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[6-8]

[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, Sweden; [4]Cytogenetics, Hôpital du Saint Sacrement, CHU de Québec, Université Laval, Department of molecular medicine, Canada; [5]Department of Clinical Cytology and Cytogenetics, Nîmes University Hospital, France; [6]Sorbonne Université, UPMC Univ Paris 06, F- 75005 Paris, France; [7]INSERM, U1138, Centre de Recherche des Cordeliers, Paris, France and [8]Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden

Correspondence: 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

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

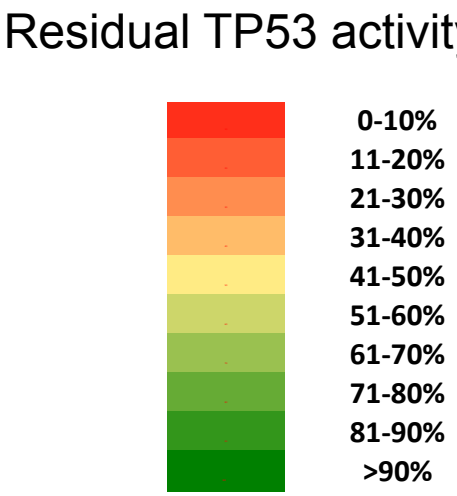
Figure S1c: analysis of AML and MDS tumors with 3 different *TP53* alterations

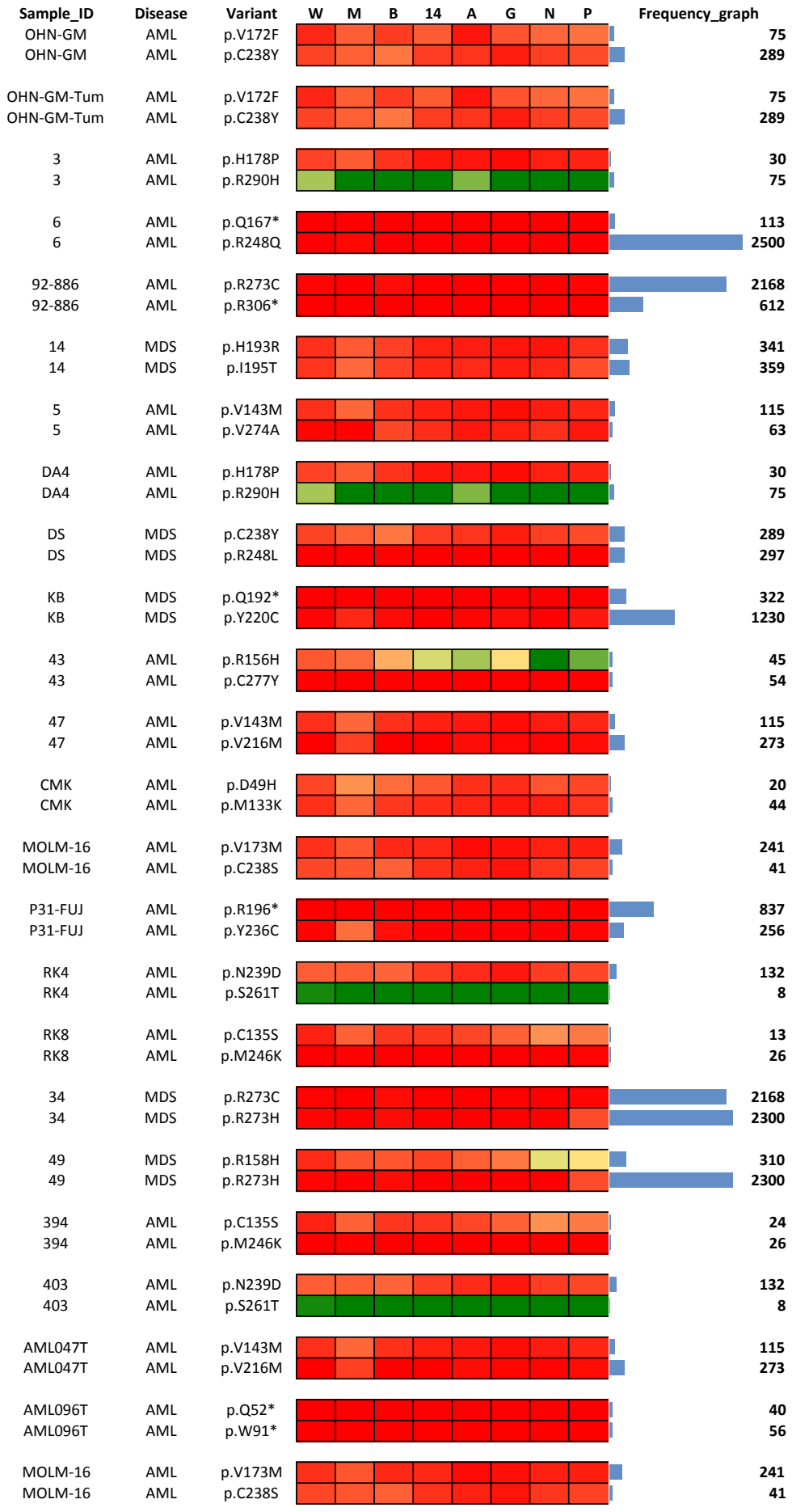
Figure S1d: analysis of AML and MDS tumors with 4 different *TP53* alterations

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

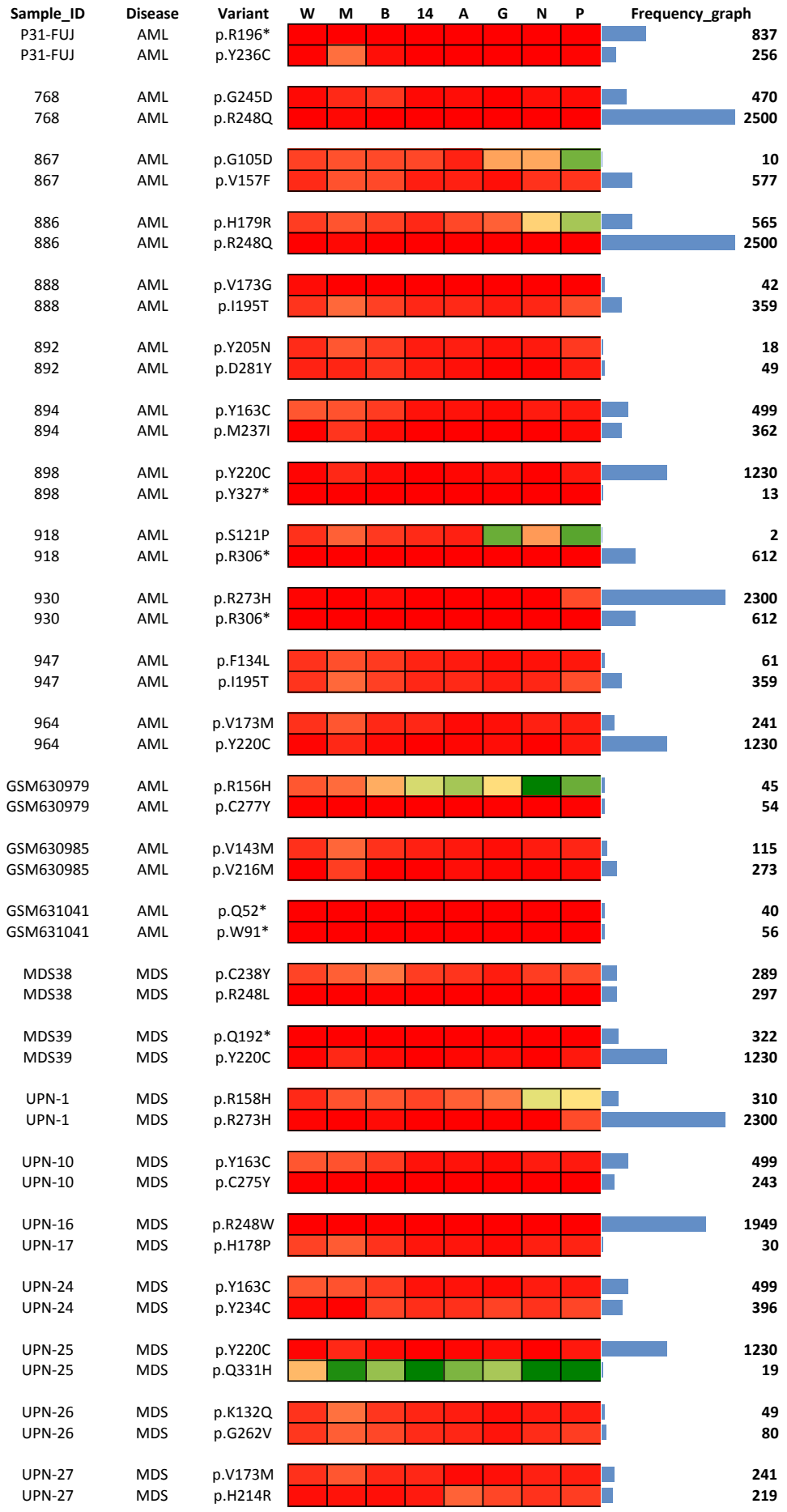
Figure S1f: analysis of AML and MDS tumors with 6 different *TP53* alterations

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.

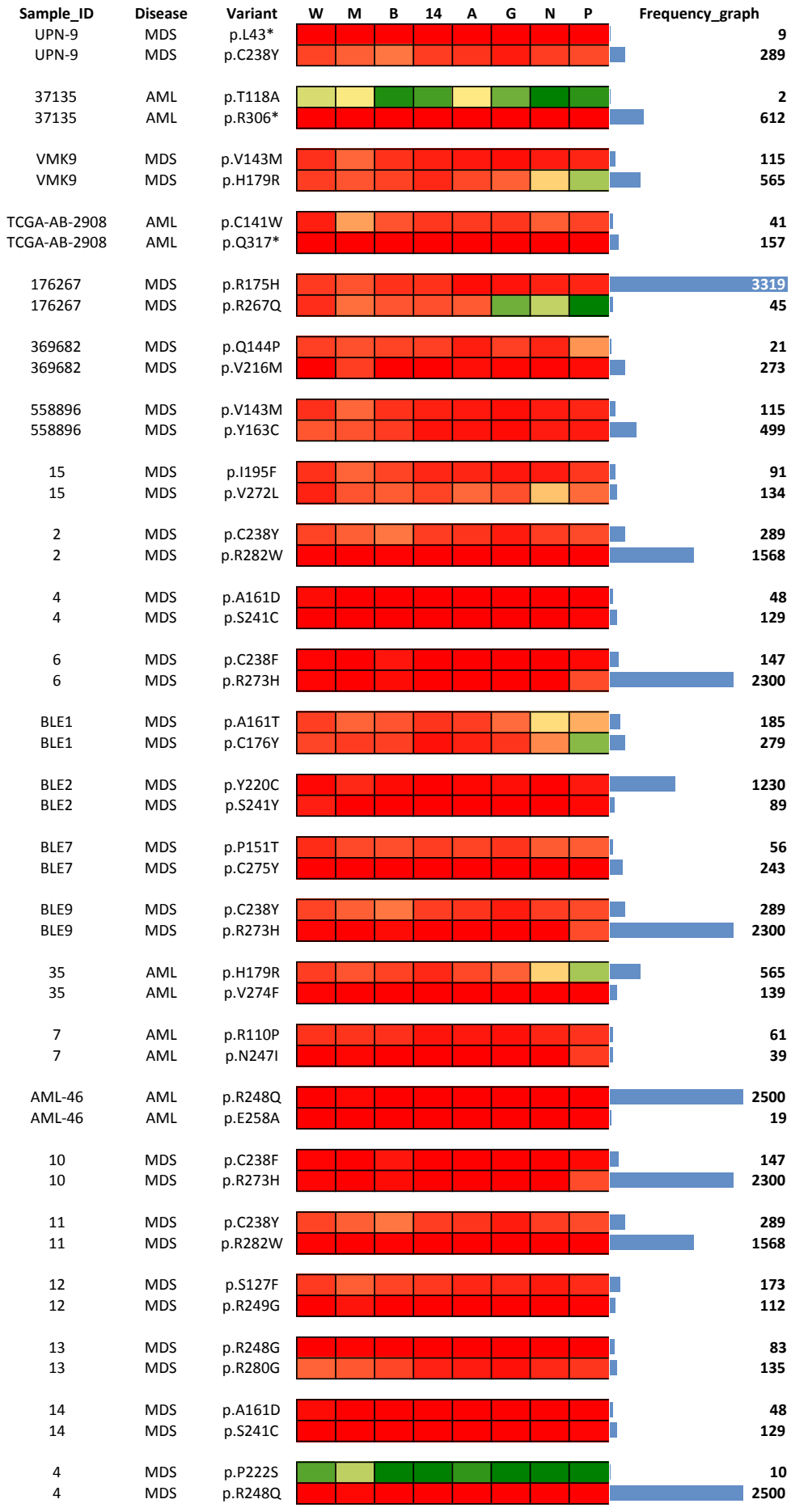




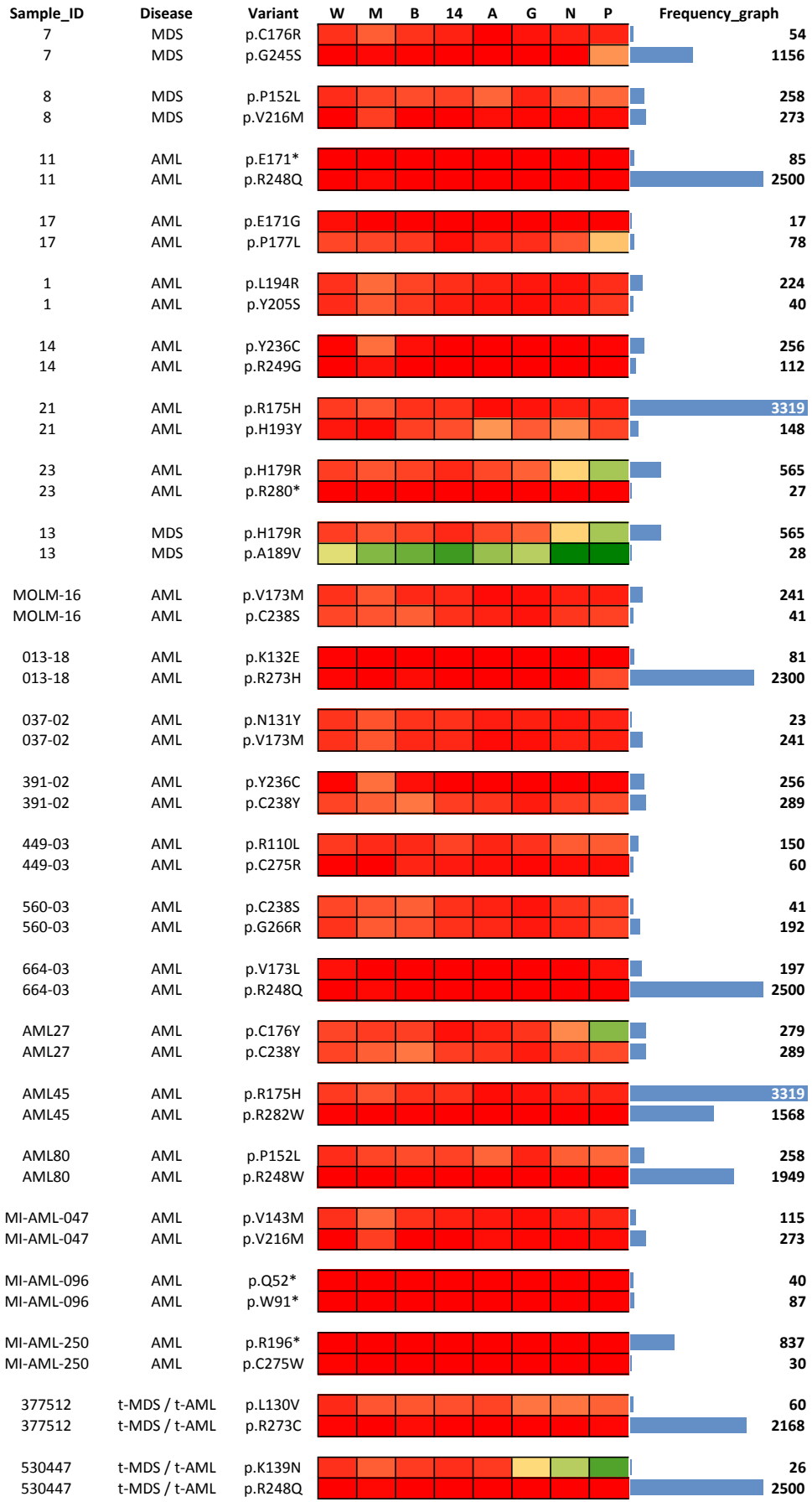
Supplemental Figure S1a (part 1)



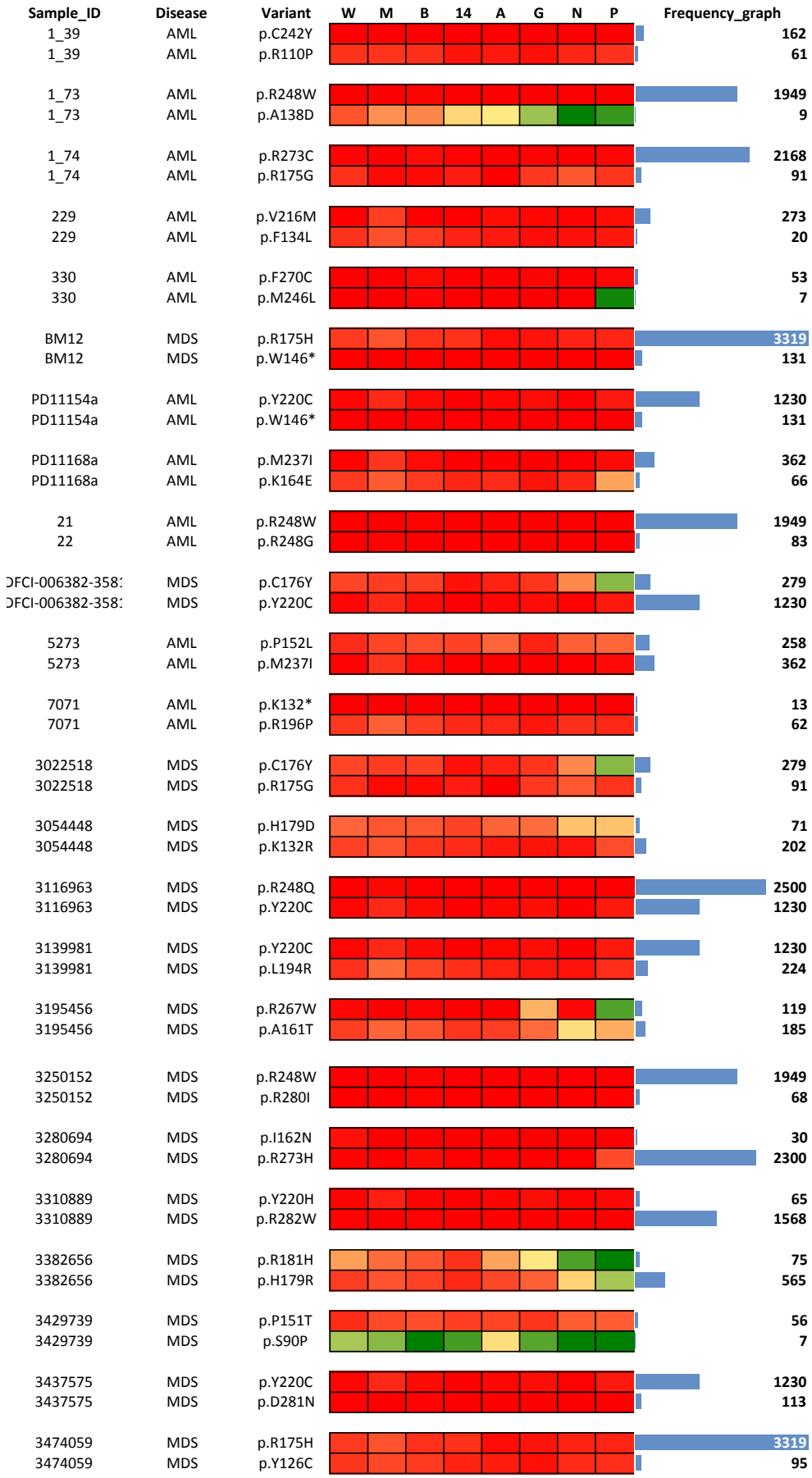
Supplemental Figure S1a (part 2)



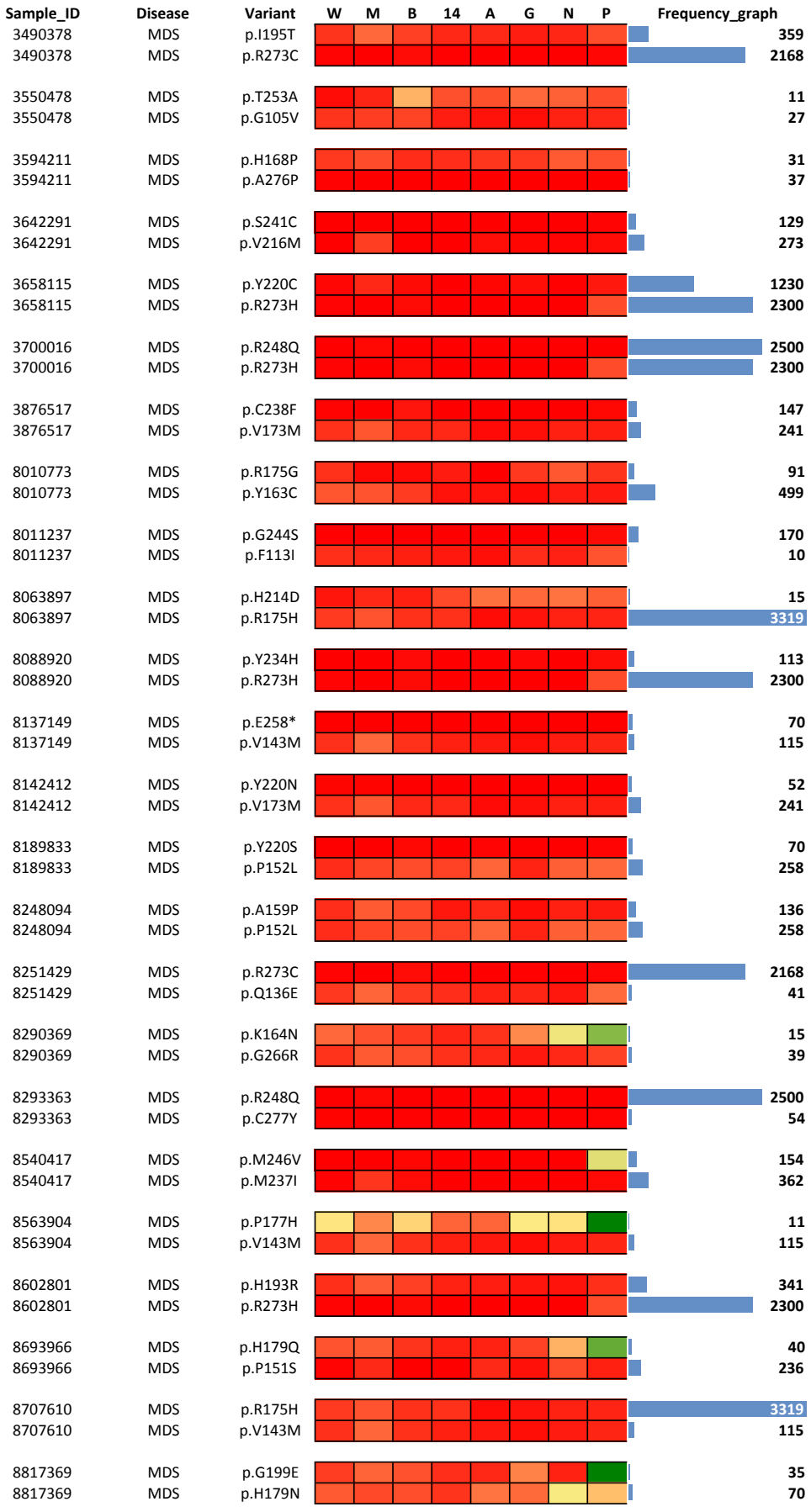
Supplemental Figure S1a (part 3)



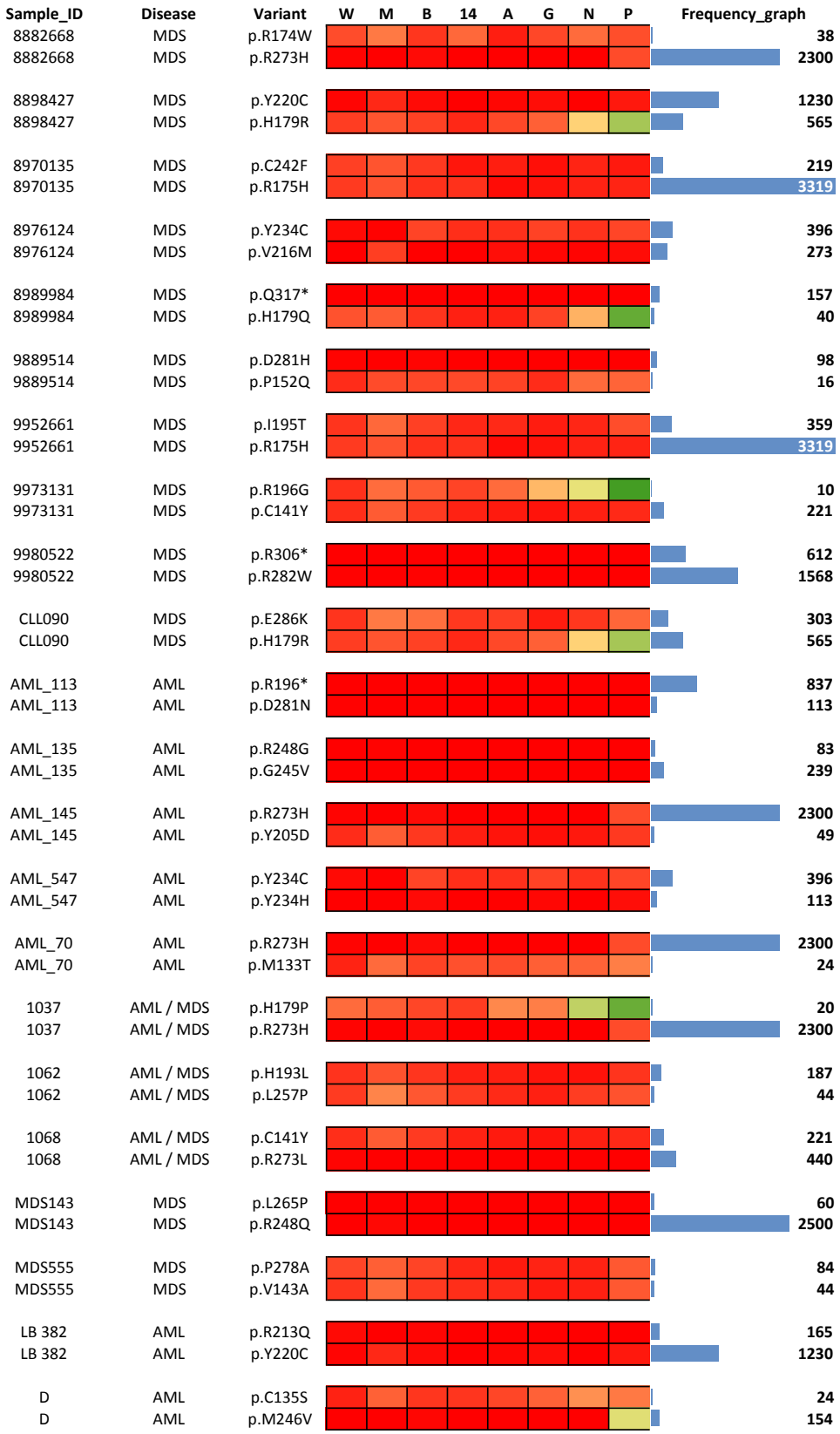
Supplemental Figure S1a (part 4)



Supplemental Figure S1a (part 5)



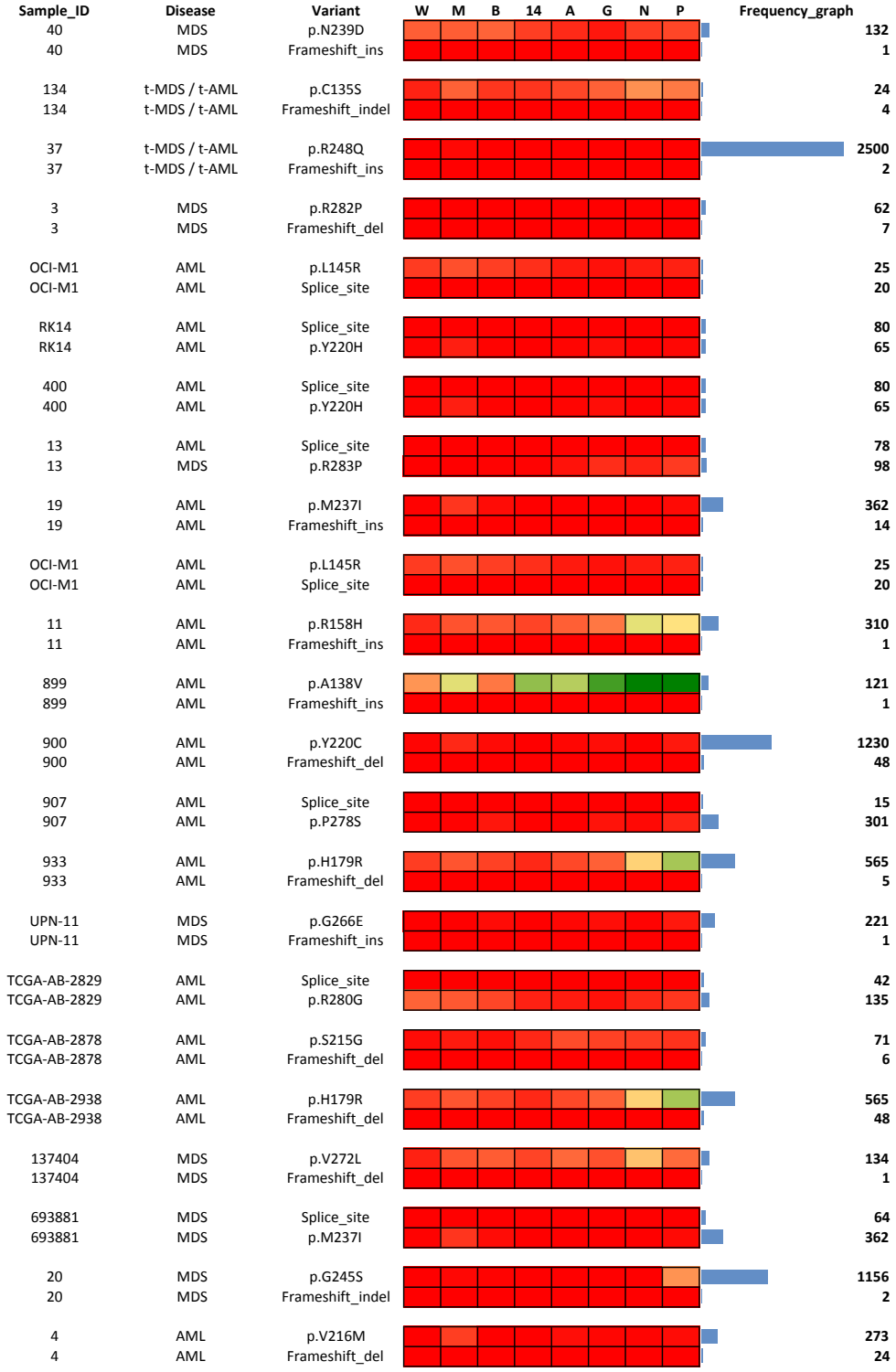
Supplemental Figure S1a (part 6)



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

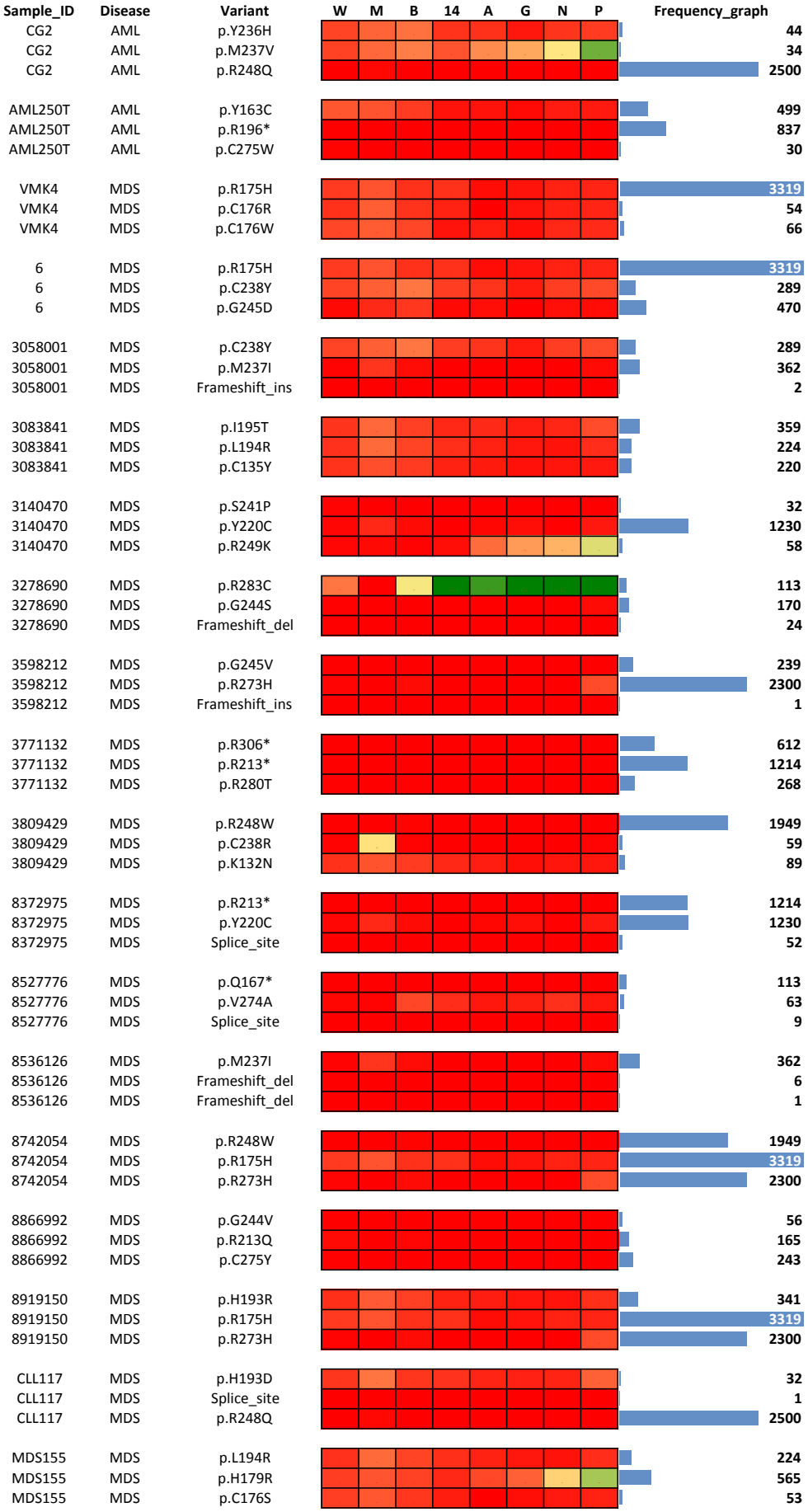
Supplemental Figure S1b (part 1)



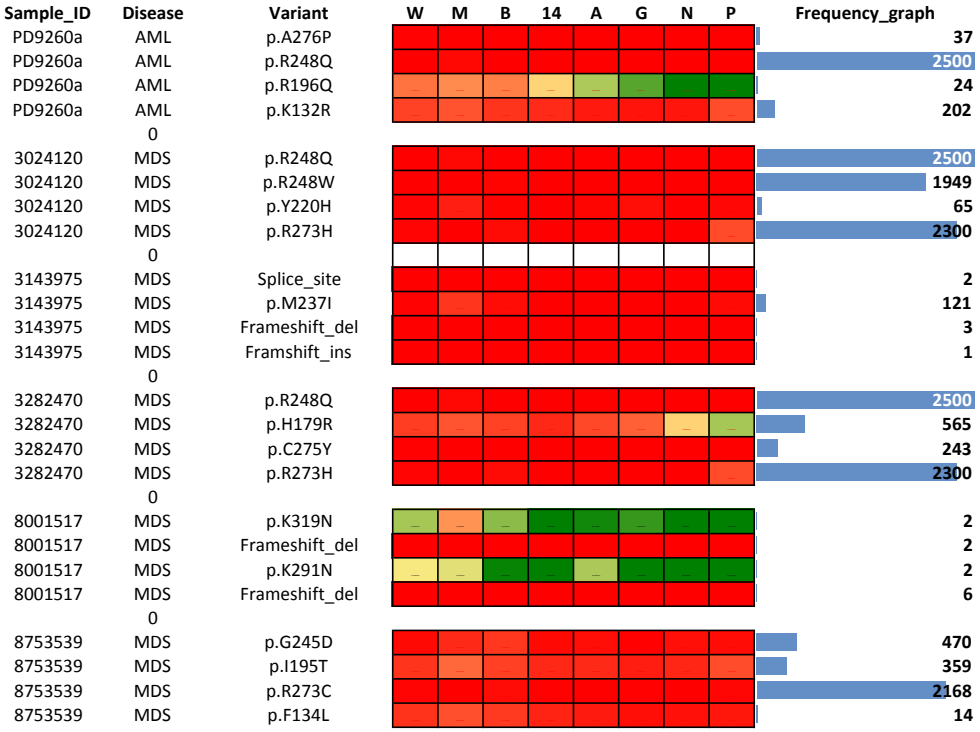
Supplemental Figure S1b (part 2)

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

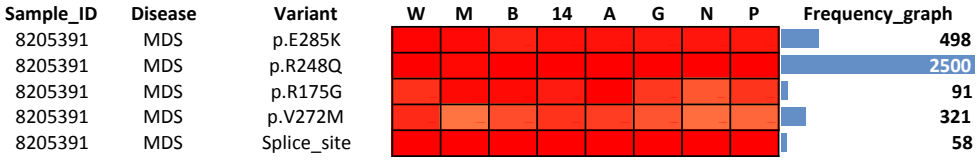
Supplemental Figure S1b (part 3)



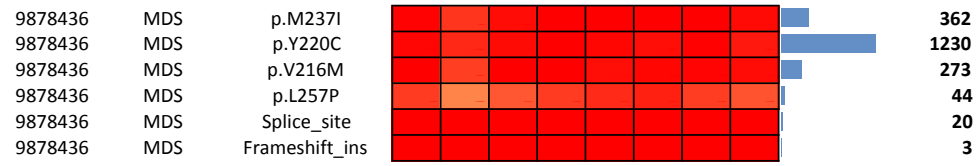
Supplemental Figure S1c



Supplemental Figure S1d

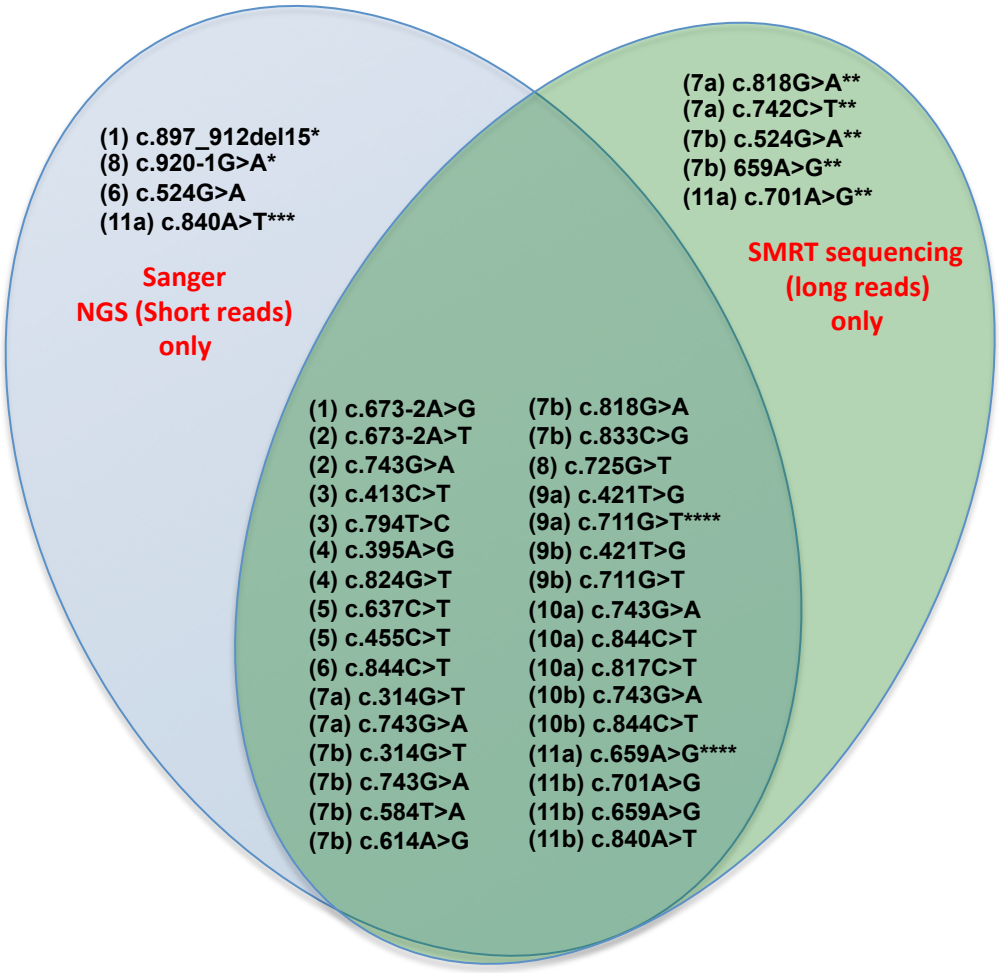


Supplemental Figure S1e



Supplemental Figure S1f

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



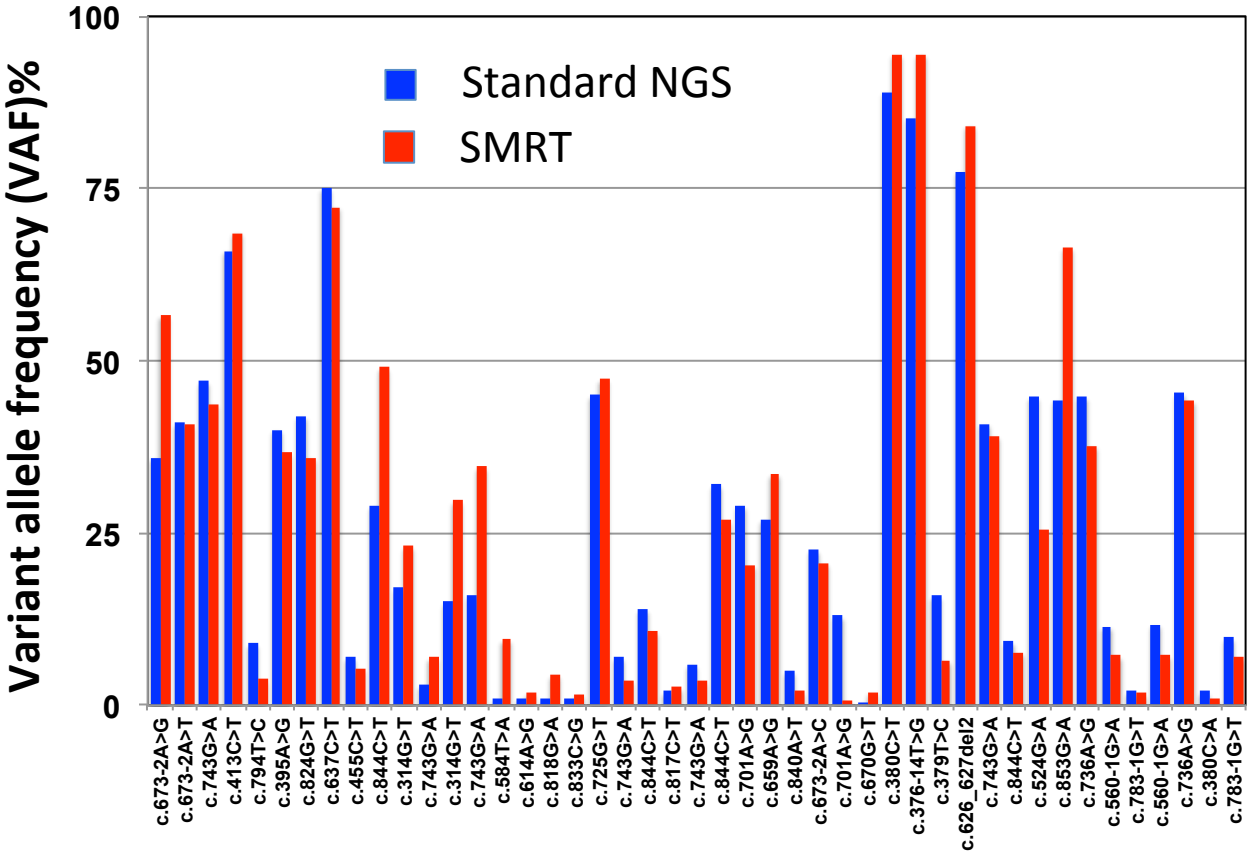
* Mutation outside the amplicon used for SMRT analysis

** 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

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

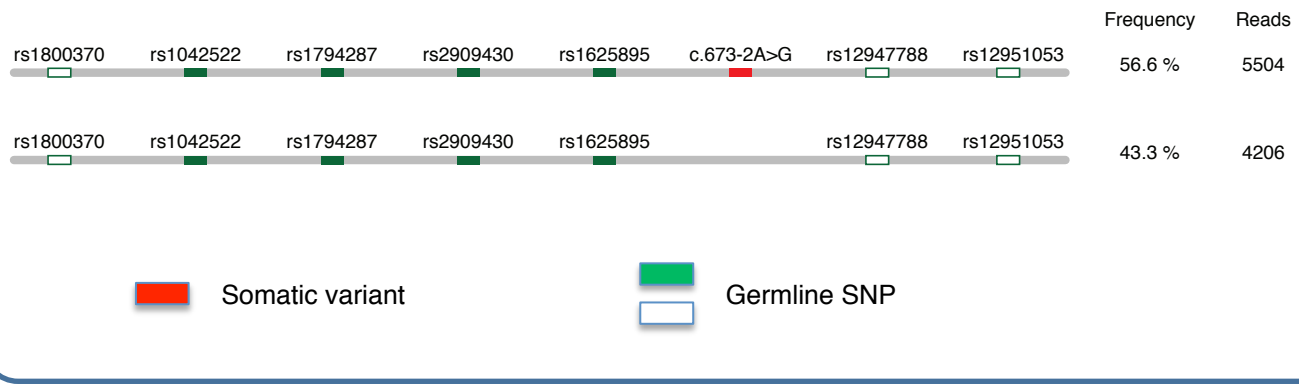


Figure S4a

Clinical information

Patient Fr2

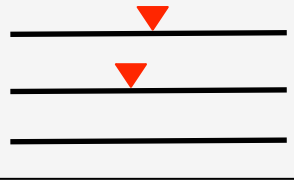
Disease: de novo MK-AML
Age: 63
Treatment: none
17p status: deletion (CGH array)

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

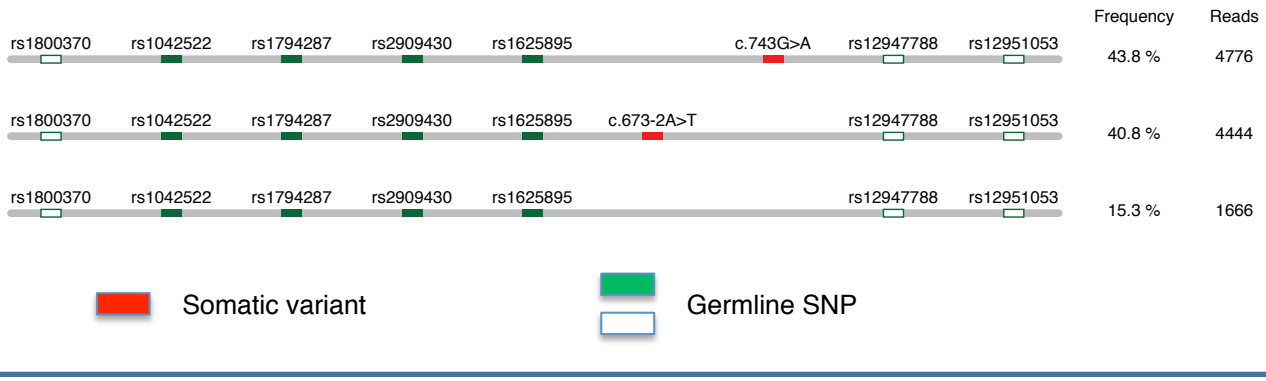


Figure S4b

Patient Fr3

Clinical information

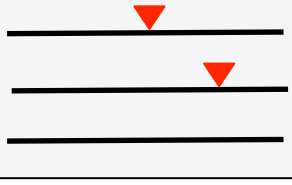
Disease: de novo MK-AML
Age: 73
Treatment: none
17p status: deletion (FISH)

Sequencing

Sanger/NGS (short reads)

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

SMRT sequencing (long reads)



Variant detected by both analyses

Haplotype

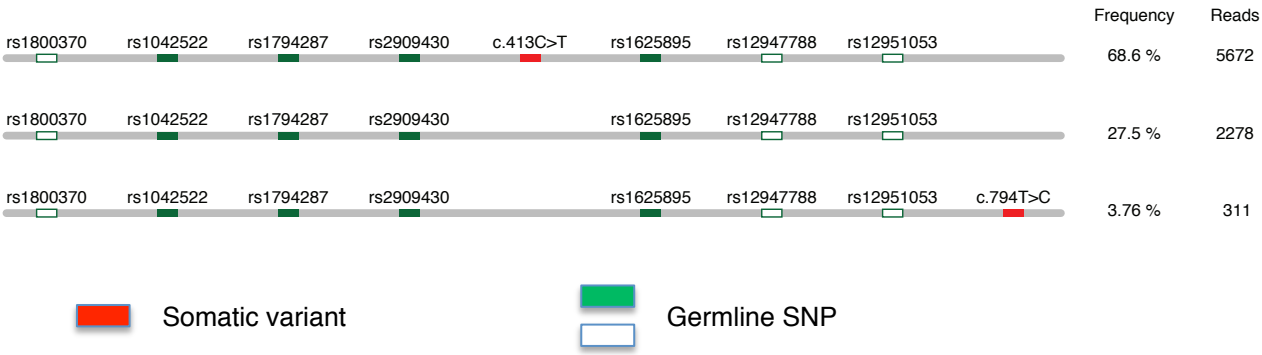


Figure S4c

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

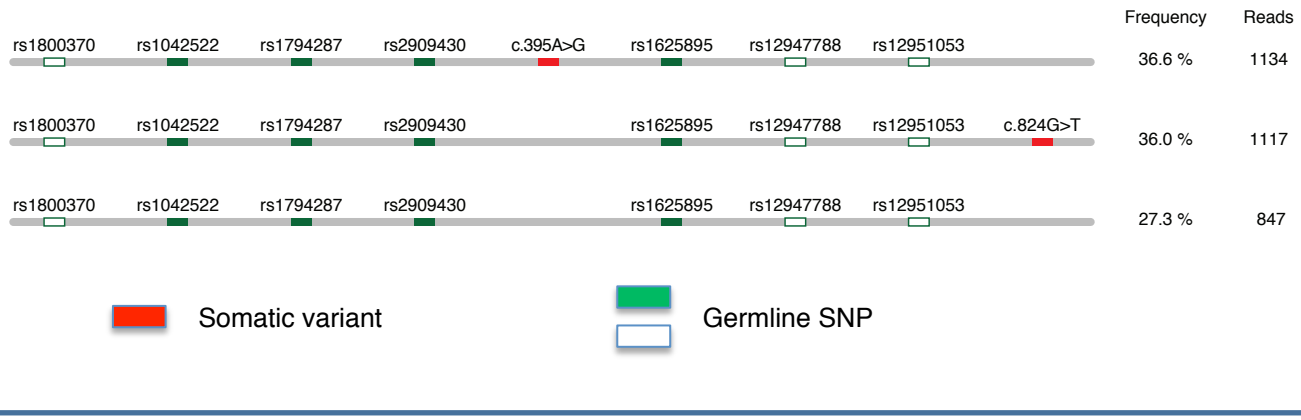


Figure S4d

Patient Fr5

Clinical information

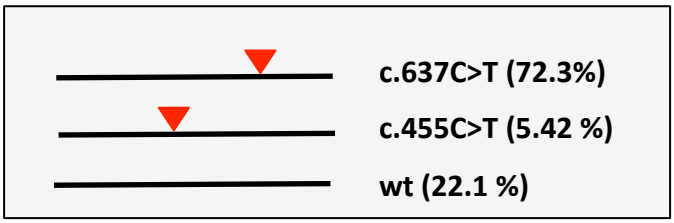
Disease: de novo MK-AML
Age: 73
Treatment: none
17p status: deletion (FISH)

Sequencing

Sanger/NGS (short reads)

▼ c.637C>T (75 %)
▼ c.455C>T (7 %)

SMRT sequencing (long reads)



▼ Variant detected by both analyses

Haplotype

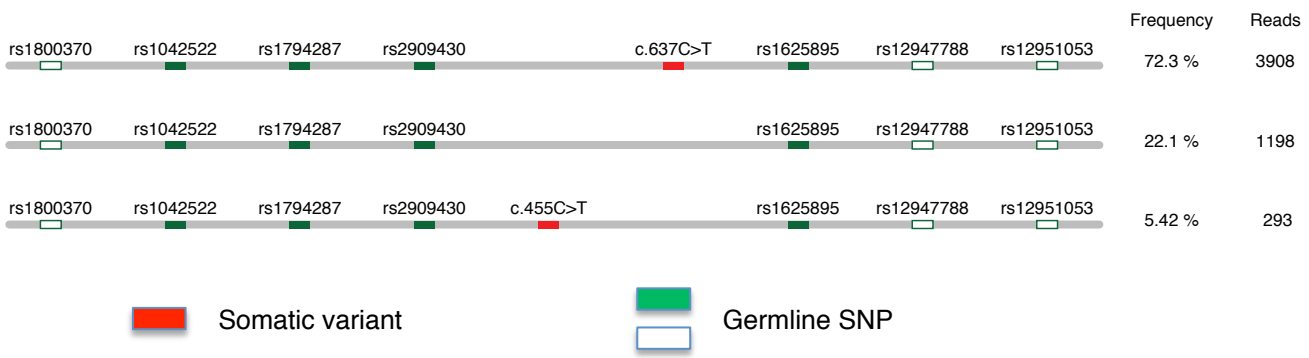


Figure S4e

Patient Fr6

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

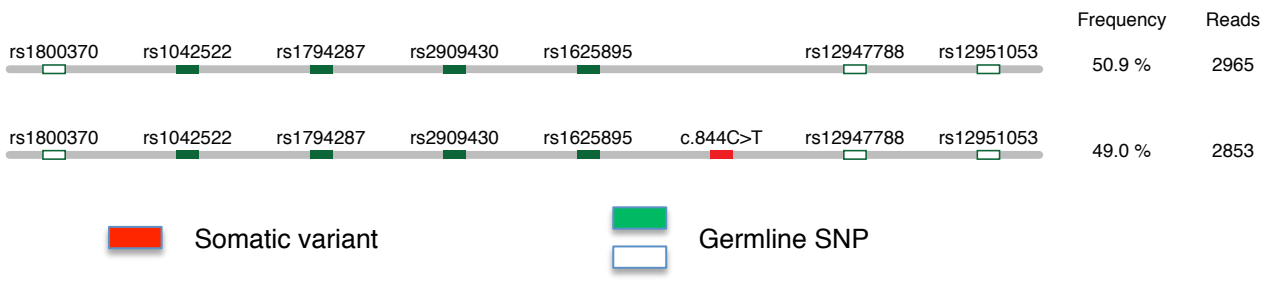


Figure S4f

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

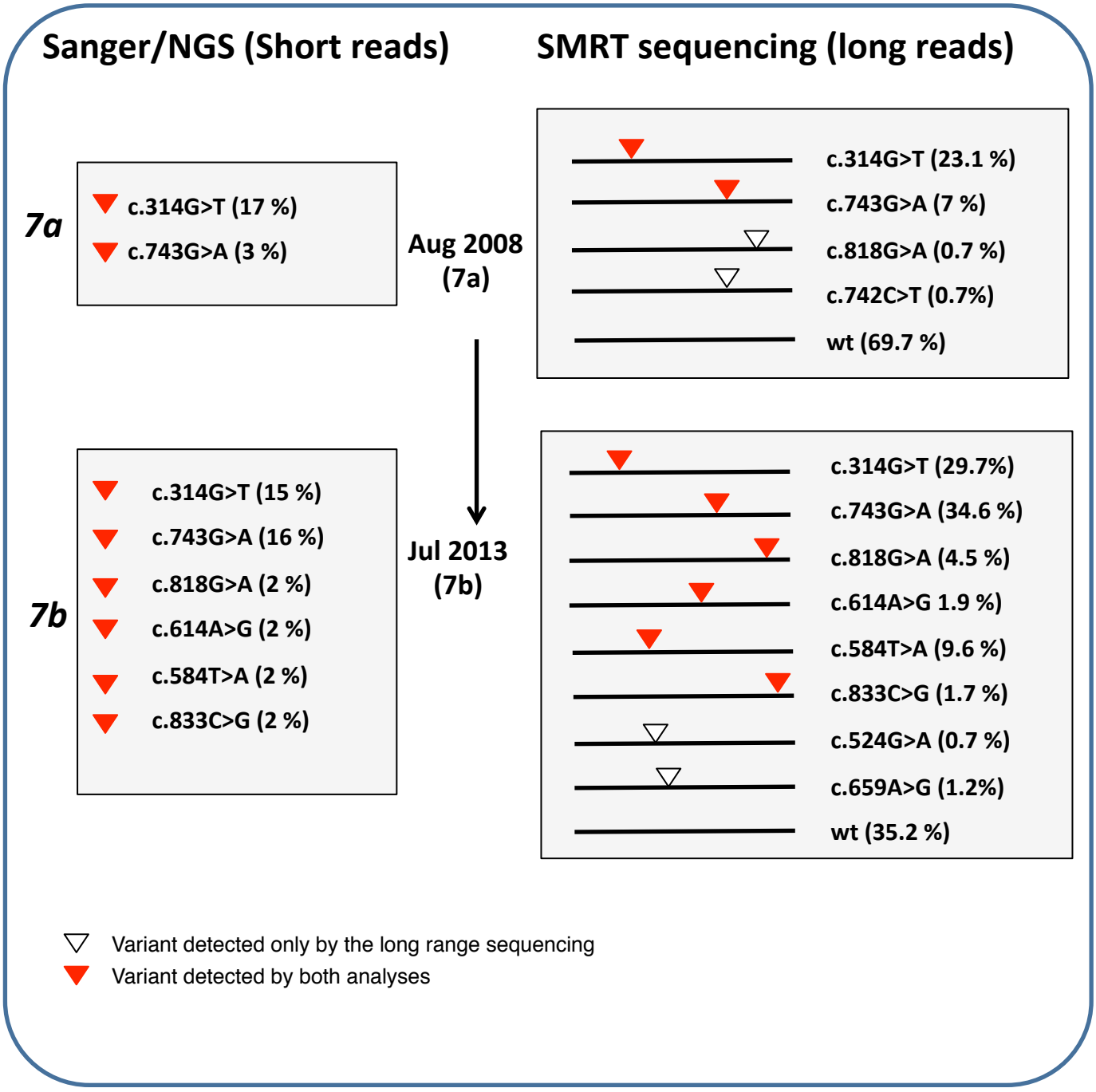


Figure S4g

Haplotype

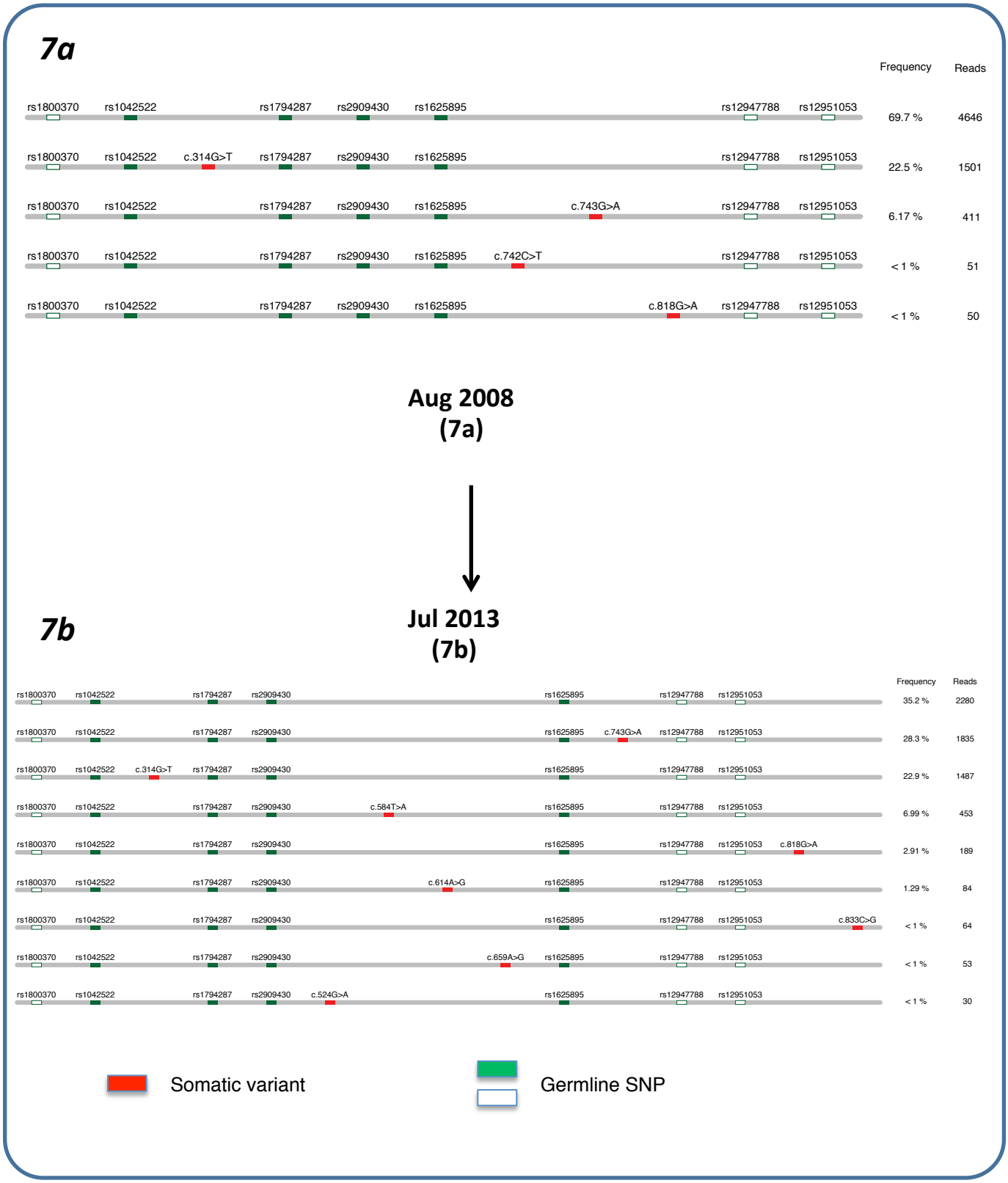


Figure S4g

Patient Fr8

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

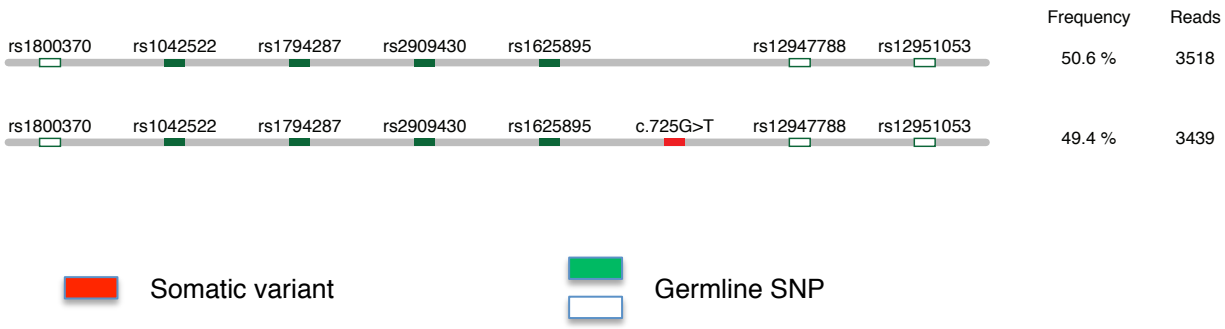


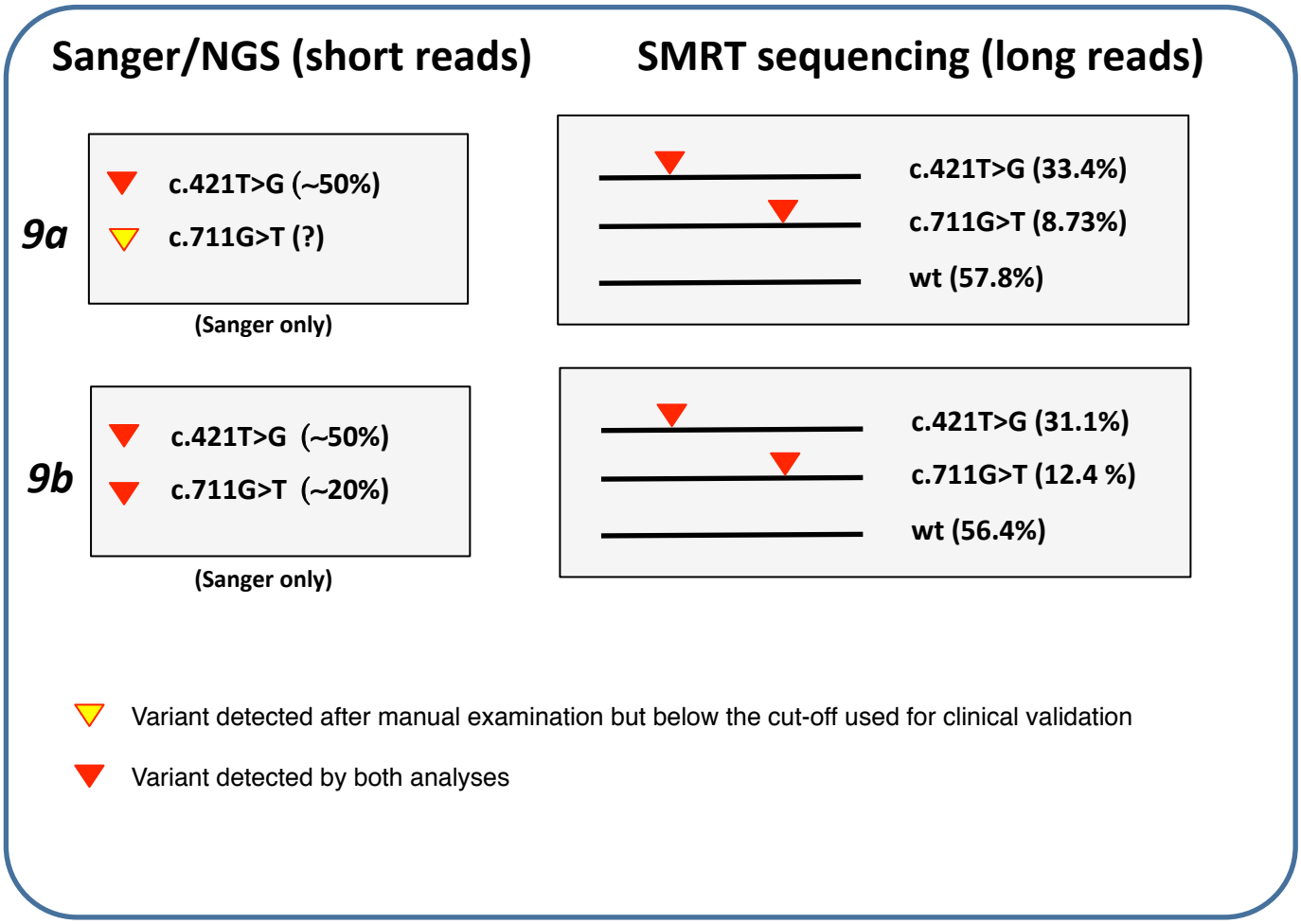
Figure S4h

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

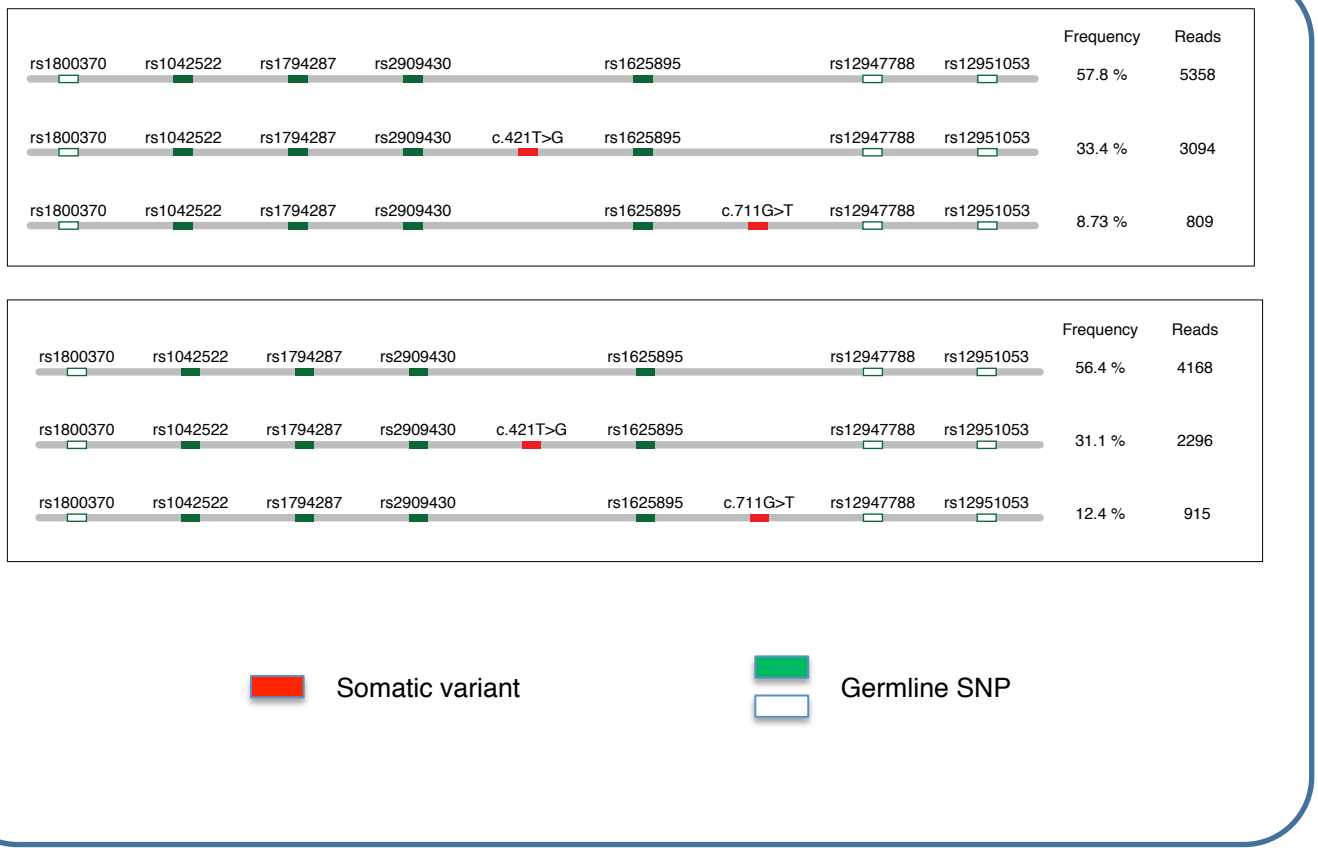


Figure S4i

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

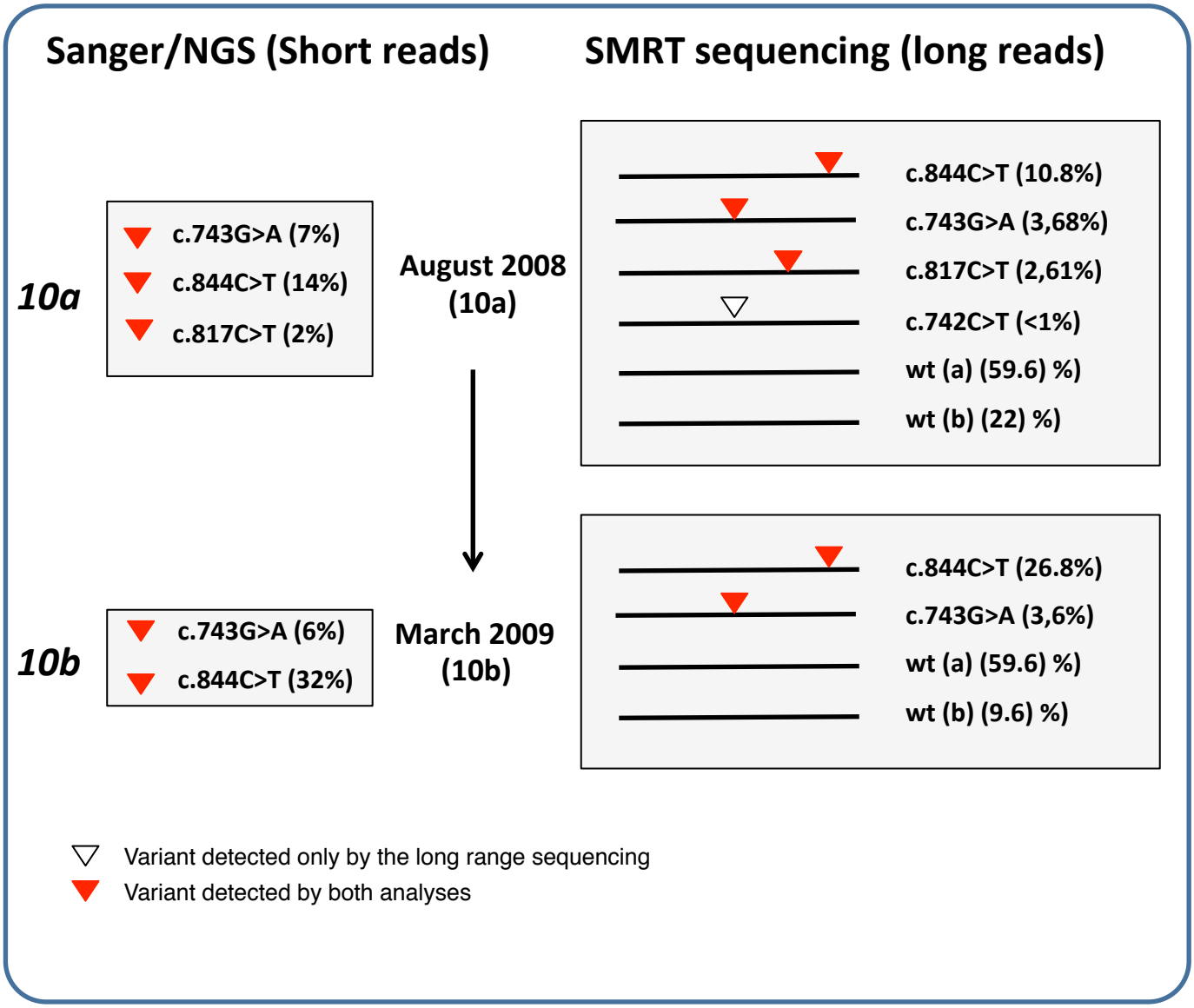


Figure S4j

Haplotype

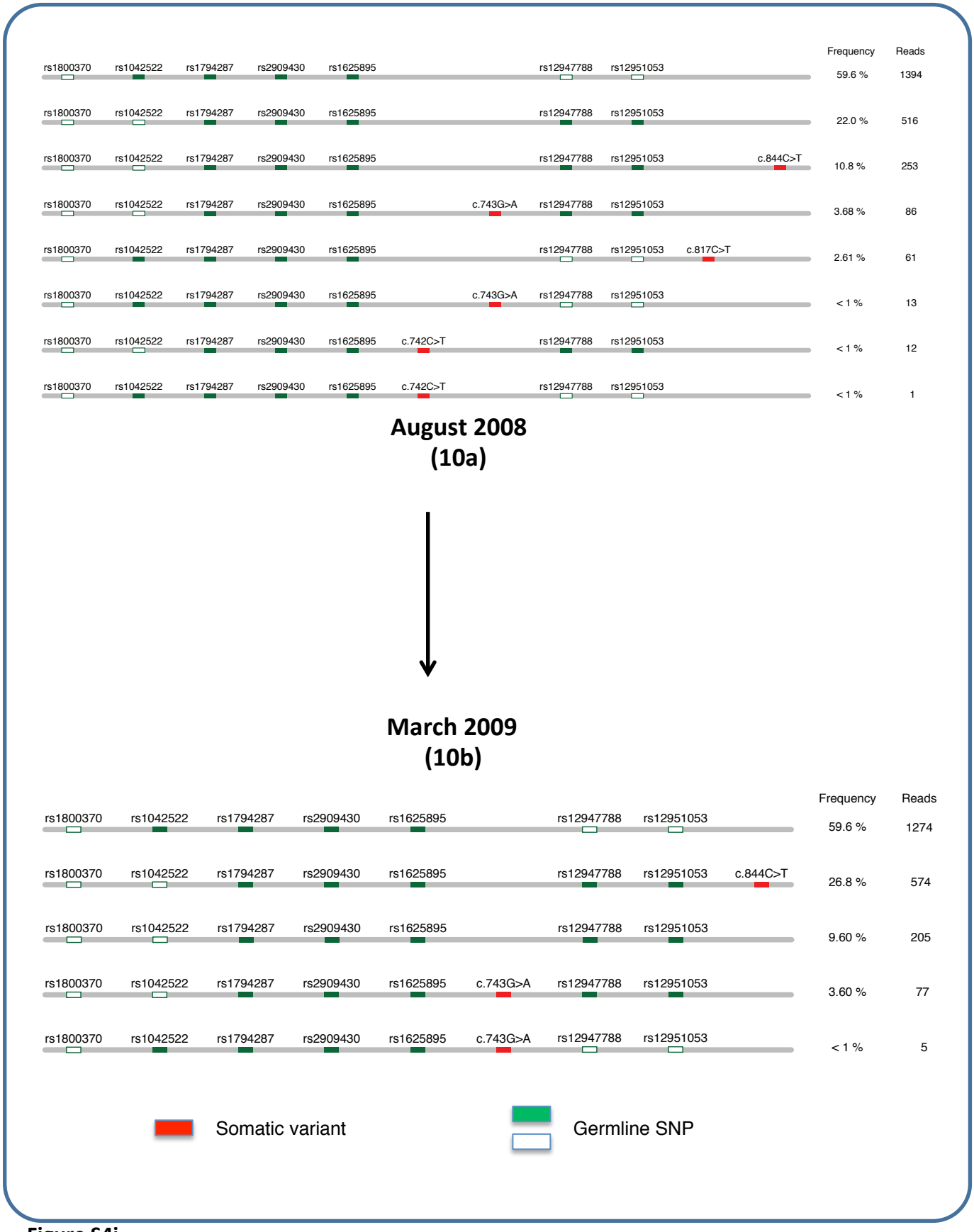


Figure S4j

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

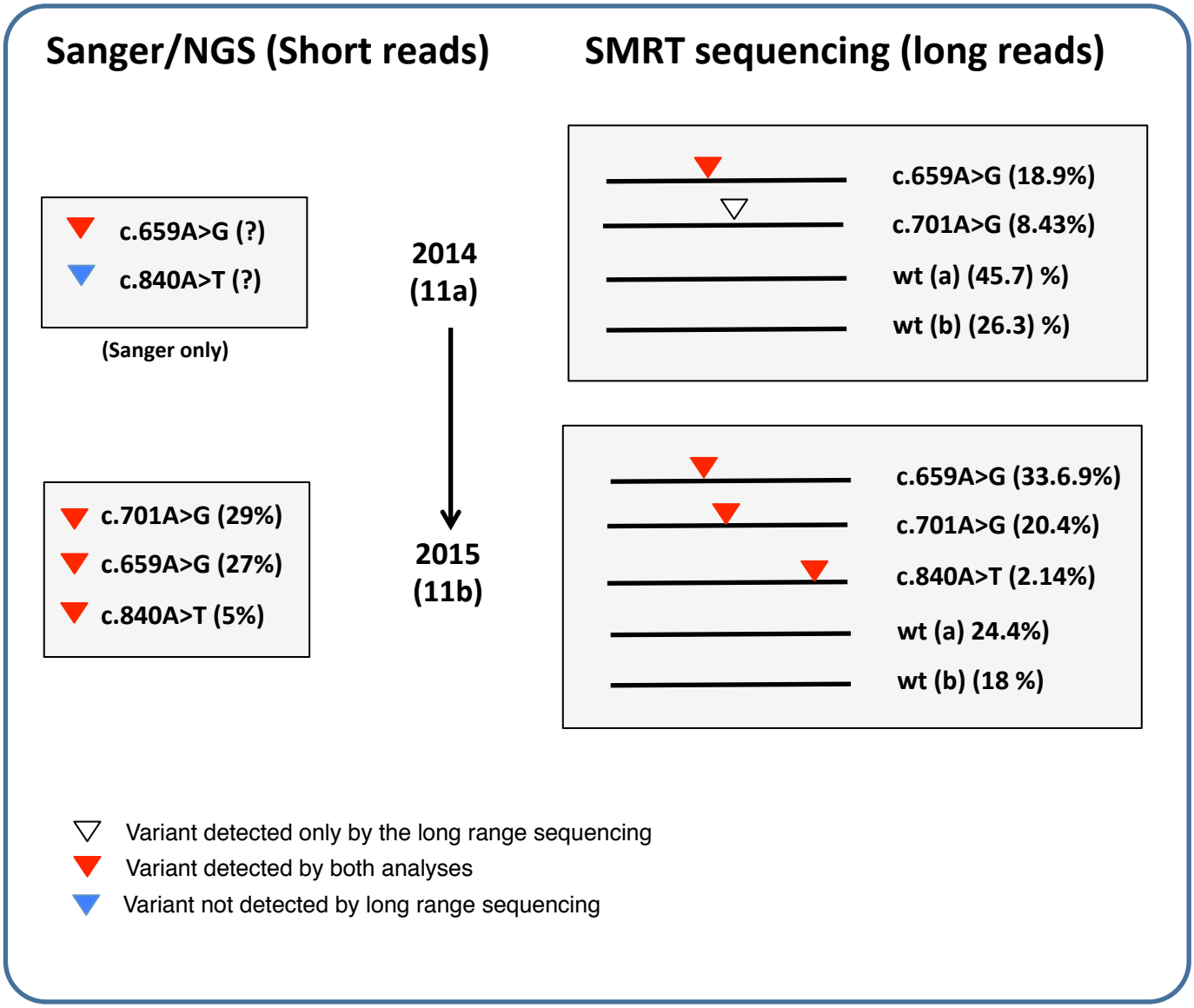


Figure S4k

Haplotype

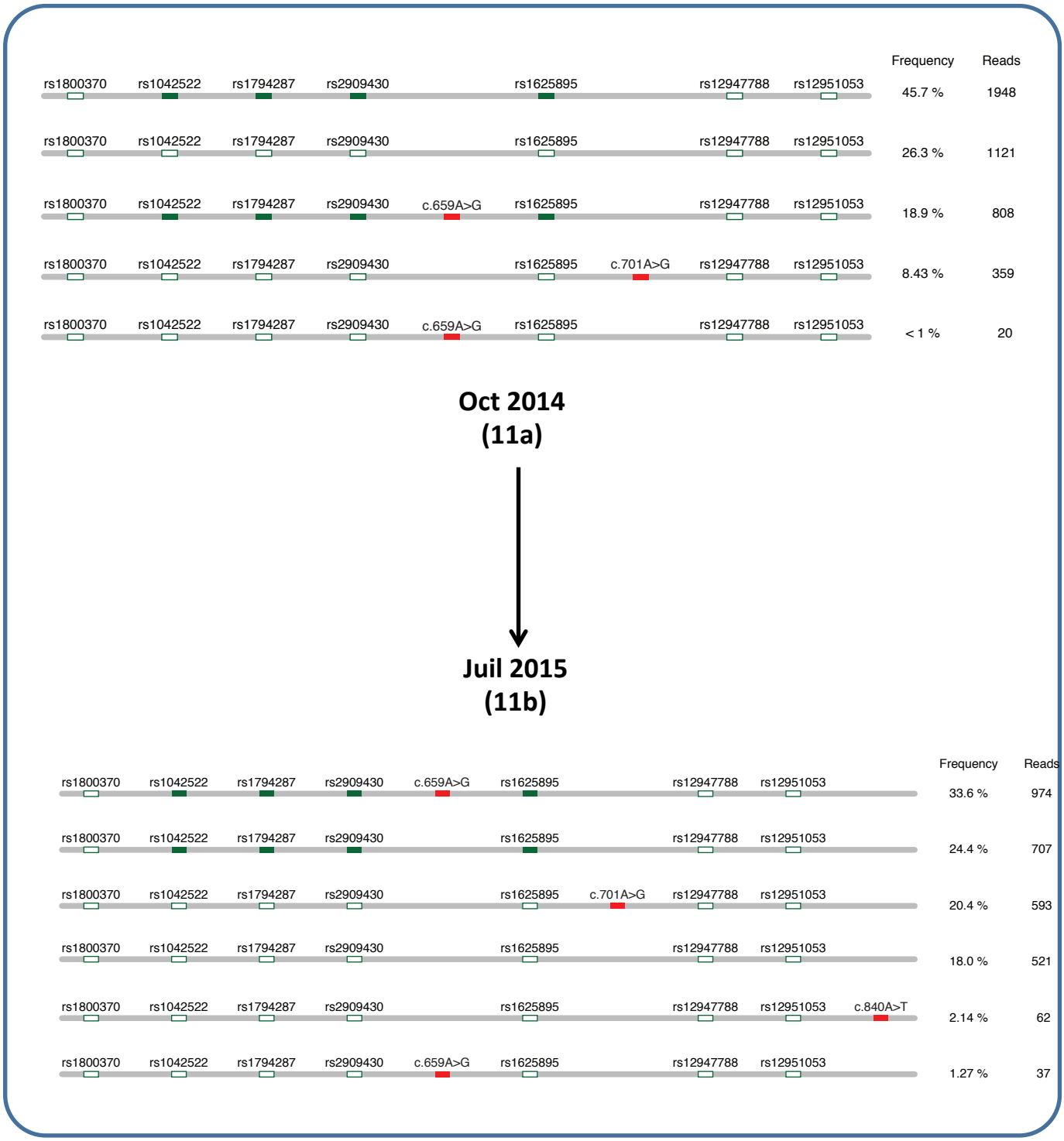
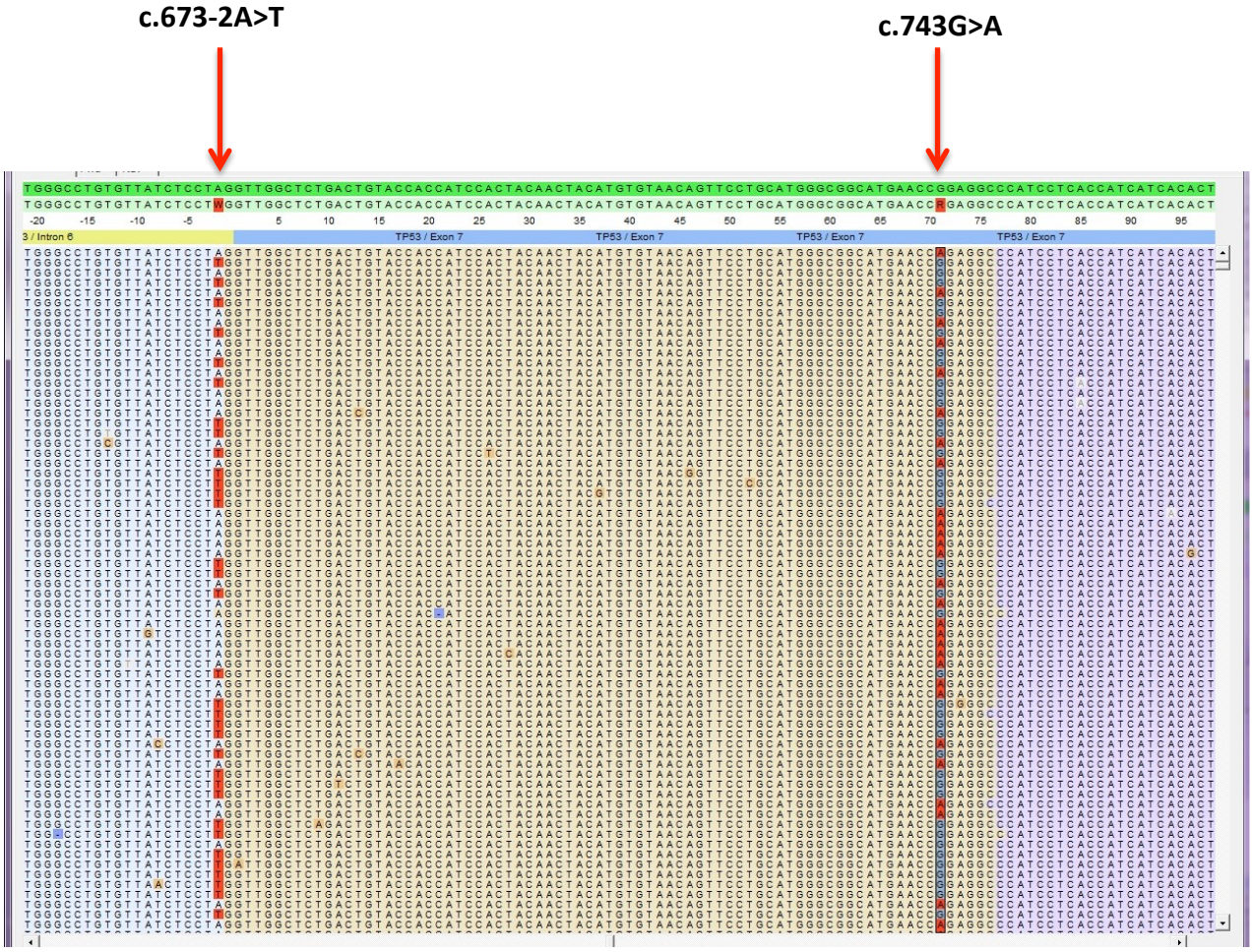


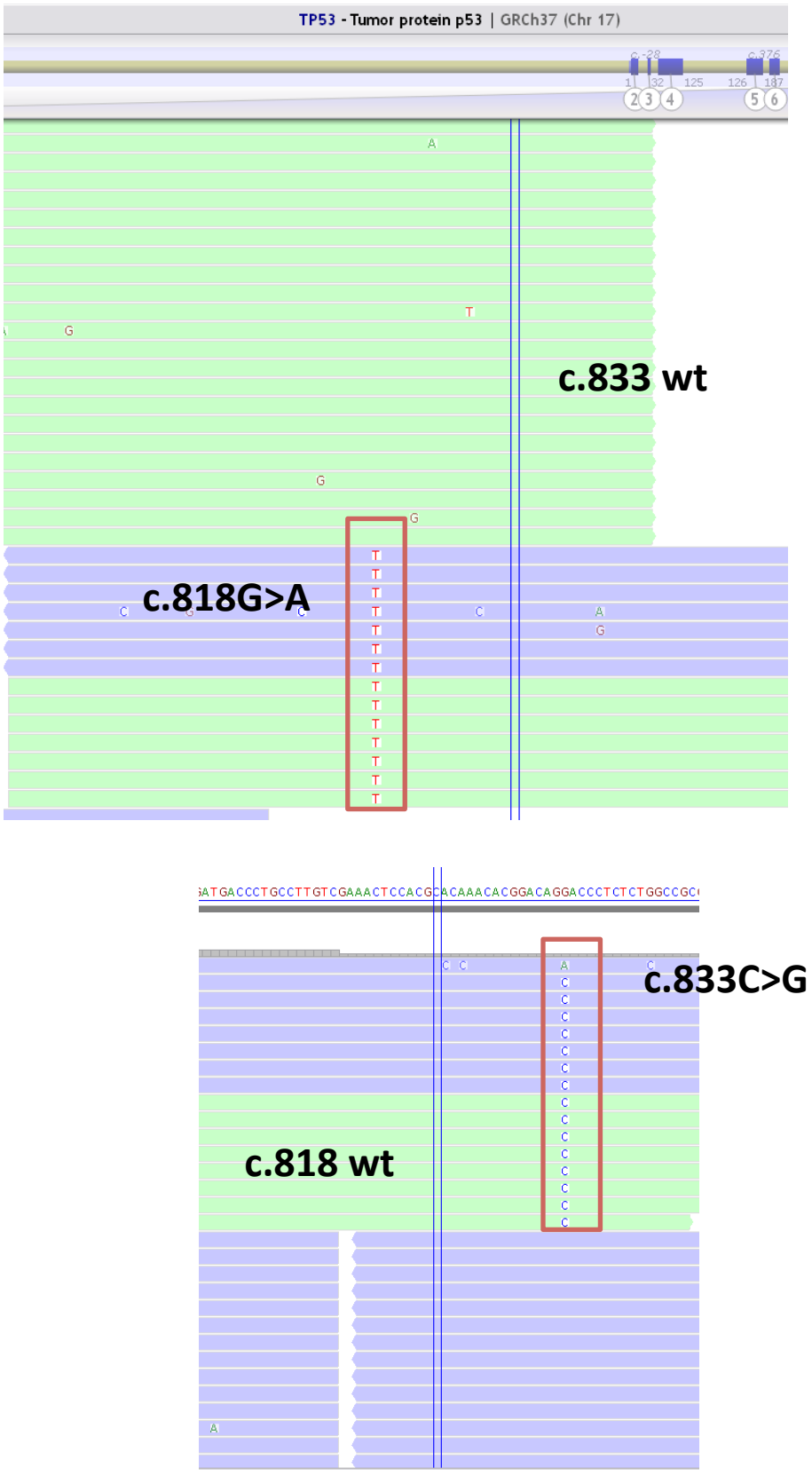
Figure S4k

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



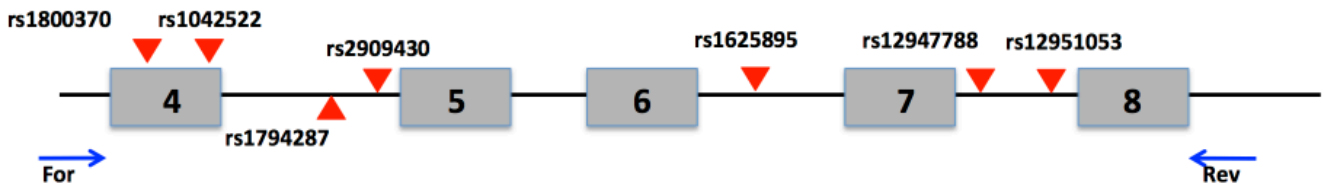
Patient Fr2

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



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