the nucleus. Several genetic lesions underpinning inherited sideroblastic anemias have been identified,[13] including loss-of-function mutations in the genes *ALAS2* (encoding delta aminolevulinate synthase 2), *ABCB7* (encoding ATP-binding cassette, subfamily B, member 7), and *SLC25A38* (solute carrier family 25, member 38). The pathogenesis of ring sideroblasts in myelodysplastic syndromes, however, remains obscure, although gene-expression studies have revealed up-regulation of genes involved in heme synthesis (including *ALAS2*) and down-regulation of *ABCB7*.[14,15]

We reasoned that the identification of recurrently mutated cancer genes in low-grade myelodysplastic syndromes could prove useful for the diagnosis of these disorders and provide new insights into the molecular pathogenesis of these syndromes.

METHODS

STUDY CONDUCT

The authors designed the study and wrote the manuscript on behalf of the Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Data were collected and analyzed by the authors from the Wellcome Trust Sanger Institute and four other authors. All authors reviewed the manuscript and vouch for the completeness and accuracy of the data collection and analysis. Genome sequence data have been deposited at the European Genome–Phenome Archive (www.ebi.ac.uk/ega) (accession number EGAS00001000089).

STUDY SAMPLES

Samples were obtained from patients with myeloid dysplastic syndromes or other cancers who provided written informed consent. Appropriate ethics-committee approval was obtained. Genomic DNA specimens were obtained from bone marrow mononuclear cells or peripheral-blood granulocytes from patients with myeloid dysplastic syndromes, and constitutional DNA samples were obtained from buccal swabs or immunomagnetically purified T cells. Myeloid dysplastic syndromes were classified according to World Health Organization (2008) categories,[11] and ring sideroblastosis was defined as more than 15% of erythroblasts containing at least 10 siderotic granules encircling more than a third of the nucleus. Laboratory data at the time of DNA sampling, as well as subsequent data on clinical outcomes, were available for 123 patients.

DNA SEQUENCING

For exome and follow-up sequencing, libraries were prepared from nonamplified tumor DNA and whole-genome–amplified constitutional DNA samples according to standard protocols[16,17] (see the Supplementary Appendix, available with the full text of this article at NEJM.org).

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

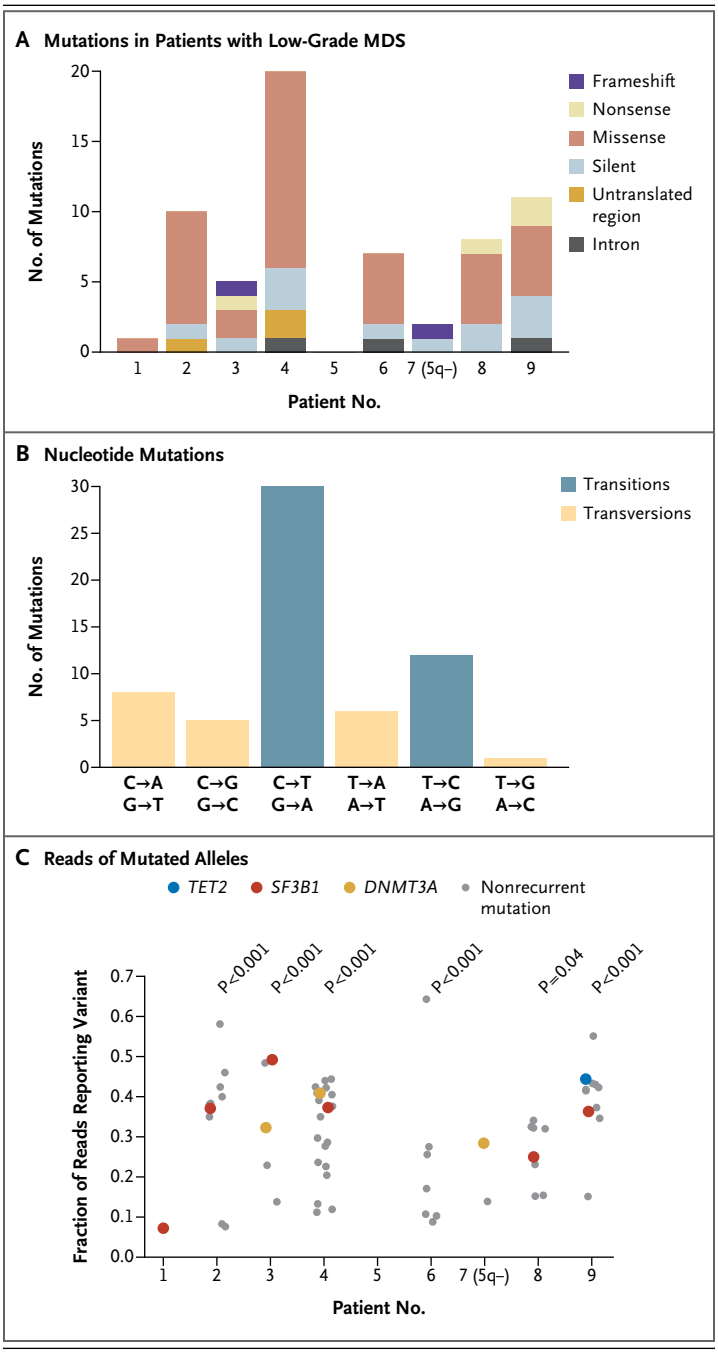


Figure 1. Exome Sequencing in Nine Patients with Low-Grade Myelodysplastic Syndromes (MDS).

Panel A shows the distribution of numbers and categories of somatically acquired point mutations among the nine patients. No mutations were found for Patient 5. Panel B shows the mutation spectrum for somatically acquired point mutations. Panel C shows the fraction of reads reporting mutated alleles from exome-sequencing data for each patient. No mutations were found for Patient 5. Mutations in known MDS genes or recurrently mutated genes identified in this screen are shown as colored points, with nonrecurrent mutations as gray points. P values were calculated with the use of chi-square tests of heterogeneity in observed allelic ratios for mutations in patients with more than two mutations.

GENE-EXPRESSION PROFILING

RNA from immunomagnetically purified CD34+ bone marrow cells was previously profiled on microarrays (U133-plus 2.0, Affymetrix),[18] and 56 patients were genotyped for *SF3B1* mutations. RNA from 12 samples in this cohort was also profiled on microarrays (SurePrint G3 Human Exon 2x400k, Agilent), according to the manufacturer's protocol.

STATISTICAL ANALYSIS

Statistical analysis was performed with the use of standard methods, as described in the Supplementary Appendix. When reported, q values denote the minimum false discovery rate at which the test may be called significant.

RESULTS

MUTATIONS IN PROTEIN-CODING GENES

In nine patients with low-grade myelodysplastic syndromes — eight who had refractory anemia with ring sideroblasts and one who had the chromosome 5q– syndrome — 64 mutations (Table 1 in the Supplementary Appendix) were found, ranging from 0 to 20 per patient (Fig. 1A). Of these mutations, 2 were frameshift insertion–deletions (indels) and 62 were substitutions; 58 were found in coding sequences, 3 in introns within 10 bp of splice junctions (but not essential splice sites), and 3 in untranslated regions. The mutation spectrum showed a predominance of transitions, especially C→T and G→A mutations (Fig. 1B). This spectrum is similar overall to those observed in colorectal, pancreatic, and brain cancers.[19,20]

Each read of a massively parallel sequencing run derives from a single molecule of genomic DNA. Thus, the proportion of sequencing reads reporting a variant allele provides a quantitative estimate of the proportion of cells in the DNA sample carrying that mutation.[17,21] In five of the nine patients, the observed proportion of reads reporting a mutant allele showed significantly greater variability than was expected on the basis of chance (Fig. 1C). For example, for Patient 3, the fraction of reads reporting each mutation

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

ranged from 32 of 65 (49%) down to 8 of 58 (14%). These data suggest that the population of malignant cells in low-grade myelodysplastic syndromes is often genetically heterogeneous, with some mutations restricted to subclones of the neoplasm, as has been described in other cancers.[17,21-24]

We identified 46 mutations that were predicted to alter the protein-coding sequence (Table 1). Of these, 44 were nonsynonymous substitutions (including 4 nonsense substitutions) and 12 were silent substitutions. Two known cancer genes had somatic mutations in the cohort. The first of these, *DNMT3A* (encoding the DNA methyltransferase 3 alpha protein), has been reported to be recurrently mutated in patients with acute myeloid leukemia and myelodysplastic syndromes[25-27] and was mutated in three of our nine patients (33%), with two frameshift indels and one missense mutation (Fig. 1 in the Supplementary Appendix). The other known cancer gene with a somatic mutation that we identified was *TET2*, which had a heterozygous substitution causing a premature stop codon, Q644*, in one patient.

RECURRENT MUTATIONS IN *SF3B1*

We identified recurrent somatic mutations in a gene that encodes a core component of the RNA splicing machinery — *SF3B1* — in six of the nine patients with myelodysplastic syndromes (Table 1, and Fig. 2 in the Supplementary Appendix). Four patients carried A→G mutations that would generate the same K700E mutation in the predicted protein, and two patients carried C→A or C→G mutations, both with a predicted H662Q protein consequence. On the basis of the proportion of reads reporting the mutant allele, the mutations all appeared to be heterozygous and present in the dominant clone of cells (Fig. 1C).

To characterize the spectrum and frequency of *SF3B1* mutations in greater detail, both in myeloid cancers and other cancers, we performed targeted resequencing of the gene in 2087 samples (Table 2, and Table 2 in the Supplementary Appendix). Among 354 patients with myelodysplastic syndromes, 72 had *SF3B1* mutations (20%). Mutations were particularly common in patients with subtypes of myelodysplastic syndromes in which ring sideroblasts are a prominent feature, with 53 of 82 patients (65%) positive for *SF3B1* changes. The subtypes represented in these patients included both refractory anemia with ring sideroblasts (with mutations found in 40 of 59 patients [68%]) and refractory cytopenia with multilineage dysplasia and ring sideroblasts (with mutations in 13 of 23 patients [57%]). Mutations in *SF3B1* were found at a lower rate in other subtypes of myelodysplastic syndromes, with mutations found in 9 of 91 patients (10%) with refractory anemia, 3 of 53 (6%) with refractory cytopenia and multilineage dysplasia, and 6 of 110 (5%) with refractory anemia and excess blasts.

SF3B1 mutations were noted in other myeloid cancers, including acute myeloid leukemia (in 3 of 57 patients [5%]), primary myelofibrosis (6 of 136 [4%]), essential thrombocythemia (6 of 189 [3%]), and chronic myelomonocytic leukemia (5 of 106 [5%]) (Table 2). *SF3B1* mutations were also seen in 1 to 5% of patients with other types of tumor (Table 2): breast cancer (in 2 of 172 patients [1%]), renal cancer (1 of 30 [3%]), chronic lymphocytic leukemia (2 of 40 [5%]), multiple myeloma (1 of 32 [3%]), and adenoid cystic carcinoma (1 of 27 [4%]). In addition, among 746 cancer cell lines,[28] we found variants in 8 (1%): melanoma (2 lines), lung cancer (1), bladder cancer (1), breast cancer (1), endometrial cancer (1), chronic myeloid leukemia (1), and teratoma (1).

The distribution of observed mutations across the gene was striking (Fig. 2). All mutations appeared to be heterozygous substitutions. No frameshift indels, splice-site mutations, or nonsense substitutions were seen. The mutations clustered in exons 12 to 15 of the gene, and 1 variant in particular, K700E, accounted for 59 of the 108 variants observed (55%; 95% confidence interval [CI], 45 to 64). Several other amino acid residues in this region were also hot spots for mutation, including E622 (5 mutations), R625 (7), H662 (7), K666 (13), and I704 (3) (Fig. 2).

Splicing of messenger RNA is carried out by the spliceosome, a complex of five small nuclear ribonucleoproteins (snRNPs) together with other proteins.[29] The SF3B1 protein is a core component of one snRNP, the U2 snRNP, which recognizes the 3′ splice site at intron–exon junctions. The *SF3B1* gene encodes a protein with an N-terminal domain involved in protein–RNA and protein–protein interactions, together with a C-terminal region consisting of 22 so-called HEAT domains (Huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast PI3-kinase TOR1). The mutations we have identified cluster most strongly in the fourth, fifth, and sixth HEAT domains

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

Table 1. Somatically Acquired Indels, Missense, and Nonsense Substitutions in Eight of the Nine Study Patients with Low-Grade Myelodysplastic Syndromes in Whom Mutations Were Found.

Patient No. and Mutation	DNA Change	Protein Change
5q–		
Patient 7		
DNMT3A	c.2247delG	p.R748fs
Refractory anemia with ring sideroblasts		
Patient 1		
SF3B1	c.1986C→A	p.H662Q
Patient 2		
IMP5	c.43C→A	p.L15I
PCDHGA1	c.2069C→T	p.T690I
TEX15	c.4018G→A	p.A1340T
COL17A1	c.2408G→A	p.G803E
SEMA6A	c.440A→G	p.Y147C
NPSR1	c.211C→G	p.L71V
SSR2	c.325A→G	p.T109A
SF3B1	c.2098A→G	p.K700E
Patient 3		
GABRA4	c.644C→T	p.P215L
DNMT3A	c.2273insA	p.V758fs
IRF1	c.195C→A	p.Y65*
SF3B1	c.2098A→G	p.K700E
Patient 4		
PREB	c.546G→C	p.K182 N
RAD52	c.359G→T	p.W120L
DNMT3A	c.1528G→A	p.G510S
CTSL1	c.742G→T	p.A248S
TMEM132C	c.2134C→T	p.R712C
TEAD1	c.35T→G	p.I12S
DMD	c.4829A→T	p.E1610V
FHL2	c.173A→T	p.D58V
ODZ3	c.5885A→G	p.D1962G
NAA16	c.1406C→T	p.T469I
SYNE2	c.8249T→A	p.L2750H
ZNF334	c.887G→A	p.S296 N
CSTF1	c.1047A→T	p.L349F
SKAP2	c.718G→C	p.V240L
SF3B1	c.2098A→G	p.K700E
Patient 6		
CENPE	c.1556C→T	p.T519I
UBA1	c.1861A→T	p.S621C
SDCCAG1	c.2764G→A	p.V922 M
CDH13	c.1559C→G	p.P520R

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

Table 1. (Continued.)

Patient No. and Mutation	DNA Change	Protein Change
Patient 8		
EFCAB8	c.595C→T	p.Q199*
SF3B1	c.1986C→G	p.H662Q
CNNM1	c.2047G→A	p.V683I
TMEM8C	c.377T→C	p.I126T
ARHGEF10L	c.994G→T	p.V332L
SLC17A8	c.944C→T	p.P315L
Patient 9		
TET2	c.1930C→T	p.Q644*
ABCG8	c.904C→T	p.Q302*
SF3B1	c.2098A→G	p.K700E
HNRNPCL1	c.163G→A	p.V55I
SLIT2	c.988G→A	p.A330T
CTNND2	c.1815G→A	p.M605I
CYP4F11	c.559G→A	p.A187T

* Asterisks denote premature stop codons. See Table 1 in the Supplementary Appendix for more details.

(Fig. 2). The structure of a multiprotein U2 snRNP subcomplex containing SF3B1 reveals that the 22 tandem helical HEAT repeats wrap in an S-shape around the outer surface of the complex.[30,31] The sixth HEAT domain falls at the hinge of this shell-like structure.

To explore the patterns of the observed amino acid substitutions, we scored the potential degree to which the missense mutations were deleterious, on the basis of a multiple alignment of HEAT domains from the Pfam database of protein families (http://pfam.sanger.ac.uk).[32,33] The scoring reveals that, as a set, the mutations are significantly less deleterious than random in silico–generated missense mutations (P<0.001) (Fig. 3A). In contrast, observed mutations in classic tumor-suppressor genes are, on average, significantly more deleterious than simulated variants; examples include *PBRM1* (encoding polybromo 1) (P=0.01) (data not shown), as well as *NF1* (encoding neurofibromin 1) (P=0.009) and *PTEN* (encoding the phosphatase and tensin homologue) (P=0.001) (Fig. 3A).[16,33] Mutations targeting one residue, R625, were predicted to be deleterious (Fig. 3A in the Supplementary Appendix); the recurrence of these mutations in seven patients indicates their probable oncogenic significance. Even when the most common K700E mutation was excluded from the analysis, the observed mutations remained significantly less deleterious than expected (P=0.01) (Fig. 3B in the Supplementary Appendix). Indeed, when the mutations in *SF3B1* are mapped onto the stacked consensus sequence[34] for the 22 HEAT domains (Fig. 4 in the Supplementary Appendix), they tend to avoid the key structural amino acids or even improve alignment with the consensus. The results are similar with other prediction algorithms, such as Polyphen and Sorting Intolerant from Tolerant (SIFT) (Table 3 in the Supplementary Appendix). These data, coupled with the absence of nonsense, splice-site, and frameshift mutations, suggest that the mutated SF3B1 protein is likely to retain structural integrity, albeit with presumably altered function.

GENE-EXPRESSION PROFILES OF MUTATED *SF3B1*

We analyzed gene-expression profiles[18] of CD34+ bone marrow cells purified from samples obtained from 56 patients with myelodysplastic syndromes, 12 (21%) of whom had *SF3B1* mutations. We used gene-set enrichment analysis[35] to identify biologic pathways and processes that showed coordinated up-regulation or down-regulation in patients with *SF3B1* mutations, after adjustment for differences due to disease subtype. With a false discovery rate of less than 10%, we identified 94 gene sets (of

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

Table 2. Variants in *SF3B1* in Patients with Myeloid or Other Cancers.

Tumor Type	Mutations		Comments
	no. of patients/total no.	*% (95% CI)*	
Myelodysplastic syndromes	72/354	20 (16–25)	Whole gene screened
Refractory anemia	9/91	10 (5–18)	
Refractory anemia with ring sideroblasts	40/59	68 (54–79)	8 Mutations were proven somatic
Refractory cytopenia with multilineage dysplasia	3/53	6 (1–16)	
Refractory cytopenia with multilineage dysplasia and ring sideroblasts	13/23	57 (35–77)	
Refractory anemia with excess blasts	6/110	6 (2–12)	
Other subtypes	1/18	6 (0–27)	
Other chronic myeloid disorders	18/629	3 (2–5)	Exons 12–15 screened
Polycythemia vera	0/95		
Essential thrombocythemia	6/189	3 (1–7)	
Primary myelofibrosis	6/136	4 (2–9)	2 Mutations were somatic
Chronic myeloid leukemia in transformation	0/53		
Atypical chronic myeloid leukemia	1/50	2 (0–11)	
Chronic myelomonocytic leukemia	5/106	5 (2–11)	
Acute myeloid leukemia	3/57	5 (1–15)	Exons 12–15 screened
Other tumors	15/1047	1 (1–2)	Exons 12–15 screened
Breast cancer	2/172	1 (0–4)	1 Mutation was proven somatic
Renal cancer	1/30	3 (0–17)	1 Mutation was proven somatic
Chronic lymphocytic leukemia	2/40	5 (1–17)	1 Mutation was proven somatic
Multiple myeloma	1/32	3 (0–16)	1 Mutation was proven somatic
Adenoid cystic carcinoma	1/27	4 (0–19)	1 Mutation was proven somatic
Cancer cell lines	8/746	1 (1–2)	2 Mutations were proven somatic

1673 screened) showing significant enrichment, all of which were down-regulated in patients with *SF3B1* mutations (Fig. 3B, and Table 4 in the Supplementary Appendix). Of the 50 most down-regulated gene sets in patients with *SF3B1* mutations, 7 involved key pathways determining mitochondrial function (Fig. 3C). Although these gene sets do partially overlap, genes involved in the mitochondrial ribosome ($q<0.001$) and in the electron transport chain ($q<0.001$) were notably down-regulated in patients with *SF3B1* mutations.

To explore whether these changes were due to abnormal messenger RNA splicing, we undertook exon-specific expression profiling by using exon microarrays in 12 patients, 6 of whom had *SF3B1* mutations. Overall, 20 genes showed differences in exon usage between patients with and those without *SF3B1* mutations (Table 5 and Fig. 5 in the Supplementary Appendix), although none of these genes has obvious relevance to myelodysplastic syndromes, and the number of genes is small relative to the thousands of genes expressed in CD34+ cells. We did not find consistent abnormalities of splicing across the transcriptome globally or specifically in genes involved in mitochondrial function in patients with *SF3B1* mutations.

Taken together, the findings on transcriptome profiling suggest that *SF3B1* mutation is associated with systematic down-regulation of essential mitochondrial gene networks. The mechanism of down-regulation is not clear, given that we did not find consistent abnormalities of splicing in patients with *SF3B1* mutations. The expression profiles were derived from undifferentiated CD34+ hematopoietic progenitor cells, in which mitochondrial ferritin first appears in patients with refractory anemia and ring sideroblasts[36]; such cells are more immature than ring sideroblasts. The implication is that transcriptional changes affecting mitochondrial pathways precede the ap-

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

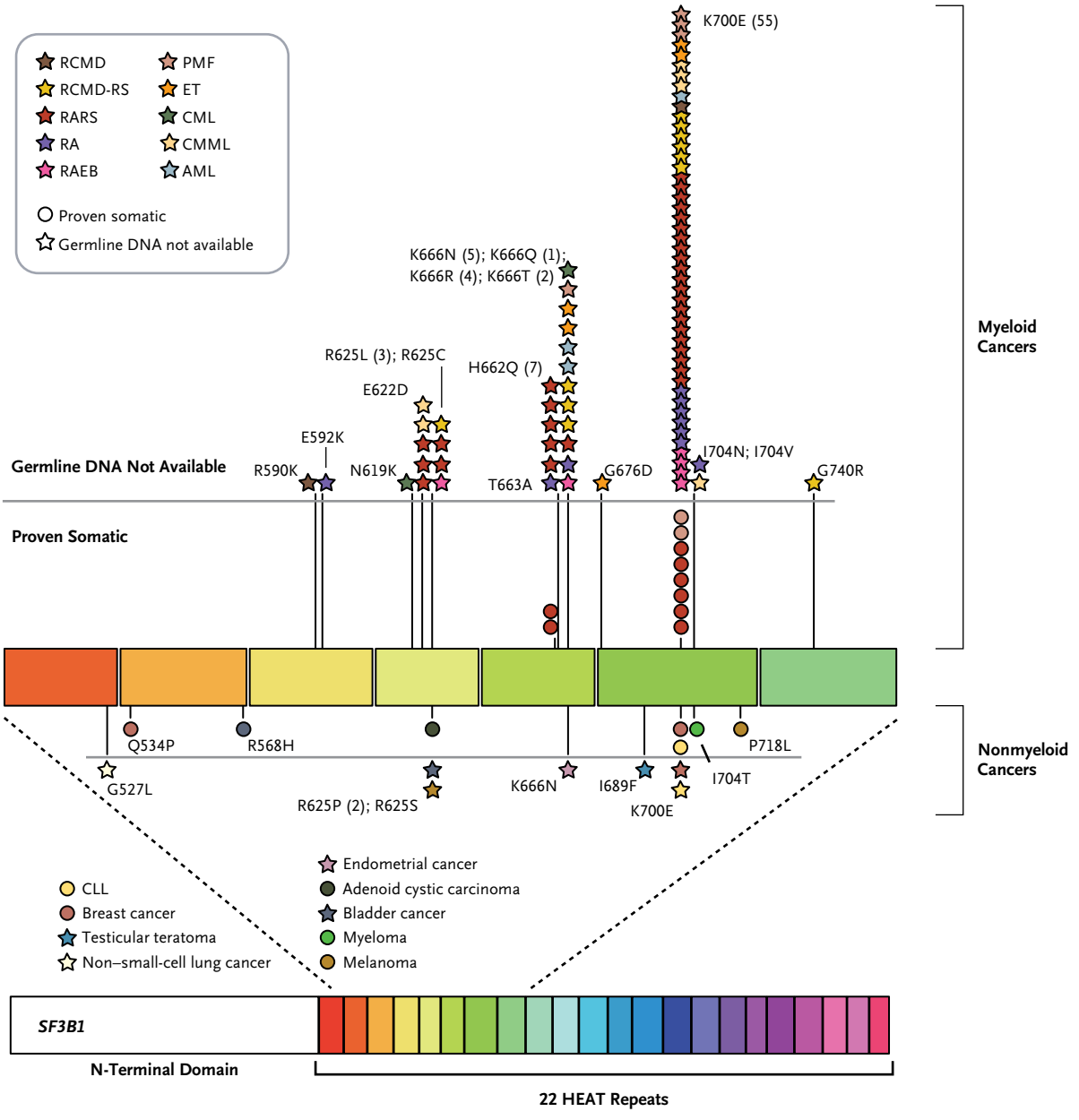


Figure 2. Distribution of Missense Mutations in *SF3B1*.

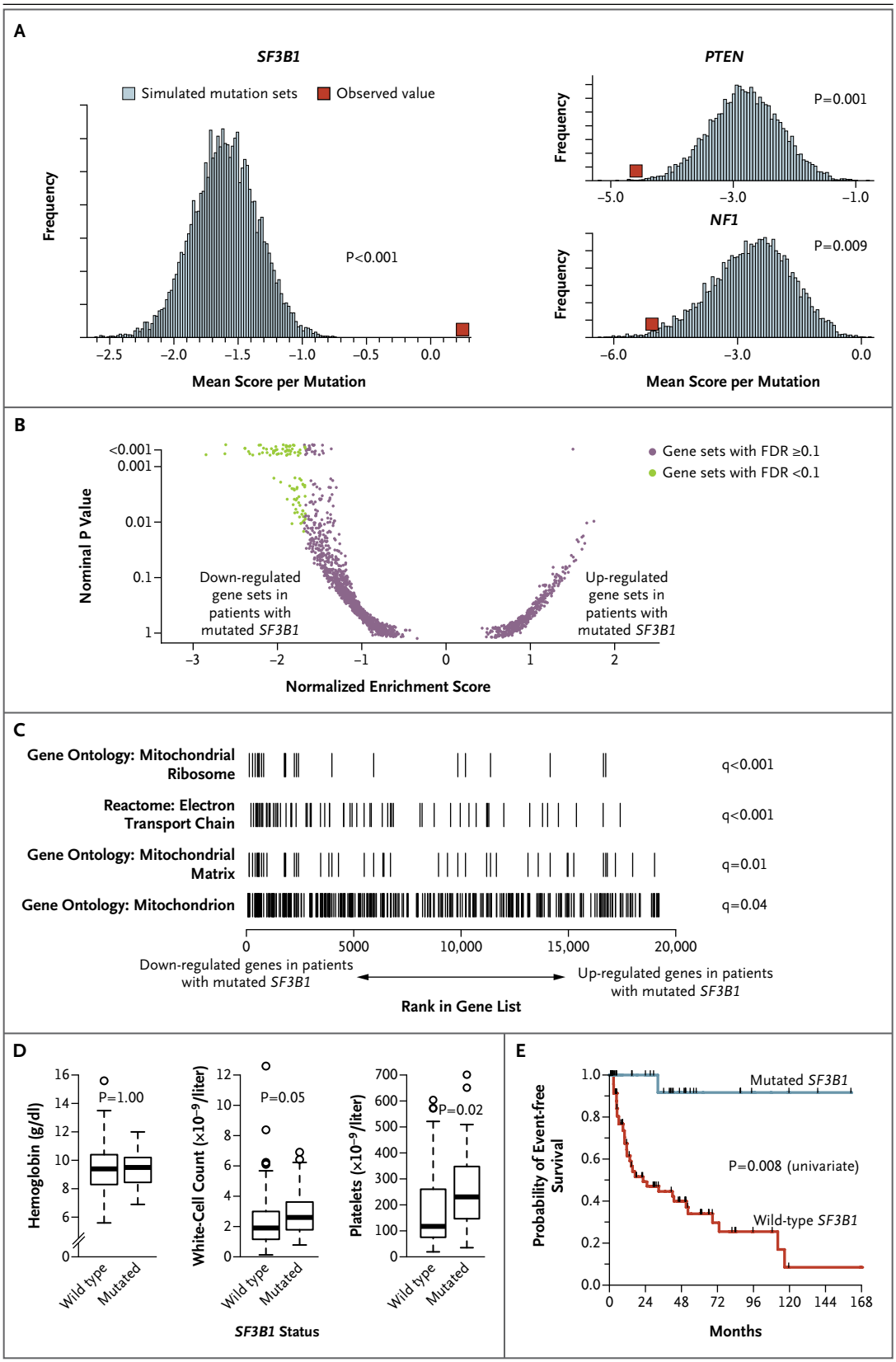
The *SF3B1* gene encodes a protein with an N-terminal domain together with a C-terminal region consisting of 22 so-called HEAT domains (Huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast PI3-kinase TOR1). Two mutations (R1041H and E491G) fell outside the second through the eighth HEAT domains and are not shown. AML denotes acute myeloid lymphoma, CLL chronic lymphocytic leukemia, CML chronic myeloid leukemia, CMML chronic myelomonocytic leukemia, ET essential thrombocythemia, PMF primary myelofibrosis, RA refractory anemia, RAEB refractory anemia with excess blasts, RARS refractory anemia with ring sideroblasts, RCMD refractory cytopenia with multilineage dysplasia, and RCMD-RS RCMD and ring sideroblasts.

pearance of iron-laden mitochondria during erythroid development and are unlikely to be merely a consequence of dysfunctional mitochondria.

CLINICAL PHENOTYPE OF *SF3B1* MUTATION

Data on clinical outcome and laboratory features at the time of DNA sampling were available for

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.



The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

Figure 3 (facing page). Modeling of Mutations and Results of Clinical Studies.

Panel A shows bars representing the mean scores for computer-generated random missense mutations. The height of the bars represents the frequency histogram of this null distribution. The red squares indicate the mean scores for the somatic mutations on exons 12 through 15 that could be scored, with higher scores representing effects predicted to be less deleterious than those of random mutations. The somatic set has a significantly higher mean score than the null set for *SF3B1* mutations (P<0.001). This is in contrast to the tumor-suppressor genes *PTEN* (encoding the phosphatase and tensin homologue) and *NF1* (encoding neurofibromin 1), where observed mutations have a significantly lower score than the null set. Panel B shows a volcano plot of the distribution of normalized enrichment scores for various gene sets according to their nominal P values. Negative enrichment scores indicate gene sets that are down-regulated in patients with an *SF3B1* mutation as compared with patients without such a mutation. Gene sets with significant enrichment and a false discovery rate (FDR) of less than 10% are shown in green. Panel C shows enrichment plots for four mitochondrial gene sets. Each vertical stripe represents the rank of a gene within the gene set, among all 19,578 genes represented on the microarray. A low rank indicates genes that are especially down-regulated among patients with *SF3B1* mutations as compared with patients without such mutations, after adjustment for disease type. The name of the gene set is shown to the left, and the q value (the minimum false discovery rate at which the test may be called significant) is shown to the right. Panel D shows box-and-whisker plots for hemoglobin levels, white-cell counts, and platelet counts in patients with myelodysplastic syndromes according to the presence or absence of *SF3B1* mutations. The central horizontal line within each box indicates the median, with the top and bottom edges of each box indicating the interquartile range. The I bars extend to 1.5 times the interquartile range, with the circles indicating outlier data for individual patients. Panel E shows Kaplan–Meier curves for event-free survival during the study period among patients with myelodysplastic syndromes, according to the presence or absence of *SF3B1* mutations.

123 patients with myelodysplastic syndromes, of whom 34 were positive for *SF3B1* mutations. *SF3B1* mutation was associated with the syndrome subtypes defined by ring sideroblasts, refractory anemia with ring sideroblasts, and refractory cytopenia with multilineage dysplasia and ring sideroblasts (P<0.001). Of the 34 patients, 8 did not have the latter two syndrome subtypes. Of these 8 patients, 2 had more than 15% ring sideroblasts and also had excess blasts (so they were considered to have refractory anemia with excess blasts), and 2 others did have ring sideroblasts but at a level of less than 15%.

As compared with patients who did not have an *SF3B1* mutation, patients with an *SF3B1* mutation had a higher median white-cell count (2.0×10^9 vs. 2.61×10^9 per liter, P=0.05) (Fig. 3D), a higher median platelet count (117×10^9 vs. 242×10^9 per liter, P=0.02) (Fig. 3D), more marked bone marrow erythroid hyperplasia (28% vs. 40% erythroblasts and a myeloid:erythroid ratio of 2.5 vs. 1.5; P=0.009 and P=0.007, respectively), and a lower proportion of bone marrow blasts (4% vs. 1%, P<0.001). However, the median hemoglobin level was the same in those with and those without a mutation (9.5 g per deciliter; P=1.00) (Fig. 3D).

In an analysis of the composite end point of leukemic progression or death, patients with an *SF3B1* mutation, as compared with those without an *SF3B1* mutation, had significantly longer overall survival (P=0.01), leukemia-free survival (P=0.05), and event-free survival (P=0.008) (Fig. 3E). After adjustment for the effects of age, sex, and karyotype, the presence of an *SF3B1* mutation was still significantly associated with longer event-free survival (hazard ratio, 0.1; 95% CI, 0.0 to 0.7; P=0.02). These findings suggest that *SF3B1* mutations are associated with relatively benign myelodysplastic syndromes characterized phenotypically by the presence of ring sideroblasts.

DISCUSSION

Recurrent mutation of the *SF3B1* gene was found in 20% of patients with myelodysplastic syndromes. Mutations were found in 65% of patients whose disease was characterized by the presence of ring sideroblasts, although the clonal dominance of mutations in blood or bone marrow granulocytic cells suggests that oncogenic effects may not be restricted to the erythroid lineage. Even among patients with other subtypes of myelodysplastic syndromes, those with *SF3B1* mutations frequently had large numbers of ring sideroblasts in their bone marrow.

The absence of frameshift, nonsense, and splice-site mutations, the lack of key structural amino acid residues as sites for mutation, and the fact that the mutations are less deleterious than expected on the basis of chance all suggest that the mutant SF3B1 protein retains structural integ-

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

rity and some function. It is increasingly recognized that initial splicing occurs as the nascent RNA molecule is being transcribed, an integrated process in which the spliceosome is in continuous cross-talk with proteins involved in the initiation, elongation, and termination phases of the transcription cycle.[29] *SF3B1* mutations could influence either splicing itself or interactions with the transcriptional complex. CD34+ cells from patients with *SF3B1* mutations show underexpression of several key biologic pathways, including those involved in mitochondrial function, although a detailed mechanistic understanding will require further biochemical studies. Mutations in pathways regulating RNA processing and protein homeostasis have been described in multiple myeloma.[37]

Refractory anemia with ring sideroblasts generally has a relatively benign clinical course.[38] We have found that patients with myelodysplastic syndromes and *SF3B1* mutations have higher neutrophil and platelet counts, fewer bone marrow blasts, and longer event-free survival than patients with these syndromes who do not have *SF3B1* mutations. The prognostic effect is independent of variables that could coexist at the time the mutations are acquired (age, sex, and cytogenetic abnormalities), indicating that *SF3B1* mutations define a benign clinical phenotype. These mutations can be readily identified in peripheral-blood DNA, whereas detection of ring sideroblasts requires bone marrow samples. We speculate that it may be feasible to identify patients who have myelodysplastic syndromes with a benign prognosis on the basis of screening for *SF3B1* mutations, without the need for an invasive bone marrow biopsy. As we piece together the genomic architecture of myelodysplastic syndromes, it may be possible to develop assays for causative driver mutations, leading to definitive diagnoses, from a single blood sample.

Supported by grants from the Wellcome Trust (077012/Z/05/Z, for the overall study, as well as WT088340MA, to Dr. Campbell), the Kay Kendall Leukaemia Fund, Leukemia Lymphoma Research (for the overall study and to Drs. Boultwood, Green, Vyas, and Wainscoat), the Adenoid Cystic Carcinoma Research Foundation, the Medical Research Council (MRC) (to Dr. Warren), the Oxford National Institutes for Health Research Biomedical Research Centre (to Drs. Boultwood, Vyas, and Wainscoat), the Swedish Cancer Society (to Dr. Hellstrom-Lindburg), the International Human Frontier Science Program Organization (to Dr. Varela), the Department of Veterans Affairs and the National Institutes of Health (R01-124929, P01-155249, P50-100007, and P01-78378, to Drs. Munshi and Anderson), the Association for International Cancer Research and the Leukemia Lymphoma Society (to Drs. Warren and Green), Associazione Italiana per la Ricerca sul Cancro (to the University of Pavia, the University of Milan Bicocca, and Dr. Cazzola), and Fondazione Cariplo (to the University of Pavia and the University of Milan Bicocca).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the National Institute for Health Research Cambridge Biomedical Research Centre (especially Drs. Anthony Bench and Wendy Erber), the Tayside Tissue Bank (especially Dr. Susan Bray), and Dr. Alessandra Pirola (Milan) for assisting with sample banking and processing; Dr. Maria Jose Calasanz of the University of Navarra, Spain, for contributing samples; Dr. Rocco Piazza for contributing clinical data; and Dr. Willem Ouwehand for discussions and a critical reading of an earlier version of the manuscript.

APPENDIX

The authors' full names and degrees are as follows: Elli Papaemmanuil, Ph.D., Mario Cazzola, M.D., Jacqueline Boultwood, Ph.D., Luca Malcovati, M.D., Paresh Vyas, F.R.C.Path., D.Phil., David Bowen, M.D., Andrea Pellagatti, Ph.D., James S. Wainscoat, M.D., Eva Hellstrom-Lindberg, M.D., Carlo Gambacorti-Passerini, M.D., Anna L. Godfrey, M.R.C.P., F.R.C.Path., Inmaculada Rapado, Ph.D., Ana Cvejic, Ph.D., Richard Rance, B.Sc., Chris McGee, M.Sc., Peter Ellis, Ph.D., Laura J. Mudie, B.Sc., Philip J. Stephens, D.Phil., Stuart McLaren, Dip.Mgmt., Charles E. Massie, Ph.D., Patrick S. Tarpey, Ph.D., Ignacio Varela, Ph.D., Serena Nik-Zainal, M.D., Helen R. Davies, Ph.D., Adam Shlien, Ph.D., David Jones, M.Sc., Keiran Raine, M.Sc., Jon Hinton, M.Sc., Adam P. Butler, B.Sc., Jon W. Teague, M.Sc., E. Joanna Baxter, Ph.D., Joannah Score, Ph.D., Anna Galli, B.Sc., Matteo G. Della Porta, M.D., Erica Travaglino, B.Sc., Michael Groves, M.I.B.M.S., Sudhir Tauro, F.R.C.Path., Nikhil C. Munshi, M.D., Kenneth C. Anderson, M.D., Ph.D., Adel El-Naggar, M.D., Andrej Fischer, M.Sc., Ville Mustonen, D.Phil., Alan J. Warren, Ph.D., F.R.C.Path., Nicholas C.P. Cross, Ph.D., F.R.C.Path., Anthony R. Green, F.R.C.Path., F.Med.Sci., P. Andrew Futreal, Ph.D., Michael R. Stratton, M.D., Ph.D., and Peter J. Campbell, F.R.A.C.P., Ph.D., for the Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium.

The authors' affiliations are as follows: the Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton (E.P., I.R., A.C., R.R., C.M., P.E., L.J.M., P.J.S., S.M., P.S.T., I.V., S.N.-Z., H.R.D., A.S., D.J., K.R., J.H., A.P.B., J.W.T., A.F., V.M., P.A.F., M.R.S., P.J.C.); the Nuffield Department of Clinical Laboratory Sciences (J.B., A.P., J.S.S.) and the Weatherall Institute of Molecular Medicine (P.V.), University of Oxford, Oxford; St. James Institute of Oncology, St. James Hospital, Leeds (D.B.); the Department of Haematology (A.L.G., A.C., C.E.M., E.J.B., A.J.W., A.R.G., P.J.C.) and Medical Research Council Laboratory of Molecular Biology (A.J.W.), University of Cambridge, and the Department of Haematology, Addenbrooke's Hospital (A.J.W., A.R.G., P.J.C.), Cambridge; the School of Medicine, University of Southampton, Southampton (J.S., N.C.P.C.); and the Division of Medical Sciences, University of Dundee, Dundee (M.G., S.T.) — all in the United Kingdom; Fondazione Istituto di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo, University of Pavia, Pavia, Italy (M.C., L.M., A.G., M.G.D.P., E.T.); the Department of Hematology, Karolinska Institute, Stockholm (E.H.-L.); the Department of Hematology, S. Gerardo Hospital, Monza and University of Milan Bicocca, Milan (C.G.-P.); Dana–Farber Cancer Institute, Harvard Medical School (N.C.M., K.C.A.), and VA Boston Healthcare System (N.C.M.) — both in Boston; M.D. Anderson Cancer Center, Houston (A.E.-N.); and Institut für Theoretische Physik, Universität zu Köln, Cologne, Germany (A.F.).

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

REFERENCES

1. Tefferi A, Vardiman JW. Myelodysplastic syndromes. N Engl J Med 2009; 361:1872-85.

2. Ebert BL, Pretz J, Bosco J, et al. Identification of RPS14 as a 5q– syndrome gene by RNA interference screen. Nature 2008;451:335-9.

3. Pellagatti A, Jadersten M, Forsblom AM, et al. Lenalidomide inhibits the malignant clone and up-regulates the SPARC gene mapping to the commonly deleted region in 5q– syndrome patients. Proc Natl Acad Sci U S A 2007;104:11406-11.

4. Grand FH, Hidalgo-Curtis CE, Ernst T, et al. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. Blood 2009;113:6182-92.

5. Sanada M, Suzuki T, Shih LY, et al. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature 2009;460:904-8.

6. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in *TET2* in myeloid cancers. N Engl J Med 2009;360:2289-301.

7. Langemeijer SM, Kuiper RP, Berends M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 2009;41:838-42.

8. Boultwood J, Perry J, Pellagatti A, et al. Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. Leukemia 2010;24:1062-5.

9. Gelsi-Boyer V, Trouplin V, Adelaide J, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol 2009;145:788-800.

10. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 2010;42:722-6.

11. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the WHO classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 2009;114:937-51.

12. Navarro I, Ruiz MA, Cabello A, et al. Classification and scoring systems in myelodysplastic syndromes: a retrospective analysis of 311 patients. Leuk Res 2006; 30:971-7.

13. Camaschella C. Hereditary sideroblastic anemias: pathophysiology, diagnosis, and treatment. Semin Hematol 2009;46:371-7.

14. Boultwood J, Pellagatti A, Nikpour M, et al. The role of the iron transporter ABCB7 in refractory anemia with ring sideroblasts. PLoS ONE 2008;3(4):e1970.

15. Pellagatti A, Cazzola M, Giagounidis AA, et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood 2006;108:337-45. [Erratum, Blood 2006;108:1128.]

16. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011; 469:539-42.

17. Campbell PJ, Pleasance ED, Stephens PJ, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci U S A 2008;105:13081-6.

18. Pellagatti A, Cazzola M, Giagounidis A, et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia 2010;24: 756-64.

19. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446: 153-8.

20. Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801-6.

21. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 2010; 464:999-1005.

22. Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature 2011;469:356-61.

23. Campbell PJ, Yachida S, Mudie LJ, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 2010;467:1109-13.

24. Notta F, Mullighan CG, Wang JC, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. Nature 2011;469:362-7. [Erratum, Nature 2010;471:254.]

25. Ewalt M, Galili NG, Mumtaz M, et al. DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia. Blood Cancer J 2011;1(3):e9.

26. Ley TJ, Ding L, Walter MJ, et al. *DNMT3A* mutations in acute myeloid leukemia. N Engl J Med 2010;363:2424-33.

27. Walter MJ, Ding L, Shen D, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia 2011;25:1153-8.

28. Bignell GR, Greenman CD, Davies H, et al. Signatures of mutation and selection in the cancer genome. Nature 2010; 463:893-8.

29. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 2009;10:741-54.

30. Golas MM, Sander B, Will CL, Lührmann R, Stark H. Molecular architecture of the multiprotein splicing factor SF3b. Science 2003;300:980-4.

31. *Idem*. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Mol Cell 2005;17:869-83.

32. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. Nucleic Acids Res 2010;38:D211-D222.

33. Fischer A, Greenman C, Mustonen V. Germline fitness-based scoring of cancer mutations. Genetics 2011;188:383-93.

34. Wang C, Chua K, Seghezzi W, Lees E, Gozani O, Reed R. Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev 1998;12: 1409-14.

35. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

36. Tehranchi R, Invernizzi R, Grandien A, et al. Aberrant mitochondrial iron distribution and maturation arrest characterize early erythroid precursors in low-risk myelodysplastic syndromes. Blood 2005;106:247-53.

37. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. Nature 2011; 471:467-72.

38. Malcovati L, Porta MG, Pascutto C, et al. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: a basis for clinical decision making. J Clin Oncol 2005;23:7594-603.

Copyright © 2011 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org at CSIC on June 19, 2012. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.