

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

Laurence Lodé,<sup>1,2</sup> Adam Ameer,<sup>3</sup> Thibault Coste,<sup>1,5</sup> Audrey Ménard,<sup>2</sup> Steven Richebourg,<sup>2,4</sup> Jean-Baptiste Gaillard,<sup>5</sup> Yannick Le Bris,<sup>2</sup> Marie-Christine Béné,<sup>2</sup> Thierry Lavabre-Bertrand<sup>5</sup> and Thierry Soussi<sup>6-8</sup>

<sup>1</sup>Hematology Laboratory, University Hospital, Montpellier, F-34000, France; <sup>2</sup>Hematology Laboratory, University Hospital, Nantes, F-44093, France; <sup>3</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden; <sup>4</sup>Cytogenetics, Hôpital du Saint Sacrement, CHU de Québec, Université Laval, Department of molecular medicine, Canada; <sup>5</sup>Department of Clinical Cytology and Cytogenetics, Nîmes University Hospital, France; <sup>6</sup>Sorbonne Université, UPMC Univ Paris 06, F- 75005 Paris, France; <sup>7</sup>INSERM, U1138, Centre de Recherche des Cordeliers, Paris, France and <sup>8</sup>Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden

Correspondence: [thierry.soussi@ki.se](mailto:thierry.soussi@ki.se)

doi:10.3324/haematol.2017.176719

**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	Page 3
SMRT sequencing of <i>TP53</i> amplicons	Page 3
Detection of SNPs and mutations in SMRT sequencing data	Page 3
Analysis of the phasing of SNPs and mutations	Page 3
<i>In silico</i> analysis of <i>TP53</i> variants in AML and MDS	Page 4

### B Supplemental Tables

**Supplemental Table S1:** frequency of patients with multiple *TP53* variants in the UMD *TP53* database.  
Page 5

**Supplemental Table S2:** frequency of tumors with multiple *TP53* variants according to the type of alterations in AML and MDS.  
Page 6

**Supplemental Table S3:** *TP53* mutations in AML and MDS patients. Supplied as an Excel file.

### C Supplemental Figures

**Supplemental Figure S1a to f:** analysis of *TP53* variant loss of function in AML and MDS tumors with multiple *TP53* mutations  
Page 7-19

**Supplemental Figure S2:** Venn diagram showing the mutational concordance of validated somatic variants based on the sequencing strategy.  
Page 20

**Supplemental Figure S3:** variant allele frequency (VAF) observed for all *TP53* variants identified by both standard NGS and SMRT methodologies  
Page 21

**Supplemental Figure S4a to k:** detailed analysis of the 11 patients included in this study.  
Page 22-37

**Supplemental Figure S5:** visualization of NGS alignment, confirming that the two variants are located on different alleles for patient Fr2.  
Page 38

**Supplemental Figure S6:** visualization of NGS alignment, confirming that the two variants are located on different alleles for patient Fr7  
Page 39

**Supplemental Figure S7:** strategy used for the analysis of *TP53* mutations  
Page 40

### D References to Supplemental Material

Page 41

## 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**).<sup>1,2</sup> 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.<sup>1</sup>, 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<sup>2</sup>. 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.<sup>3</sup>

### **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' plugin (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<sup>3</sup>. 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.<sup>4,5</sup> 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.<sup>6</sup> 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.

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

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

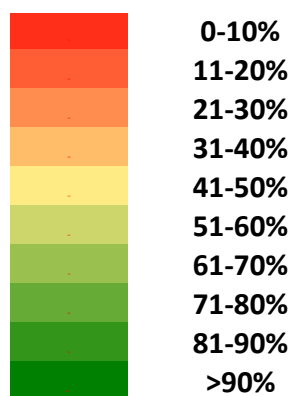
**Figure S1d:** analysis of AML and MDS tumors with 4 different *TP53* alterations

**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- $\sigma$  (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.

### Residual TP53 activity





Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
OHN-GM	AML	p.V172F	■	■	■	■	■	■	■	■	75
OHN-GM	AML	p.C238Y	■	■	■	■	■	■	■	■	289
OHN-GM-Tum	AML	p.V172F	■	■	■	■	■	■	■	■	75
OHN-GM-Tum	AML	p.C238Y	■	■	■	■	■	■	■	■	289
3	AML	p.H178P	■	■	■	■	■	■	■	■	30
3	AML	p.R290H	■	■	■	■	■	■	■	■	75
6	AML	p.Q167*	■	■	■	■	■	■	■	■	113
6	AML	p.R248Q	■	■	■	■	■	■	■	■	2500
92-886	AML	p.R273C	■	■	■	■	■	■	■	■	2168
92-886	AML	p.R306*	■	■	■	■	■	■	■	■	612
14	MDS	p.H193R	■	■	■	■	■	■	■	■	341
14	MDS	p.I195T	■	■	■	■	■	■	■	■	359
5	AML	p.V143M	■	■	■	■	■	■	■	■	115
5	AML	p.V274A	■	■	■	■	■	■	■	■	63
DA4	AML	p.H178P	■	■	■	■	■	■	■	■	30
DA4	AML	p.R290H	■	■	■	■	■	■	■	■	75
DS	MDS	p.C238Y	■	■	■	■	■	■	■	■	289
DS	MDS	p.R248L	■	■	■	■	■	■	■	■	297
KB	MDS	p.Q192*	■	■	■	■	■	■	■	■	322
KB	MDS	p.Y220C	■	■	■	■	■	■	■	■	1230
43	AML	p.R156H	■	■	■	■	■	■	■	■	45
43	AML	p.C277Y	■	■	■	■	■	■	■	■	54
47	AML	p.V143M	■	■	■	■	■	■	■	■	115
47	AML	p.V216M	■	■	■	■	■	■	■	■	273
CMK	AML	p.D49H	■	■	■	■	■	■	■	■	20
CMK	AML	p.M133K	■	■	■	■	■	■	■	■	44
MOLM-16	AML	p.V173M	■	■	■	■	■	■	■	■	241
MOLM-16	AML	p.C238S	■	■	■	■	■	■	■	■	41
P31-FUJ	AML	p.R196*	■	■	■	■	■	■	■	■	837
P31-FUJ	AML	p.Y236C	■	■	■	■	■	■	■	■	256
RK4	AML	p.N239D	■	■	■	■	■	■	■	■	132
RK4	AML	p.S261T	■	■	■	■	■	■	■	■	8
RK8	AML	p.C135S	■	■	■	■	■	■	■	■	13
RK8	AML	p.M246K	■	■	■	■	■	■	■	■	26
34	MDS	p.R273C	■	■	■	■	■	■	■	■	2168
34	MDS	p.R273H	■	■	■	■	■	■	■	■	2300
49	MDS	p.R158H	■	■	■	■	■	■	■	■	310
49	MDS	p.R273H	■	■	■	■	■	■	■	■	2300
394	AML	p.C135S	■	■	■	■	■	■	■	■	24
394	AML	p.M246K	■	■	■	■	■	■	■	■	26
403	AML	p.N239D	■	■	■	■	■	■	■	■	132
403	AML	p.S261T	■	■	■	■	■	■	■	■	8
AML047T	AML	p.V143M	■	■	■	■	■	■	■	■	115
AML047T	AML	p.V216M	■	■	■	■	■	■	■	■	273
AML096T	AML	p.Q52*	■	■	■	■	■	■	■	■	40
AML096T	AML	p.W91*	■	■	■	■	■	■	■	■	56
MOLM-16	AML	p.V173M	■	■	■	■	■	■	■	■	241
MOLM-16	AML	p.C238S	■	■	■	■	■	■	■	■	41

Supplemental Figure S1a (part 1)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph	
P31-FUJ	AML	p.R196*	■	■	■	■	■	■	■	■		837
P31-FUJ	AML	p.Y236C	■	■	■	■	■	■	■	■		256
768	AML	p.G245D	■	■	■	■	■	■	■	■		470
768	AML	p.R248Q	■	■	■	■	■	■	■	■		2500
867	AML	p.G105D	■	■	■	■	■	■	■	■		10
867	AML	p.V157F	■	■	■	■	■	■	■	■		577
886	AML	p.H179R	■	■	■	■	■	■	■	■		565
886	AML	p.R248Q	■	■	■	■	■	■	■	■		2500
888	AML	p.V173G	■	■	■	■	■	■	■	■		42
888	AML	p.I195T	■	■	■	■	■	■	■	■		359
892	AML	p.Y205N	■	■	■	■	■	■	■	■		18
892	AML	p.D281Y	■	■	■	■	■	■	■	■		49
894	AML	p.Y163C	■	■	■	■	■	■	■	■		499
894	AML	p.M237I	■	■	■	■	■	■	■	■		362
898	AML	p.Y220C	■	■	■	■	■	■	■	■		1230
898	AML	p.Y327*	■	■	■	■	■	■	■	■		13
918	AML	p.S121P	■	■	■	■	■	■	■	■		2
918	AML	p.R306*	■	■	■	■	■	■	■	■		612
930	AML	p.R273H	■	■	■	■	■	■	■	■		2300
930	AML	p.R306*	■	■	■	■	■	■	■	■		612
947	AML	p.F134L	■	■	■	■	■	■	■	■		61
947	AML	p.I195T	■	■	■	■	■	■	■	■		359
964	AML	p.V173M	■	■	■	■	■	■	■	■		241
964	AML	p.Y220C	■	■	■	■	■	■	■	■		1230
GSM630979	AML	p.R156H	■	■	■	■	■	■	■	■		45
GSM630979	AML	p.C277Y	■	■	■	■	■	■	■	■		54
GSM630985	AML	p.V143M	■	■	■	■	■	■	■	■		115
GSM630985	AML	p.V216M	■	■	■	■	■	■	■	■		273
GSM631041	AML	p.Q52*	■	■	■	■	■	■	■	■		40
GSM631041	AML	p.W91*	■	■	■	■	■	■	■	■		56
MDS38	MDS	p.C238Y	■	■	■	■	■	■	■	■		289
MDS38	MDS	p.R248L	■	■	■	■	■	■	■	■		297
MDS39	MDS	p.Q192*	■	■	■	■	■	■	■	■		322
MDS39	MDS	p.Y220C	■	■	■	■	■	■	■	■		1230
UPN-1	MDS	p.R158H	■	■	■	■	■	■	■	■		310
UPN-1	MDS	p.R273H	■	■	■	■	■	■	■	■		2300
UPN-10	MDS	p.Y163C	■	■	■	■	■	■	■	■		499
UPN-10	MDS	p.C275Y	■	■	■	■	■	■	■	■		243
UPN-16	MDS	p.R248W	■	■	■	■	■	■	■	■		1949
UPN-17	MDS	p.H178P	■	■	■	■	■	■	■	■		30
UPN-24	MDS	p.Y163C	■	■	■	■	■	■	■	■		499
UPN-24	MDS	p.Y234C	■	■	■	■	■	■	■	■		396
UPN-25	MDS	p.Y220C	■	■	■	■	■	■	■	■		1230
UPN-25	MDS	p.Q331H	■	■	■	■	■	■	■	■		19
UPN-26	MDS	p.K132Q	■	■	■	■	■	■	■	■		49
UPN-26	MDS	p.G262V	■	■	■	■	■	■	■	■		80
UPN-27	MDS	p.V173M	■	■	■	■	■	■	■	■		241
UPN-27	MDS	p.H214R	■	■	■	■	■	■	■	■		219

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)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
7	MDS	p.C176R									54
7	MDS	p.G245S									1156
8	MDS	p.P152L									258
8	MDS	p.V216M									273
11	AML	p.E171*									85
11	AML	p.R248Q									2500
17	AML	p.E171G									17
17	AML	p.P177L									78
1	AML	p.L194R									224
1	AML	p.Y205S									40
14	AML	p.Y236C									256
14	AML	p.R249G									112
21	AML	p.R175H									3319
21	AML	p.H193Y									148
23	AML	p.H179R									565
23	AML	p.R280*									27
13	MDS	p.H179R									565
13	MDS	p.A189V									28
MOLM-16	AML	p.V173M									241
MOLM-16	AML	p.C238S									41
013-18	AML	p.K132E									81
013-18	AML	p.R273H									2300
037-02	AML	p.N131Y									23
037-02	AML	p.V173M									241
391-02	AML	p.Y236C									256
391-02	AML	p.C238Y									289
449-03	AML	p.R110L									150
449-03	AML	p.C275R									60
560-03	AML	p.C238S									41
560-03	AML	p.G266R									192
664-03	AML	p.V173L									197
664-03	AML	p.R248Q									2500
AML27	AML	p.C176Y									279
AML27	AML	p.C238Y									289
AML45	AML	p.R175H									3319
AML45	AML	p.R282W									1568
AML80	AML	p.P152L									258
AML80	AML	p.R248W									1949
MI-AML-047	AML	p.V143M									115
MI-AML-047	AML	p.V216M									273
MI-AML-096	AML	p.Q52*									40
MI-AML-096	AML	p.W91*									87
MI-AML-250	AML	p.R196*									837
MI-AML-250	AML	p.C275W									30
377512	t-MDS / t-AML	p.L130V									60
377512	t-MDS / t-AML	p.R273C									2168
530447	t-MDS / t-AML	p.K139N									26
530447	t-MDS / t-AML	p.R248Q									2500

Supplemental Figure S1a (part 4)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
1_39	AML	p.C242Y	■	■	■	■	■	■	■	■	162
1_39	AML	p.R110P	■	■	■	■	■	■	■	■	61
1_73	AML	p.R248W	■	■	■	■	■	■	■	■	1949
1_73	AML	p.A138D	■	■	■	■	■	■	■	■	9
1_74	AML	p.R273C	■	■	■	■	■	■	■	■	2168
1_74	AML	p.R175G	■	■	■	■	■	■	■	■	91
229	AML	p.V216M	■	■	■	■	■	■	■	■	273
229	AML	p.F134L	■	■	■	■	■	■	■	■	20
330	AML	p.F270C	■	■	■	■	■	■	■	■	53
330	AML	p.M246L	■	■	■	■	■	■	■	■	7
BM12	MDS	p.R175H	■	■	■	■	■	■	■	■	3319
BM12	MDS	p.W146*	■	■	■	■	■	■	■	■	131
PD11154a	AML	p.Y220C	■	■	■	■	■	■	■	■	1230
PD11154a	AML	p.W146*	■	■	■	■	■	■	■	■	131
PD11168a	AML	p.M237I	■	■	■	■	■	■	■	■	362
PD11168a	AML	p.K164E	■	■	■	■	■	■	■	■	66
21	AML	p.R248W	■	■	■	■	■	■	■	■	1949
22	AML	p.R248G	■	■	■	■	■	■	■	■	83
DFCI-006382-358:	MDS	p.C176Y	■	■	■	■	■	■	■	■	279
DFCI-006382-358:	MDS	p.Y220C	■	■	■	■	■	■	■	■	1230
5273	AML	p.P152L	■	■	■	■	■	■	■	■	258
5273	AML	p.M237I	■	■	■	■	■	■	■	■	362
7071	AML	p.K132*	■	■	■	■	■	■	■	■	13
7071	AML	p.R196P	■	■	■	■	■	■	■	■	62
3022518	MDS	p.C176Y	■	■	■	■	■	■	■	■	279
3022518	MDS	p.R175G	■	■	■	■	■	■	■	■	91
3054448	MDS	p.H179D	■	■	■	■	■	■	■	■	71
3054448	MDS	p.K132R	■	■	■	■	■	■	■	■	202
3116963	MDS	p.R248Q	■	■	■	■	■	■	■	■	2500
3116963	MDS	p.Y220C	■	■	■	■	■	■	■	■	1230
3139981	MDS	p.Y220C	■	■	■	■	■	■	■	■	1230
3139981	MDS	p.L194R	■	■	■	■	■	■	■	■	224
3195456	MDS	p.R267W	■	■	■	■	■	■	■	■	119
3195456	MDS	p.A161T	■	■	■	■	■	■	■	■	185
3250152	MDS	p.R248W	■	■	■	■	■	■	■	■	1949
3250152	MDS	p.R280I	■	■	■	■	■	■	■	■	68
3280694	MDS	p.I162N	■	■	■	■	■	■	■	■	30
3280694	MDS	p.R273H	■	■	■	■	■	■	■	■	2300
3310889	MDS	p.Y220H	■	■	■	■	■	■	■	■	65
3310889	MDS	p.R282W	■	■	■	■	■	■	■	■	1568
3382656	MDS	p.R181H	■	■	■	■	■	■	■	■	75
3382656	MDS	p.H179R	■	■	■	■	■	■	■	■	565
3429739	MDS	p.P151T	■	■	■	■	■	■	■	■	56
3429739	MDS	p.S90P	■	■	■	■	■	■	■	■	7
3437575	MDS	p.Y220C	■	■	■	■	■	■	■	■	1230
3437575	MDS	p.D281N	■	■	■	■	■	■	■	■	113
3474059	MDS	p.R175H	■	■	■	■	■	■	■	■	3319
3474059	MDS	p.Y126C	■	■	■	■	■	■	■	■	95

Supplemental Figure S1a (part 5)

Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
3490378	MDS	p.I195T									 359
3490378	MDS	p.R273C									 2168
3550478	MDS	p.T253A									 11
3550478	MDS	p.G105V									 27
3594211	MDS	p.H168P									 31
3594211	MDS	p.A276P									 37
3642291	MDS	p.S241C									 129
3642291	MDS	p.V216M									 273
3658115	MDS	p.Y220C									 1230
3658115	MDS	p.R273H									 2300
3700016	MDS	p.R248Q									 2500
3700016	MDS	p.R273H									 2300
3876517	MDS	p.C238F									 147
3876517	MDS	p.V173M									 241
8010773	MDS	p.R175G									 91
8010773	MDS	p.Y163C									 499
8011237	MDS	p.G244S									 170
8011237	MDS	p.F113I									 10
8063897	MDS	p.H214D									 15
8063897	MDS	p.R175H									 3319
8088920	MDS	p.Y234H									 113
8088920	MDS	p.R273H									 2300
8137149	MDS	p.E258*									 70
8137149	MDS	p.V143M									 115
8142412	MDS	p.Y220N									 52
8142412	MDS	p.V173M									 241
8189833	MDS	p.Y220S									 70
8189833	MDS	p.P152L									 258
8248094	MDS	p.A159P									 136
8248094	MDS	p.P152L									 258
8251429	MDS	p.R273C									 2168
8251429	MDS	p.Q136E									 41
8290369	MDS	p.K164N									 15
8290369	MDS	p.G266R									 39
8293363	MDS	p.R248Q									 2500
8293363	MDS	p.C277Y									 54
8540417	MDS	p.M246V									 154
8540417	MDS	p.M237I									 362
8563904	MDS	p.P177H									 11
8563904	MDS	p.V143M									 115
8602801	MDS	p.H193R									 341
8602801	MDS	p.R273H									 2300
8693966	MDS	p.H179Q									 40
8693966	MDS	p.P151S									 236
8707610	MDS	p.R175H									 3319
8707610	MDS	p.V143M									 115
8817369	MDS	p.G199E									 35
8817369	MDS	p.H179N									 70

## Supplemental Figure S1a (part 6)

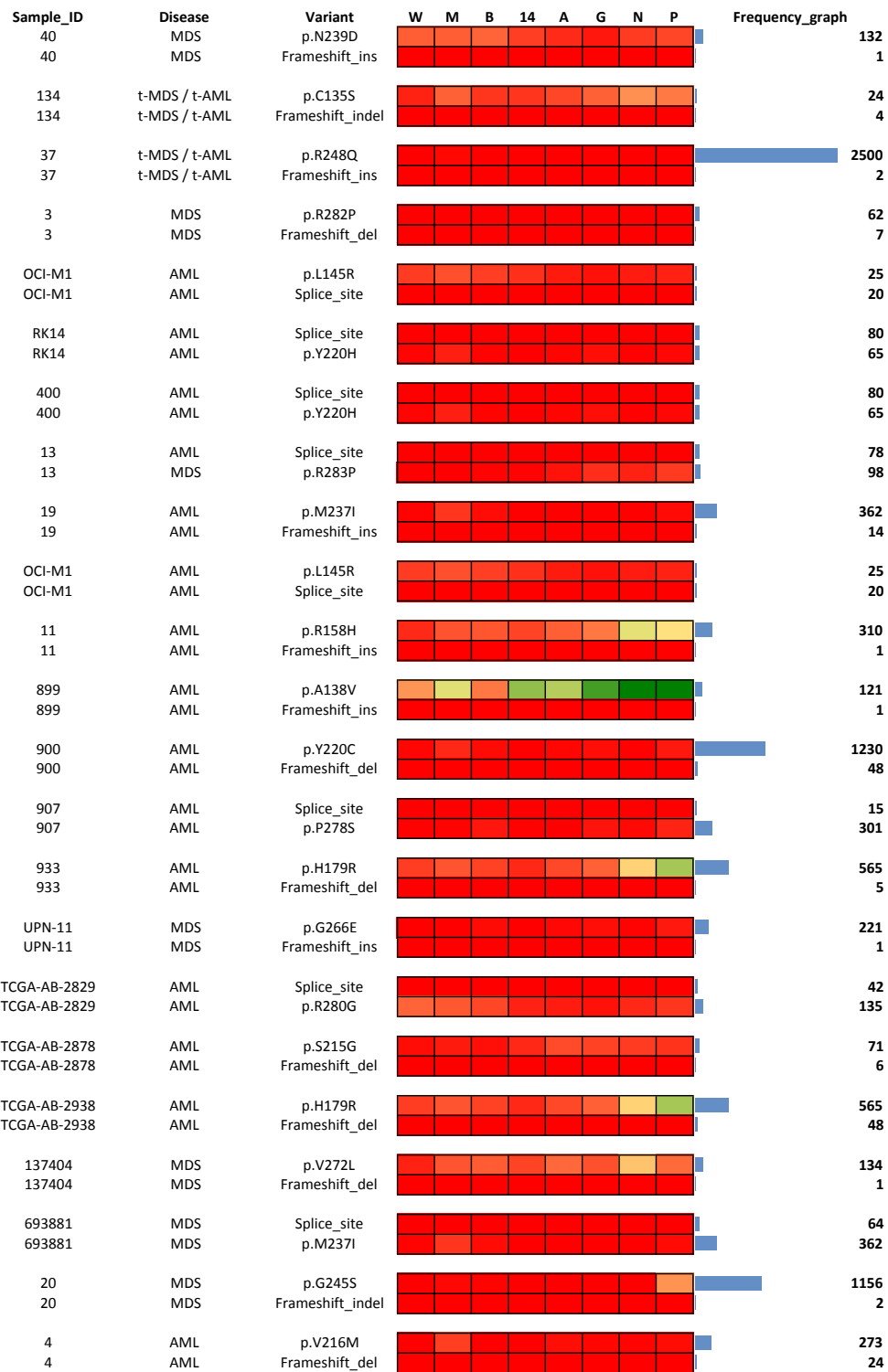
Sample_ID	Disease	Variant	W	M	B	14	A	G	N	P	Frequency_graph
8882668	MDS	p.R174W									38
8882668	MDS	p.R273H									2300
8898427	MDS	p.Y220C									1230
8898427	MDS	p.H179R									565
8970135	MDS	p.C242F									219
8970135	MDS	p.R175H									3319
8976124	MDS	p.Y234C									396
8976124	MDS	p.V216M									273
8989984	MDS	p.Q317*									157
8989984	MDS	p.H179Q									40
9889514	MDS	p.D281H									98
9889514	MDS	p.P152Q									16
9952661	MDS	p.I195T									359
9952661	MDS	p.R175H									3319
9973131	MDS	p.R196G									10
9973131	MDS	p.C141Y									221
9980522	MDS	p.R306*									612
9980522	MDS	p.R282W									1568
CLL090	MDS	p.E286K									303
CLL090	MDS	p.H179R									565
AML_113	AML	p.R196*									837
AML_113	AML	p.D281N									113
AML_135	AML	p.R248G									83
AML_135	AML	p.G245V									239
AML_145	AML	p.R273H									2300
AML_145	AML	p.Y205D									49
AML_547	AML	p.Y234C									396
AML_547	AML	p.Y234H									113
AML_70	AML	p.R273H									2300
AML_70	AML	p.M133T									24
1037	AML / MDS	p.H179P									20
1037	AML / MDS	p.R273H									2300
1062	AML / MDS	p.H193L									187
1062	AML / MDS	p.L257P									44
1068	AML / MDS	p.C141Y									221
1068	AML / MDS	p.R273L									440
MDS143	MDS	p.L265P									60
MDS143	MDS	p.R248Q									2500
MDS555	MDS	p.P278A									84
MDS555	MDS	p.V143A									44
LB 382	AML	p.R213Q									165
LB 382	AML	p.Y220C									1230
D	AML	p.C135S									24
D	AML	p.M246V									154

### 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)

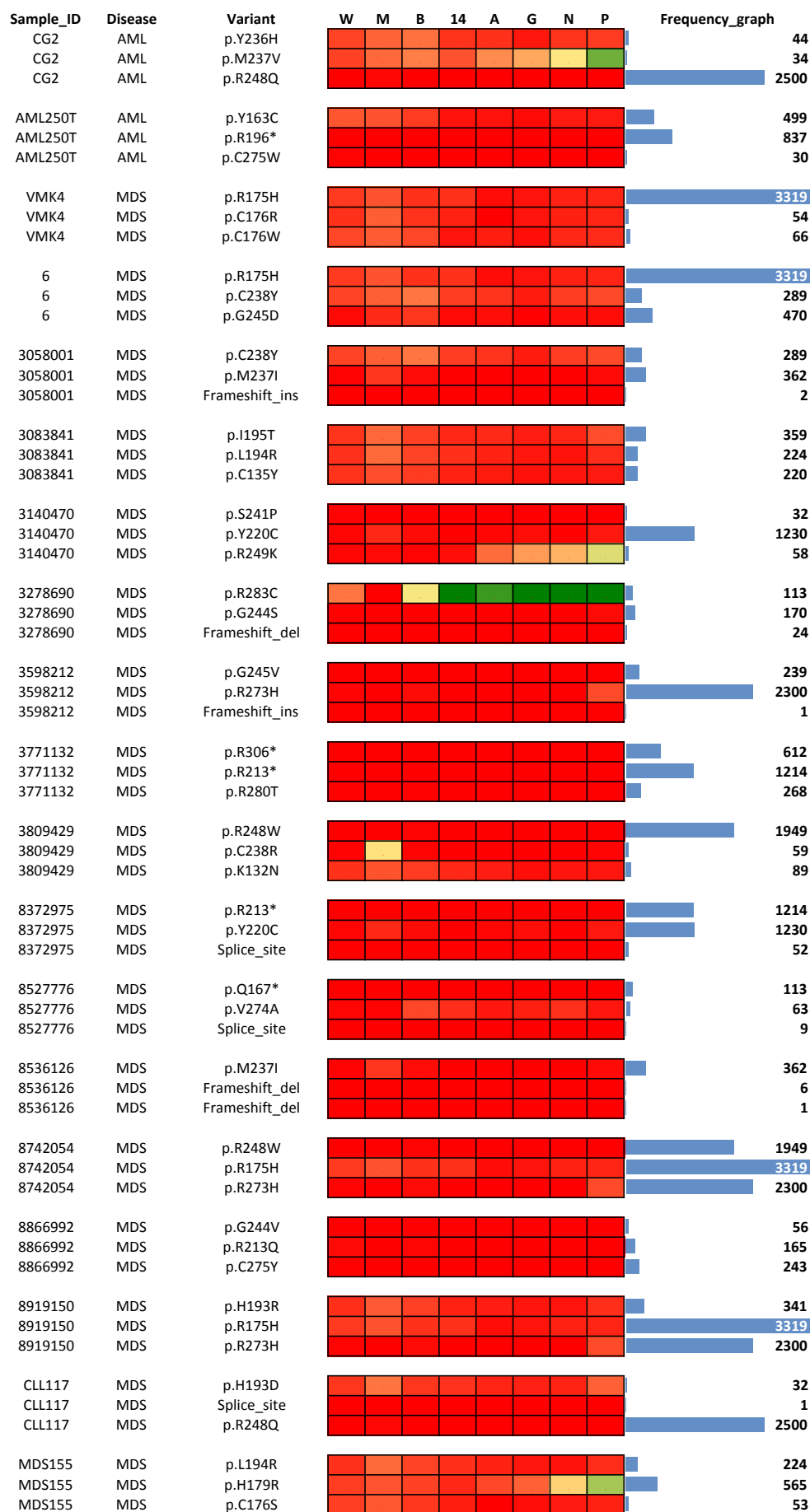




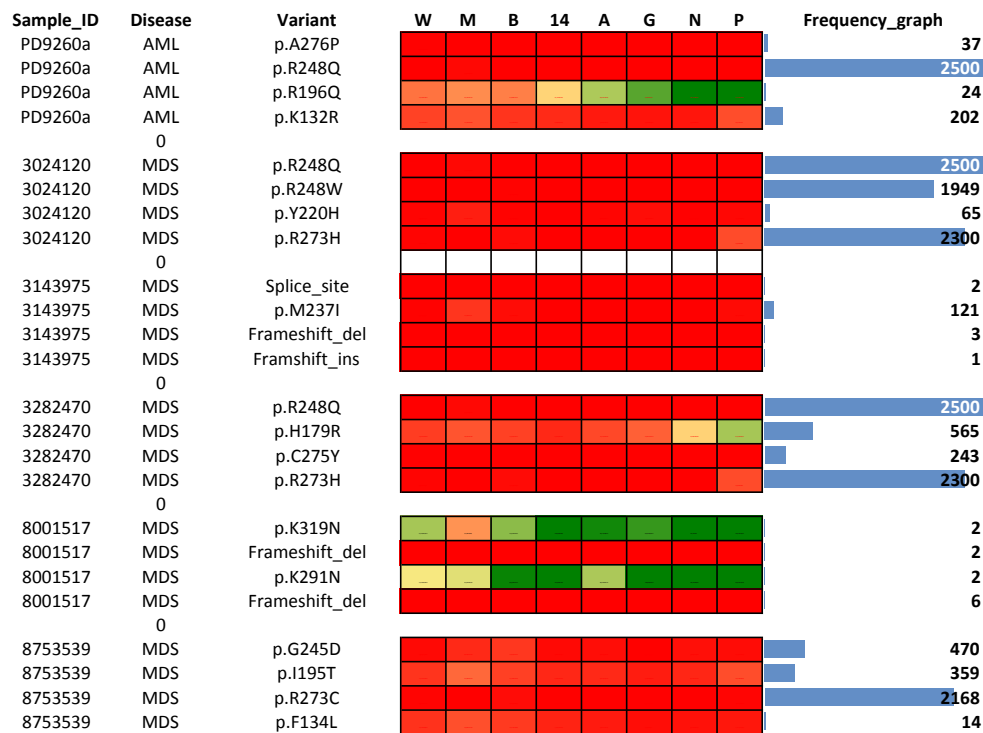
Supplemental Figure S1b (part 2)

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

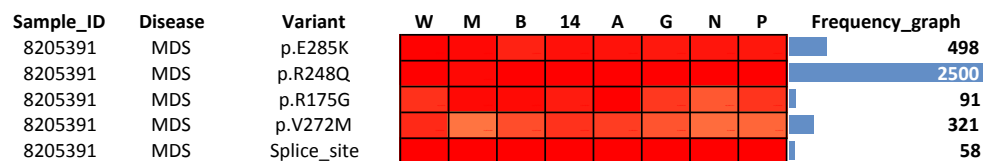
### Supplemental Figure S1b (part 3)



Supplemental Figure S1c



### Supplemental Figure S1d

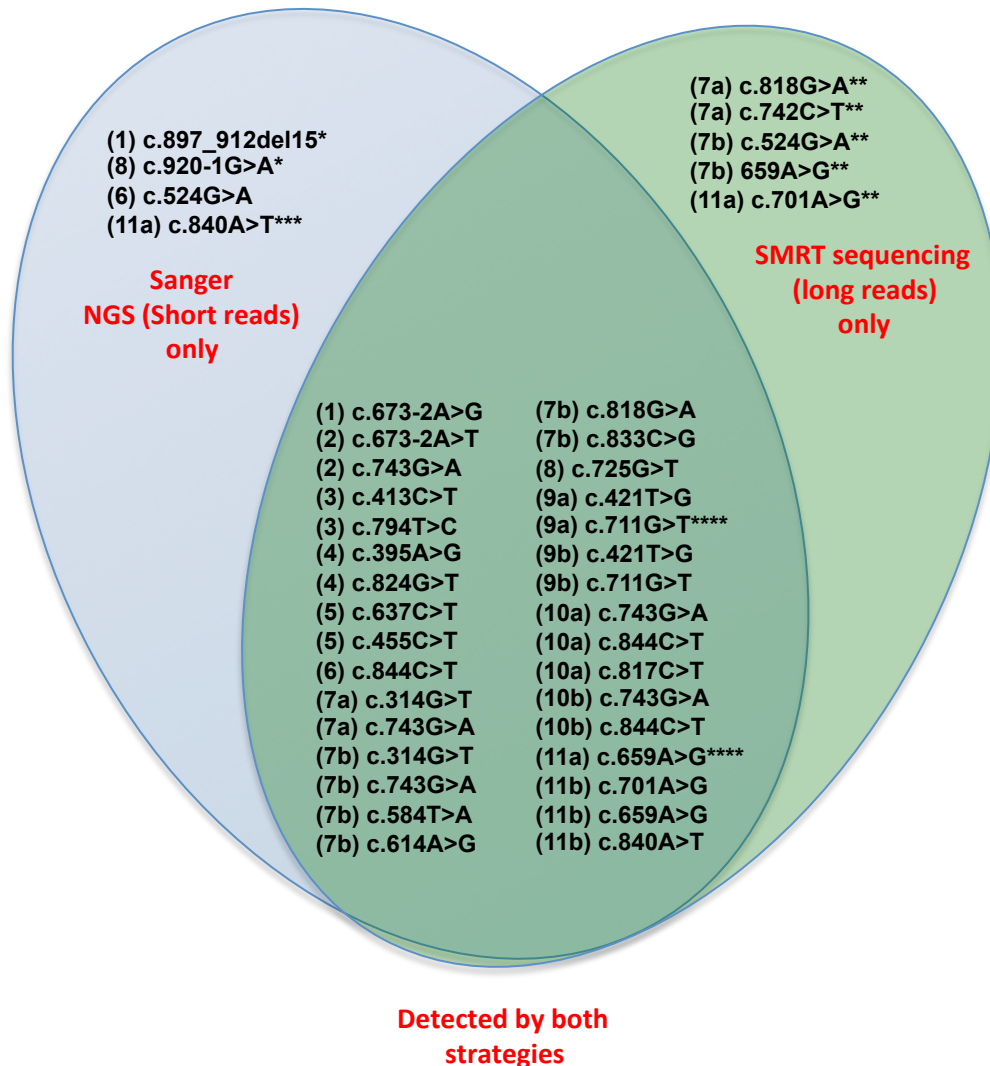


### Supplemental Figure S1e



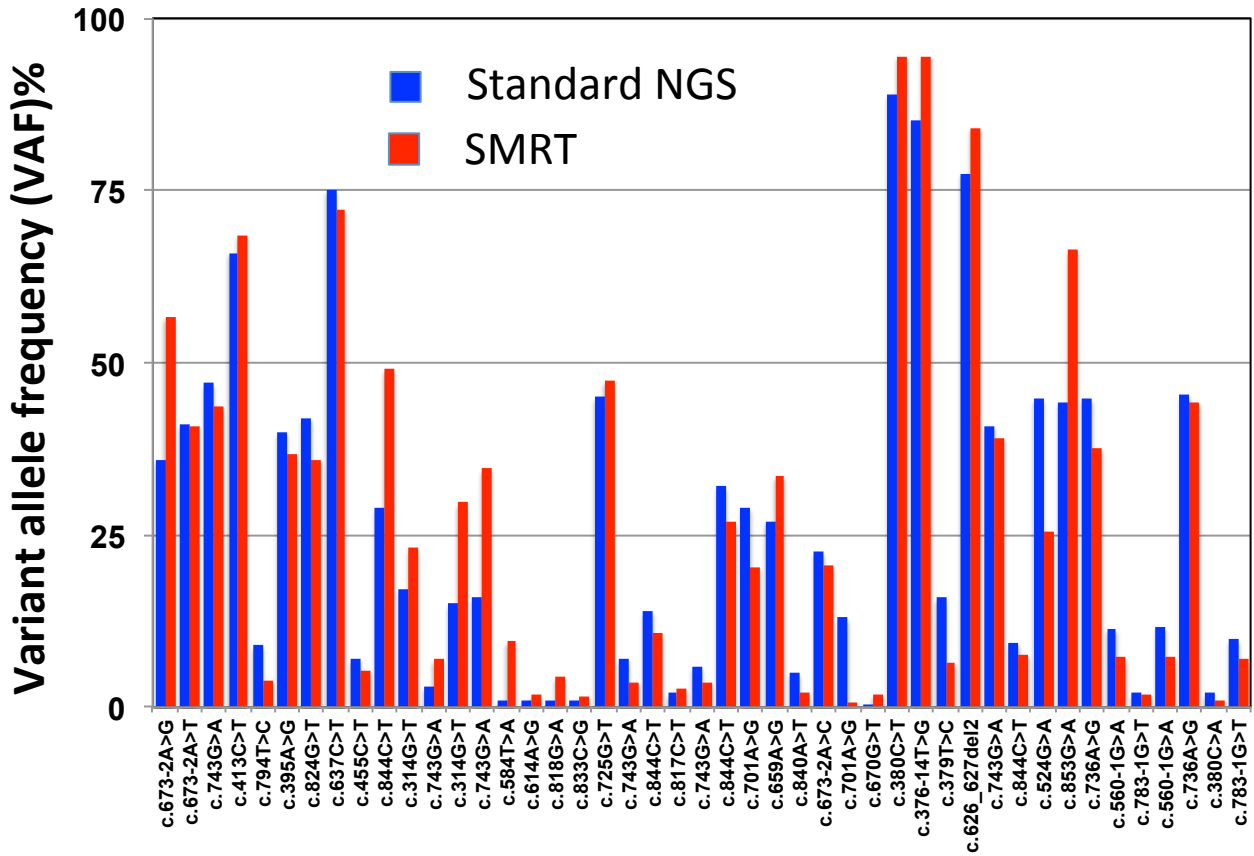
### Supplemental Figure S1f

**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 : *TP53* variants outside the amplicon used for the long range sequencing  
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

**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

rs1800370	rs1042522	rs1794287	rs2909430	rs1625895	c.673-2A>G	rs12947788	rs12951053	Frequency	Reads
■	■	■	■	■	■	■	■	56.6 %	5504
■	■	■	■	■		■	■	43.3 %	4206

■ Somatic variant

■ Germline SNP

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

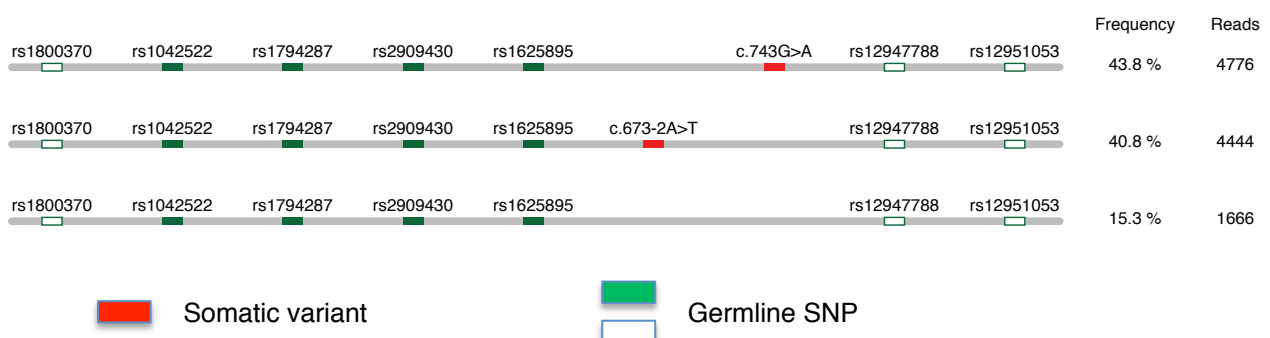


Figure S4b

## 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

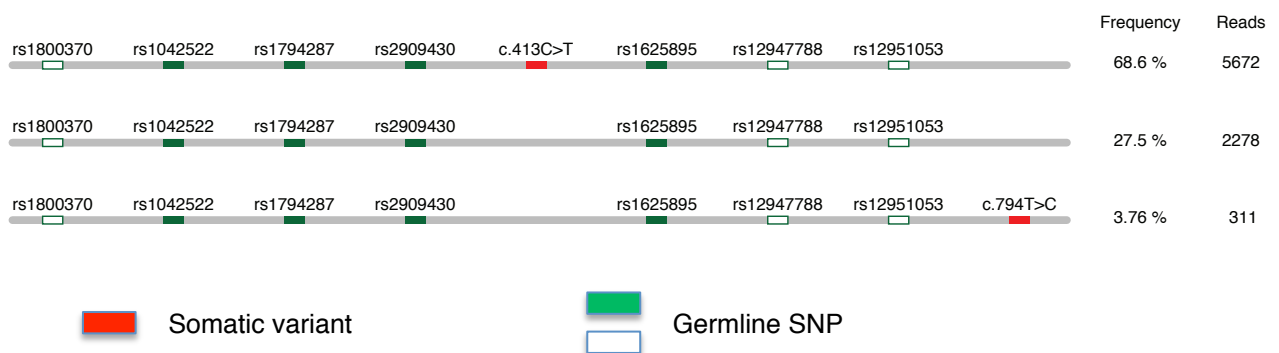


Figure S4c

## 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

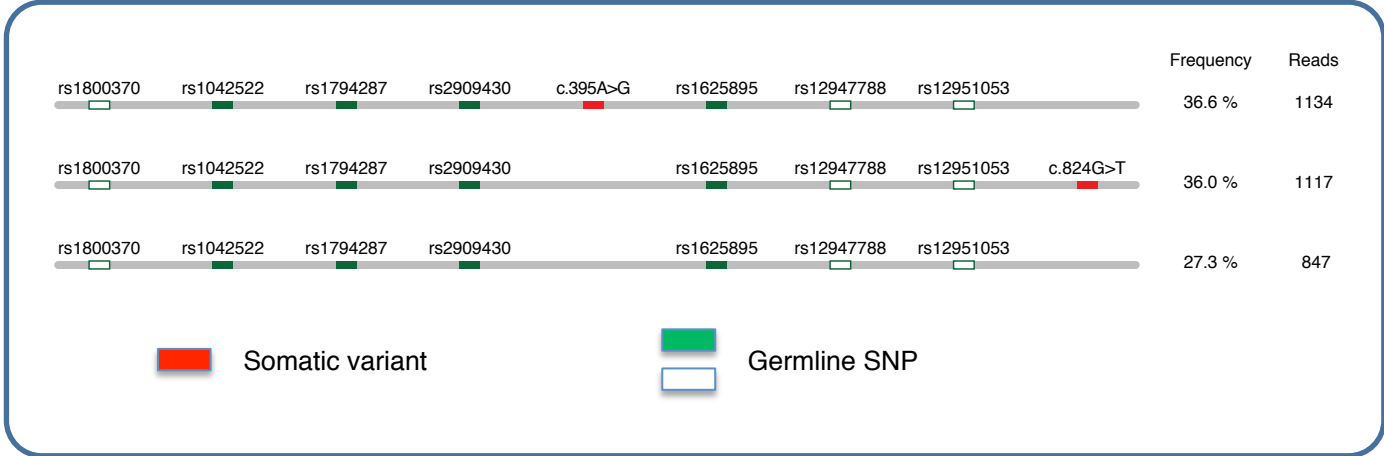


Figure S4d

## Clinical information

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

## Sequencing

### Sanger/NGS (short reads)

▼ c.637C>T (75 %)  
 ▼ c.455C>T (7 %)

### SMRT sequencing (long reads)

▼ c.637C>T (72.3%)  
 ▼ c.455C>T (5.42 %)  
 wt (22.1 %)

▼ Variant detected by both analyses

## Haplotype

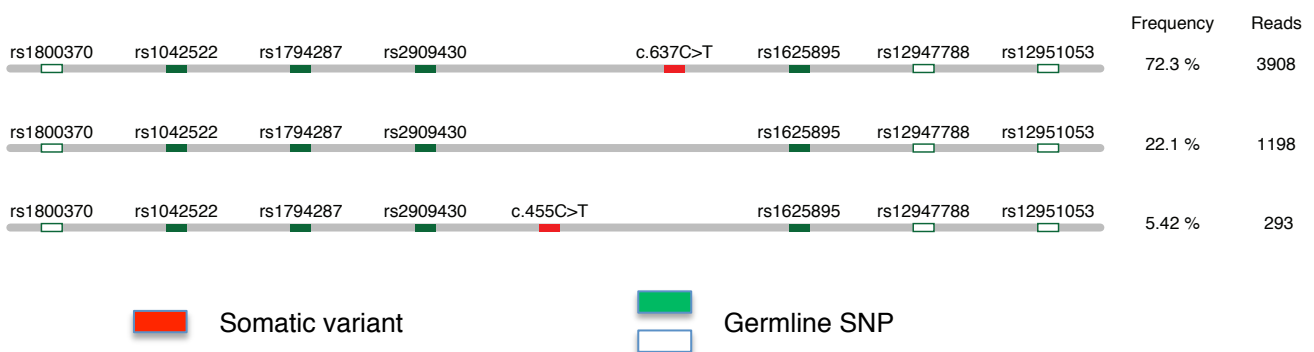


Figure S4e

## Clinical information

**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

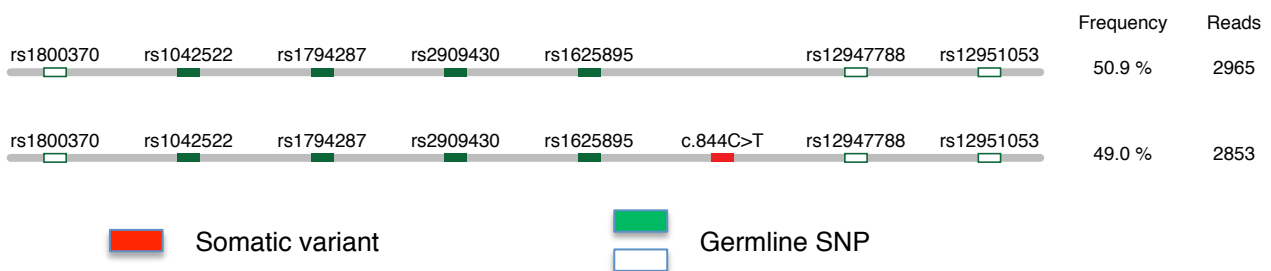


Figure S4f

## 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

### Sanger/NGS (Short reads)

### SMRT sequencing (long reads)

7a

▼ c.314G>T (17 %)  
 ▼ c.743G>A (3 %)

Aug 2008  
(7a)

▼ c.314G>T (23.1 %)  
 ▼ c.743G>A (7 %)  
 ▽ c.818G>A (0.7 %)  
 ▽ c.742C>T (0.7%)  
 wt (69.7 %)

7b

▼ 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  
(7b)

▼ 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 %)

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

Figure S4g

# Haplotype

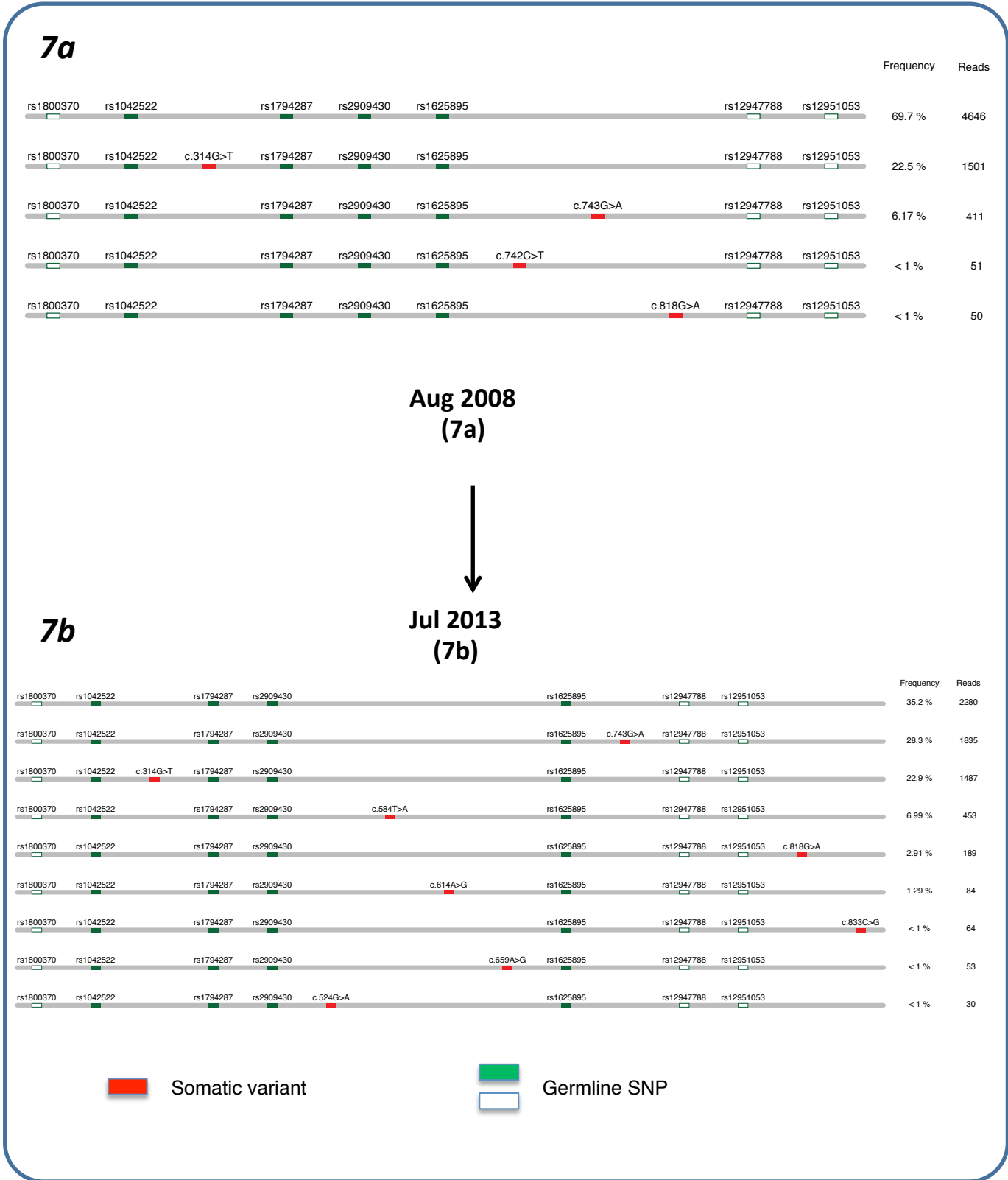


Figure S4g

## 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

rs1800370	rs1042522	rs1794287	rs2909430	rs1625895		rs12947788	rs12951053	Frequency	Reads
Green	Green	Green	Green	Green		Green	Green	50.6 %	3518
Green	Green	Green	Green	Green	c.725G>T	Green	Green	49.4 %	3439

Red box Somatic variant

Green box Germline SNP

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

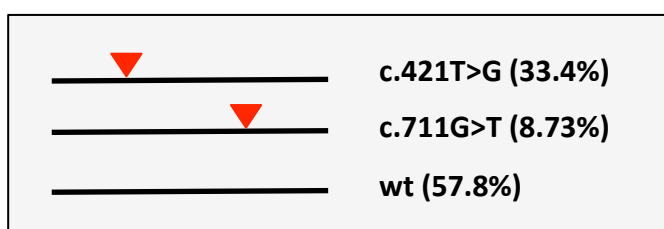
### Sanger/NGS (short reads)

### SMRT sequencing (long reads)

9a

▼ c.421T>G (~50%)  
▼ c.711G>T (?)

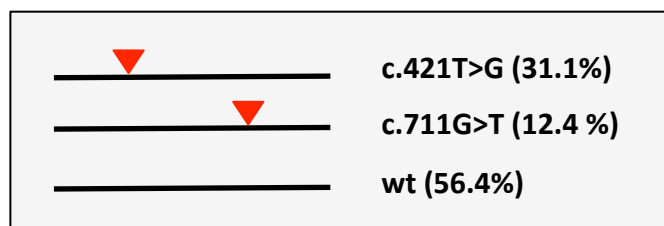
(Sanger only)



9b

▼ c.421T>G (~50%)  
▼ c.711G>T (~20%)

(Sanger only)

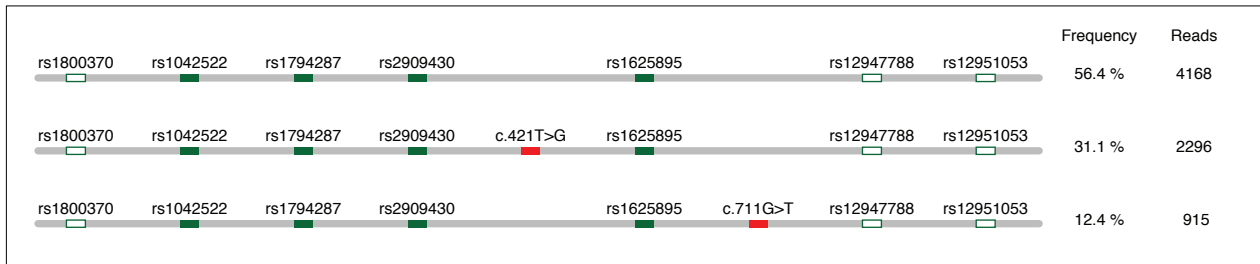
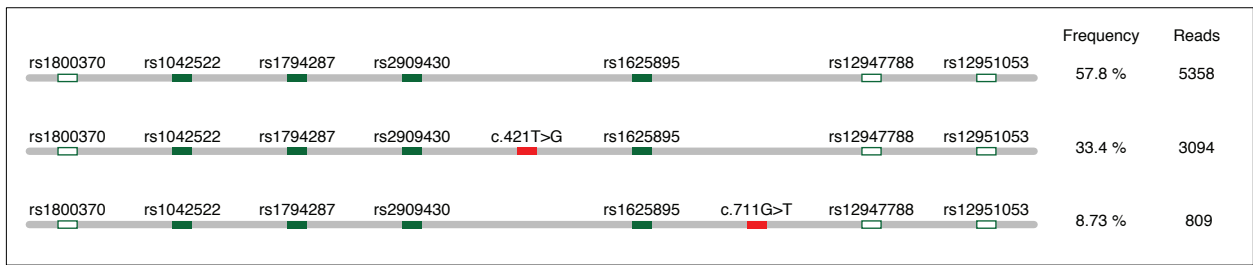


- ▼ Variant detected after manual examination but below the cut-off used for clinical validation
- ▼ Variant detected by both analyses

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

Figure S4i

# Haplotype



Somatic variant
  Germline SNP

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

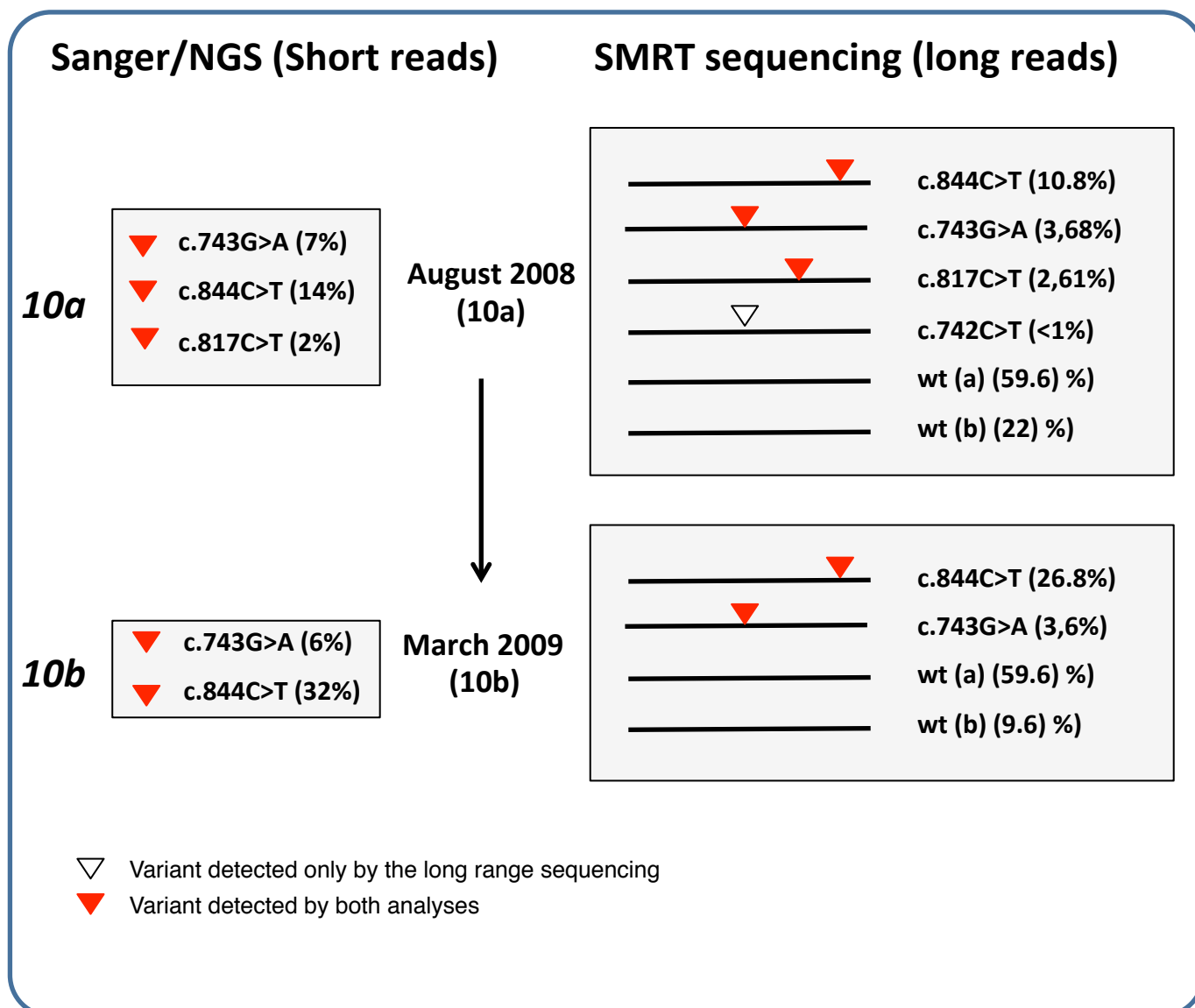


Figure S4j

## Haplotype




August 2008  
(10a)



March 2009  
(10b)



 Somatic variant

 Germline SNP  


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

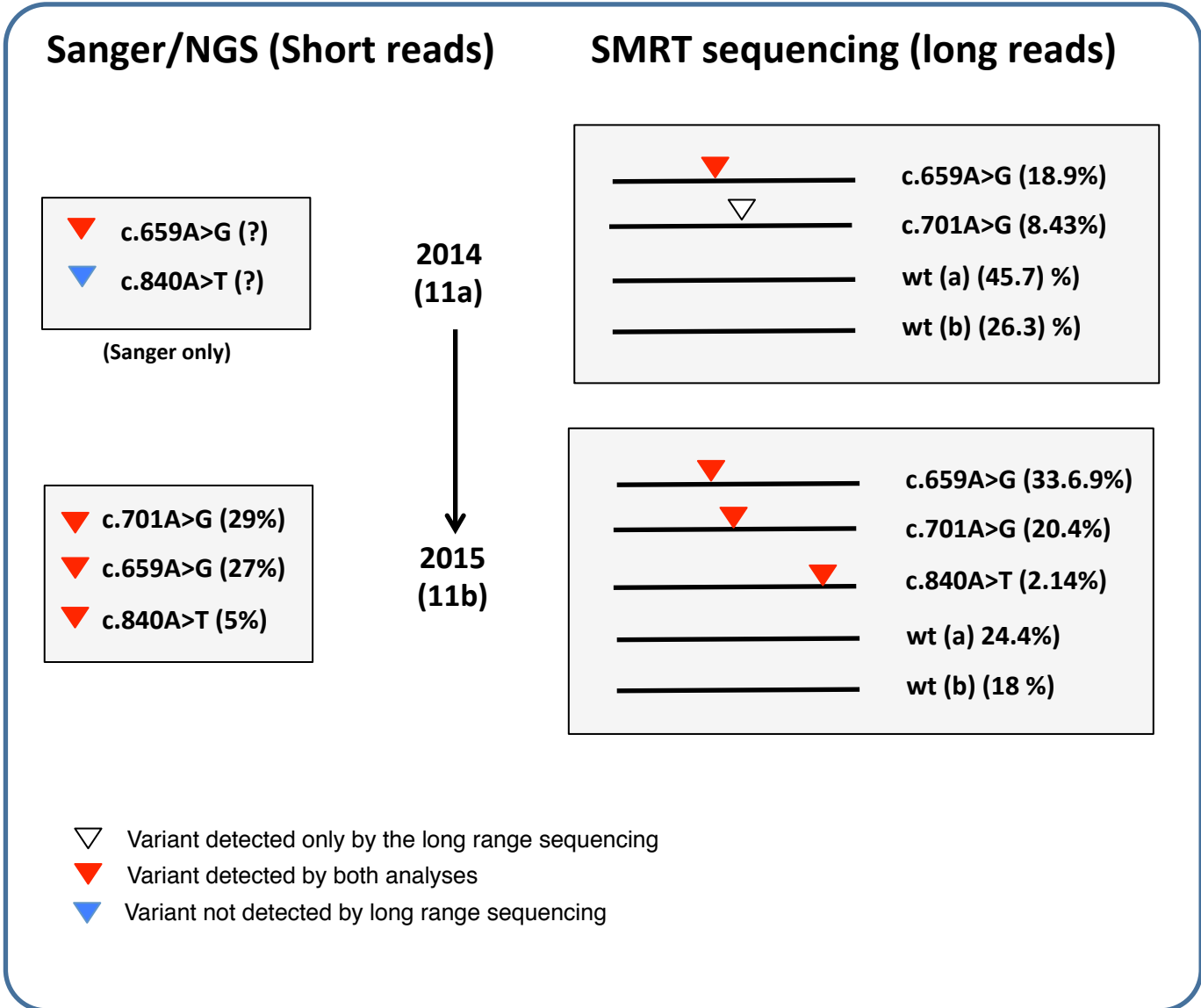
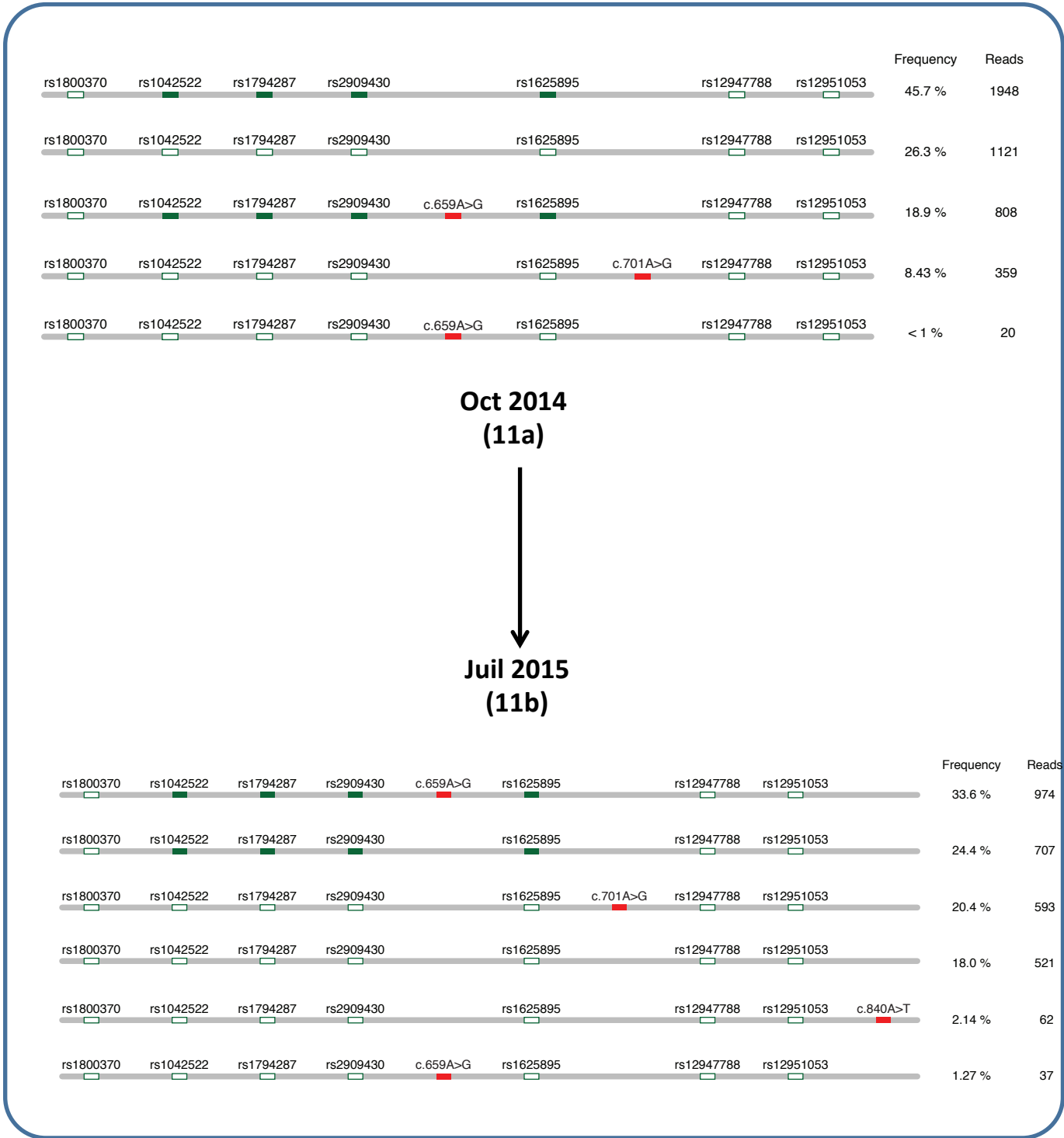


Figure S4k

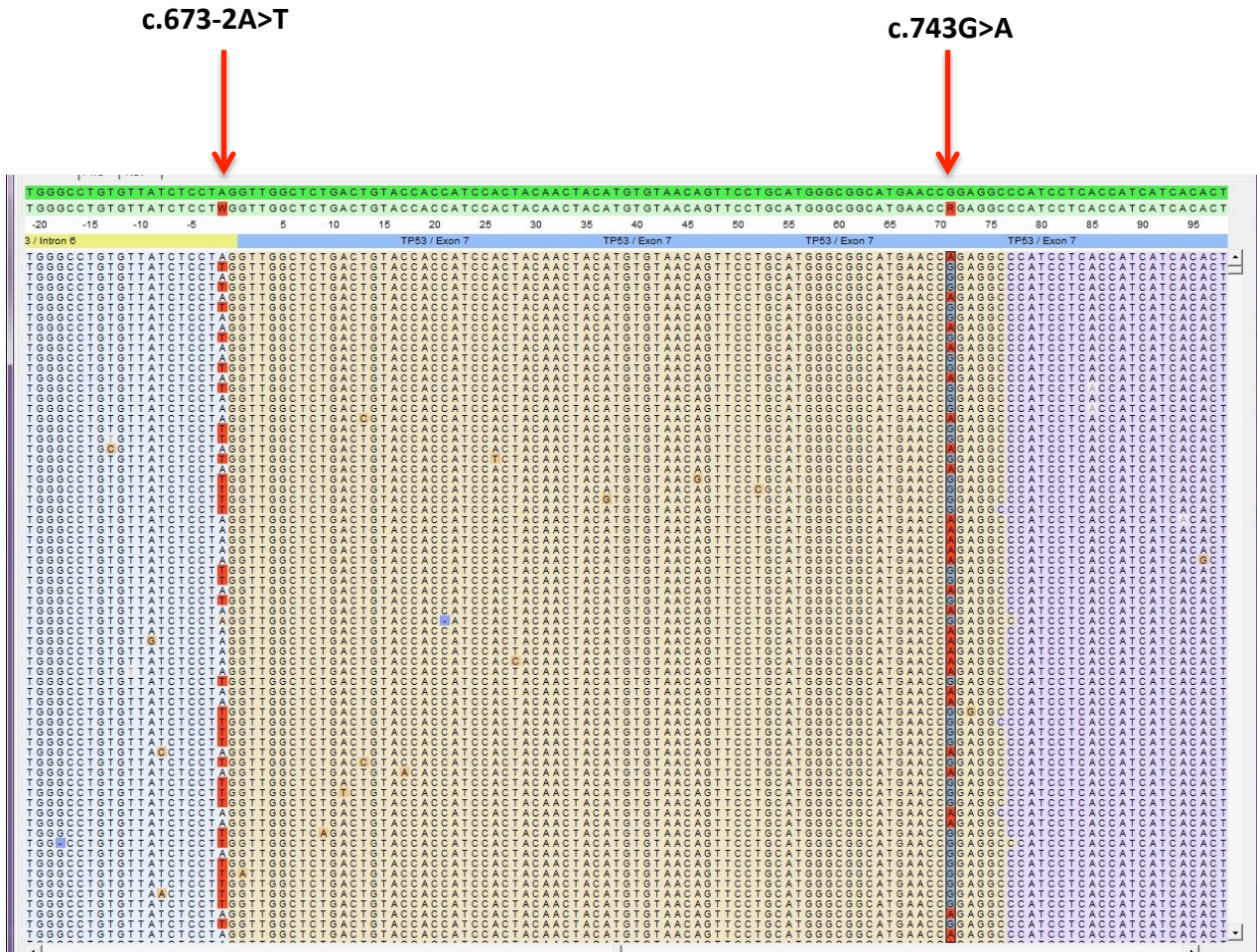
## Haplotype



Somatic variant      Germline SNP

Figure S4k

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



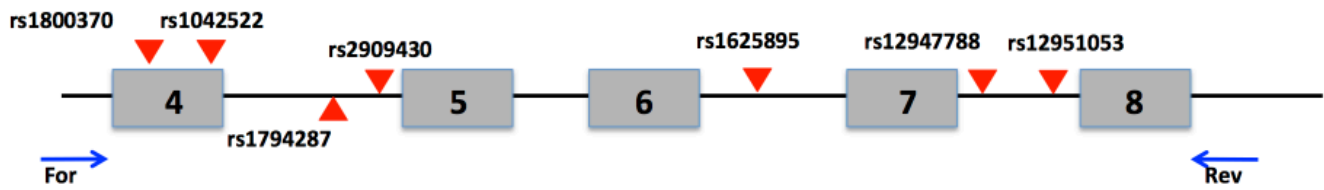
Patient Fr2





## Supplemental Figure 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' cctgctctctgactgctct 3'	7579626-7579607
Reverse primer	5' tacctcgcttagtgctcct 3	7577035-7577016