Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple TP53 alterations

Laurence Lodé, Adam Ameur, Thibault Coste, Audrey Ménard, Steven Richebourg, Jean-Baptiste Gaillard, Yannick Le Bris, Marie-Christine Béné, Thierry Lavabre-Bertrand, Thierry Soussi

EUROPEAN HEMATOLOGY ASSOCIATION
haematologica
Journal of the European Hematology Association

Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple TP53 alterations

by Laurence Lodé, Adam Ameur, Thibault Coste, Audrey Ménard, Steven Richebourg, Jean Baptiste Gaillard, Yannick Le Bris, Marie Christine Béné, Thierry Lavabre-Bertrand, and Thierry Soussi

Haematologica 2017 [Epub ahead of print]

Citation: Lodé L, Ameur A, Coste T, Ménard A, Richebourg S, Gaillard JB, Le Bris Y, Béné MC, Lavabre-Bertrand T, and Soussi T. Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple TP53 alterations. Haematologica. 2017; 102:xxx doi:10.3324/haematol.2017.176719

Letter to the editor:

Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple *TP53* alterations

Laurence Lodé [1,2], Adam Ameur [3], Thibault Coste[1,5], Audrey Ménard [2], Steven Richebourg [2,4], Jean-Baptiste Gaillard [5], Yannick Le Bris [2], Marie-Christine Béné [2], Thierry Lavabre-Bertrand [5] and Thierry Soussi [6-9]

Affiliations:

1 - Hematology Laboratory, University Hospital, Montpellier, F-34000, France
2 - Hematology Laboratory, University Hospital, Nantes, F-44093, France
3- Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.
4 - Cytogenetics, Hôpital du Saint Sacrement, CHU de Québec, Université Laval, Department of molecular medicine, Quebec, Canada
5 - Department of Clinical Cytology and Cytogenetics, Nîmes University Hospital, Nîmes, France.

6 - Sorbonne Université, UPMC Univ Paris 06, F- 75005 Paris, France
7 - INSERM, U1138, Centre de Recherche des Cordeliers, Paris, France
8 - Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden

9: Correspondence: thierry.soussi@ki.se

Correspondence: Thierry Soussi, Sorbonne Université, UPMC Univ Paris 06, F- 75005 Paris, France or Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden
thierry.soussi@ki.se

Keywords: *TP53* mutation; AML; MDS; Next Generation Sequencing; MDS

L.L. and A.A. contributed equally to this study.

Although the frequency of *TP53* mutations in hematologic malignancies is low, these mutations have a high clinical relevance and are usually associated with poor prognosis. Somatic *TP53* mutations have been detected in up to 73.3% of cases of acute myeloid leukemia (AML) with complex karyotype and 18.9% of AML with other unfavorable cytogenetic risk factors. [1] AML with *TP53* mutations, and/or chromosomal aneuploidy has been defined as a distinct AML subtype. In low-risk myelodysplastic syndromes (MDS), TP53 mutations occur at an early disease stage and predict disease progression. [2] *TP53* mutation diagnosis is now part of the revised European LeukemiaNet (ELN) guidelines.[3,4]

The use of next generation sequencing (NGS), particularly ultra-deep sequencing, has led to the discovery that patients with either MDS or AML (either *de novo*, secondary or therapy-related) present multiple *TP53* mutations, suggesting that several *TP53* independent clones may co-exist. [5] Patient follow-up also reveals a highly dynamic evolution of these mutations during disease progression in treated and untreated patients. [6,7] This observation is in line with the recent recognition that human tumors harbor an extensive genetic intratumoral heterogeneity. [8]. These findings will likely have implications for therapy and biomarker discovery and determination of genetic complexity is becoming part of clinical decision-making processes in the age of precision medicine.

In this report, *in silico* analysis of the UMD *TP53* database showed that *TP53* variants detected in patients with multiple *TP53* alterations are fully oncogenic. Furthermore, using long-range single-molecule real-time (SMRT) sequencing on AML and MDS patients harboring multiple *TP53* mutations, we showed that all of these variants are localized on different subclones, emphasizing the considerable tumor heterogeneity in these patients.

The 2017 release of the UMD_*TP53* database contains the mutation status of 75,448 patients, including 922 cases of AML and 899 cases of MDS (**Online Supplemental Table S1)**. [9,10]. Among these patients, 158 MDS cases (22.3%) and 99 AML cases (13%) harbor more than one *TP53* variant in their tumors, higher than the rate observed in solid tumors **(Online Supplemental Table S1)**. Chronic Lymphocytic Leukemia (CLL) patients also harbor a high frequency of tumors with multiple *TP53* mutations. This feature has been observed increasingly frequently over recent years with the advent of deep sequencing techniques. Whether or not all of the multiple variants identified in these patients are truly deleterious or comprise a mix of driving and passenger mutations has never been addressed. The UMD_*TP53* database includes quantitative functional data for all *TP53* missense variants and can therefore be used to determine whether patients with multiple *TP53* mutations frequently harbor non-deleterious *TP53* variants. Analysis of the 257 AML and MDS cases with more than one *TP53* variant showed that the majority (98%) of these variants are true deleterious *TP53* mutations with complete loss of function and not simply random passenger mutations co-selected during tumor progression (**Online Supplemental Table S2 and Online Supplemental Figure S1a to e)**. A few non-deleterious variants have been identified, but they are likely very rare non-somatic polymorphisms.

To further demonstrate the presence and the dynamics of multiple independent tumor clones in AML and MDS, we have developed a novel, third-generation single-molecule real-time (SMRT) sequencing assay using the Pacific Biosciences platform with long-read lengths that span the most frequently mutated region of the *TP53* gene. Sanger sequencing cannot be used to define the allelic distribution of multiple *TP53* variants. This is also true for standard NGS if the two variants are more than 200 base pairs apart. On

the other hand, SMRT analysis can be used to phase mutations located multiple kilobases apart directly from sequencing reads.

Eleven patients harboring multiple *TP53* mutations in their tumors were enrolled. For 3 patients, sequential samples were available to assess the evolution of the various variants. The *TP53* status of these patients was already defined according to stringent clinical criteria using either Sanger sequencing or standard NGS **(Online Supplemental Material and Online Supplemental Table S3)**. *In silico* analysis of all these variants using the UMD *TP53* database showed that they were true deleterious *TP53* mutations that have already been described in various types of cancer **(Online Supplemental Table S1)**.

The majority of mutations detected for clinical evaluation were readily identified by SMRT, except for 2 variants that were not included in the amplicon used for analysis. SMRT identified 5 mutations that were not identified by clinical analysis **(Online Supplemental Table S3 and Online Supplemental Figure S2)**. Manual examination of the sequencing data performed for clinical analysis confirmed that 2 of these mutations were detected at a frequency below the cut-off used for the analysis **(Table 1).** Most of the remaining mutations detected by SMRT were present at a very low frequency **(Online Supplemental Figure S2)**. The Variant allele frequency (VAF) observed for each variant detected by the two analyses was remarkably similar, despite being performed in different centers according to very different methodologies **(Online Supplemental Figure S3)**.

Our analysis shows that all oncogenic *TP53* variants were located in different alleles **(Table 1 and Online Supplemental Figures S4a to S4k).** For two samples, patient Fr7, sample 7b and patient Fr2, the close proximity of two *TP53* variants allowed analysis of the alignment obtained after standard NGS and confirmed that these mutations were carried by different alleles **(Online Supplemental Figures S5 and S6).** For 2 samples, the allelic distribution was also confirmed by the observation that the different *TP53* variants were associated with different *TP53* haplotypes **(Online Supplemental Figure S4j and k, patients Fr10 and Fr11)**.

Figure 1 shows a typical result observed for two samples collected 5 years apart from a patient with multiple *TP53* mutations. Of note, the diagnostic sample was negative with standard NGS and was therefore not used for SMRT analysis. In the first sample analysed by SMRT, clinical analysis identified two pathogenic *TP53* mutations confirmed by SMRT. SMRT analysis also identified two novel *TP53* mutations at very low frequency and showed that the 4 variants were distributed in different *TP53* molecules. The two novel variants were readily identified in the second sample collected 5 years later. New *TP53* variants were also identified by both methodologies (2 variants) and 2 additional variants were found at low frequency by SMRT **(Figure 1)**. All these variants, carried by different *TP53* molecules, were true driver mutations already identified in multiple tumor types, as shown by their high frequency in the *TP53* database. This dynamic evolution of the various subclones can also lead to the elimination of certain subclones, as shown for patient Fr10 with the disappearance of *TP53* variants **(Online Supplemental Figure S4J)**.

Using both *in silico* analysis and SMRT sequencing, we demonstrate that the presence of multiple subclones with different *TP53* variants is a common feature in AML and MDS. All *TP53* variants detected in MDS and AML patients by SMRT sequencing are true, physically independent *TP53* variants, confirming the results of indirect computational studies currently used to infer cancer heterogeneity. It is highly likely that each TP53 variant belongs to independent subclone arising from a wild-type TP53 founder clone. The observation of multiple subclones with different *TP53* variants in these patients suggests the occurrence of a specific genetic background in the founder clones that requires *TP53* inactivation for further progression. All of these subclones present a highly dynamic evolution, but it remains to be determined whether this evolution is driven by treatment, a

natural characteristic of the tumor or both. A recent study on 1,514 MDS patients after stem-cell transplantation showed that 283 patients (19%) had at least one oncogenic *TP53* mutation and a poor overall survival. [11] One hundred two (36%)of these patients had more than one *TP53* variant (range 2-6). It is likely that the use of a sensitive methodology for DNA sequencing will reveal that tumors with multiple *TP53* variants constitute a general feature raising potential problems for treatment options. Finally, in this report, we demonstrate the efficiency of SMRT sequencing for the analysis of complex samples. The rapid progress in NGS, combining longer reads, increased sensitivity and decreased costs, will allow investigation of the whole sequence of clinically relevant genes in a single analysis. Long-read RNA-seq analysis could also be used to address the issue of *TP53* alternative spliced transcripts that have already been described to be of clinical interest in AML.[12] .

Acknowledgments: This work was supported by Radiumhemmets Forskningsfonder and the Swedish Cancer Society (Cancerfonden) to TS. SMRT sequencing was performed by the National Genomics Infrastructure (NGI) hosted by SciLifeLab Uppsala. The authors are most grateful to the Molecular Hematology team and to the IRCNA Tumor Bank (CHU de Nantes, Institut de Cancérologie de l'Ouest, Saint-Herblain F44800, France) for their assistance.

Contribution: T.S. and L.L. designed the research; T.C., A.M., S.R., J.B.G., Y.L., M.C.B. and T.L.B. were actively involved in patient care or routine diagnostic procedures; A.A. performed the SMRT sequence and bioinformatics analysis; all authors reviewed and approved the manuscript

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

Correspondence: Thierry Soussi, Sorbonne Université, UPMC Univ Paris 06, F- 75005 Paris, France or Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden
thierry.soussi@ki.se

ORCID profiles: orcid.org/0000-0001-8184-3293

References

1. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-2221.
2. Jadersten M, Saft L, Smith A, et al. TP53 Mutations in Low-Risk Myelodysplastic Syndromes With del(5q) Predict Disease Progression. J Clin Oncol. 2011;29(15):1971-1979.
3. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-447.
4. Malcovati L, Hellström-Lindberg E, Bowen D, et al. Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. Blood. 2013;122(17):2943-2964.
5. Stengel A, Kern W, Haferlach T, Meggendorfer M, Fasan A, Haferlach C. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. Leukemia. 2017;31(3):705-711.
6. Mossner M, Jann JC, Nowak D, et al. Prevalence, clonal dynamics and clinical impact of TP53 mutations in patients with myelodysplastic syndrome with isolated deletion (5q) treated with lenalidomide: results from a prospective multicenter study of the german MDS study group (GMDS). Leukemia. 2016;30(9):1956-1959.
7. da Silva-Coelho P, Kroeze LI, Yoshida K, et al. Clonal evolution in myelodysplastic syndromes. Nat Commun. 2017;815099.
8. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323-334.
9. Leroy B, Ballinger ML, Baran-Marszak F, et al. Recommended Guidelines for Validation, Quality Control, and Reporting of TP53 Variants in Clinical Practice. Cancer Res. 2017;77(6):1250-1260.
10. Leroy B, Anderson M, Soussi T. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum Mutat. 2014;35(6):672-688.
11. Lindsley RC, Saber W, Mar BG, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. N Engl J Med. 2017;376(6):536-547.
12. Anensen N, Hjelle SM, Van Belle W, et al. Correlation analysis of p53 protein isoforms with NPM1/FLT3 mutations and therapy response in acute myeloid leukemia. Oncogene. 2012;31(12):1533-1545.

Patient	Sample	Sex/Age (diagnosis)	Hematologic malignancy[1]	Treatment	*TP53* alteration (clinical)[2]	*TP53* alteration (SMRT)[3]	*TP53* Mutation[4]	Allelic distribution[5]
Fr1	Fr1	M/77	De novo MK-AML	No (at diagnosis, before treatment)	Yes	Yes	c.673-2A>G	NR
					Yes	No[6]	c.897_912del15	NR
Fr2	Fr2	M/63	De novo MK-AML	No (at diagnosis, before treatment)	Yes	Yes	c.673-2A>T	Yes
					Yes	Yes	c.743G>A	Yes
Fr3	Fr3	F/73	De novo MK-AML	No (at diagnosis, before treatment)	Yes	Yes	c.413C>T	Yes
					Yes	Yes	c.794T>C	Yes
Fr4	Fr4	F/78	De novo MK-AML	No (at diagnosis, before treatment)	Yes	Yes	c.395A>G	Yes
					Yes	Yes	c.824G>T	Yes
Fr5	Fr5	F/73	De novo MK-AML	No (at diagnosis, before treatment)	Yes	Yes	c.637C>T	Yes
					Yes	Yes	c.455C>T	Yes
Fr6	Fr6	F/75	s-AML (post LR-MDS del5q)	Yes (Lenalidomide)	Yes	No	c.524G>A	NR
					Yes	Yes	c.844C>T	NR
Fr7[7]	Fr7a	F/73	LR-MDS del5q	Yes (Lenalidomide)	Yes	Yes	c.314G>T	Yes
					Yes	Yes	c.743G>A	Yes
					No	Yes	c.818G>A	Yes
					No	Yes	c.742C>T	Yes
	Fr7b				Yes	Yes	c.314G>T	Yes
					Yes	Yes	c.743G>A	Yes
					Yes	Yes	c.584T>A	Yes
					Yes	Yes	c.614A>G	Yes

1 MK-AML: Monosomal karyotype AML; LR-MDS: Low-risk-MDS

2 Identification of *TP53* mutations for clinical analysis using conventional Sanger sequencing (VAF cut-off: 10-15%)or standard NGS (VAF cut-off: 1%).

3 Identification of *TP53* mutations using SMRT sequencing

4 *TP53* variant description using the NM_000546.5 reference. A full description of the mutations and their consequences is presented in supplementary Table S1.

5 Yes: mutations are located on different alleles; NR: not relevant.

6 Mutation outside the amplicon used for SMRT analysis

7 Samples Fr7a and Fr7b were taken at an interval of 59 months

					Yes	Yes	c.818G>A	Yes
					Yes	Yes	c.833C>G	Yes
					No	Yes	c.524G>A	Yes
					No	Yes	c.659A>G	Yes
Fr8	Fr8	F/72	s-AML (post LR-MDS del5q)	Yes (Lenalidomide)	Yes	Yes	c.725G>T	NR
					Yes	No	c.920-1G>A	NR
Fr9	Fr9a[8]	M/76	s-AML (post LR-MDS del5q)	Yes (Lenalidomide)	Yes	Yes	c.421T>G	Yes
					Yes[9]	Yes	c.711G>T	Yes
	Fr9b		s-AML (post LR-MDS del5q)		Yes	Yes	c.421T>G	Yes
					Yes	Yes	c.711G>T	Yes
Fr10[10]	Fr10a	M/69	LR-MDS del5q	Yes (Lenalidomide)	Yes	Yes	c.743G>A	Yes
					Yes	Yes	c.844C>T	Yes
					Yes	Yes	c.817C>T	Yes
	Fr10b		s-AML (post LR-MDS del5q)	Yes (Lenalidomide)	Yes	Yes	c.743G>A	Yes
					Yes	Yes	c.844C>T	Yes
Fr11[11]	Fr11a	F/85	LR-MDS del5q	Yes (Lenalidomide)	Yes[7]	Yes	c.659A>G	Yes
					Yes[12]	Yes	c.840A>T	NR
					No	Yes	c.701A>G	Yes
	Fr11b	F/85	LR-MDS del5q	Yes (Lenalidomide)	Yes	Yes	c.701A>G	Yes
					Yes	Yes	c.659A>G	Yes
					Yes	Yes	c.840A>T	Yes

Table 1: Patient characteristics. More information is available in **Supplemental Table S1**.

[8] Sample 9a and 9: DNA extracted from a frozen pellet from whole blood leukocytes or a cytogenetic pellet from bone marrow, respectively (same sampling date)

[9] Mutation detected by SMRT and identified at very low frequency by reviewing the standard NGS data

[10] Sample Fr10a and Fr10b were taken at an interval of 7 months

[11] Sample Fr11a and Fr11b were taken at an interval of 9 months

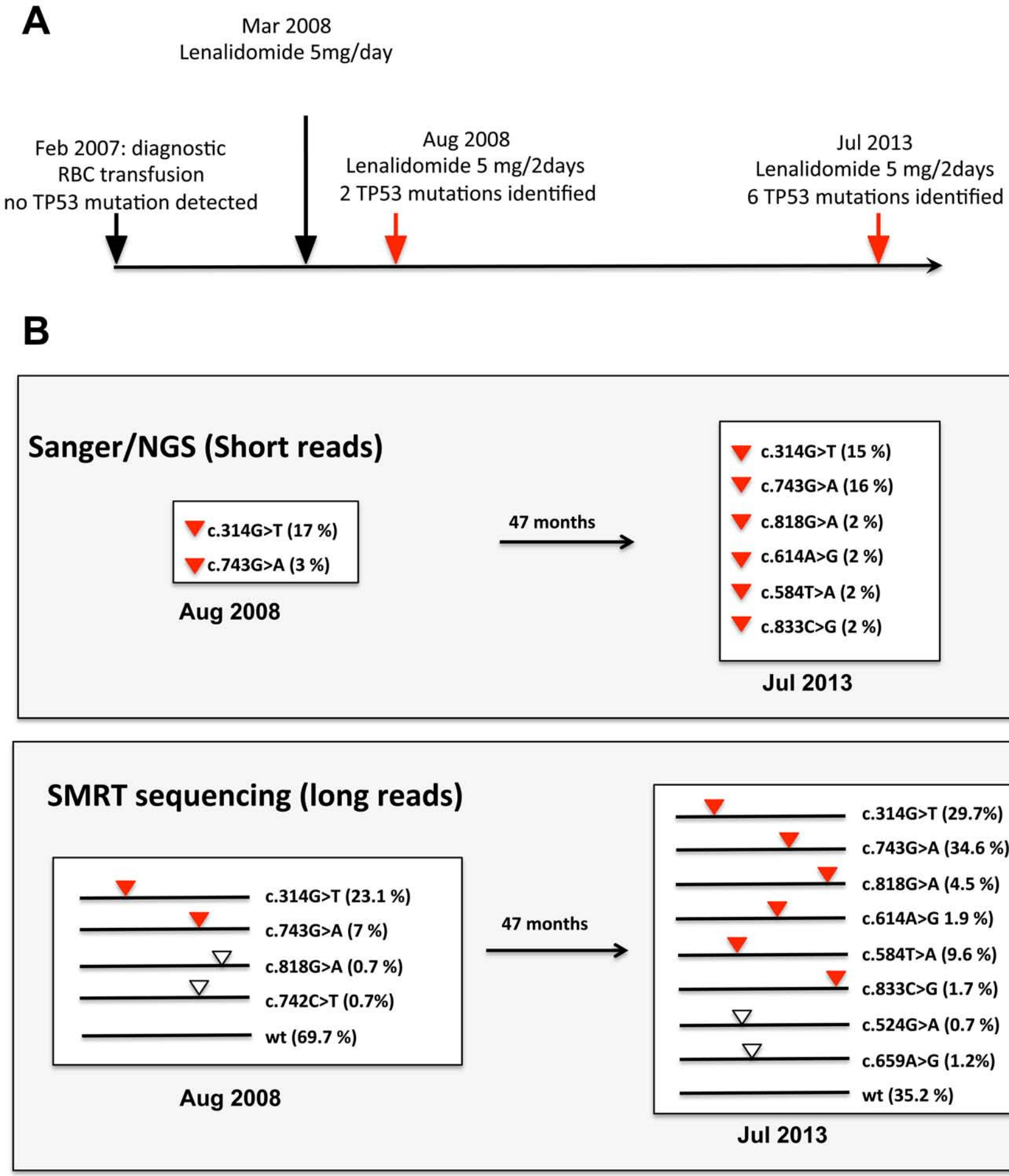
[12] Mutation detected at high frequency in the second sample and identified at low frequency by reviewing the data of the first sample.

Figure 1: Clinical course and TP53 mutation analysis on patient Fr7

A: Patient Fr7 was diagnosed in February 2007 with lower risk MDS with del5q with no *TP53* mutations detected at a cut-off of 1%. One year after initiation of red blood cell (RBC) transfusion for anemia, patient Fr7 was treated with lenalidomide at a dosage of 5 mg/day, but rapidly experienced major adverse effects, leading to reduction of the dosage to 5 mg every other day (or less in a context of poor adherence). Five months after onset of therapy, 2 *TP53* mutations were identified together with improvement of anemia. Five years later, in July 2013 (and until the present time), patient Fr7 was still alive and with no disease progression to secondary AML despite growth of at least 6 TP53-mutant clones, suggesting possible clonal equilibrium due to competition between the numerous mutant clones.

B: Sanger sequencing and/or standard NGS analysis is shown in the upper part with 2 and 6 mutations in sample 7a and 7b, respectively. No allelic distribution can be inferred from this type of analysis. SMRT sequencing (lower part) provides an accurate picture of the allelic distribution of each *TP53* variant, as well as the remaining wt allele. The frequencies of the 9 different alleles are shown in brackets.

Red triangle: *TP53* variants identified by both types of analysis. White triangle: *TP53* variants detected only by SMRT sequencing.


A
Mar 2008
Lenalidomide 5mg/day
Feb 2007: diagnostic
RBC transfusion
no TP53 mutation detected
Aug 2008
Lenalidomide 5 mg/2days
2 TP53 mutations identified
Jul 2013
Lenalidomide 5 mg/2days
6 TP53 mutations identified
B
Sanger/NGS (Short reads)
c.314G>T (17 %)
c.743G>A (3 %)
Aug 2008
47 months
c.314G>T (15 %)
c.743G>A (16 %)
c.818G>A (2 %)
c.614A>G (2 %)
c.584T>A (2 %)
c.833C>G (2 %)
Jul 2013
SMRT sequencing (long reads)
c.314G>T (23.1 %)
c.743G>A (7 %)
c.818G>A (0.7 %)
c.742C>T (0.7%)
wt (69.7 %)
Aug 2008
47 months
c.314G>T (29.7%)
c.743G>A (34.6 %)
c.818G>A (4.5 %)
c.614A>G 1.9 %)
c.584T>A (9.6 %)
c.833C>G (1.7 %)
c.524G>A (0.7 %)
c.659A>G (1.2%)
wt (35.2 %)
Jul 2013

Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple *TP53* alterations

Laurence Lodé, Adam Ameur, Thibault Coste, Audrey Ménard, Steven Richebourg, Jean-Baptiste Gaillard, Yannick Le Bris, Marie-Christine Béné, Thierry Lavabre-Bertrand and Thierry Soussi

SUPPLEMENTAL MATERIAL

Table of contents

A Supplemental Methods

Patients and Samples

This retrospective study was conducted on 11 patients shown to harbor multiple *TP53* mutations in their tumors (**Table 1 and Supplemental Table S3)**. Six patients were diagnosed with refractory anemia between 1996 and 2010 and a diagnosis of lower risk MDS with del(5q) was confirmed at Nantes University Hospital (5 patients) or Nîmes University Hospital (1 patient) by conventional cytogenetics and/or FISH analysis. Four patients experienced disease progression to secondary acute myeloid leukemia (s-AML). The remaining five patients were diagnosed with poor-risk *de novo* monosomal karyotype acute myeloid leukemia (MK-AML) at Nîmes University Hospital between 1998 and 2011 **(Supplemental Table S3)**. Multiple samples obtained during the course of the disease were available for 4 of the 11 patients and were sequential for 3 patients. *TP53* status was established by two certified ***TP53*** centers either by Sanger sequencing, standard NGS or both, using stringent criteria specific for clinical analysis **(Table 1 and Supplemental Table S2)**. [1,2]
Peripheral blood or bone marrow samples were stored as frozen cell pellets or cytogenetic pellets before therapy for MK-AML and before and/or after initiation of lenalidomide therapy for Lower Risk MDS del5q. *TP53* short-read libraries were prepared using either primer plates from the IRON-II study network, sequenced on a GS-Junior (Roche, Basel, Switzerland) with data processed as described by Kohlmann et al. [1], or a custom TSCA design (Illumina, San Diego, CA, USA) sequenced on a MiSeq (Illumina, San Diego, CA, USA), in which case, reads were processed with a homemade analysis pipeline including VarScan v2.3.6 or an analysis pipeline provided by Illumina including the MiSeqReporter Suite and Variant Studio annotation tool. With a depth of coverage between 800X and 9,500X (after resequencing of selected samples), a minimum of 10 mutated bidirectional reads were taken into account allowing a VAF detection threshold of 1% to 2%. The VAF detection threshold was set to a lower limit of >1% for bidirectional reads, according to a recent study investigating the assay's lower limit of detection [2]. Mutations with VAF>10% allowed cross-validation with Sanger Sequencing using a VAF detection threshold of 10-15%.

SMRT sequencing of TP53 amplicons

A 2.8 Kb amplicon that encompasses exons 4 to 8 was used for SMRT analysis (**Supplemental Figure S7)**. This region includes the majority of the mutations detected in the patients, as well as some common *TP53* SNP useful for phasing the various mutations on the two alleles.
The *TP53* amplicons (11 patients, 15 samples) underwent DNA damage repair and end-repair before ligation of hairpin adaptors to generate SMRTbell™ libraries for circular consensus sequencing**.** Libraries were then subjected to exo treatment and PB AMPure bead wash procedures for clean-up. Each library was loaded onto one SMRTcell™ and sequenced on the PacBio RS II instrument using C4 chemistry, P6 polymerase and a 240-min movie time. [3]

Detection of SNPs and mutations in SMRT sequencing data

SNPs and mutations were identified by a two-step procedure. First, the 'Minor and Compound Variants' plug-in (v2.3.0 of SMRT Analysis) was executed on each sample. This resulted in a total of 84 positive variants in all 25 samples. We then performed a more stringent analysis of each of these mutations by counting the number of reference/alternative alleles occurring in the CCS read using a 20 bp window surrounding each mutation. This counting-based method is a sensitive approach to determine exact mutation frequencies, as previously demonstrated [3]. Forty-three of the original 84 variants were detected at a frequency of at least 0.5% in at least one sample.

Analysis of the phasing of SNPs and mutations

Custom R scripts were used to determine the clonal composition of mutations and SNPs. We counted the number of CCS reads comprising all possible combinations of reference/alternative variants and obtained a read count for all different *TP53* molecules present in each sample. To remove any chimeric molecules introduced by aberrations in the PCR step, we first determined the phasing of homozygous and heterozygous SNPs from the information provided by the molecule with the highest read count, and then filtered out all molecules discordant with this SNP phasing pattern. Next, we removed any remaining molecules that could be explained by a single jump between different molecules during PCR, i.e. molecules with a phasing pattern that can be created by concatenation of two other molecules with higher read counts.

In silico analysis of *TP53* variants in AML and MDS

The 2017 release of the *TP53* mutation database contains 82,134 *TP53* mutations, from 75,448 patients including those from 1,821 cases with AML or MDS. [4,5]. The database includes records for each tumor, indicating the number and description of each variant.
The database also includes functional data for most missense mutations. Residual transactivating activity for WAF, MDM2, BAX, 14-3-3-s, AIP, GADD45, NOXA and P53R2 promoters was originally published by Kato et al [6]. The residual transcriptional activity of mutant p53 was always compared to wild-type p53 for the same promoter (%).
For nonsense and splice variants as well as indels (*TP53* null), this value was set to 0 as no *TP53* protein is generally expressed.

Supplemental Table S1: Frequency of patients with multiple *TP53* variants in the UMD *TP53* database.

Cancer type	MM	SM	Total	MM Frequency
Acute myeloid leukemia	99	664	763	13.00
Myelodysplastic syndrome	158	550	708	22.30
Chronic lymphocytic leukemia	181	1742	1923	9.41
Head and Neck SCC	454	4385	4839	9.40
Lung (NSCLC)	400	7142	7542	5.30
Colorectal carcinoma	420	7794	8214	5.10
Gastric carcinoma	85	1542	1627	5.20
Ovarian carcinoma	120	4223	4343	2.70
Pancreatic carcinoma	35	1686	1721	2.03

Analysis was performed using the 2017 release of the UMD *TP53* database (82,134 *TP53* mutations and 75,448 patients). For each cancer type, the number of patients with either one (SM) or more than 1 (MM) *TP53* variants is reported.

Supplemental Table S2: Frequency of patients with multiple *TP53* variants according to the type of alterations in AML and MDS .

Tumors with multiple TP53 variants	Number
2 single nucleotide substitutions	161
1 single nucleotide substitution and 1 null variant	70
3 different variants (all types)	19
4 different variants (all types)	5
5 different varints (all types)	1
6 differents variants (all types)	1

Two hundred fifty-seven patients with MDS and AML carry multiple *TP53* mutations (**Supplemental Table S1).** Two different nucleotide substitutions were detected in 161 tumors (62%), while, in 70 patients, the second event was a *TP53* null event (either splice or indel variants).

Supplemental Figure S1 a to f: analysis of *TP53* variant loss of function in AML and MDS tumors with multiple *TP53* mutations

Figure S1a: analysis of AML and MDS tumors with two different *TP53* single nucleotide substitutions

Figure S1b: analysis of AML and MDS tumors with one *TP53* single nucleotide substitution and one frameshift TP53 variant

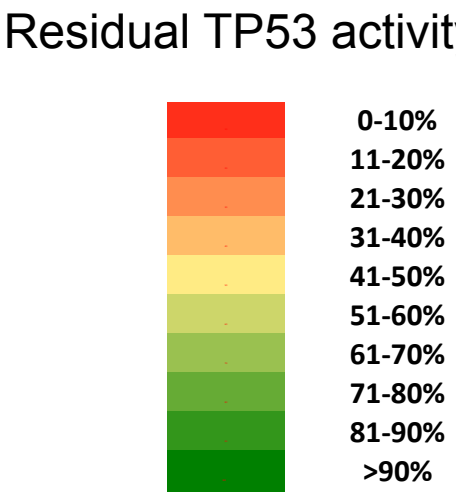
Figure S1c: analysis of AML and MDS tumors with 3 different *TP53* alterations

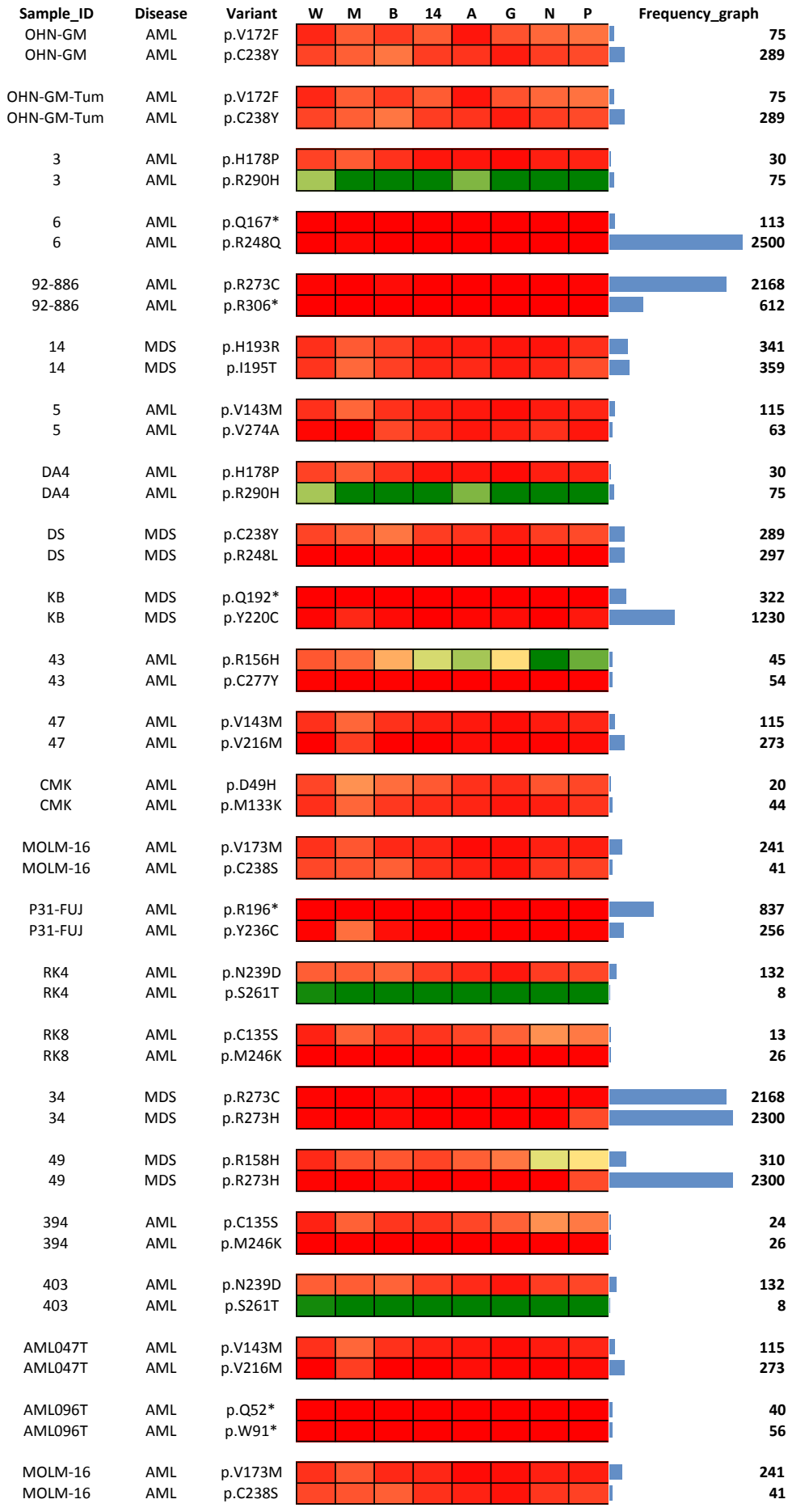
Figure S1d: analysis of AML and MDS tumors with 4 different *TP53* alterations

Figure S1e: analysis of AML and MDS tumors with 5 different *TP53* alterations

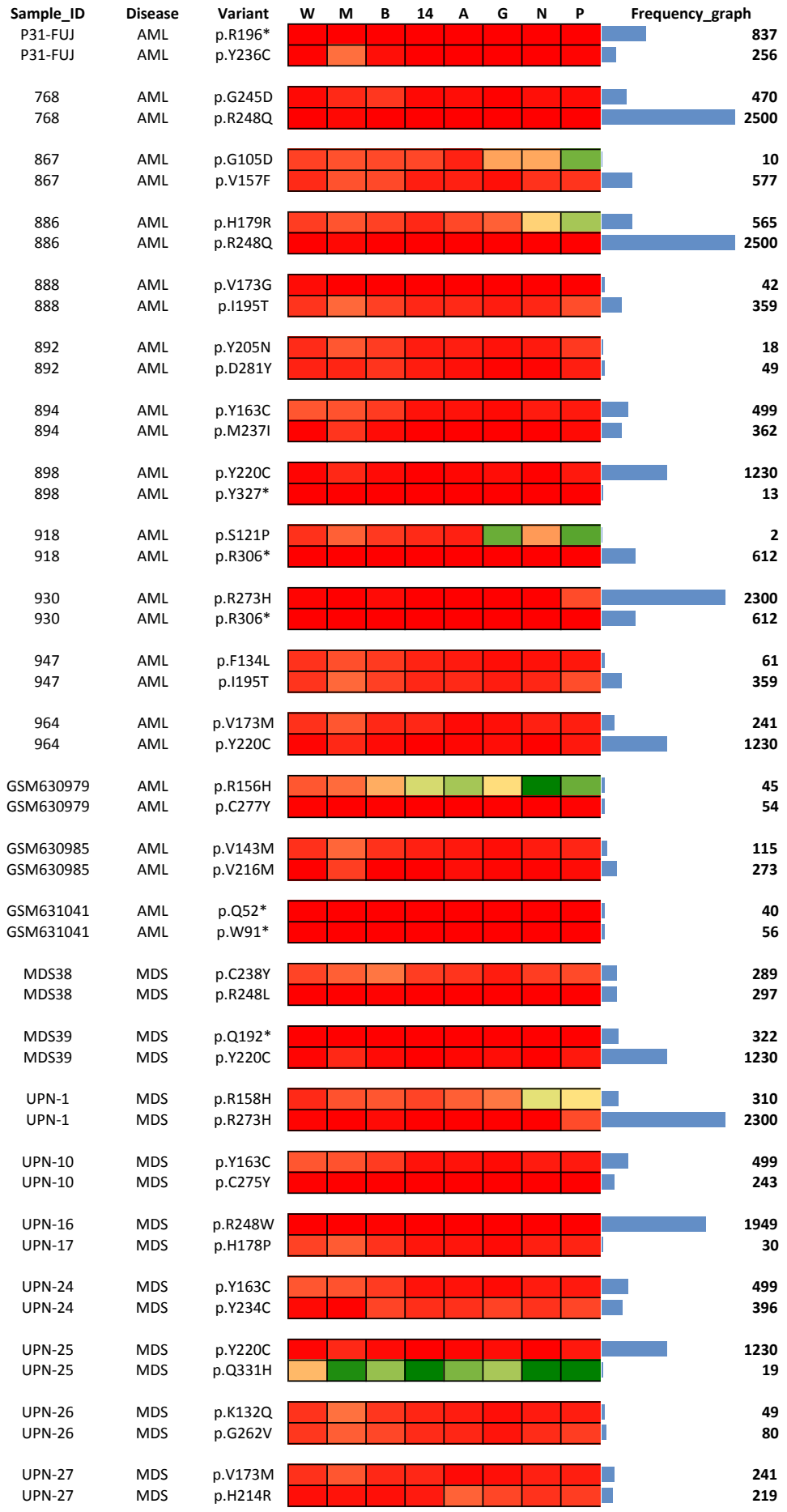
Figure S1f: analysis of AML and MDS tumors with 6 different *TP53* alterations

Residual transactivating activity for WAF (W), MDM2 (M), BAX (B), 14-3-3-σ (14), AIP (A), GADD45 (G), NOXA (N) and P53R2 (P) ranges from 0 (red) to 100% (green). The frequency of the variant in the database is shown as both a bar and a number in the right part of the Figure.





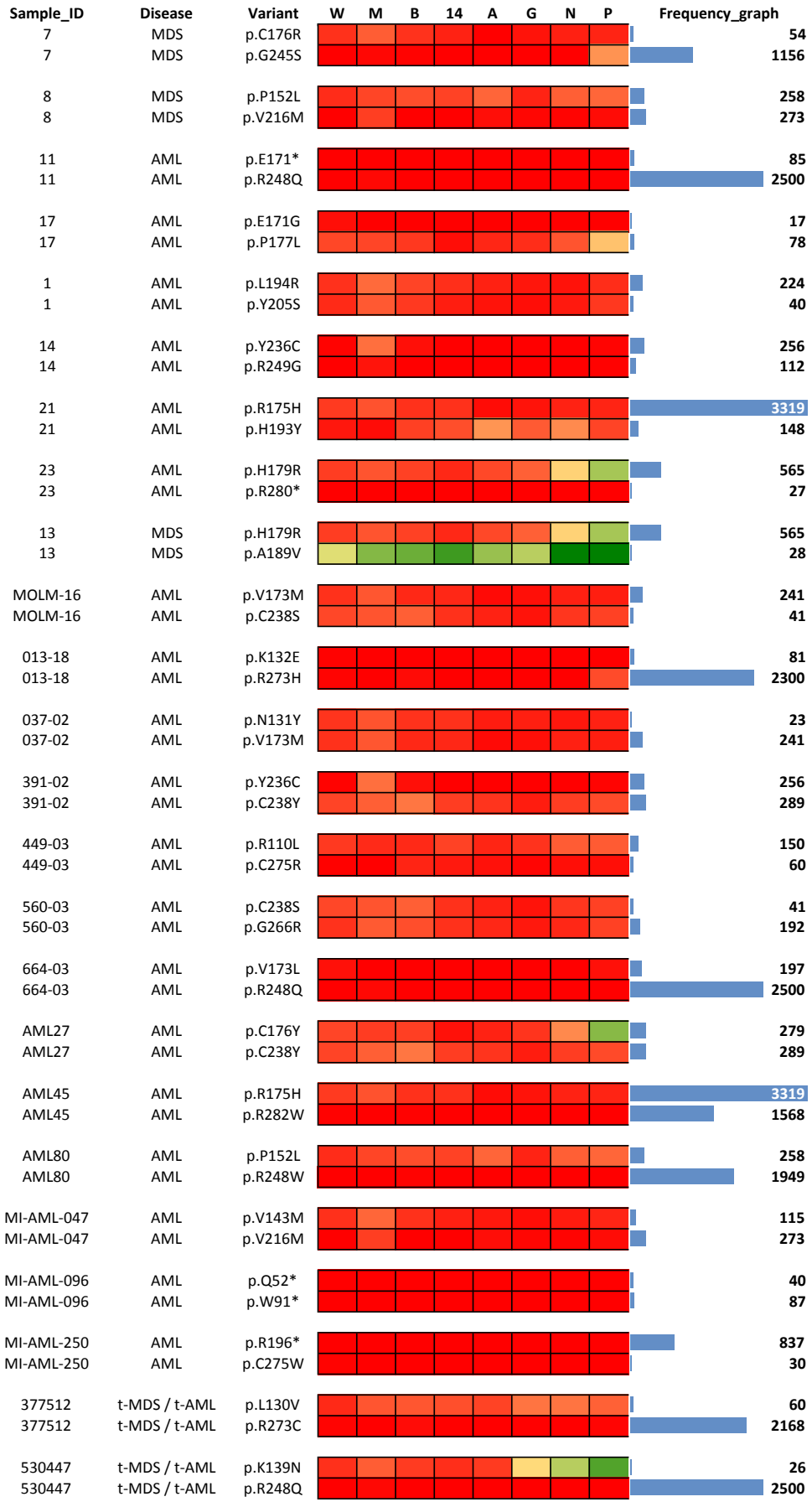
Supplemental Figure S1a (part 1)



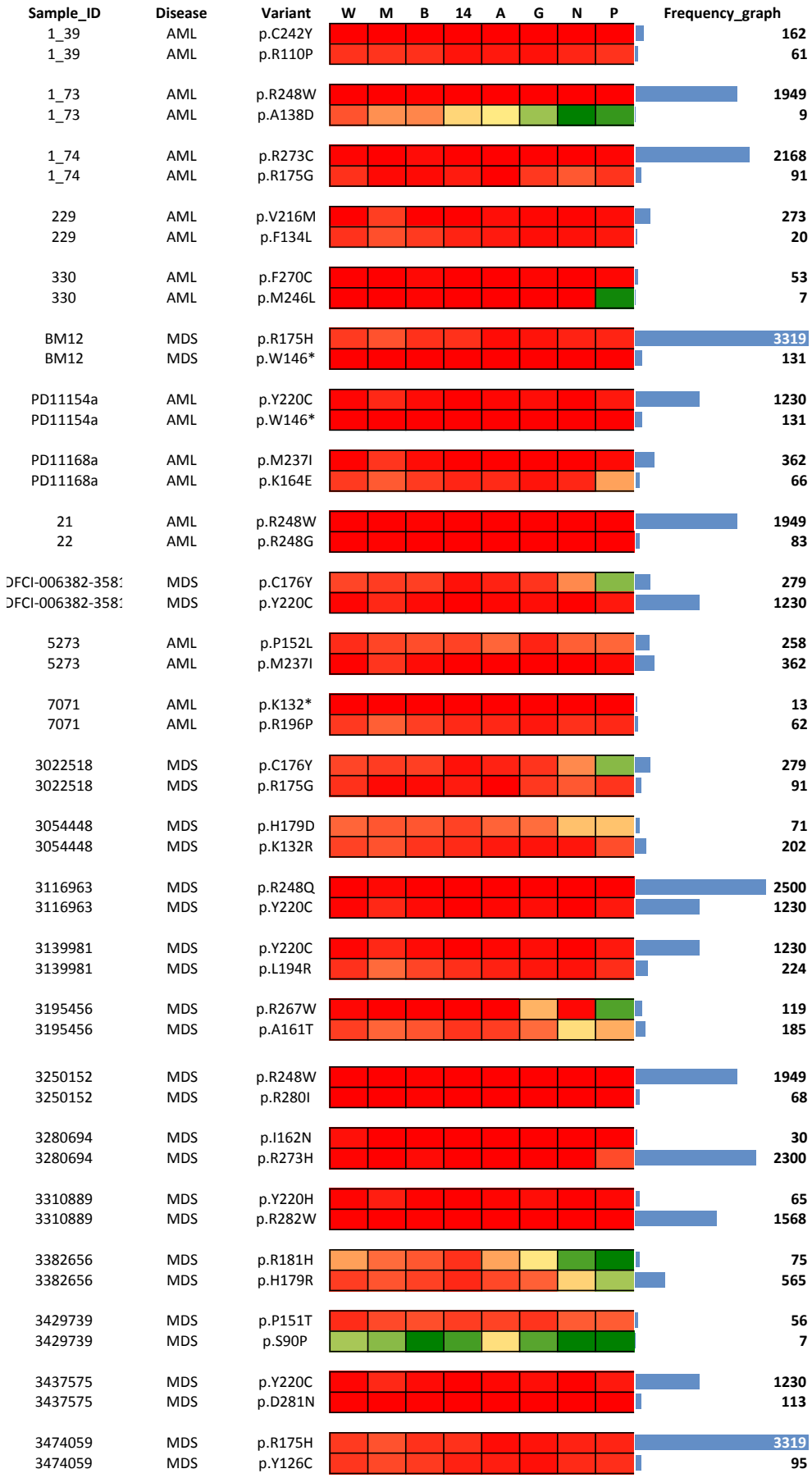
Supplemental Figure S1a (part 2)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
UPN-9	MDS	p.L43*									9
UPN-9	MDS	p.C238Y									289
37135	AML	p.T118A									2
37135	AML	p.R306*									612
VMK9	MDS	p.V143M									115
VMK9	MDS	p.H179R									565
TCGA-AB-2908	AML	p.C141W									41
TCGA-AB-2908	AML	p.Q317*									157
176267	MDS	p.R175H									3319
176267	MDS	p.R267Q									45
369682	MDS	p.Q144P									21
369682	MDS	p.V216M									273
558896	MDS	p.V143M									115
558896	MDS	p.Y163C									499
15	MDS	p.I195F									91
15	MDS	p.V272L									134
2	MDS	p.C238Y									289
2	MDS	p.R282W									1568
4	MDS	p.A161D									48
4	MDS	p.S241C									129
6	MDS	p.C238F									147
6	MDS	p.R273H									2300
BLE1	MDS	p.A161T									185
BLE1	MDS	p.C176Y									279
BLE2	MDS	p.Y220C									1230
BLE2	MDS	p.S241Y									89
BLE7	MDS	p.P151T									56
BLE7	MDS	p.C275Y									243
BLE9	MDS	p.C238Y									289
BLE9	MDS	p.R273H									2300
35	AML	p.H179R									565
35	AML	p.V274F									139
7	AML	p.R110P									61
7	AML	p.N247I									39
AML-46	AML	p.R248Q									2500
AML-46	AML	p.E258A									19
10	MDS	p.C238F									147
10	MDS	p.R273H									2300
11	MDS	p.C238Y									289
11	MDS	p.R282W									1568
12	MDS	p.S127F									173
12	MDS	p.R249G									112
13	MDS	p.R248G									83
13	MDS	p.R280G									135
14	MDS	p.A161D									48
14	MDS	p.S241C									129
4	MDS	p.P222S									10
4	MDS	p.R248Q									2500

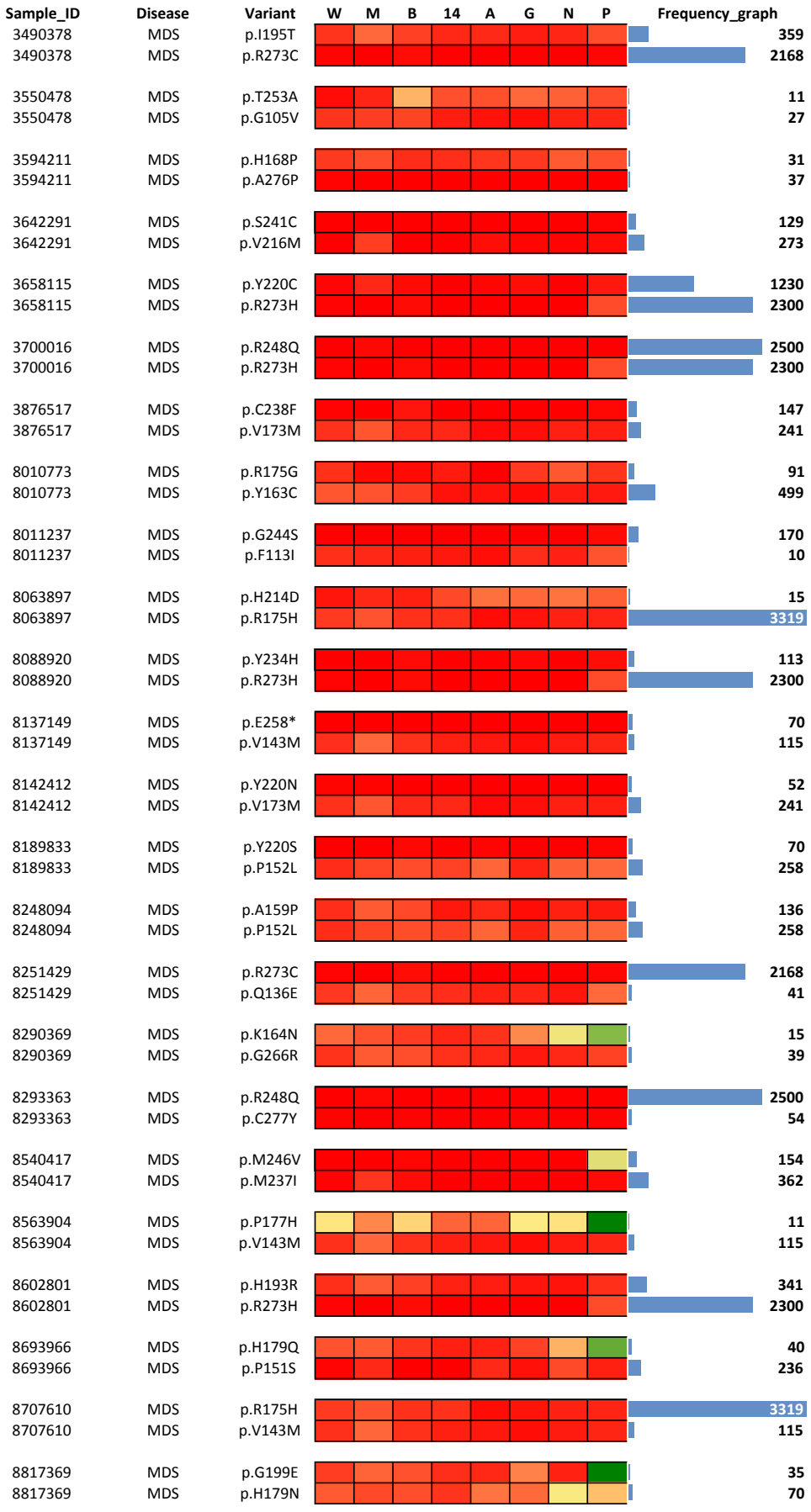
Supplemental Figure S1a (part 3)



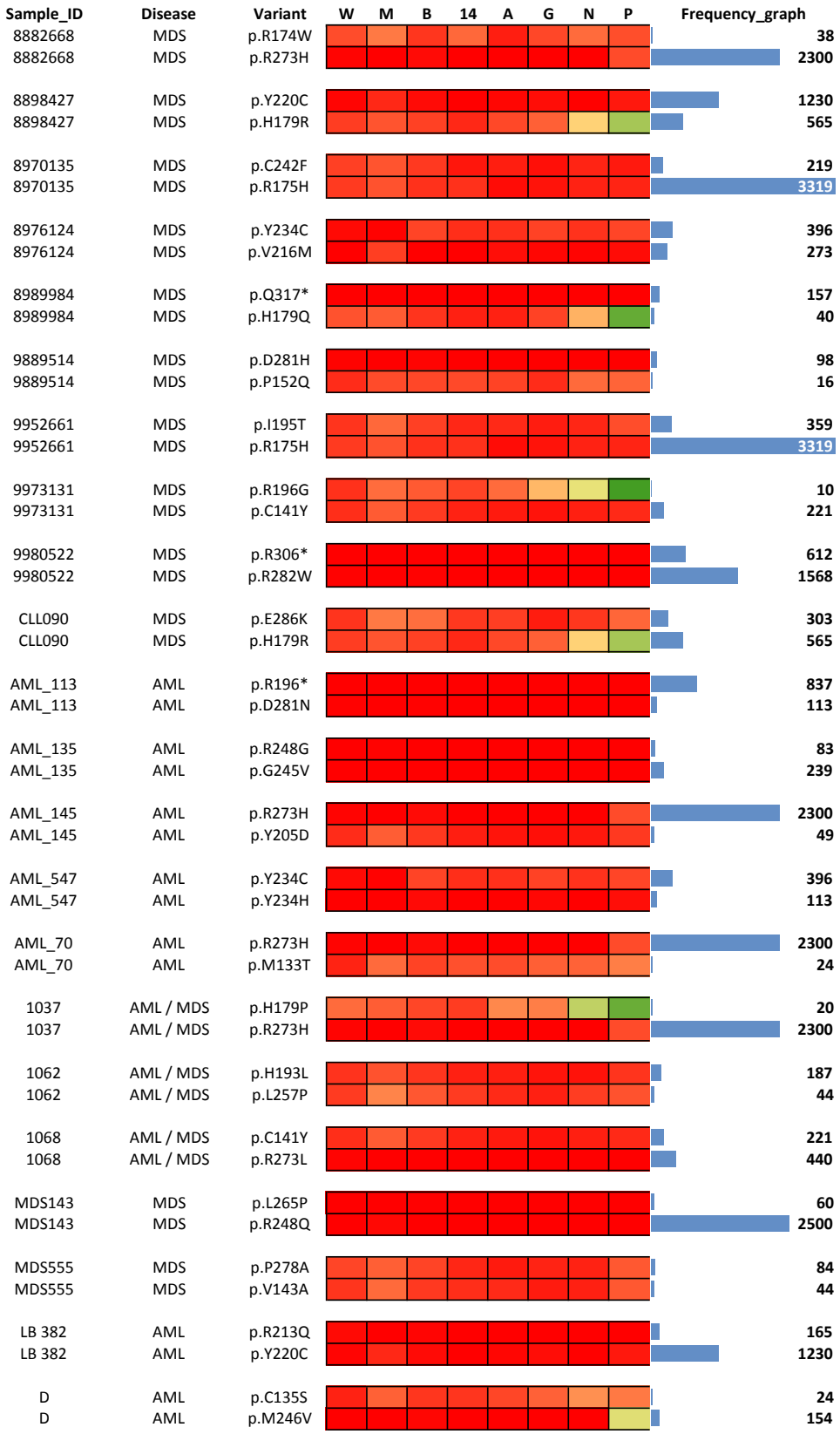
Supplemental Figure S1a (part 4)



Supplemental Figure S1a (part 5)



Supplemental Figure S1a (part 6)



Supplemental Figure S1a (part 7)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
3709100	MDS	p.E171G									17
3709100	MDS	Splice_site									107
3751801	MDS	p.E258D									20
3751801	MDS	Splice_site									28
7803089	MDS	Splice_site									118
7803089	MDS	Splice_site									20
7884287	MDS	p.Q104*									85
7884287	MDS	Splice_site									54
7891530	MDS	p.V272E									34
7891530	MDS	Splice_site									55
7939626	MDS	Splice_site									118
7939626	MDS	Frameshift_del									8
8040244	MDS	p.V272L									36
8040244	MDS	Splice_site									38
8223840	MDS	p.V173M									241
8223840	MDS	Splice_site									80
8238788	MDS	p.R282W									1568
8238788	MDS	Splice_site									18
8590527	MDS	p.R213*									1214
8590527	MDS	Splice_site									18
8592327	MDS	p.R280K									206
8592327	MDS	Splice_site									44
8610838	MDS	p.V173M									241
8610838	MDS	Splice_site									58
8677482	MDS	Splice_site									78
8677482	MDS	Splice_site									80
8706141	MDS	p.R248W									1949
8706141	MDS	Splice_site									69
8809836	MDS	p.M246V									154
8809836	MDS	Splice_site									14
8921623	MDS	p.R306*									612
8921623	MDS	Splice_site									118
9889870	MDS	Splice_site									56
9889870	MDS	Frameshift_del									4
9978666	MDS	p.I232F									27
9978666	MDS	Splice_site									55
CLL125	MDS	Frameshift_del									173
CLL125	MDS	Frameshift_del									73
AML_500	AML	p.R282W									1568
AML_500	AML	Splice_site									7
1072	AML / MDS	p.C124R									13
1072	AML / MDS	Splice_site									11
1085	AML / MDS	p.P250L									150
1085	AML / MDS	Splice_site									17
800684	AML / MDS	Splice_site									56
800684	AML / MDS	p.A161T									185

Supplemental Figure S1b (part 1)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
40	MDS	p.N239D									132
40	MDS	Frameshift_ins									1
134	t-MDS / t-AML	p.C135S									24
134	t-MDS / t-AML	Frameshift_indel									4
37	t-MDS / t-AML	p.R248Q									2500
37	t-MDS / t-AML	Frameshift_ins									2
3	MDS	p.R282P									62
3	MDS	Frameshift_del									7
OCI-M1	AML	p.L145R									25
OCI-M1	AML	Splice_site									20
RK14	AML	Splice_site									80
RK14	AML	p.Y220H									65
400	AML	Splice_site									80
400	AML	p.Y220H									65
13	AML	Splice_site									78
13	MDS	p.R283P									98
19	AML	p.M237I									362
19	AML	Frameshift_ins									14
OCI-M1	AML	p.L145R									25
OCI-M1	AML	Splice_site									20
11	AML	p.R158H									310
11	AML	Frameshift_ins									1
899	AML	p.A138V									121
899	AML	Frameshift_ins									1
900	AML	p.Y220C									1230
900	AML	Frameshift_del									48
907	AML	Splice_site									15
907	AML	p.P278S									301
933	AML	p.H179R									565
933	AML	Frameshift_del									5
UPN-11	MDS	p.G266E									221
UPN-11	MDS	Frameshift_ins									1
TCGA-AB-2829	AML	Splice_site									42
TCGA-AB-2829	AML	p.R280G									135
TCGA-AB-2878	AML	p.S215G									71
TCGA-AB-2878	AML	Frameshift_del									6
TCGA-AB-2938	AML	p.H179R									565
TCGA-AB-2938	AML	Frameshift_del									48
137404	MDS	p.V272L									134
137404	MDS	Frameshift_del									1
693881	MDS	Splice_site									64
693881	MDS	p.M237I									362
20	MDS	p.G245S									1156
20	MDS	Frameshift_indel									2
4	AML	p.V216M									273
4	AML	Frameshift_del									24

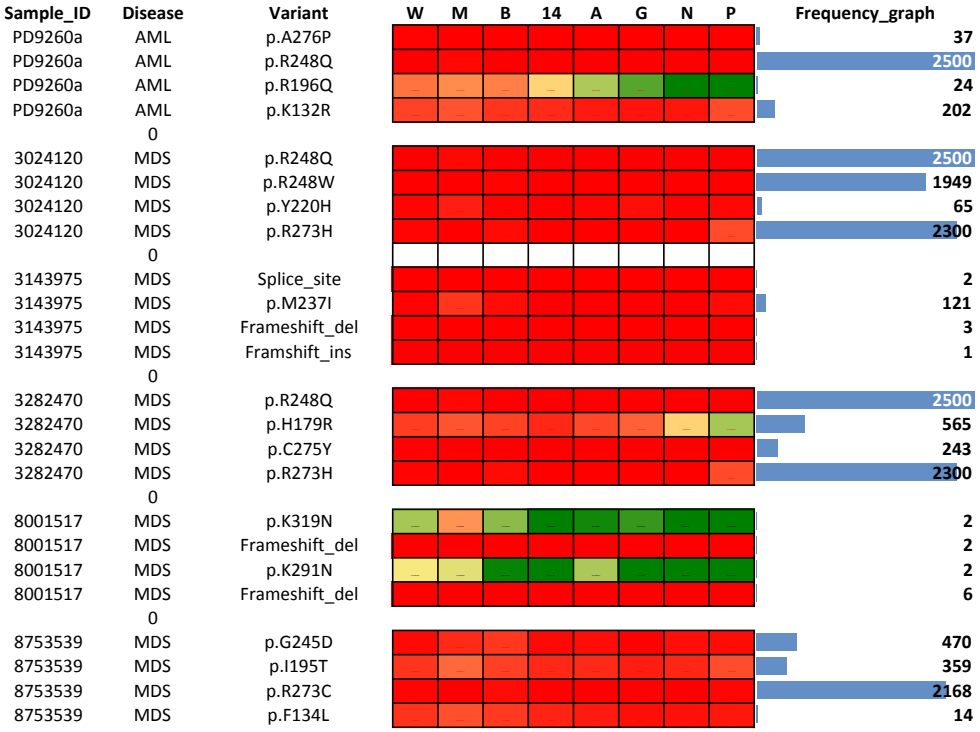
Supplemental Figure S1b (part 2)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
002-27	AML	p.V172G									9
002-27	AML	Frameshift_del									7
363-01	AML	Splice_site									9
363-01	AML	p.C275Y									243
198041	t-MDS / t-AML	p.Y163H									76
198041	t-MDS / t-AML	Frameshift_del									2
400992	t-MDS / t-AML	p.R273C									2168
400992	t-MDS / t-AML	Frameshift_del									3
433687	t-MDS / t-AML	p.R306*									612
433687	t-MDS / t-AML	Frameshift_del									32
837334	t-MDS / t-AML	p.G112R									1
837334	t-MDS / t-AML	Frameshift_ins									1
889867	t-MDS / t-AML	p.L265P									60
889867	t-MDS / t-AML	Frameshift_del									34
1_42	AML	Splice_site									118
1_42	AML	Frameshift_del									1
1_54	AML	p.C135F									188
1_54	AML	Splice_site									80
5_110	AML	p.V216M									273
5_110	AML	Splice_site									55
PD11178a	AML	p.G262V									80
PD11178a	AML	Frameshift_del									2
PD11213a	AML	p.S215R									56
PD11213a	AML	Splice_site									80
PD11215a	AML	Frameshift_del									21
PD11215a	AML	Frameshift_ins									1
PD9312a	AML	Splice_site									42
PD9312a	AML	p.R175H									3319
3115973	MDS	Splice_site									80
3115973	MDS	Splice_site									28
3157019	MDS	p.G244C									144
3157019	MDS	Frameshift_ins									3
3329749	MDS	p.R273H									2300
3329749	MDS	Frameshift_ins									1
3333824	MDS	p.Y103*									14
3333824	MDS	Splice_site									107
3431933	MDS	Splice_site									60
3431933	MDS	Frameshift_del									1
3468465	MDS	p.G245D									470
3468465	MDS	Splice_site									25
3490509	MDS	Frameshift_del									7
3490509	MDS	Frameshift_ins									3
3556079	MDS	p.A159P									136
3556079	MDS	Frameshift_ins									1
3586308	MDS	p.Q165*									144
3586308	MDS	Splice_site									53
3668658	MDS	Splice_site									14
3668658	MDS	Splice_site									44

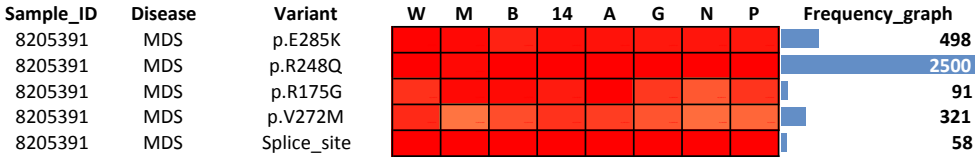
Supplemental Figure S1b (part 3)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
CG2	AML	p.Y236H									44
CG2	AML	p.M237V									34
CG2	AML	p.R248Q									2500
AML250T	AML	p.Y163C									499
AML250T	AML	p.R196*									837
AML250T	AML	p.C275W									30
VMK4	MDS	p.R175H									3319
VMK4	MDS	p.C176R									54
VMK4	MDS	p.C176W									66
6	MDS	p.R175H									3319
6	MDS	p.C238Y									289
6	MDS	p.G245D									470
3058001	MDS	p.C238Y									289
3058001	MDS	p.M237I									362
3058001	MDS	Frameshift_ins									2
3083841	MDS	p.I195T									359
3083841	MDS	p.L194R									224
3083841	MDS	p.C135Y									220
3140470	MDS	p.S241P									32
3140470	MDS	p.Y220C									1230
3140470	MDS	p.R249K									58
3278690	MDS	p.R283C									113
3278690	MDS	p.G244S									170
3278690	MDS	Frameshift_del									24
3598212	MDS	p.G245V									239
3598212	MDS	p.R273H									2300
3598212	MDS	Frameshift_ins									1
3771132	MDS	p.R306*									612
3771132	MDS	p.R213*									1214
3771132	MDS	p.R280T									268
3809429	MDS	p.R248W									1949
3809429	MDS	p.C238R									59
3809429	MDS	p.K132N									89
8372975	MDS	p.R213*									1214
8372975	MDS	p.Y220C									1230
8372975	MDS	Splice_site									52
8527776	MDS	p.Q167*									113
8527776	MDS	p.V274A									63
8527776	MDS	Splice_site									9
8536126	MDS	p.M237I									362
8536126	MDS	Frameshift_del									6
8536126	MDS	Frameshift_del									1
8742054	MDS	p.R248W									1949
8742054	MDS	p.R175H									3319
8742054	MDS	p.R273H									2300
8866992	MDS	p.G244V									56
8866992	MDS	p.R213Q									165
8866992	MDS	p.C275Y									243
8919150	MDS	p.H193R									341
8919150	MDS	p.R175H									3319
8919150	MDS	p.R273H									2300
CLL117	MDS	p.H193D									32
CLL117	MDS	Splice_site									1
CLL117	MDS	p.R248Q									2500
MDS155	MDS	p.L194R									224
MDS155	MDS	p.H179R									565
MDS155	MDS	p.C176S									53

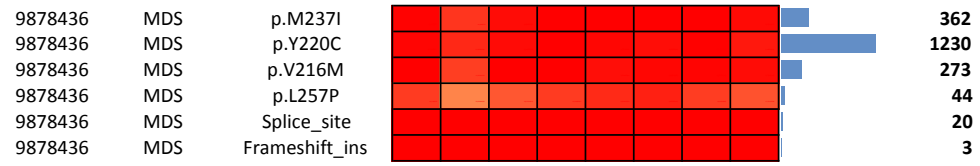
Supplemental Figure S1c



Supplemental Figure S1d

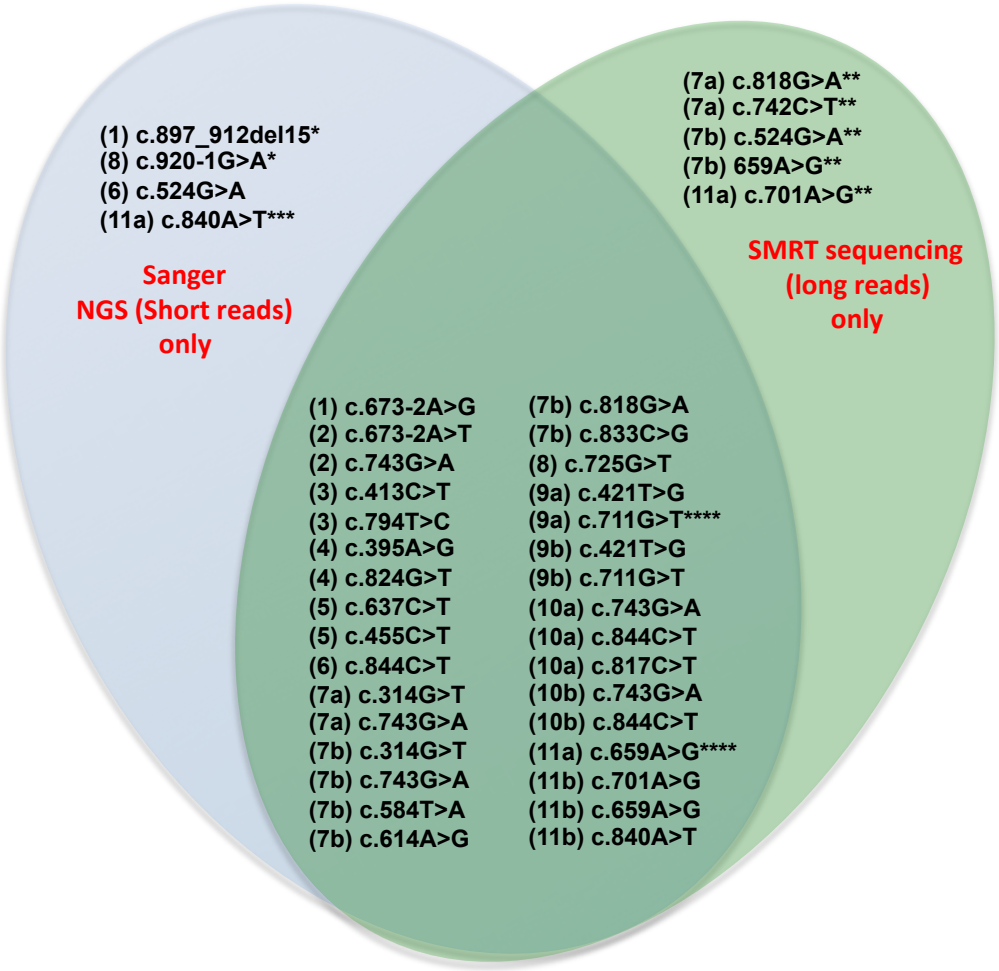


Supplemental Figure S1e



Supplemental Figure S1f

Supplemental Figure S2: Venn diagram showing the mutational concordance of validated somatic variants based on the sequencing strategy



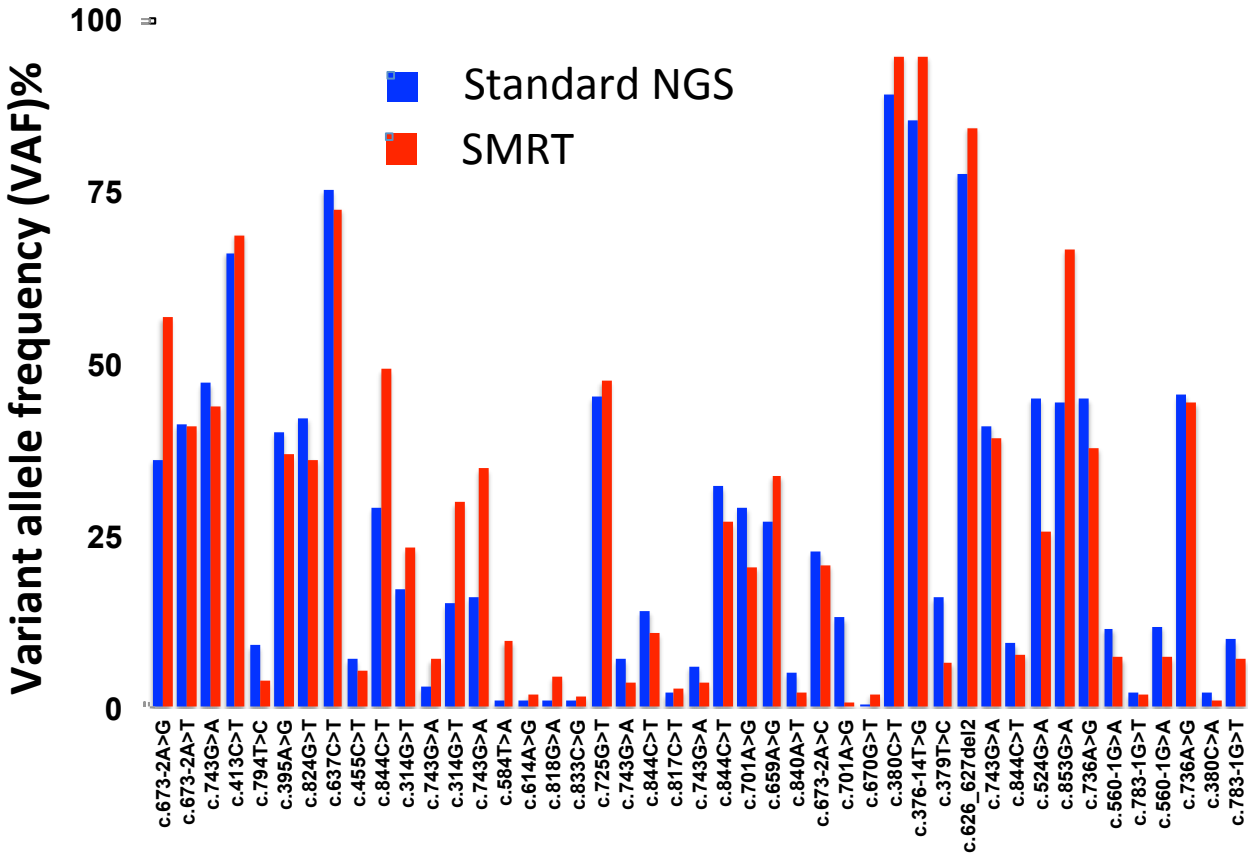
* Mutation outside the amplicon used for SMRT analysis

** Mutation detected at very low frequency by SMRT (1-2%)

*** Mutation detected at high frequency in the recurrent sample and identified at low frequency after manual examination of the primary sample.

**** Mutation detected by SMRT and identified at very low frequency by manual examination of the short-read NGS data.

Supplemental Figure S3: Variant allele frequency (VAF) observed for all *TP53* variants identified by both classical NGS and SMRT methodologies



Supplemental Figure S4a to k: Detailed analysis of the 11 patients included in this study.

For each patient, 3 sections are available i.e. clinical information, sequencing and haplotype

Clinical information: this section includes age, disease information, treatment and 17p status

Sequencing: Sanger sequencing and/or standard NGS analysis is shown in the left part. No allelic distribution can be inferred from this type of analysis. SMRT sequencing (right part) provides an accurate picture of the allelic distribution of each *TP53* variant, as well as the remaining wt allele. The frequencies of the different alleles are shown in brackets.

Red triangle:	*TP53* variants identified by both types of analysis.
White triangle:	*TP53* variants detected only by SMRT sequencing.
Yellow triangle:	*TP53* Variants detected after manual examination but below the cut-off used for clinical validation
Black triangle : sequencing	*TP53* variants outside the amplicon used for the long range
Blue triangle:	*TP53* variants not detected by long range sequencing.

Haplotype: allelic distribution of all *TP53* variants (germline and somatic) according to the SMRT analysis
Somatic *TP53* variants are shown in red. Biallelic germline variants (SNP) are shown in white (allele 1) and green (allele 2) to make a distinction for heterozygote cases (see cases Fr10).

Clinical information

Patient Fr1

Disease: de novo MK-AML
Age: 77
Treatment: none (diagnosis)
17p status: no deletion

Sequencing

Sanger/NGS (short reads)

- c.673-2A>G (36%)
- c.897_912del15 (61%)

SMRT sequencing (long reads)

- c.673-2A>G (56.6 %)
- wt (43.3 %)

- Variant outside the amplicon used for the long range sequencing
- Variant detected by both analyses

Haplotype

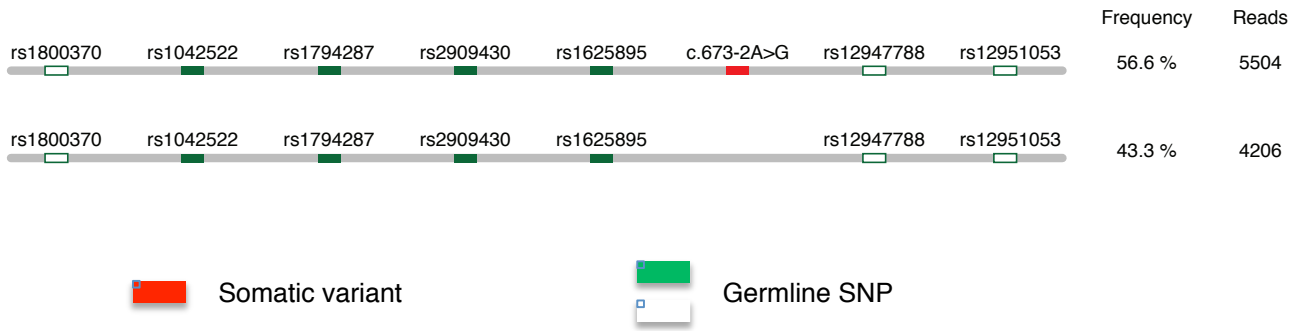


Figure S4a

Clinical information

Patient Fr2

Disease: de novo MK-AML
Age: 63
Treatment: none
17p status: deletion (CGH array)

Sequencing

Sanger/NGS (short reads)

c.743G>A (47 %)
c.673-2A>T (41 %)

SMRT sequencing (long reads)

c.743G>A (43.8 %)
c.673-2A>T (40.8 %)
wt (15.3 %)

Variant detected by both analyses

Haplotype

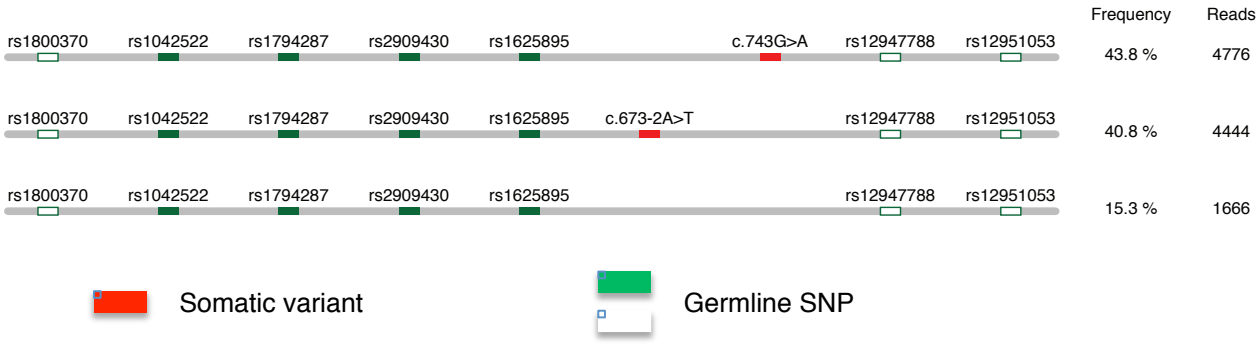


Figure S4b

Patient Fr3

Clinical information

Disease: de novo MK-AML
Age: 73
Treatment: none
17p status: deletion (FISH)

Sequencing

Sanger/NGS (short reads)

c.413C>T (66%)
c.794T>C (9 %)

SMRT sequencing (long reads)

c.413C>T (68.6%)
c.794T>C (3.76%)
wt (27.5%)

Variant detected by both analyses

Haplotype

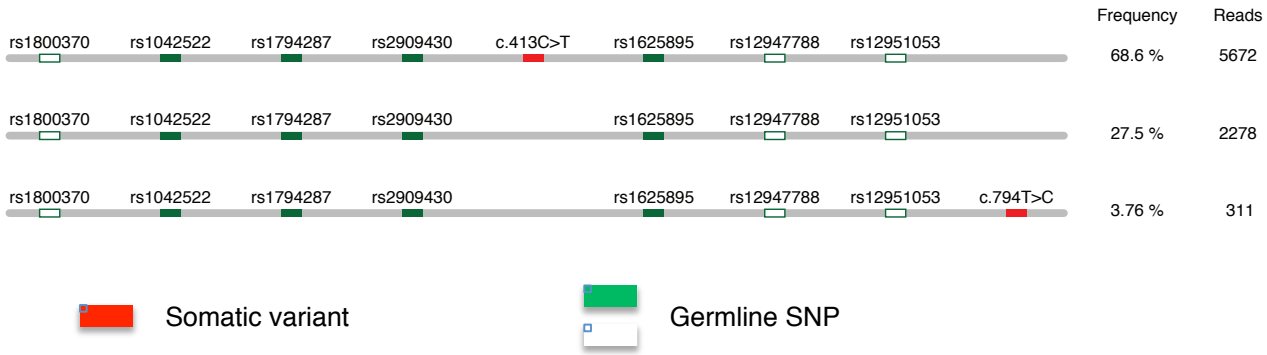


Figure S4c

Patient Fr4

Clinical information

Disease: de novo MK-AML
Age: 78
Treatment: none
17p status: no deletion (FISH)

Sequencing

Sanger/NGS (short reads)

c.842G>T (42%)
c.395A>G (40%)

SMRT sequencing (long reads)

c.842G>T (36%)
c.395A>G (36.6%)
wt (27.3 %)

Variant detected by both analyses

Haplotype

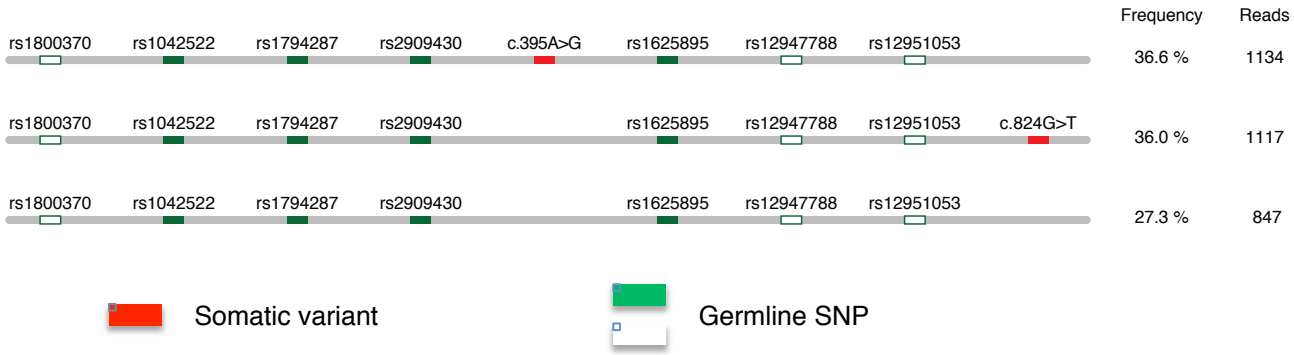


Figure S4d

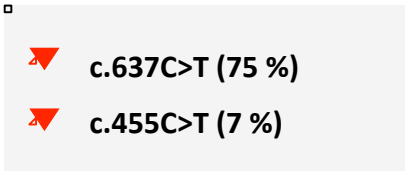
Patient Fr5

Clinical information

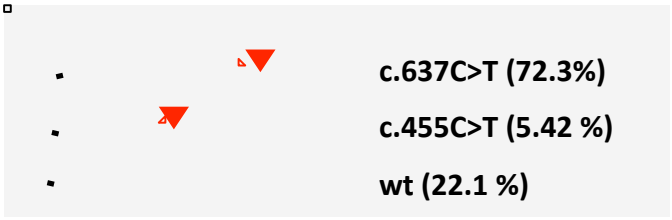
Disease: de novo MK-AML
Age: 73
Treatment: none
17p status: deletion (FISH)

Sequencing

Sanger/NGS (short reads)



SMRT sequencing (long reads)



Variant detected by both analyses

Haplotype

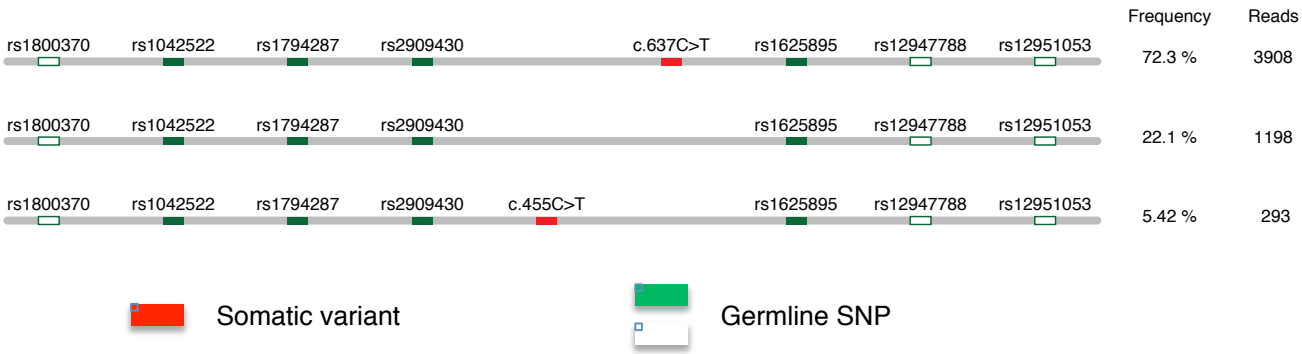


Figure S4e

Clinical information

Patient Fr6

Disease: s-AML (post LR-MDS del5q)
Age: 75
Treatment: Lenalidomide
17p status: no deletion (karyotype)

Sequencing

Sanger/NGS (short reads)

c.524G>A (24%)
c.844C>T (29%)

SMRT sequencing (long reads)

c.844C>T (49%)
wt (50.9 %)

Variant detected by both analyses
Variant not detected by long range sequencing

Haplotype

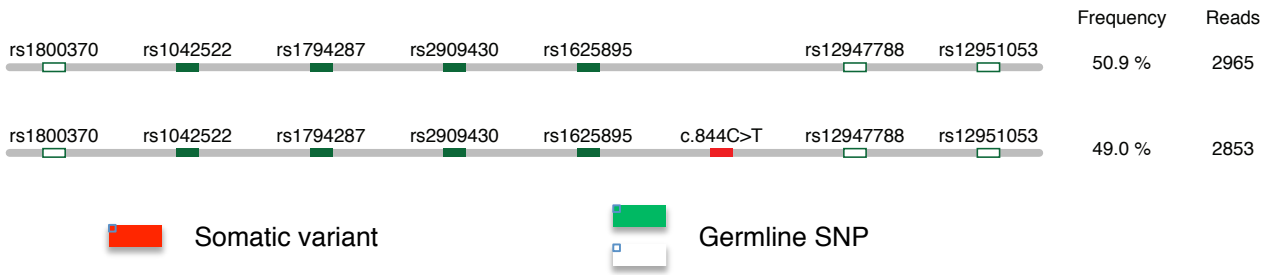


Figure S4f

Clinical information

Patient Fr7
Sample 7a and 7b

Disease: LR-MDS del5q
Age: 73
Treatment sample 7a: Lenalidomide
sample 7b: Lenalidomide
17p status: no deletion

Sequencing

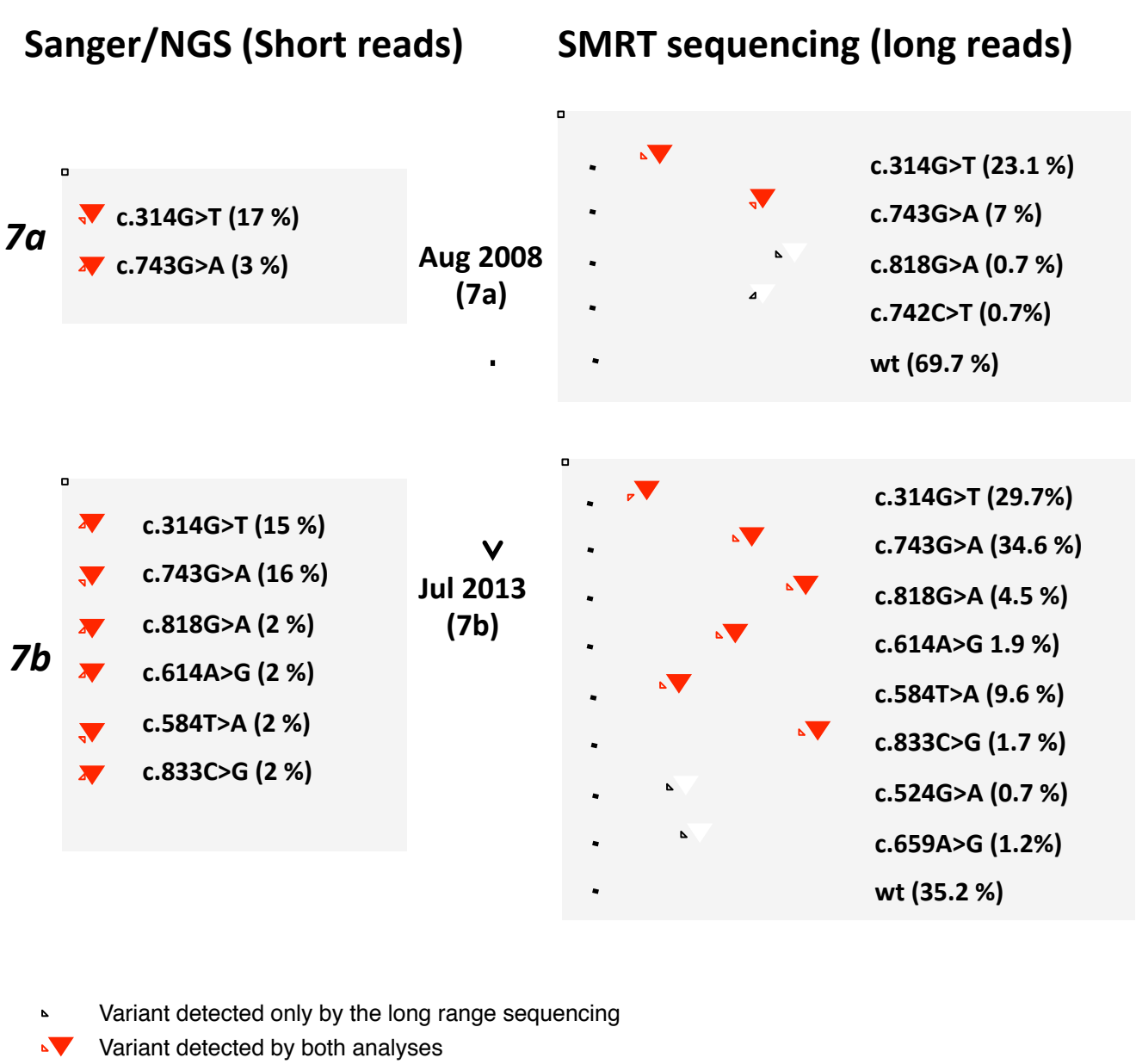
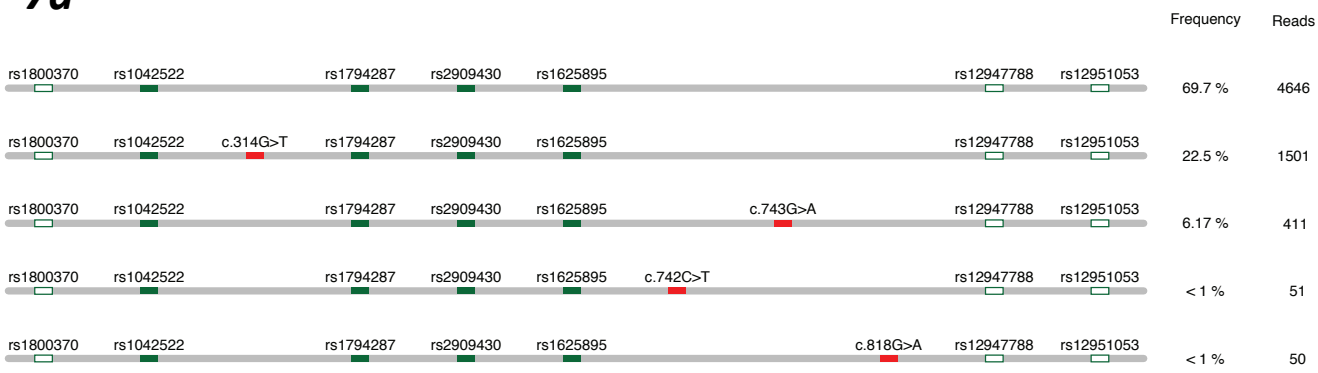


Figure S4g

Haplotype

7a



Aug 2008
(7a)

.

Jul 2013
(7b)

7b

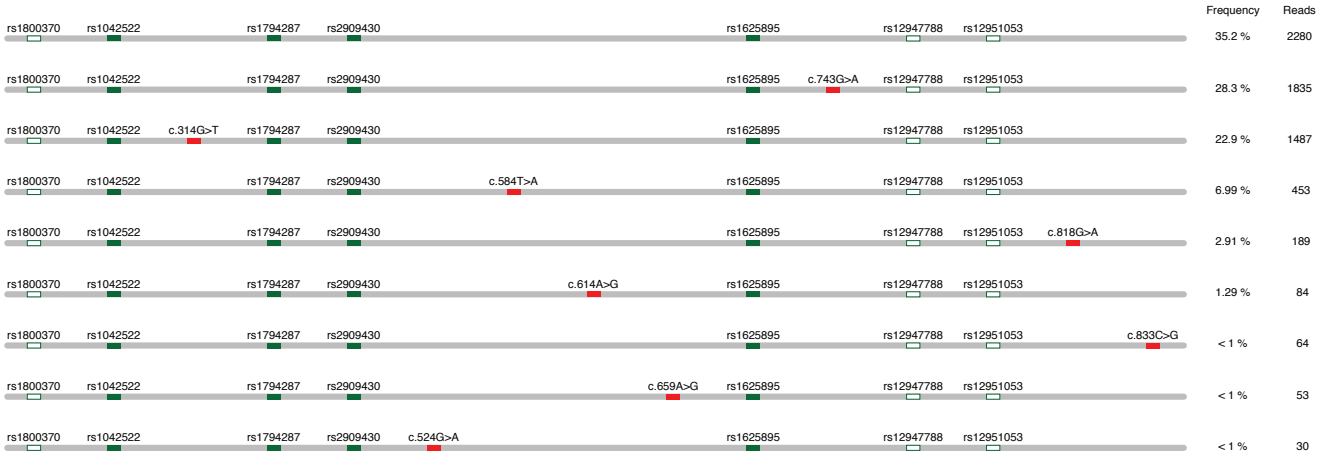


Figure S4g

Clinical information

Disease: s-AML (post LR-MDS del5q)
Age: 72
Treatment: Lenalidomide
17p status: deletion

Sequencing

Sanger/NGS (short reads)

- c.725G>T (45%)
- c.920-1G>A (37%)

SMRT sequencing (long reads)

- c.725G>T (49.4%)
- wt (50.6%)

Variant outside the amplicon used for the long range sequencing

Variant detected by both analyses

Haplotype

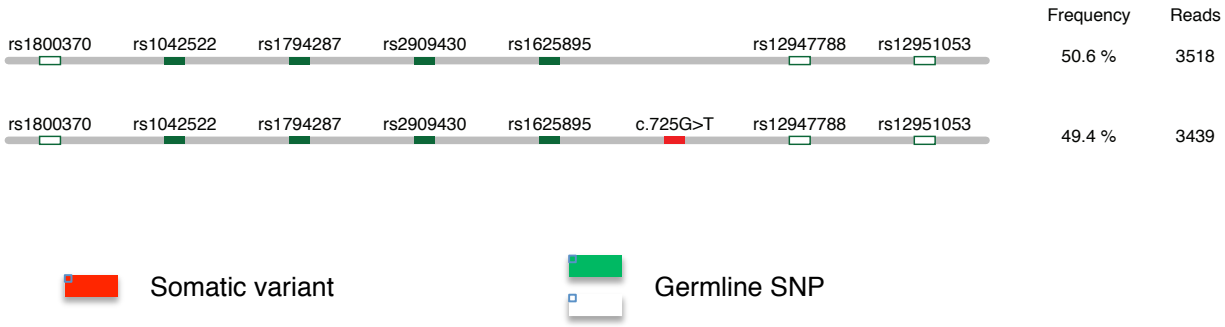


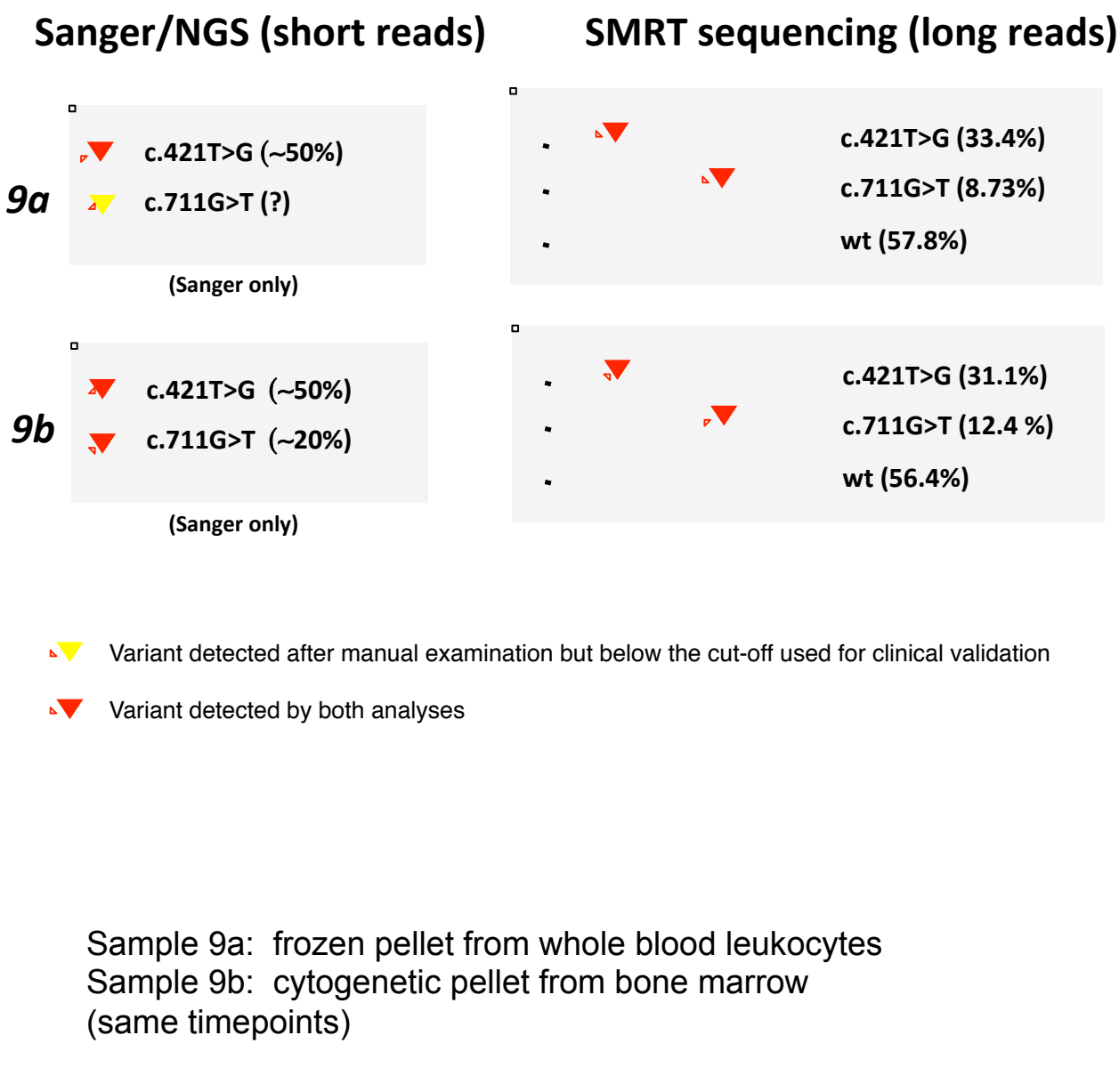
Figure S4h

Clinical information

Patient Fr9
Sample 9a and 9b

Disease: s-AML (post LR-MDS del5q)
Age: 76
Treatment: Lenalidomide
17p status: deletion (partial)

Sequencing



Sample 9a: frozen pellet from whole blood leukocytes
Sample 9b: cytogenetic pellet from bone marrow
(same timepoints)

Figure S4i

Haplotype

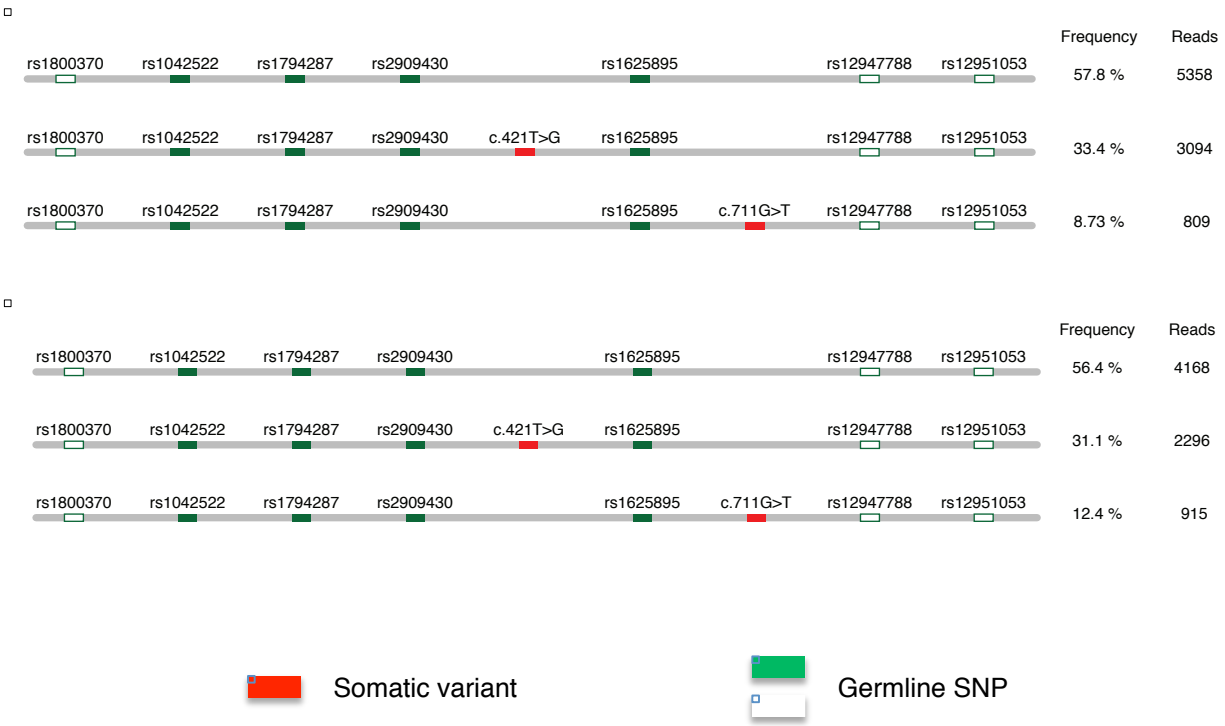


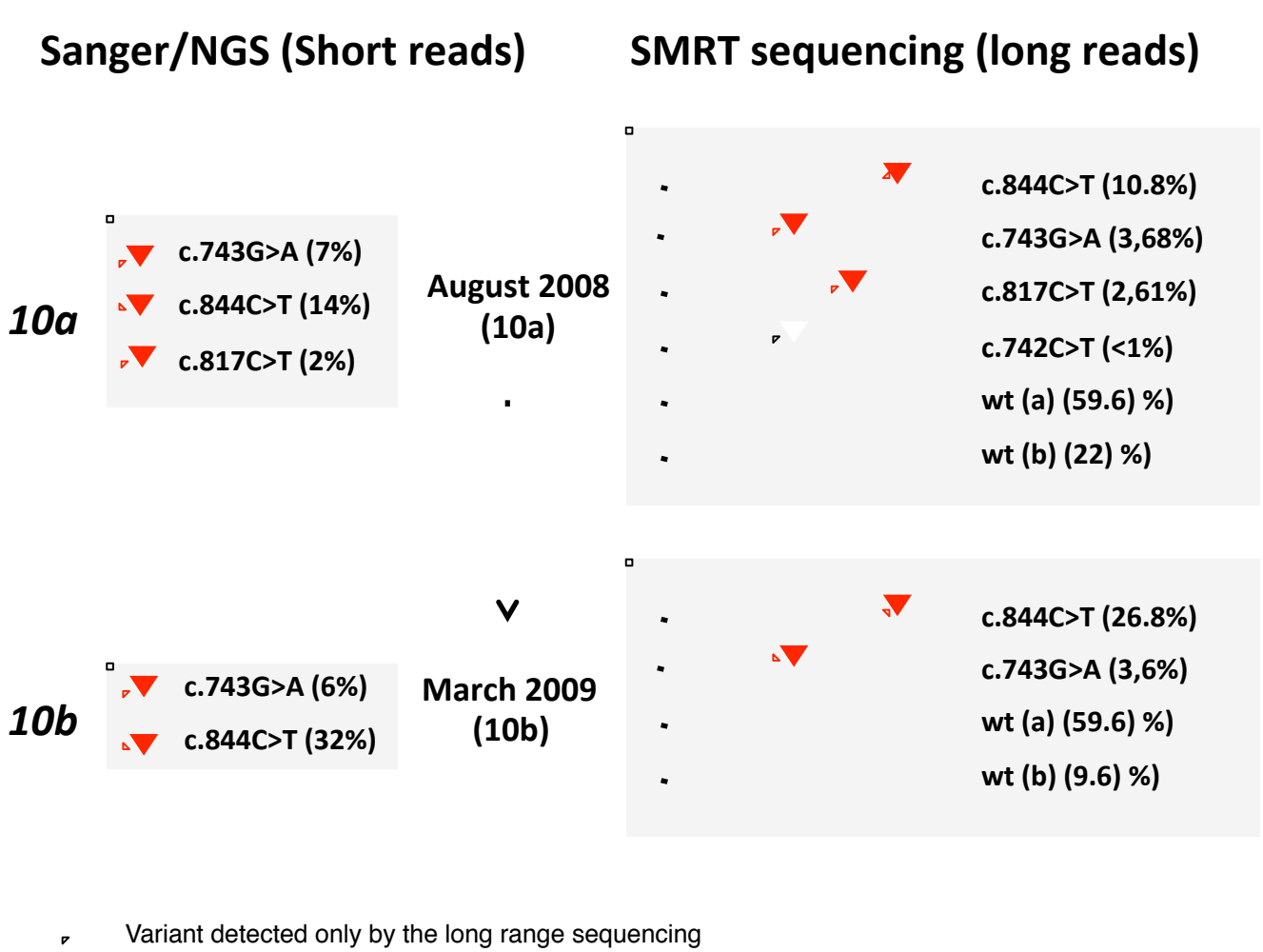
Figure S4i

Clinical information

Patient Fr10
Sample 10a and 10b

Disease: sample 10a: LR-MDS del5q
sample 10b: s-AML (post LR-MDS del5q)
Age: 69
Treatment sample 10a: Lenalidomide
sample 10b: Lenalidomide
17p status: no deletion (karyotype)

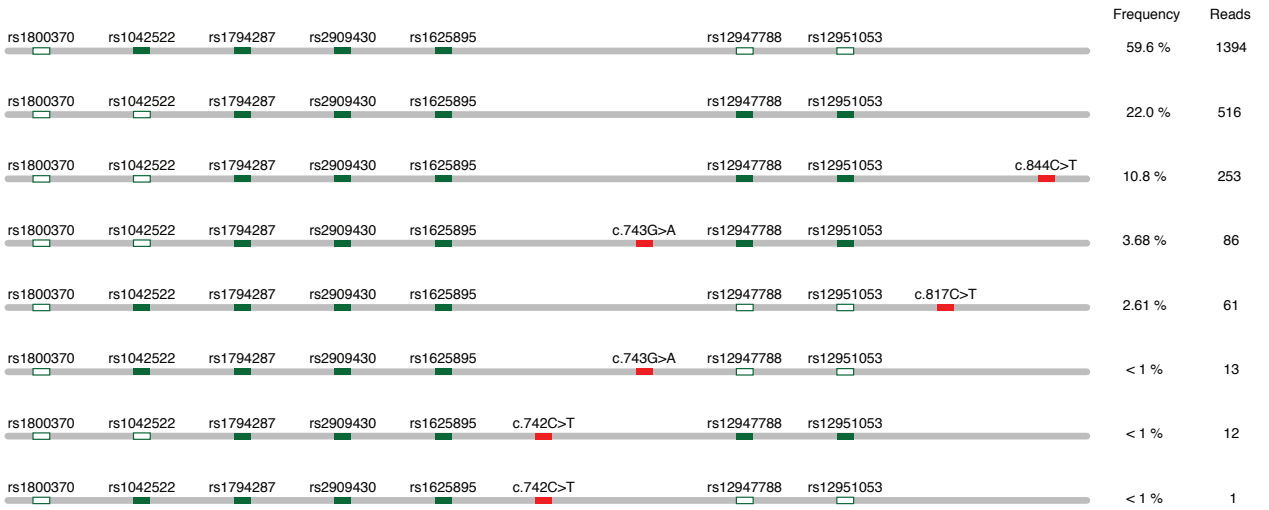
Sequencing



Variant detected only by the long range sequencing
Variant detected by both analyses

Figure S4j

Haplotype



August 2008
(10a)

.

March 2009
(10b)

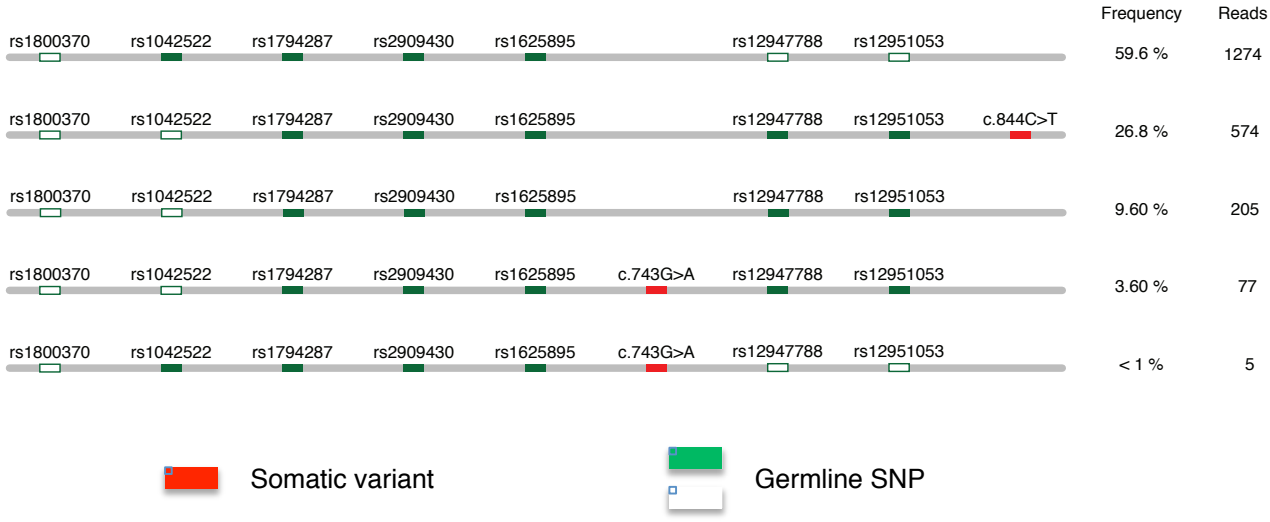


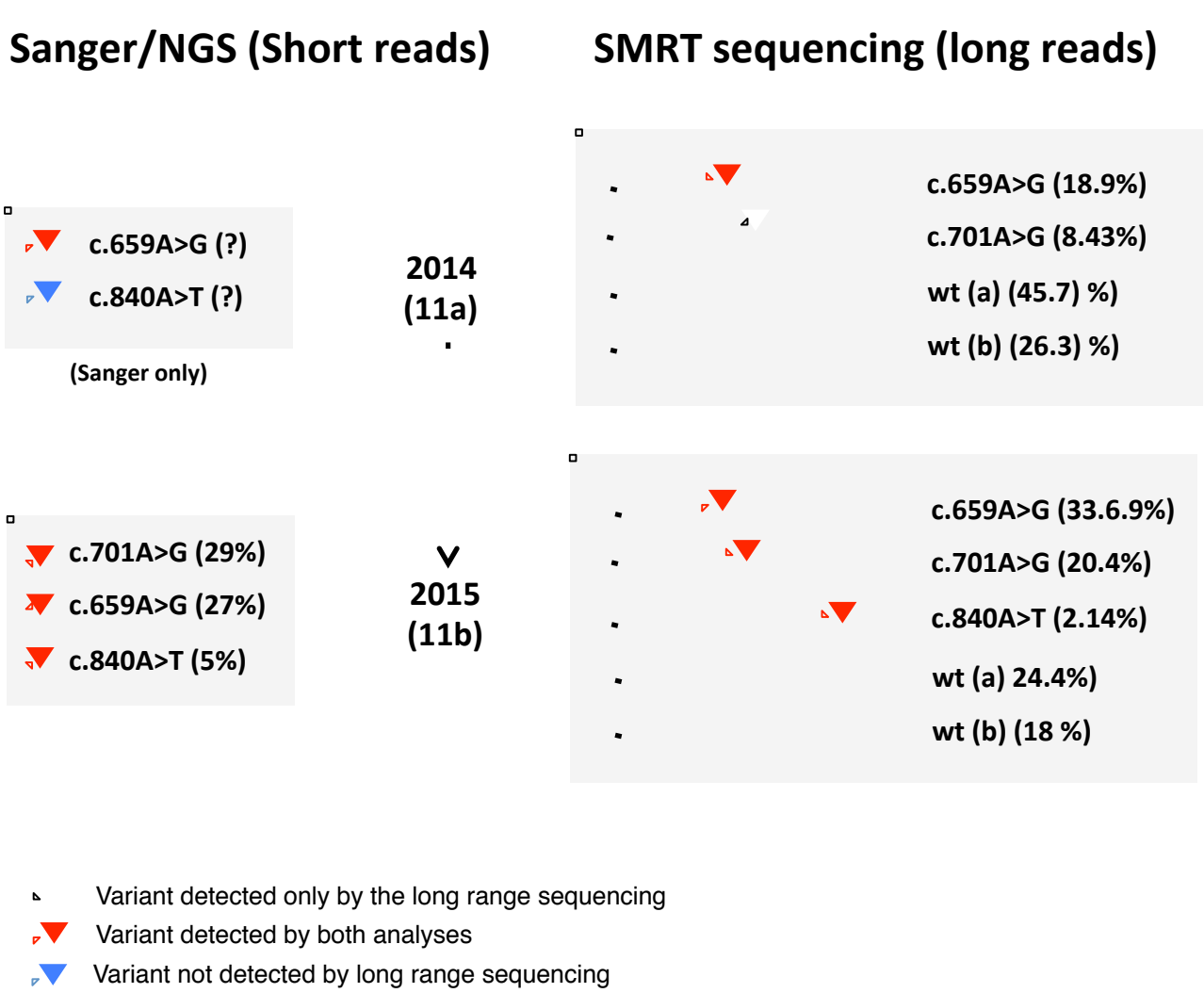
Figure S4j

Clinical information

Patient Fr11
Sample 11a and 11b

Disease: sample 1: LR-MDS del5q
sample 2: LR-MDS del5q)
Age: 85
Treatment sample 11a: Lenalidomide
sample 11b: Lenalidomide
17p status: no deletion (karyotype)

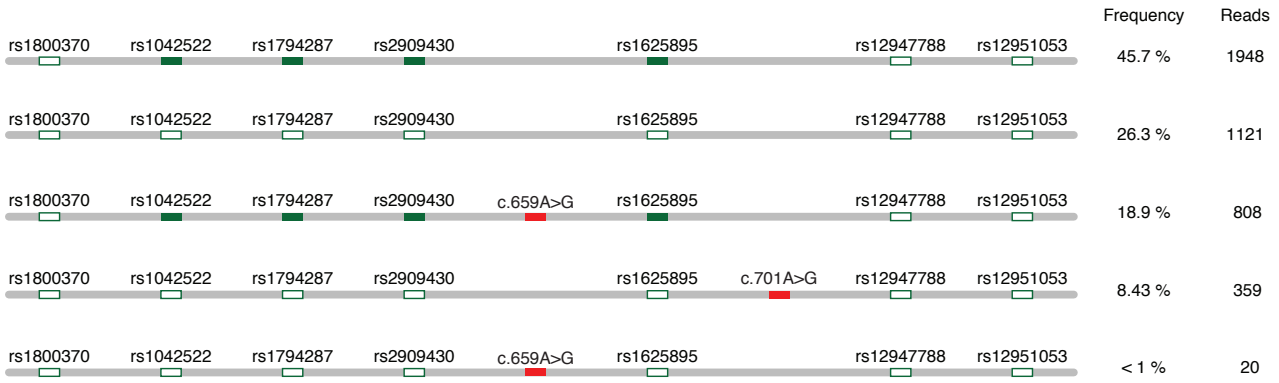
Sequencing



Variant detected only by the long range sequencing
Variant detected by both analyses
Variant not detected by long range sequencing

Figure S4k

Haplotype



Oct 2014
(11a)

Juil 2015
(11b)

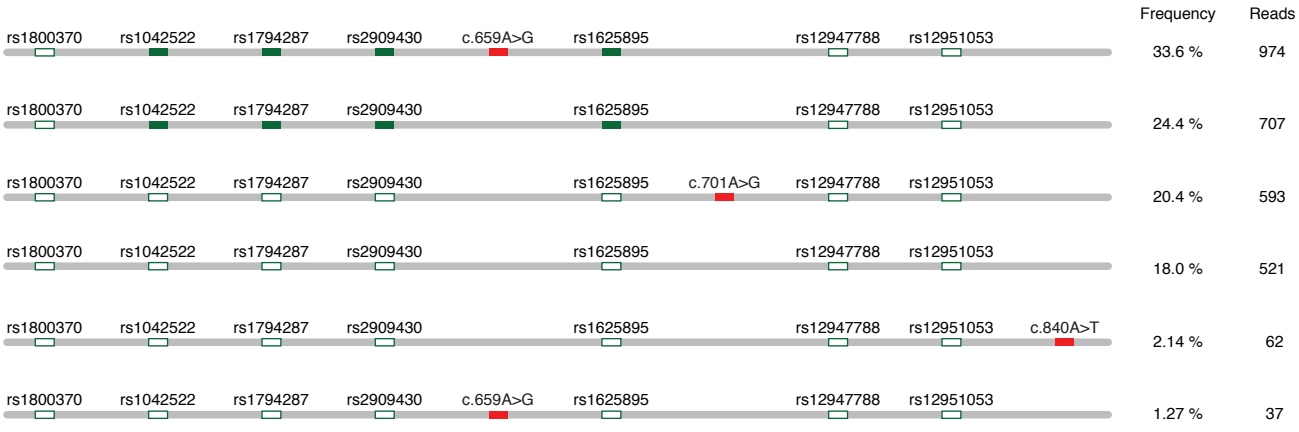
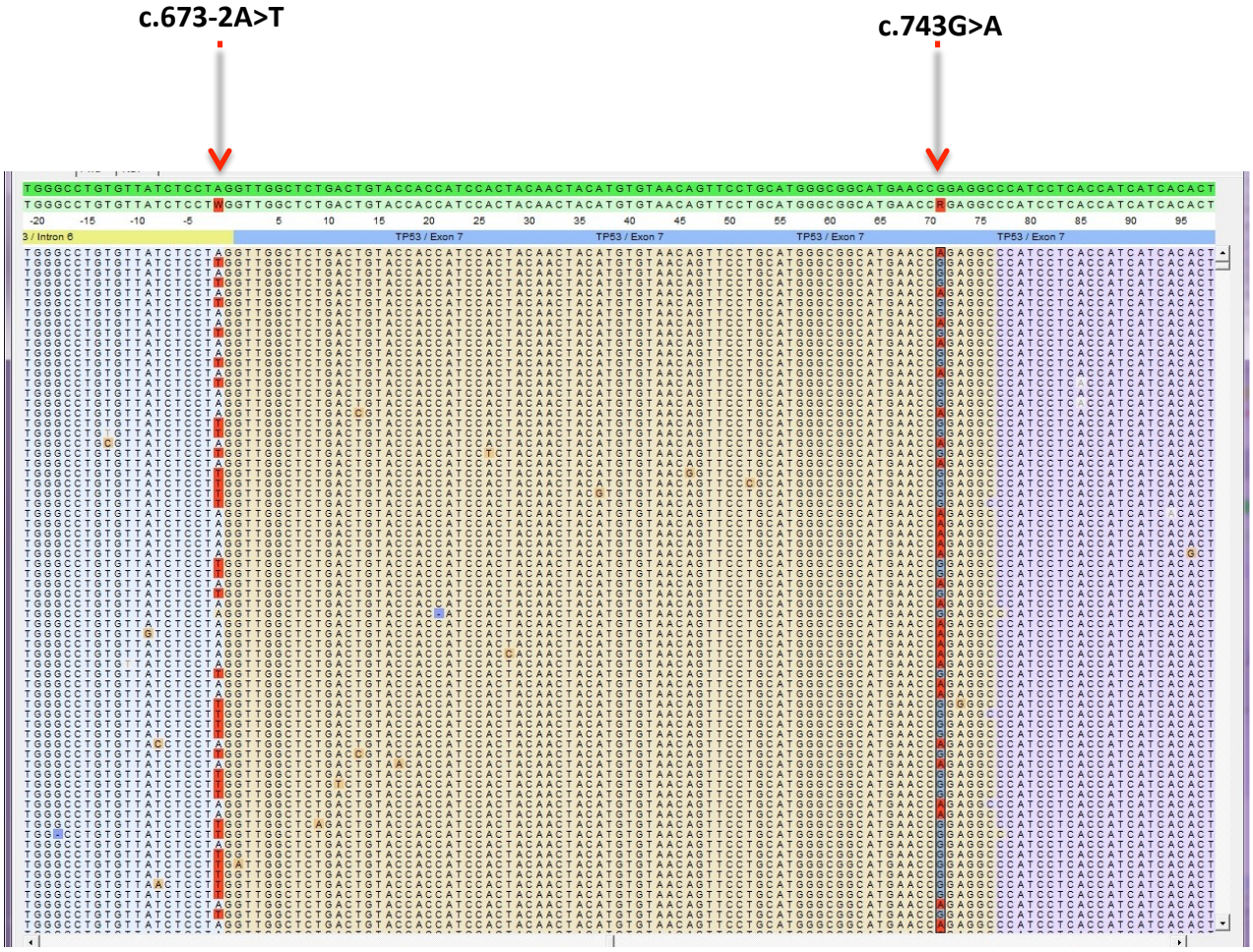


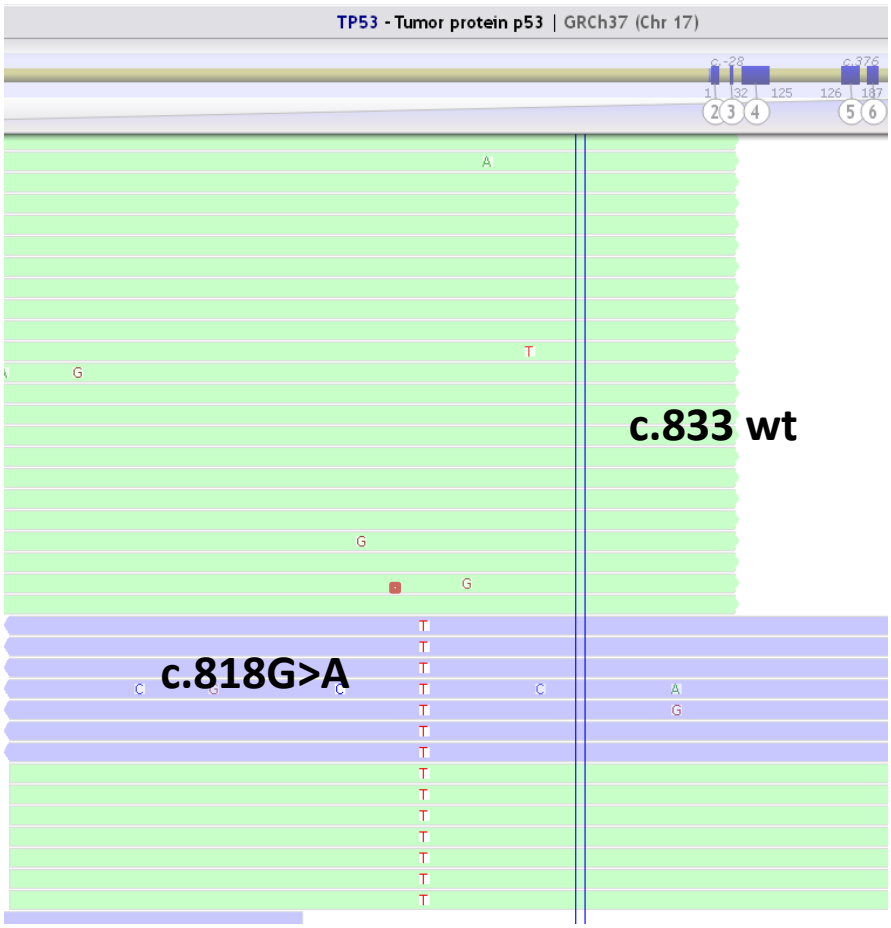
Figure S4k

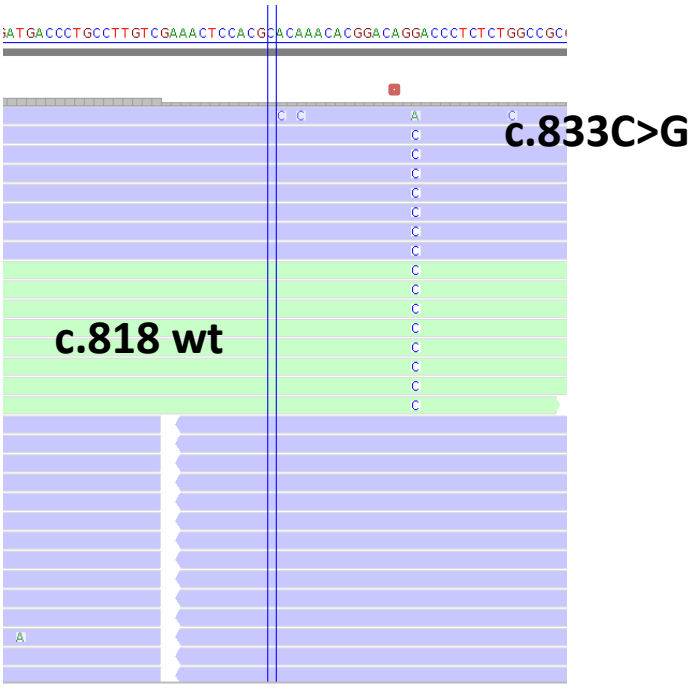
Supplemental Figure S5: Visualization of NGS alignment, confirming that the two variants are located on different alleles for patient Fr2.



Patient Fr2

Supplemental Figure S6: Visualization of NGS alignment, confirming that the two variants are located on different alleles for patient Fr7

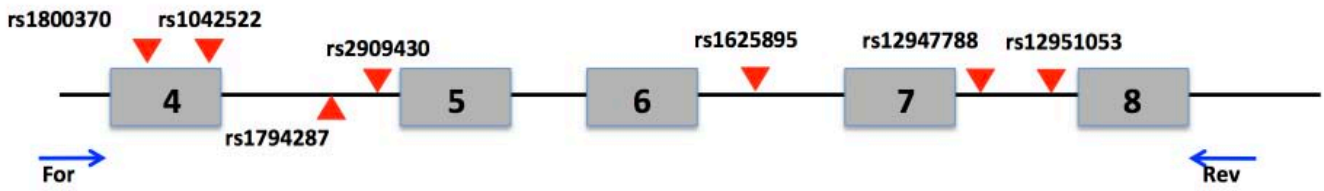




Patient Fr7
Sample 7b

Supplemental Figure S7: strategy used for the analysis of *TP53* mutations

A



B

SNP	cDNA_variant	Genomic_variant (HG19)	Protein variant
rs1800370	c.108G>A	chr17:g.7579579G>A	p.P36=
rs1042522	c.215C>G	chr17:g.7579472C>G	p.P72R
rs1794287	c.376-283T>C	chr17:g.7578837T>C	p.(=)
rs2909430	c.376-91G>A	chr17:g.7578645G>A	p.(=)
rs1625895	c.672+62A>C	chr17:g.7578115A>C	p.(=)
rs12947788	c.782+72C>T	chr17:g.7577427C>T	p.(=)
rs12951053	c.782+92T>G	chr17:g.7577407T>G	p.(=)

C

	Sequence	Coordinates (HG19)
Forward primer	5' cctggtcctctgactgctct 3'	7579626-7579607
Reverse primer	5' tacctcgcttagtgctccct 3	7577035-7577016