Targeted re-sequencing analysis of 25 genes commonly mutated in myeloid disorders in del(5q) myelodysplastic syndromes

Marta Fernandez-Mercado,[1]* Adam Burns,[2]* Andrea Pellagatti,[1] Aristoteles Giagounidis,[3] Ulrich Germing,[4] Xabier Agirre,[5] Felipe Prosper,[5] Carlo Aul,[3] Sally Killick,[6] James S. Wainscoat,[1] Anna Schuh,[2¥] and Jacqueline Boultwood[1¥]

[1]LLR Molecular Haematology Unit, NDCLS, RDM, John Radcliffe Hospital, Oxford, UK; [2]NIHR Biomedical Research Centre, Oxford, UK; [3]Medizinische Klinik II, St Johannes Hospital, Duisburg, Germany; [4]Department of Hematology, Oncology and Clinical Immunology, Heinrich-Heine-Universität, Düsseldorf, Germany; [5]Division of Cancer and Area of Cell Therapy and Haematology Service, Foundation for Applied Medical Research, Clínica Universitaria, Universidad de Navarra, Pamplona, Spain; and [6]Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK

ABSTRACT

Interstitial deletion of chromosome 5q is the most common chromosomal abnormality in myelodysplastic syndromes. The catalogue of genes involved in the molecular pathogenesis of myelodysplastic syndromes is rapidly expanding and next-generation sequencing technology allows detection of these mutations at great depth. Here we describe the design, validation and application of a targeted next-generation sequencing approach to simultaneously screen 25 genes mutated in myeloid malignancies. We used this method alongside single nucleotide polymorphism-array technology to characterize the mutational and cytogenetic profile of 43 cases of early or advanced del(5q) myelodysplastic syndromes. A total of 29 mutations were detected in our cohort. Overall, 45% of early and 66.7% of advanced cases had at least one mutation. Genes with the highest mutation frequency among advanced cases were *TP53* and *ASXL1* (25% of patients each). These showed a lower mutation frequency in cases of 5q- syndrome (4.5% and 13.6%, respectively), suggesting a role in disease progression in del(5q) myelodysplastic syndromes. Fifty-two percent of mutations identified were in genes involved in epigenetic regulation (*ASXL1, TET2, DNMT3A* and *JAK2*). Six mutations had allele frequencies <20%, likely below the detection limit of traditional sequencing methods. Genomic array data showed that cases of advanced del(5q) myelodysplastic syndrome had a complex background of cytogenetic aberrations, often encompassing genes involved in myeloid disorders. Our study is the first to investigate the molecular pathogenesis of early and advanced del(5q) myelodysplastic syndromes using next-generation sequencing technology on a large panel of genes frequently mutated in myeloid malignancies, further illuminating the molecular landscape of del(5q) myelodysplastic syndromes.

©Ferrata Storti Foundation

Introduction

The myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic stem cell malignancies that are characterized by ineffective hematopoiesis resulting in peripheral cytopenias, and typically a hypercellular bone marrow. The MDS are pre-leukemic conditions showing frequent progression (in approximately 40% of patients) to acute myeloid leukemia (AML). In the early stages of the disease, apoptosis of bone marrow precursor cells prevails, but in more advanced disease increased proliferation of immature blasts occurs.[1] About 50% of MDS exhibit acquired genomic abnormalities detected by conventional cytogenetic banding techniques. Recent molecular investigations have revealed additional genetic abnormalities in MDS, including microdeletions and loss of heterozygosity due to acquired uniparental disomy (UPD).[2]

Interstitial deletion within the long arm of chromosome 5 [del(5q)] is one of the most frequent cytogenetic abnormalities observed in myeloid malignancies, occurring in approximately 10-20% of patients with *de novo* MDS[3] and in a similar proportion of patients with *de novo* AML.[4] In *de novo* MDS the del(5q) occurs either in isolation or together with other karyotypic abnormalities. Although the 5q- is a good prognostic indicator when found in isolation,[5] this is not the case when the 5q- is part of a complex karyotype.[6] In a large MDS database, del(5q) was reported as an isolated abnormality in 14% of patients with clonal abnormalities, in 5% with one other abnormality, and in 11% with a complex karyotype.[6] The median overall survival in these groups was 80, 47 and 7 months, respectively.[6] These findings are consistent with the general notion that the total number of cytogenetic changes found is an independent factor that can allow for the stratification of cohorts of patients into prognostic subgroups.

The 5q- syndrome is the most distinct of all the MDS and is characterized by isolated del(5q), severe macrocytic anemia, frequent thrombocytosis, female predominance, and a lower risk of progression to AML.[7] Patients with the 5q- syndrome have one of the best outcomes of any MDS subgroup,[8]

©2013 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2013.086686
The online version of this article has a Supplementary Appendix.
*MFM and AB contributed equally to this manuscript; ¥JB and AS were co-senior authors
Manuscript received on February 22, 2013. Manuscript accepted on July 3, 2013.
Correspondence: jacqueline.boultwood@ndcls.ox.ac.uk

with a relatively long survival often of several years.[7,8] While a small number of gene mutations have been reported in the 5q- syndrome, including mutations of *TP53* and *JAK2*,[9,10] the molecular landscape of this disease remains to be fully elucidated. Approximately 10% of patients with the 5q- syndrome show transformation to AML,[7] but the genetic aberrations that drive this process have not been fully determined.

The International Prognostic Scoring System (IPSS)[11] and its revised version[8] are based upon karyotypic abnormalities as well as on morphological data. Recently, a new and more comprehensive cytogenetic scoring system has been developed, which allows for a refined cytogenetic risk prediction.[12] Nevertheless, the heterogeneous clinical outcome observed within the karyotypically and morphologically-defined groups in the IPSS indicates that it may be possible to refine the cytogenetic classification by using additional markers. The catalogue of genes that play a role in the molecular pathogenesis of MDS is rapidly expanding, and includes *TET2*, *SF3B1*, *EZH2* and *ASXL1*.[13-16] Unraveling the genetic complexity of MDS promises to elucidate the pathophysiology of this disease, refine the taxonomy and prognostic scoring systems, and provide novel therapeutic targets.

Technological advances in DNA sequencing provide an important tool to analyze heterogeneous cancer samples. Massively parallel sequencing enables the analysis of independent, clonal, DNA molecules[17] and offers the opportunity to adjust the balance between breadth and depth of such assays to identify a wide variety of potentially critical DNA changes in tumors. Broad approaches, such as whole genome and whole exome sequencing, have been used to discover new cancer gene mutations[15,18,19] or to study clonal evolution.[20] In particular, several such studies in MDS have identified recurrently mutated genes and novel pathways involved in pathogenesis, such as those encoding splicing factors.[15,19] However, these genome-wide approaches are still expensive and have relatively low sensitivity. In contrast, a more targeted sequencing approach aimed at detecting selected recurrent mutations in MDS allows for cost-effective and fast sequencing at the profound depth required for accurate characterization of heterogeneous cancer samples.

Here we describe the design, validation and application of a targeted next-generation sequencing approach using a bench-top platform to simultaneously screen 25 genes mutated in a range of myeloid malignancies. We used this method to characterize the mutational profile of a cohort of 43 MDS cases with del(5q).

Methods

Patients' samples

Test cohort

Nine MDS samples, with known mutations detected by Sanger sequencing, pyrosequencing or amplification refractory mutation system polymerase chain reaction (ARMS-PCR) analysis in at least one of our target genes, were selected to validate our gene panel. The nine samples chosen contained a total of 13 variants across eight genes and included missense (n=4), nonsense (n=1) and frameshift (n=8) mutations (Table 1).

del(5q) myelodysplastic syndrome cohort

Samples from 43 untreated MDS cases harboring a del(5q) were selected for mutational screening (mean age 66.0 years; range, 24-88). These included 22 patients with 5q- syndrome, nine cases with refractory anemia (RA) with additional karyotypic abnormalities, and 12 cases with advanced MDS [defined as having an increased number of blasts and including 11 RA with excess of blasts (RAEB), and 1 chronic myelomonocytic leukemia (CMML) in transformation]. All karyotypes were determined by conventional G-banding. This study was approved by the ethics committees of the institutes involved and informed consent was obtained.

DNA extraction

Genomic DNA was isolated by phenol-chloroform extraction from peripheral blood neutrophils obtained using Histopaque (Sigma-Aldrich) and pelleted after hypotonic lysis of erythrocytes. The purity of the neutrophil populations was high, >95%, as assessed by standard morphology on Wright-Giemsa-stained cytospin preparations.

Targeted re-sequencing

We designed a TruSeq Custom Amplicon panel (TSCA, Illumina), targeting 25 genes mutated in various myeloid malignancies (Table 2). The panel was developed using the online DesignStudio pipeline (http://designstudio.illumina.com, Illumina), and covers a total of 46,604 bp with 322 amplicons. In genes with well-defined mutational hotspots only these regions were targeted; otherwise the entire coding sequence of the gene was sequenced. Libraries prepared from 250 ng DNA were subjected to 250 bp paired-end sequencing.

Protein sequences resulting from detected DNA-sequence changes were predicted using the insilico.ehu.es on-line tool,[21] and Alamut Software (Interactive Biosoftware, San Diego, CA, USA). PolyPhen-2 v2.2.2, also on-line, was used to predict the functional effect of variant calls (Polymorphism Phenotyping v2, *http://genetics.bwh.harvard.edu/pph2/*).[22]

FLT3 *internal tandem duplication fragment analysis*

Thirty-three samples in the test cohort with sufficient DNA were screened for internal tandem duplications in the *FLT3* gene (*FLT3*-ITD) using conventional fragment analysis.[23] These comprised 20 cases of 5q- syndrome, five cases of del(5q) RA with additional karyotypic abnormalities and eight cases of advanced del(5q) MDS.

Genomic array profiling

Single nucleotide polymorphism array data were available from a previously published study[2] for 33 of the 43 samples included in the targeted sequencing analysis. Those data allowed us to identify cryptic copy number changes and regions of UPD.

Results

Quality of the myelodysplastic syndrome del(5q) MiSeq sample run

The number of clusters that passed the quality filter was over 100,000 for the majority of the samples (40/43, 93%) (*Online Supplementary Figure S1*). Paired-end MiSeq sequencing produced more than 2.2 Gb of sequence data with 91% of reads higher than the quality threshold of Q30, exceeding the expected minimums of 2 Gb and 75%, respectively. The average depth of coverage across all samples was >390x, with 98% of cases (42/43) over 250x, 91% (39/43) over 300x and 49% (21/43) ≥400x. The overall sensitivity of the assay and its background noise were

©Ferrata Storti Foundation 2013

Table 1A. Summary of mutations present in test samples used for the TSCA panel validation. All Q-score values were generated by GATK through the BaseSpace pipeline, with the exception of the TEST009 variant, which was generated by Platypus.

Sample ID	Gene	Mutation	Position	Q-score	Depth of coverage (x)	Frequency
TEST001	*ASXL1*	c.1925het_insA; p.G643RfsX13	Chr20:31,022,442	99	239	60%
TEST001	*EZH2*	p.L98Ifs*28	Chr7:148,529,801	99	895	32%
TEST001	*EZH2*	p.Q250X	Chr7:148,523,705	99	437	42%
TEST001	*NRAS*	p.G12D	Chr1:115,258,747	99	955	32%
TEST001	*RUNX1*	p.S141X	Chr21:36,252,940	99	361	31%
TEST002	*ASXL1*	c.1748G>GA;p.W583X	Chr20:31,022,263	99	159	28%
TEST003	*IDH2*	c.13775G>GA; p.R140Q	Chr15:90,631,934	99	1382	51%
TEST003	*NPM1*	----/TCTG	Chr5:170,837,548	99	836	23%
TEST003	*FLT3*	109bp insertion	N/A	Detected visually only		
TEST004	*NPM1*	----/TCTG	Chr5:170,837,548	99	734	25%
TEST004	*FLT3*	64bp insertion	N/A	Detected visually only		
TEST005	*NPM1*	----/TCTG	Chr5:170,837,548	99	527	15%
TEST006	*DNMT3A*	c.2648C>CT;p.R882V	Chr2:25,457,242	99	250	44%
TEST007	*KIT*	c.2447A>AG;p.D816V	Chr4:55,599,321	99	248	39%
TEST008	*NPM1*	----/TCTG	Chr5:170,837,548	99	875	20%
TEST008	*IDH1*	c.6694C>CT; p.R132C	Chr4:209,113,114	99	1390	48%
TEST009	*TP53*	TGTACATGGCCATGGCGCGG / T	Chr17:7,578,441	200	731	95%

Table 1B. Summary of additional mutations found in the test samples.

Sample ID	Gene	Mutation	Position	Q-score	Depth of coverage (x)	Frequency
TEST001	*TET2*	p.C1464X	Chr4:106,193,930	99	423	47%
TEST005	*TET2*	p.L1258Afs*10	Chr4:106,164,903	99	2968	25%
TEST006	*NPM1*	G/GTCTG	Chr5:170,837,547	99	999	18%
TEST007	*RUNX1*	p.L71Sfs*24	Chr21:36,259,199	99	395	51%
TEST007	*SF3B1*	p.K700E	Chr2:198,266,834	99	2053	45%
TEST008	*FLT3*	p.I836del	Chr13:28,592,636	99	2271	48%

©Ferrata Storti Foundation 2013

estimated at 1-3% (*Online Supplementary Information* and *Online Supplementary Table S1*).

Validation of the myeloid gene panel

In order to compare the accuracy and sensitivity of our TSCA assay against standard methods of mutation screening (Sanger sequencing, pyrosequencing, fragment analysis), we re-screened nine test samples (Table 1A), containing 17 variants across 11 genes (*ASXL1*, *DNMT3A*, *EZH2*, *FLT3*, *IDH1*, *IDH2*, *KIT*, *NPM1*, *NRAS*, *RUNX1* and *TP53*).

Using the BaseSpace data analysis pipeline, we were able to successfully identify five missense, three nonsense and seven frameshift mutations in our validation cohort (15 out of 17, 88.2%). In particular, short indel (insertions/deletions) mutations in both *ASXL1* and *NPM1* (1 bp and 5 bp, respectively) were correctly identified by the BaseSpace analysis software. Analysis of the TEST009 aligned reads in the Integrative Genomics Viewer (IGV, Broad Institute), revealed a dramatically reduced read depth of 30x across *TP53*, compared to >1000x in other samples, suggesting that there was a failure to align the reads to the reference sequence. The TEST009 sequence data were therefore submitted to a second alignment and variant calling pipeline (Stampy/Platypus[24,25]), which successfully identified a 19 bp deletion (*Online Supplementary Figure S2*). Two 109 bp and 64 bp *FLT3*-ITD in samples TEST003 and TEST004, respectively, were only called after visual inspection of the un-aligned data for reads matching part of the *FLT3* target sequence. The presence of *FLT3*-ITD was subsequently confirmed by fragment analysis.

In addition to the known control mutations, we identified six mutations affecting five genes in five samples (Table 1B). One of these mutations, the C1464X variant in TEST001, was visible in earlier Sanger sequencing traces; however, at the time, the variant had not been called as it was within the range of background noise (*Online Supplementary Figure S3*). All other additional mutations were confirmed by Sanger sequencing and fragment analysis (*Online Supplementary Figure S4* and *Online Supplementary Table S2*).

Mutations detected in del(5q) cases

Highly purified peripheral blood neutrophil DNA samples from 43 MDS cases harboring a del(5q) were subjected to mutational screening using the 25-gene panel described above. A total of 4036 variant calls were detected by a combination of BaseSpace, Stampy/Platypus and visual inspection of the *FLT3* locus. Of these, all non-synonymous variant calls with a COSMIC ID (i.e. recorded in the Catalogue of Somatic Mutations in Cancer[26]) were considered relevant. We also included in the analysis all

Table 2. List of genes targeted for enrichment in the TSCA library.

Gene	Location	Chromosomal Coordinates	Targeted Exons
ASXL1	20q11.21	chr20:30946147-31027122	12
ATRX	Xq21.1	chrX:76,760,356-77,041,719	8-10 (ADD domain) 17-31 (Helicase domain)
CBL	11q23.3	chr11:119076986-119178859	8-9 (ring finger domain and linker sequence)
CBLB	3q13.11	chr3:105377109-105587887	9, 10
CBLC	19q13.32	chr19:45281126-45303903	9, 10
DNMT3A	2p23.3	chr2:25455830-25564784	23
ETV6/TEL	12p13.2	chr12:11802788-12048325	All 8 exons
EZH2	7q36.1	chr7:148504464-148581441	2-20
FLT3	13q12.2	chr13:28577411-28674729	14, 15 (JM and TK1 domains) 20 (D835)
IDH1	2q34	chr2:209100953-209119806	4
IDH2	15q26.1	chr15:90627212-90645708	4
JAK2	9p24.1	chr9:4985245-5128183	12, 14
KIT	4q12	chr4:55524095-55606881	2, 8-11, 13, 17
MPL	1p34.2	chr1:43803475-43820135	10
NPM1	5q35.1	chr5:170814708-170837888	12
NRAS	1p13.2	chr1:115247085-115259515	2, 3
PDGFRA	4q12	chr4:55095264-55164412	12, 14, 18
RUNX1	21q22.12	chr21:36193574-36260987	3-8
SF3B1	2q33.1	chr2:198256698-198299771	15, 16
SRSF2	17q25.1	chr17:74730197-74733493	1
TET2	4q24	chr4:106067842-106200960	3-11
TP53	17p13.1	chr17:7571720-7590868	4-9
U2AF1	21q22.3	chr21:44513066-44527688	2, 6
WT1	11p13	chr11:32409322-32457081	7, 9 (Cys-His zinc finger domains)
ZRSR2	Xp22.2	chrX:15808574-15841382	All 11 exons

©Ferrata Storti Foundation 2013

non-synonymous variant calls not found in either COSMIC or the dbSNP database (build 135). A total of 29 non-synonymous variants were called in ten different genes: seven affecting *TP53*, six *ASXL1*, five *TET2*, two *CBL*, two *DNMT3A*, two *SF3B1*, two *JAK2*, one *U2AF1*, one *RUNX1* and one *WT1* (Table 3, Figure 1, *Online Supplementary Table S3*). In addition, 21 synonymous variants with a COSMIC ID were found in five different genes (10 *PDGFRA*, 5 *IDH1*, 3 *cKIT*, 2 *FLT3* and 1 *TP53*) (*Online Supplementary Table S4*).

Distribution of the non-synonymous mutations among disease subgroups

A total of 29 mutations were detected in our cohort of 43 del(5q) MDS cases. Twelve of 29 mutations were found in nine of the 22 5q- syndrome cases (45.0%) (Table 3). Five mutations affected four of the nine cases of del(5q) RA (44.0%) with additional cytogenetic aberrations (Table 3). The more advanced del(5q) MDS cases presented a higher proportion of sequence changes: 12 variant calls were found in eight of the 12 advanced del(5q) cases (66.7%) (Table 3). The genes with the highest mutation frequency in this cohort were *TP53* (3/12 patients, 25%; 5 mutations in total as 2 patients had 2 *TP53* mutations) and *ASXL1* (3/12, 25%). The mutation frequency for these two genes was lower in 5q- syndrome cases (*TP53* 1/22, 4.5%; *ASXL1* 3/22, 13.6%).

Other mutations were identified in 5q- syndrome cases (3 *TET2*, 2 *SF3B1*, 1 *DNMT3A*, 1 *RUNX1* and 1 *WT1*), in del(5q) RA with additional cytogenetic abnormalities (1 additional *TP53*, 1 *CBL*, 1 *DNMT3A*, 1 *U2AF1* and 1 *JAK2*) and in advanced del(5q) MDS cases (2 *TET2*, 1 *CBL* and 1 *JAK2*). It is of note that six of the mutations detected in this study present variant frequencies lower than 20%, which are likely to be below the level of detection of Sanger sequencing.[27,28] These low frequency mutations were found in the following genes: two in *TET2*, one in *ASXL1*, one in *DNMT3A*, one in *JAK2* and one in *SF3B1* (Figure 1, *Online Supplementary Table S3*).

These data show that a number of different gene mutations occur in patients with the 5q- syndrome and that advanced del(5q) MDS cases display a greater mutation frequency than early del(5q) MDS cases, with mutations of *TP53* and *ASXL1* genes being the most frequent.

Co-occurring mutations: analysis of clonality and timing of mutation acquisition

Clonal evolution has been documented as MDS transforms to AML,[29] and when *de novo* AML relapses after initial chemotherapy.[30] The proportion of sequencing reads reporting a given mutation can be used to estimate the fraction of tumor cells carrying that mutation, and to iden-

tify whether mutations are clonal (present in all tumor cells) or subclonal (present in a fraction of tumor cells).[31] This estimation needs to take into account copy number and loss of heterozygosity data. Five cases in our cohort showed mutations in more than one gene. Whole genome array data were available for all of them.[2] The genes with co-occurring mutations were *ASXL1*, *WT1*, *SF3B1*, *TET2*, *DNMT3A*, *JAK2* and *CBL* (Figures 1 and 2).

In two cases (1 5q- syndrome, MDS08, and 1 CMML, MDS42) two mutations were present at similar allele frequencies, *ASXL1* (44.7%) and *WT1* (49.0%) in the 5q- syndrome case, and *ASXL1* (45.4%) and *CBL* (96.0%) - the latter within a region of UPD - in the CMML case. This is suggestive of a dominant clonal population of cells. In this scenario, it is not possible to determine the temporal order of mutations. A third case (MDS29, a del(5q) RA with additional karyotypic abnormalities) had a *DNMT3A* mutation at a variant allele frequency of ~44%, and a *JAK2* mutation at a frequency of ~7%. Since the copy number showed these to have occurred in diploid regions without any loss of heterozygosity, the fraction of cells carrying the mutations would be ~88% and ~14%, respectively. On this basis, we could not infer if the *JAK2* mutation was subclonal to the cells carrying the *DNMT3A* mutation or if, on the contrary, it represented an independent clone. However, assuming that each mutation occurred only once during tumor evolution, it is possible to suggest that the *DNMT3A* mutation occurred earlier than the *JAK2* one in the disease course. Similarly, the fourth case (MDS12, a 5q- syndrome case) had ~80% of cells carrying a *SF3B1* mutation, ~20% with *ASXL1* mutation and ~10% with *TET2* mutation. We can therefore suggest that the *SF3B1* mutation occurred before the *ASXL1* or *TET2* mutation. The variant allele fractions for *ASXL1* and *TET2* could be consistent with either the *TET2* mutation being subclonal to the *ASXL1* mutation or on a separate branch of the phylogenetic tree, so we cannot establish the timing of those two mutations relative to each other. Finally, the fifth case [MDS43, a del(5q) RAEB case] had ~86% of cells with an *ASXL1* mutation and ~55% with a *TET2* mutation. In this case, it was clear that the *TET2* mutation was subclonal to the *ASXL1* mutation and must have occurred later.

Copy number changes and uniparental disomy analysis

Thirty-three [18 5q- syndrome, 6 del(5q) RA with additional cytogenetic abnormalities, and 9 cases of advanced del(5q) MDS] of the 43 del(5q) MDS samples included in the targeted sequencing analysis had been previously analyzed by single nucleotide polymorphism-arrays to identify cryptic copy number changes and regions of UPD (defined as continuous stretches of homozygous single nucleotide polymorphism calls >2 Mb without copy number loss).[2]

The results of the analysis are listed in *Online Supplementary Table S5*. The del(5q) was characterized in all 33 cases. Copy number changes in addition to the del(5q) were observed in six of nine cases of advanced MDS (66.7%) and four of six cases of del(5q) RA with additional cytogenetic abnormalities (66.7%), but in only four of 18 5q- syndrome cases (22.2%). In the 5q- syndrome group, 31 regions of UPD were identified in 17 of 18 patients. All other cases included in this study showed regions of UPD, six regions in all six cases of del(5q) RA with additional cytogenetic aberrations, and 17 in all nine advanced del(5q) cases.[2]

Table 3. Summary of non-synonymous variant calls with a COSMIC ID or not present in dbSNP.

	5q- syndrome (n=22)	RA del(5q) with additional karyotypic abnormalities (n=9)	Advanced del(5q) cases (n=12)
	N. of mutations (%)	N. of mutations (%)	N. of mutations (%)
TP53	1 (4.5)	1 (11.1)	5 (41.7)*
ASXL1	3 (13.6)	0	3 (25.0)
TET2	3 (13.6)	0	2 (16.7)
JAK2^{V617F}	0	1/8 (12.5)	1/8 (12.5)
CBL	0	1 (11.1)	1 (8.3)
DNMT3A	1 (4.5)	1 (11.1)	0
U2AF1	0	1 (11.1)	0
SF3B1	2 (9.1)	0	0
RUNX1	1 (4.5)	0	0
WT1	1 (4.5)	0	0
Total n. of mutations	12	5	12
Patients presenting at least one mutation	9 (45.0)	4 (44.4)	8 (66.7)

* *Two patients had two* TP53 *mutations.*

A proportion of the regions affected by copy number loss encompassed genes that are part of our TSCA gene panel. In advanced del(5q) MDS cases, these were *EZH2*, *NPM1*, *ETV6*, *ASXL1* and *TP53* (Figure 1, *Online Supplementary Table S6*). Additional regions of cytogenetic loss encompassed *CBL* and *ETV6* in two different cases of del(5q) RA with additional cytogenetic aberrations (Figure 1, *Online Supplementary Table S6*). The only *DNMT3A* loss was seen in a case of 5q- syndrome. In the one case [a del(5q) RAEB] with cytogenetic loss encompassing *TP53*, the remaining copy had a missense mutation (R273H), predicted to be damaging to the function of the protein (*Online Supplementary Table S3*).

These results show that advanced del(5q) MDS cases display a more complex landscape of cytogenetic aberrations, both karyotypically evident and cryptic. These regions often contain genes involved in myeloid disease.

Discussion

In this study, we sought to validate an Illumina-based targeted next-generation sequencing platform to simultaneously screen 25 genes relevant to myeloid malignancies for mutations. Once validated, we aimed to use this gene panel to characterize the mutational profile of a cohort of 43 cases of MDS with del(5q), in the context of additional molecular and high-density genomic array data. The prevalence of the mutations detected in complex DNA samples has typically been limited to approximately 20% using Sanger sequencing.[27,28] The development of specific mutation enrichment or detection strategies has greatly increased this sensitivity.[32,33] In keeping with the improved power of mutation detection of next-generation sequencing over traditional sequencing techniques, we identified previously undetected mutations in the validation cohort (that comprised 9 test samples containing 17 variants across 11 genes) in addition to the previously known

©Ferrata Storti Foundation 2013

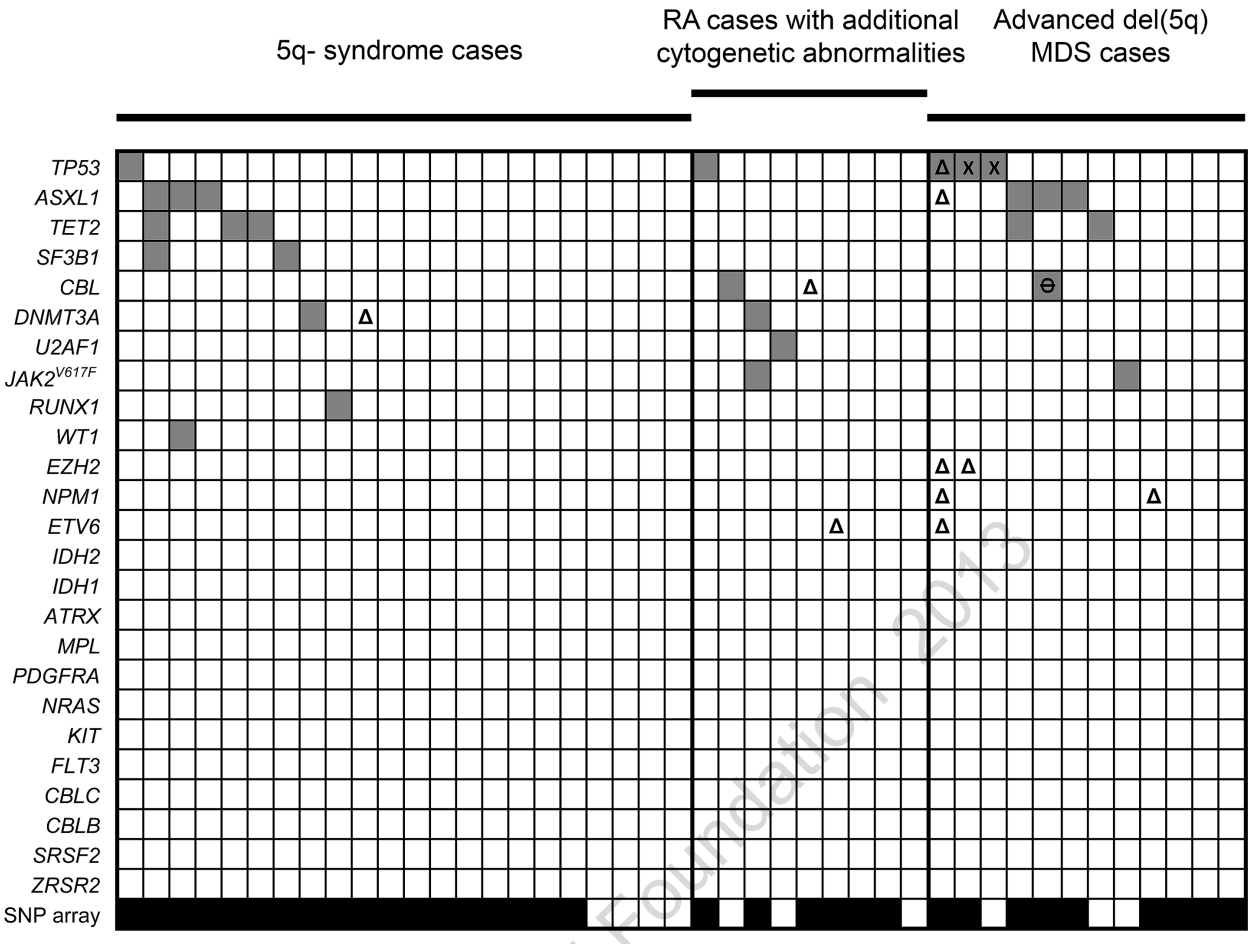


Figure 1. Mutations, deletions and loss of heterozygosity in 25 genes analyzed in del(5q) MDS samples. Columns show results for each of the 43 cases analyzed. Gray boxes indicate mutated cases. Black boxes mark samples for which single nucleotide polymorphism-array (SNP) data were available. X: double mutant. Δ: gene encompassed within a region of cytogenetic loss. Θ: gene encompassed within a region of UPD.

©Ferrata Storti Foundation 2013

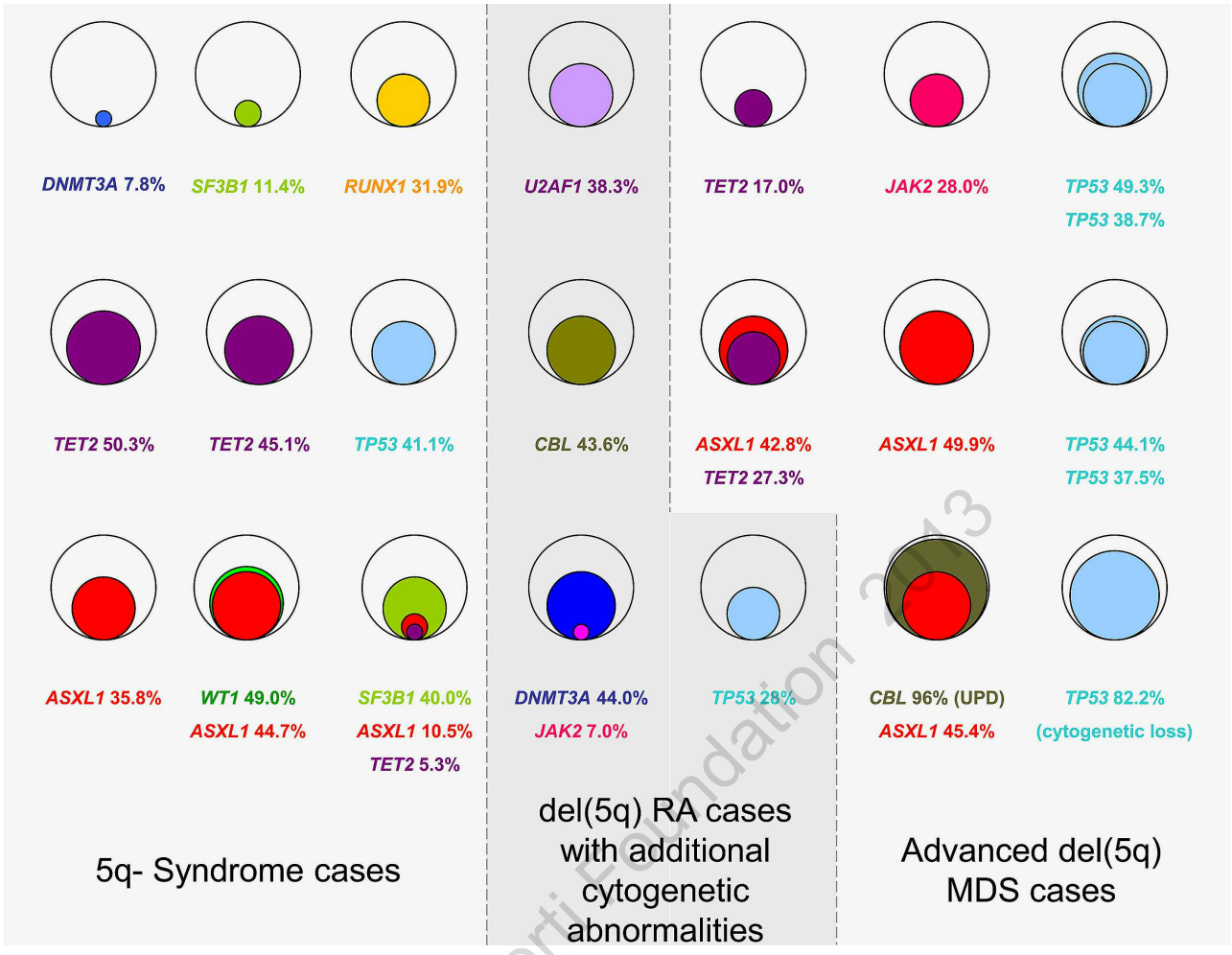
mutations in these samples.

Using BaseSpace and Stampy/Platypus[24,25] analysis software, we were able to successfully identify all point mutations, short indels and deletions included in the validation cohort. However, *FLT3*-ITD variants that consist of patient-specific sequence duplications were amplified and sequenced, but were not identified using either bio-informatics pipeline and therefore had to be visually identified. This highlights the need for further refinements to commercially available analysis pipelines before their use in routine clinical practice. The non-alignment of these reads is largely a function of the comparative size of the insertion or deletion compared to the absolute read length. We are hopeful that in the future longer read lengths, in combination with improvements to the alignment algorithms, will greatly increase the ability to detect these important mutations.

Once our panel had been successfully validated, we applied it to study the mutational profile of a series of MDS cases with del(5q). Gene mutation screening in del(5q) MDS has been performed in previous studies, but most of these studies focused on a limited number of genes, and mainly employed traditional sequencing methods.[9,34-40] Other studies have investigated a larger number of genes[41-43] but did not specifically focus on MDS cases with del(5q). To our knowledge, the present work is the first attempt to screen a large number of genes using a targeted next-generation sequencing approach in both early and advanced del(5q) MDS.

A total of 29 mutations were detected in our cohort of 43 del(5q) MDS cases. Overall, 45% of the cases of 5q- syndrome and 44% of cases of del(5q) RA with additional cytogenetic aberrations had at least one mutation. The more advanced del(5q) cases showed a higher proportion of mutated cases, and 66.7% presented at least one mutation. The genes with the highest mutation frequency among advanced cases were *TP53* and *ASXL1* (25% of patients each). The mutation frequency for these two genes was lower in cases of 5q- syndrome (*TP53* 4.5%, *ASXL1* 13.6%). We therefore confirmed in our del(5q) cohort the observation made by our group and others that *TP53* mutations occur predominantly in MDS with complex karyotype.[9,44,45] The increased incidence of *TP53* and *ASXL1* mutations in advanced del(5q) cases in our present study suggests that these abnormalities may play a role in disease progression in del(5q) MDS. These data are consistent with those of a recent study that showed that *TP53* mutations were associated with disease progression in del(5q) MDS.[43]

The 5q- syndrome is widely considered to be relatively



©Ferrata Storti Foundation 2013

Figure 2. Mutant allele frequencies in individual del(5q) MDS samples. The area of each colored circle indicates the allele frequency of the given mutation. The text under the circles lists the frequency and nature of each mutation in order of decreasing allele frequency.

genetically stable compared to other MDS subtypes, on the basis of molecular studies (including genomic array data analysis).[2] This is reflected in its relatively good prognosis.[8,11] Previous studies have shown mutations in a limited number of genes, including *TP53*, *JAK2* and *ASXL1*[9,10,34,46] in this subtype of MDS. The incidence of *JAK2* and *ASXL1* mutations is ~6%.[37] Here, we show that over 40% of patients with the 5q- syndrome in fact harbor a gene mutation, including mutations in *TET2*, *SF3B1*, *RUNX1*, *WT1* and *ASXL1*.

The *SF3B1* mutations detected in this study were identified in cases of 5q- syndrome, and not in the other two groups of patients with del(5q), which are karyotypically or morphologically defined by more advanced disease. *SF3B1* mutations have been associated with a relatively benign disease course.[15,43,47] It has been suggested that multipotent hematopoietic stem cells initially attain a splicing factor mutation as a founding genetic lesion, and subsequently acquire additional mutations that drive their malignant transformation.[43,48] This is consistent with our finding in one 5q- syndrome case with a high *SF3B1* mutant allele frequency and two other mutations (*ASXL1* and *TET2*) with lower mutant allele frequencies.

The present study shows that a high proportion of genes involved in the epigenetic regulation of the cell (*TET2*, *ASXL1*, *DNMT3A* and *JAK2*) are affected by either mutations or cytogenetic losses in del(5q) MDS cases: 15 of 29 genes with non-synonymous mutations (51.7%) and 4 of 10 genes in regions affected by cytogenetic loss (40%) were epigenetic regulators. This observation is consistent with a recent report of mutations in a large cohort of MDS patients (n=117), in which the authors also found 80 mutations in genes predicted to affect the epigenetic regulation of the cell in half of the cohort (52% of cases).[43] Genome-wide methylation analysis on a subset of cases with and without mutations in epigenetic factors did not highlight a specific DNA methylation profile associated with these mutations (*Online Supplementary Information* and *Online Supplementary Figure S5*).

A total of six of the mutations detected in this study present variant frequencies below the level of detection of Sanger sequencing, which is estimated to be around 15-20%.[27,28] For example one patient with the 5q- syndrome showed a *DNMT3A* mutant allele frequency of 7.8% and another case a *SF3B1* mutant allele frequency of 11.4%. Sanger sequencing has been the gold standard for sequenc-

ing for many years, and the vast majority of sequencing studies published to date have used this technology. It is likely that previous studies underestimated the prevalence of mutations in MDS. This has recently been illustrated by Jadersten *et al.*[34] who used next-generation sequencing to reveal *TP53* mutations (median clone size 11%) in nearly 20% of low-risk MDS patients with del(5q). Our data support the hypothesis that the prevalence of mutations in del(5q) MDS may have also been underestimated for other genes. Here, we have shown that genes involved in the epigenetic regulation of the cell frequently harbor low-frequency mutations in del(5q) MDS, not detectable by means of Sanger sequencing. This has previously been demonstrated for *TET2* in MDS and CMML.[49]

The proportion of variant reads can be used to determine the order of occurrence of multiple mutations and thereby infer the clonal evolution from early stages of the disease. Interestingly, *ASXL1* was one of the genes involved in four of the five cases with two or more mutations, with a lower variant frequency than the other co-mutated genes in three of the cases, suggesting that mutation of *ASXL1* represented a later event in the disease course in these cases. Our analysis of clonality was based on a small number of single cases with multiple mutations. Studies with a similar sensitivity involving larger MDS cohorts will certainly help to establish the phylogenetic structure of tumor evolution.

In summary, we have successfully developed and validated a panel that allows for the screening of 25 genes frequently mutated in myeloid malignancies. The present study on del(5q) MDS has shown that a number of gene mutations occur in patients with the 5q- syndrome, and that >40% of patients with this low-risk MDS subtype harbor at least one gene mutation. A higher percentage of mutations was found among the more advanced cases of del(5q) MDS, with *TP53* and *ASXL1* being the more frequently mutated genes. Our study is the first to investigate and compare the molecular pathogenesis of early and advanced del(5q) MDS using targeted next-generation sequencing technology on a large panel of genes frequently mutated in myeloid malignancies.

Acknowledgments

This work was funded by Leukaemia and Lymphoma Research (UK), and by the Oxford Partnership Comprehensive Biomedical Research Centre, with funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health. The authors would like to thank the patients who accepted to participate in this study, all staff in their laboratories for their technical assistance, and all the physicians who referred patients' material.

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

©Ferrata Storti Foundation

References

1. Heaney ML, Golde DW. Myelodysplasia. N Engl J Med. 1999;340(21):1649-60.
2. Wang L, Fidler C, Nadig N, Giagounidis A, Della Porta MG, Malcovati L, et al. Genome-wide analysis of copy number changes and loss of heterozygosity in myelodysplastic syndrome with del(5q) using high-density single nucleotide polymorphism arrays. Haematologica. 2008;93 (7):994-1000.
3. Bernasconi P, Klersy C, Boni M, Cavigliano PM, Calatroni S, Giardini I, et al. Incidence and prognostic significance of karyotype abnormalities in de novo primary myelodysplastic syndromes: a study on 331 patients from a single institution. Leukemia. 2005;19(8):1424-31.
4. Johansson B, Harrison C. Acute myeloid Leukemia. In: Heim S, Mitelman F, eds. Cancer Cytogenetics (3rd edn). Hoboken, NJ, 2009:45-139.
5. Giagounidis AA, Germing U, Haase S, Hildebrandt B, Schlegelberger B, Schoch C, et al. Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. Leukemia. 2004;18 (1):113-9.
6. Haase D, Germing U, Schanz J, Pfeilstocker M, Nosslinger T, Hildebrandt B, et al. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. Blood. 2007;110(13):4385-95.
7. Boultwood J, Pellagatti A, McKenzie AN, Wainscoat JS. Advances in the 5q- syndrome. Blood. 2010;116(26):5803-11.
8. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012; 120(12):2454-65.
9. Fidler C, Watkins F, Bowen DT, Littlewood TJ, Wainscoat JS, Boultwood J. NRAS, FLT3 and TP53 mutations in patients with myelodysplastic syndrome and a del(5q). Haematologica. 2004;89(7):865-6.
10. Wong KF, Wong WS, Siu LL, Lau TC, Chan NP. JAK2 V617F mutation is associated with 5q- syndrome in Chinese. Leuk Lymphoma. 2009;50(8):1333-5.
11. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997; 89(6):2079-88.
12. Schanz J, Tuchler H, Sole F, Mallo M, Luno E, Cervera J, et al. New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. J Clin Oncol. 2012;30(8):820-9.
13. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet. 2010;42(8):722-6.
14. Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccia N, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol. 2009;145(6):788-800.
15. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011;365(15):1384-95.
16. Tefferi A, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, Patnaik MM, et al. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. Leukemia. 2009;23(7):1343-5.
17. Druley TE, Vallania FL, Wegner DJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods. 2009;6(4):263-5.
18. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469(7331):539-42.
19. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478(7367): 64-9.
20. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010;464(7291):999-1005.
21. Bikandi J, San Millan R, Rementeria A, Garaizar J. In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. Bioinformatics. 2004;20(5):798-9.
22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit 7.20.
23. Murphy KM, Levis M, Hafez MJ, Geiger T,

Cooper LC, Smith BD, et al. Detection of FLT3 internal tandem duplication and D835 mutations by a multiplex polymerase chain reaction and capillary electrophoresis assay. J Mol Diagn. 2003;5(2):96-102.
24. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011;21(6):936-9.
25. Rimmer A, Mathieson I, Lunter G, McVean G. Platypus: An Integrated Variant Caller (www.well.ox.ac.uk/platypus). 2012.
26. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39(Database issue):D945-50.
27. Bar-Eli M, Ahuja H, Gonzalez-Cadavid N, Foti A, Cline MJ. Analysis of N-RAS exon-1 mutations in myelodysplastic syndromes by polymerase chain reaction and direct sequencing. Blood. 1989;73(1):281-3.
28. Collins SJ, Howard M, Andrews DF, Agura E, Radich J. Rare occurrence of N-ras point mutations in Philadelphia chromosome positive chronic myeloid leukemia. Blood. 1989;73(4):1028-32.
29. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med. 2012;366(12):1090-8.
30. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-10.
31. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994-1007.
32. Li M, Diehl F, Dressman D, Vogelstein B, Kinzler KW. BEAMing up for detection and quantification of rare sequence variants. Nat Methods. 2006;3(2):95-7.
33. Su Z, Dias-Santagata D, Duke M, Hutchinson K, Lin YL, Borger DR, et al. A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer. J Mol Diagn. 2011;13(1):74-84.
34. Jadersten M, Saft L, Smith A, Kulasekararaj A, Pomplun S, Gohring G, et al. TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. J Clin Oncol. 2011;29(15):1971-9.
35. Jerez A, Gondek LP, Jankowska AM, Makishima H, Przychodzen B, Tiu RV, et al. Topography, clinical, and genomic correlates of 5q myeloid malignancies revisited. J Clin Oncol. 2012;30(12):1343-9.
36. Pardanani A, Patnaik MM, Lasho TL, Mai M, Knudson RA, Finke C, et al. Recurrent IDH mutations in high-risk myelodysplastic syndrome or acute myeloid leukemia with isolated del(5q). Leukemia. 2010;24(7):1370-2.
37. Patnaik MM, Lasho TL, Finke CM, Gangat N, Caramazza D, Holtan SG, et al. WHO-defined 'myelodysplastic syndrome with isolated del(5q)' in 88 consecutive patients: survival data, leukemic transformation rates and prevalence of JAK2, MPL and IDH mutations. Leukemia. 2010;24(7):1283-9.
38. Patnaik MM, Lasho TL, Finke CM, Knudson RA, Ketterling RP, Chen D, et al. Isolated del(5q) in myeloid malignancies: clinicopathologic and molecular features in 143 consecutive patients. Am J Hematol. 2011;86(5):393-8.
39. Sebaa A, Ades L, Baran-Marzack F, Mozziconacci MJ, Penther D, Dobbelstein S, et al. Incidence of 17p deletions and TP53 mutation in myelodysplastic syndrome and acute myeloid leukemia with 5q deletion. Genes Chromosomes Cancer. 2012;51(12):1086-92.
40. Sokol L, Caceres G, Rocha K, Stockero KJ, Dewald DW, List AF. JAK2(V617F) mutation in myelodysplastic syndrome (MDS) with del(5q) arises in genetically discordant clones. Leuk Res. 2010;34(6):821-3.
41. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011;364(26):2496-506.
42. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood. 2012;119(14):3211-8.
43. Mian SA, Smith AE, Kulasekararaj AG, Kizilors A, Mohamedali AM, Lea NC, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica. 2013;98(7):1058-66.
44. Jonveaux P, Fenaux P, Quiquandon I, Pignon JM, Lai JL, Loucheux-Lefebvre MH, et al. Mutations in the p53 gene in myelodysplastic syndromes. Oncogene. 1991;6(12):2243-7.
45. Lai JL, Preudhomme C, Zandecki M, Flactif M, Vanrumbeke M, Lepelley P, et al. Myelodysplastic syndromes and acute myeloid leukemia with 17p deletion. An entity characterized by specific dysgranulopoiesis and a high incidence of P53 mutations. Leukemia. 1995;9(3):370-81.
46. Boultwood J, Perry J, Pellagatti A, Fernandez-Mercado M, Fernandez-Santamaria C, Calasanz MJ, et al. Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. Leukemia. 2010;24(5):1062-5.
47. Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Della Porta MG, Pascutto C, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood. 2011;118(24):6239-46.
48. Cazzola M, Rossi M, Malcovati L. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood. 2013;121(2):260-9.
49. Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gaken J, Lea NC, et al. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. Blood. 2010;116(19):3923-32.

©Ferrata Storti Foundation 2013