

Title

The broad spectrum of *TP53* mutations in CLL: evidence of multiclonality and novel mutation hotspots.

Running title (60 max)

Spectrum of *TP53* mutations in CLL

Lazarian et al.

Supplementary materials

Supplementary Figure S1: Genetic landscape of *TP53* alterations.

Supplementary Figure S2: High frequency of AT>GC transition in CLL.

Supplementary Figure S3: Copy-neutral LOH (CN-LOH) is frequent in CLL and can be validated by SNP analysis.

Supplementary Figure S4: Identification of *TP53* mutation hotspots in CLL in codon 234.

Supplementary Figure S5: Multiple *TP53* variants in codon 234 in a same patient are associated with CLB exposure.

Supplementary Figure S6: Variant NM_000546_c.626_627del (NP_000537_p.Arg209LysfsTer6) is a hotspot mutation in CLL.

Supplementary Figure S7: Frequency of splice mutations in the *TP53* gene among the various types of cancer.

Supplementary Figure S8: CLL hotspot mutation in the splice acceptor signal of intron 6.

Supplementary Figure S9: Wide variant distribution in the number of *TP53* variants in CLL patients.

Supplementary Figure S10 to 10p: *TP53* mutations in FILO-cohort polymutated patients are in different alleles.

Supplementary figure S11 A to E: Long-range sequencing of CLL patients shows that *TP53* mutations in multi-mutated patients are in different alleles.

Supplementary Figure S12: Multiple *TP53* variants are found at the same frequency in patients with or without 17p deletion.

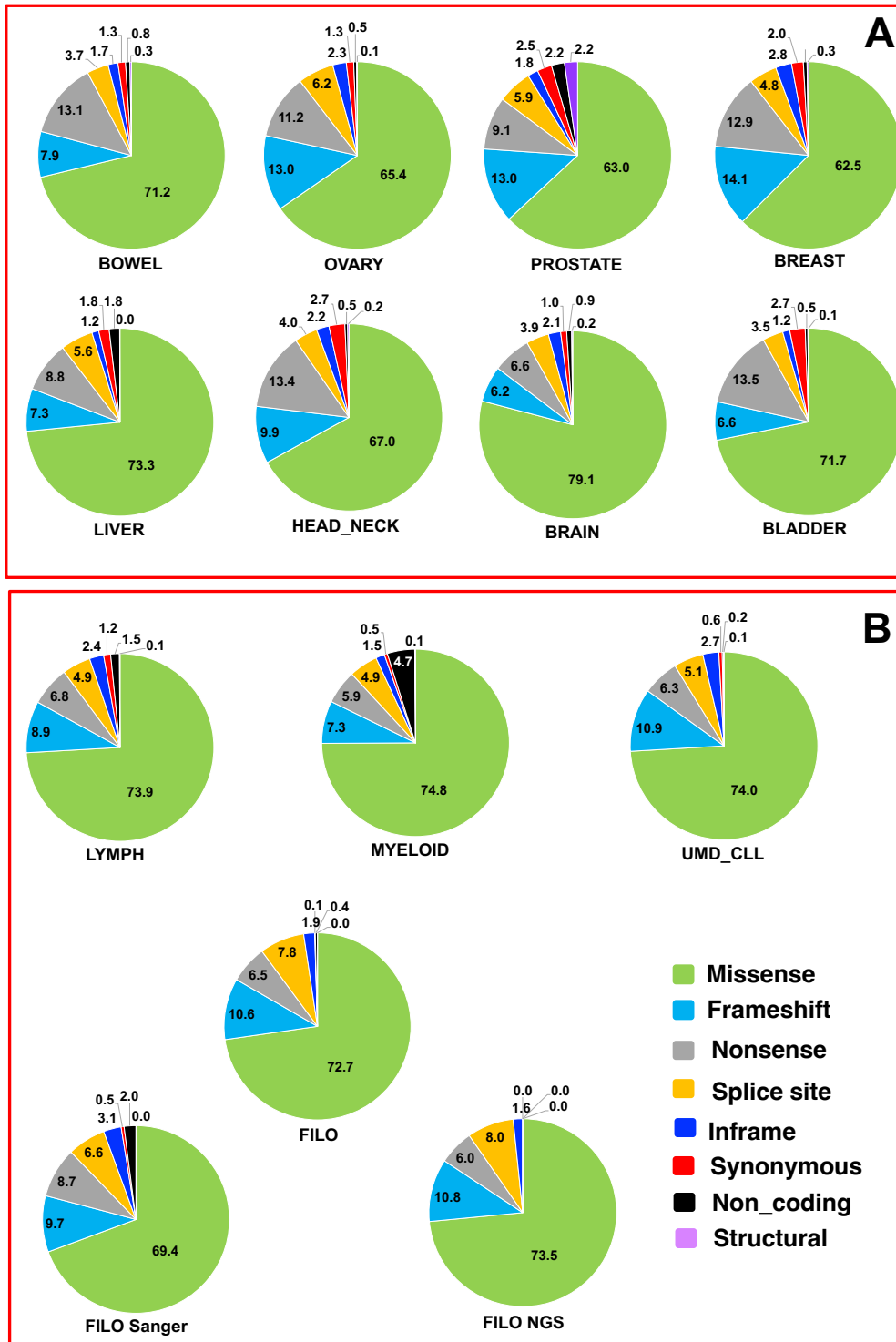
Supplementary Figure S13: Distribution of *TP53* hotspot variants in CLL patients.

Supplementary Table S1: Patient characteristics.

Supplementary Table S2: Mutations summary.

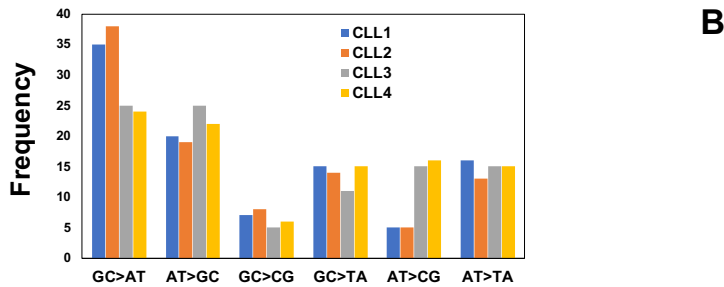
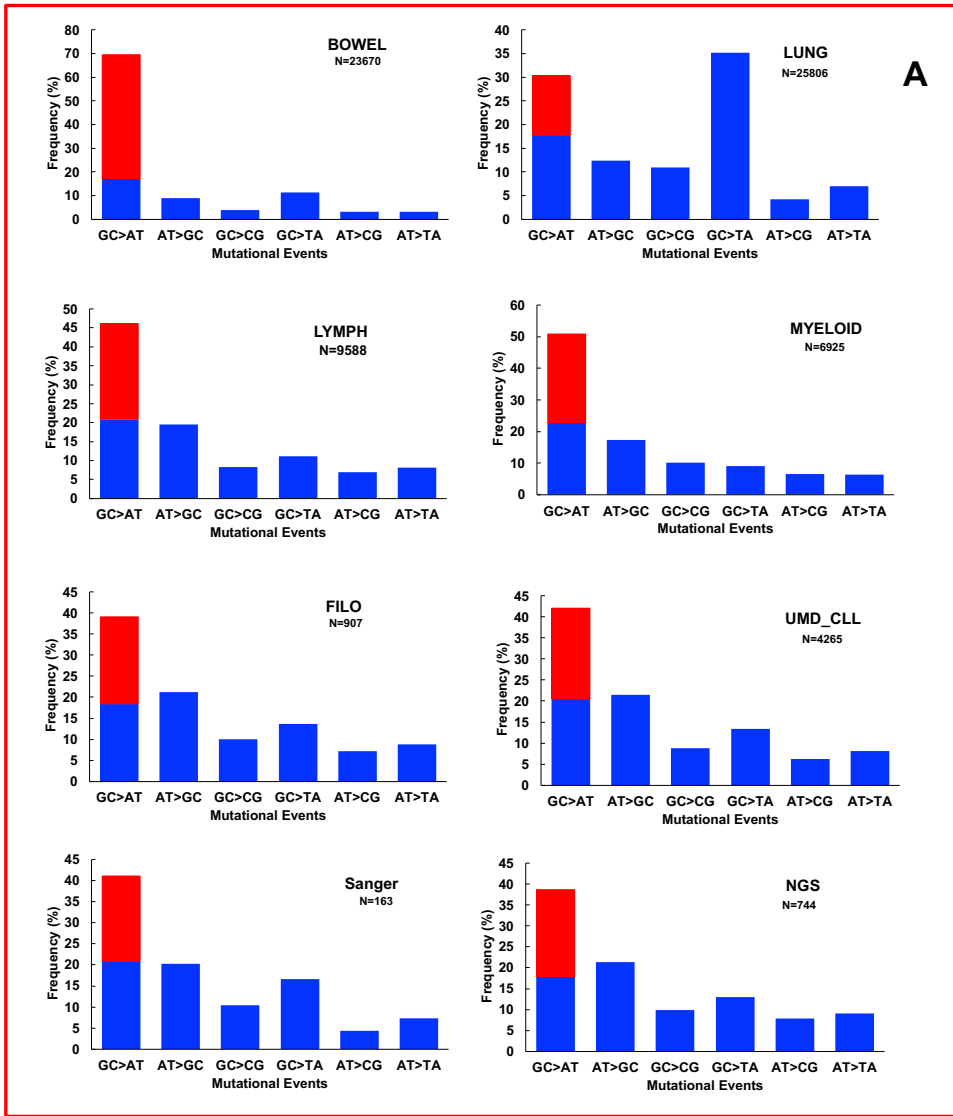
Supplementary Table S3: Mutation description. (Provided as an Excel file)

Supplementary Table S4: Benign polymorphisms identified in CLL patients included in UMD.



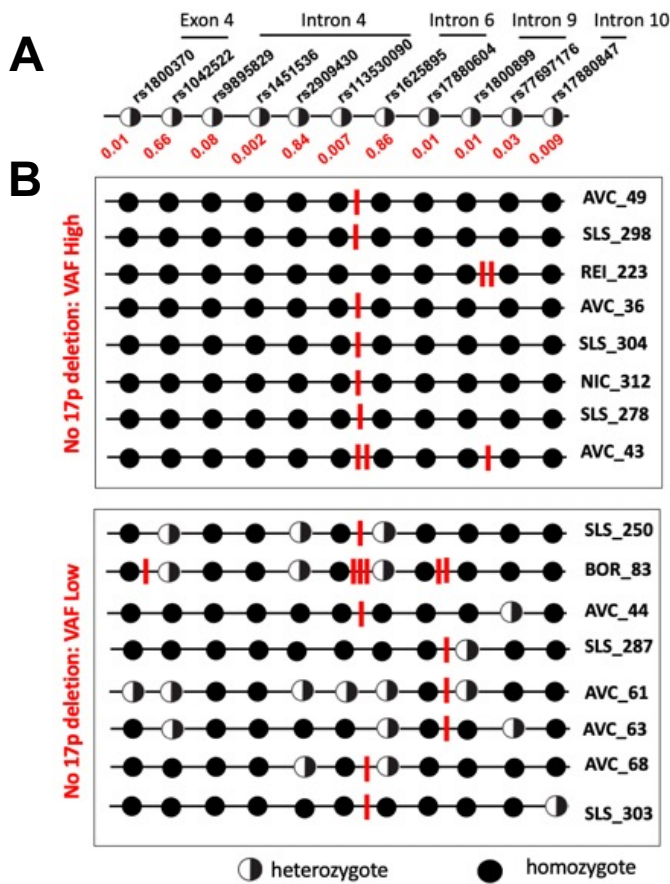
Supplementary Figure S1: Genetic landscape of *TP53* alterations.

A: Data from UMD_TP53 (R2, 2022) for solid tumors. **B:** Data from UMD_TP53 (R2, 2022) for hematological neoplasia. LYMPH and MYELOID: all lymphoid and myeloid neoplasms respectively. UMD_CLL: all chronic lymphocytic leukemia except FILO data. FILO: data from the whole FILO cohort or sequenced via Sanger (FILO Sanger or NGS 'FILO NGS'). Organ site definition for all tumors in UMD was performed using the OncoTree classification system {Kundra et al., 2021, #201830}.



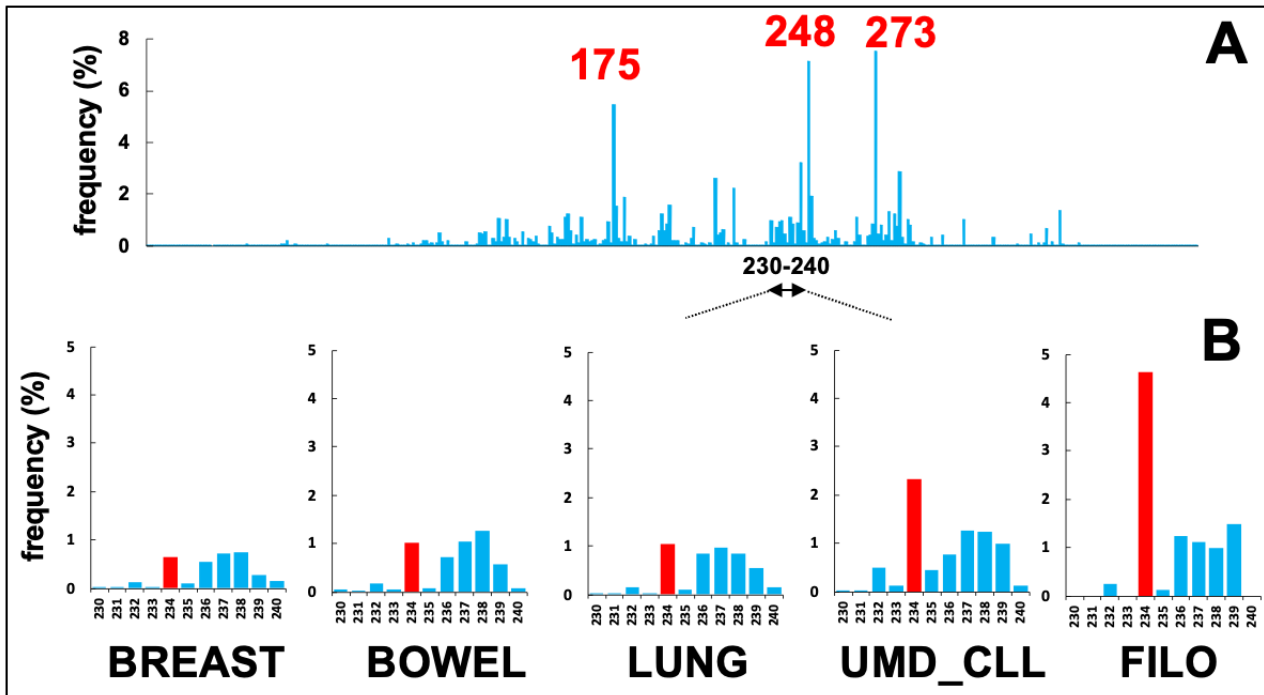
Supplementary Figure S2: High frequency of AT>GC transition in CLL.

A: Mutation spectra were obtained from the UMD_TP53 database for solid tumors (BOWEL or LUNG) or lymphoid and myeloid neoplasm (LYMPH or MYELOID). **B:** Profile of somatic mutations in four CLL patients (CLL1 to 4) analyzed by whole genome sequencing {Punkte et al., 2011, #62374}. Organ site definition for all tumors in UMD was performed using the OncoTree classification system {Kundra et al., 2021, #201830}.



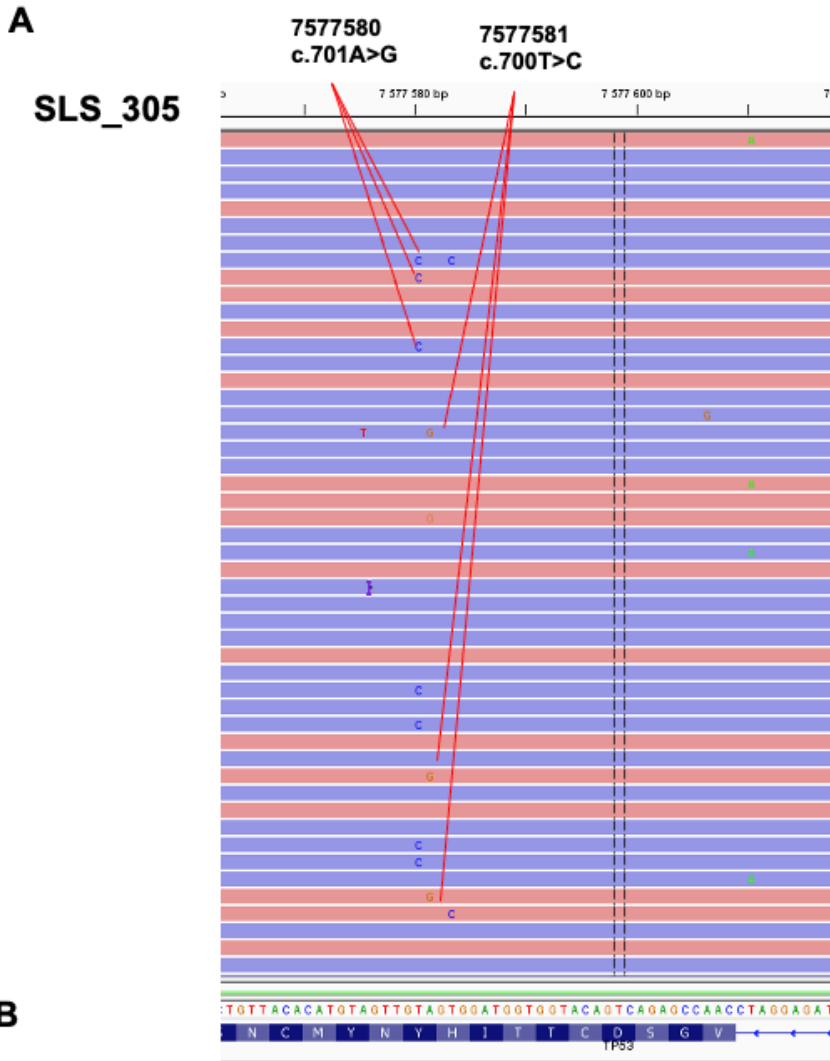
Supplementary Figure S3: Copy-neutral LOH (CN-LOH) is frequent in CLL and can be validated by SNP analysis.

A: Distribution of 11 *TP53* SNPs located in the region covered by NGS sequencing. For each SNP, the population frequency is shown in red. **B:** Haplotype of patients without 17p deletion with high (top) or low (bottom) VAF for *TP53* mutation. For each patient, the approximate position of the *TP53* variant is shown as a red bar.



Supplementary Figure S4: Identification of *TP53* mutation hotspots in CLL in codon 234.

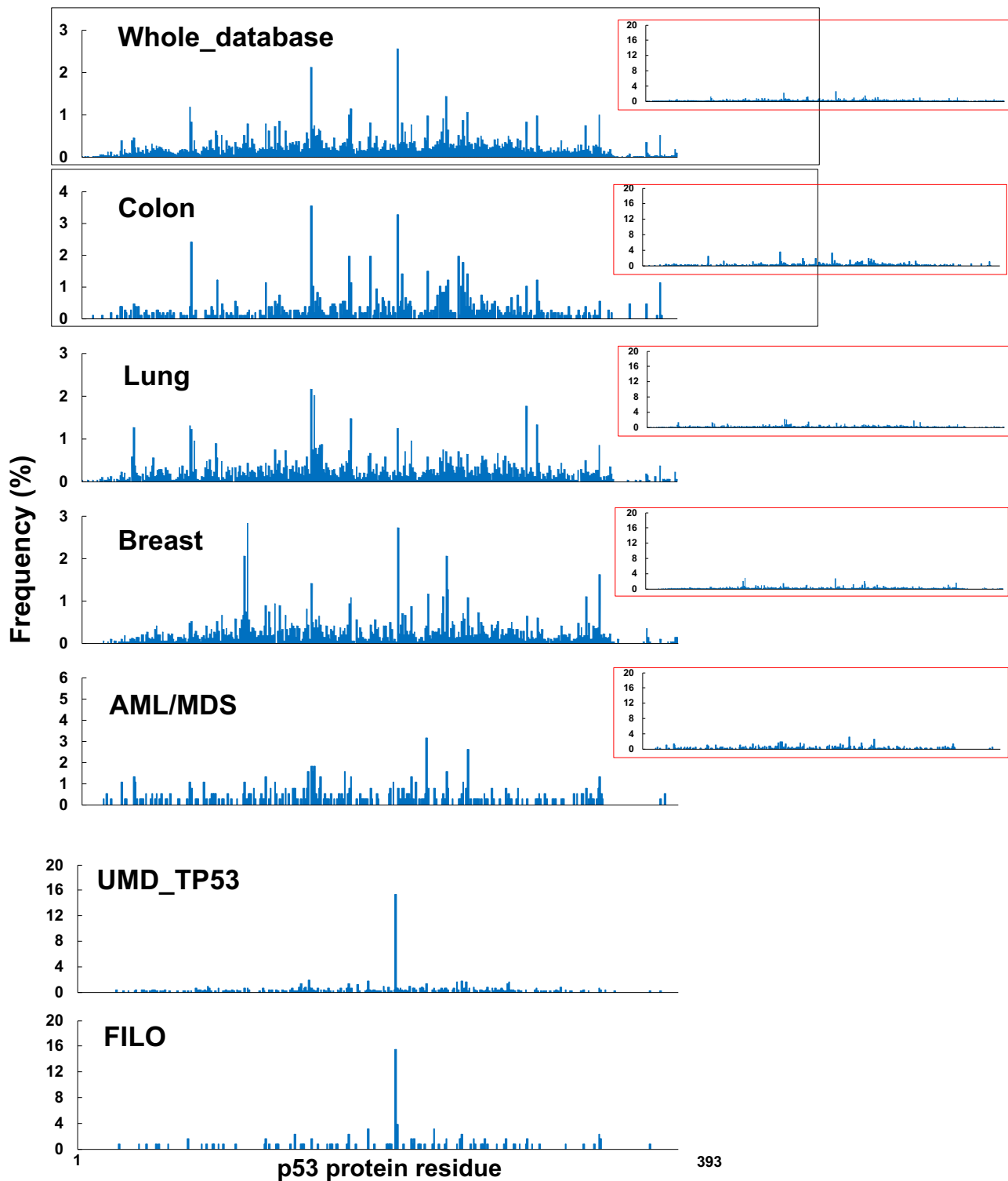
A: Distribution of mutations at each codon of the *TP53* protein. Only single nucleotide substitutions are analyzed. Classical *TP53* hotspot mutations are indicated in red with codons 175, 248 and 273 found in every type of cancer. **B:** Close up from figure A for codons 230 to 240 for different types of cancer. Data were updated from the work of Lazarian *et al.* with new patients from the FILO cohort {Lazarian *et al.*, 2022, #194534}.



LRG_321t1	TP53alpha	Disease_name	Sample_ID	Complexity	Authors	Year	Journal	Treatment
c.701A>G	p.Tyr234Cys	CLL	SLS_305	MM	Present work	NR	NR	Multiple lines of treatment including CLB
c.700T>C	p.Tyr234His	CLL	SLS_305	MM	Present work	NR	NR	Multiple lines of treatment including CLB
c.701A>G	p.Tyr234Cys	CLL	nAVC_129	MM	Present work	NR	NR	Multiple lines of treatment including CLB
c.700T>C	p.Tyr234His	CLL	nAVC_129	MM	Present work	NR	NR	Multiple lines of treatment including CLB
c.701A>G	p.Tyr234Cys	CLL	CLL-NHLBI-0011	MM	Landau DA, et al.	2017	Nat Commun	Treated; no specific information
c.701A>C	p.Tyr234Ser	CLL	CLL-NHLBI-0011	MM	Landau DA, et al.	2018	Nat Commun	Treated; no specific information
c.701A>G	p.Tyr234Cys	CLL	8	MM	Malcikova J, S et al.	2015	Leukemia	Multiple lines of treatment including CLB
c.701A>C	p.Tyr234Ser	CLL	8	MM	Malcikova J, S et al.	2015	Leukemia	Multiple lines of treatment including CLB
c.701A>G	p.Tyr234Cys	CLL	485	MM	Malcikova J, S et al.	2015	Leukemia	Multiple lines of treatment including CLB
c.701A>C	p.Tyr234Ser	CLL	485	MM	Malcikova J, S et al.	2015	Leukemia	Multiple lines of treatment including CLB
c.700T>C	p.Tyr234His	CLL	485	MM	Malcikova J, S et al.	2015	Leukemia	Multiple lines of treatment including CLB
c.701A>G	p.Tyr234Cys	AML	AML_547	DMU	Metzeler KH, et al.	2016	Blood	Multiple lines of treatment: no alkylating agents
c.700T>C	p.Tyr234His	AML	AML_547	DMU	Metzeler KH, et al.	2016	Blood	Multiple lines of treatment: no alkylating agents

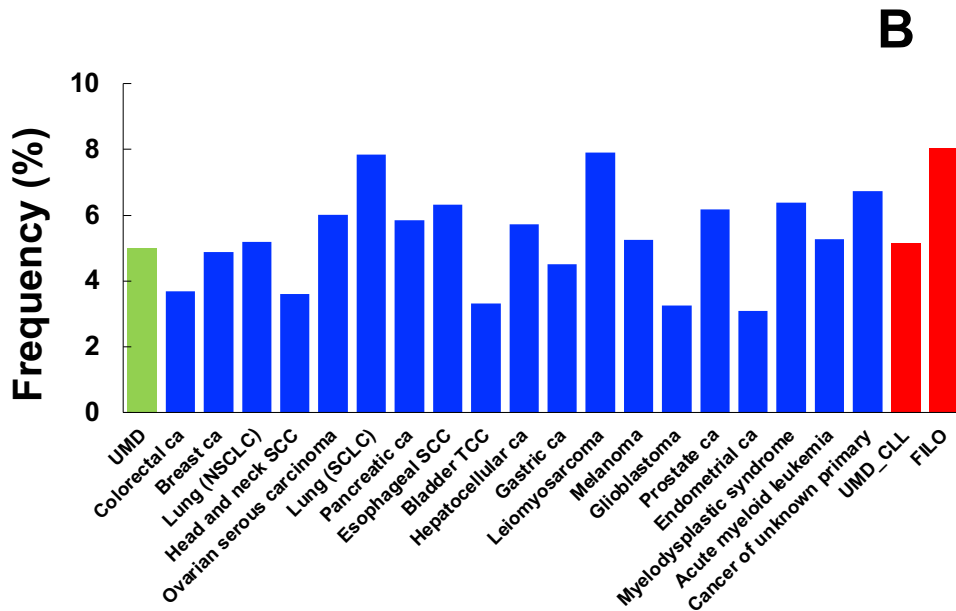
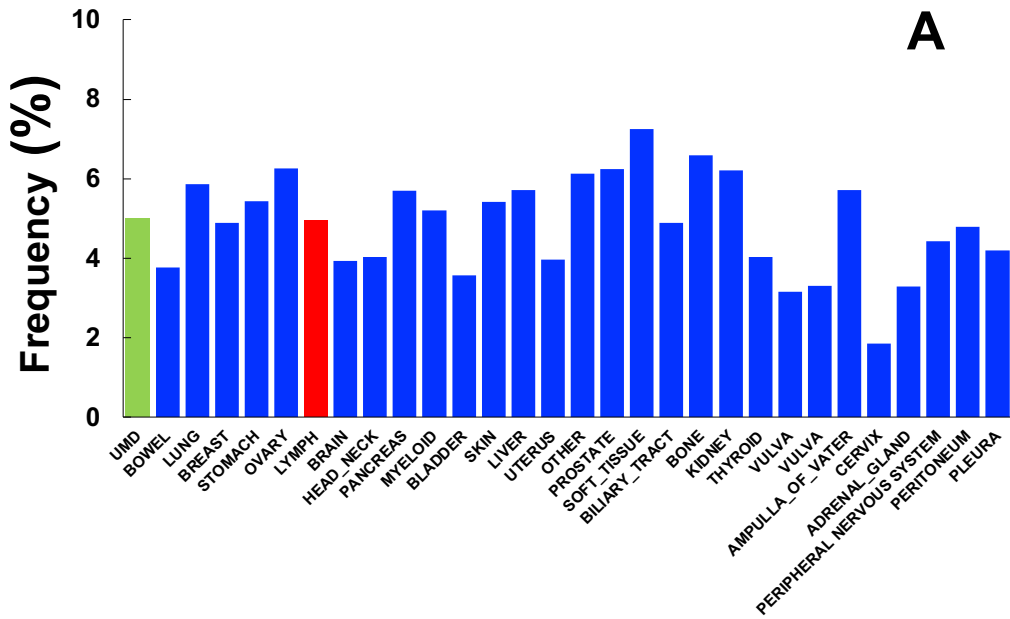
Supplementary Figure S5: Multiple *TP53* variants in codon 234 in a same patient are associated with CLB exposure.

A: IGV plot demonstrates the phasing of patient SLS_205 with multiple *TP53* variants in codon 234. Individual DNA sequencing reads harboring mutations within the same exon and no more than 50 nucleotides apart were analyzed. **B:** Details for the six patients harboring multiple mutations in codon 234 included in the entire UMD_TP53 database.



Supplementary Figure S6: Variant NM_000546_c.626_627del (NP_000537_p.Arg209LysfsTer6) is a hotspot mutation in CLL.

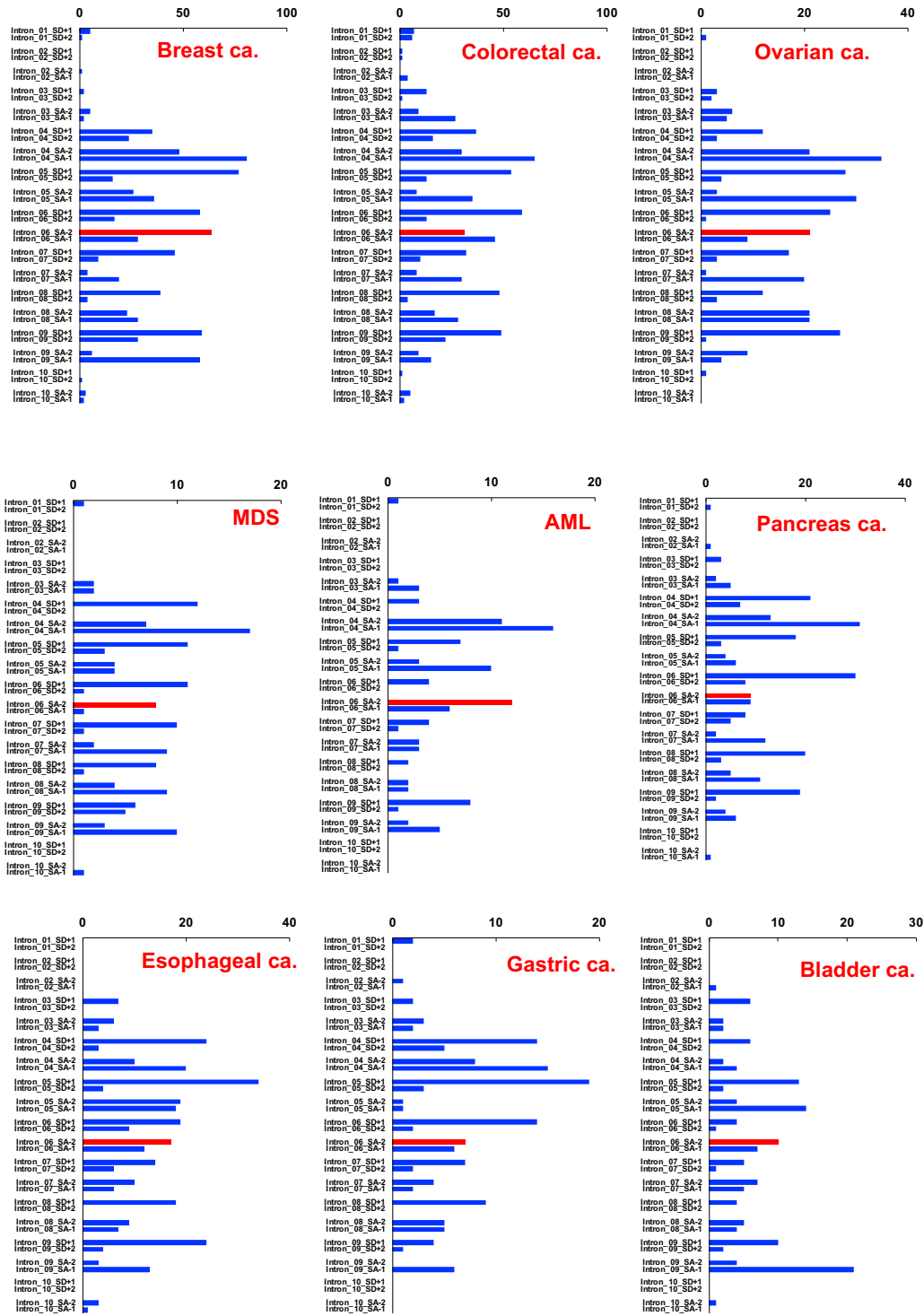
Distribution of frameshift mutations at each codon of the *TP53* protein in various types of cancer. Only insertions and deletions are analyzed. Scales for the distributions are not standardized. The insets for non-chronic lymphocytic leukemia have been standardized to chronic lymphocytic leukemia data to emphasize the high frequency of variant NM_000546_c.626_627del. AML/MDS: acute myeloid leukemia and myelodysplastic syndrome. Lung: NSCLC and SCLC.



Supplementary Figure S7: Frequency of splice mutations in the *TP53* gene among the various types of cancer.

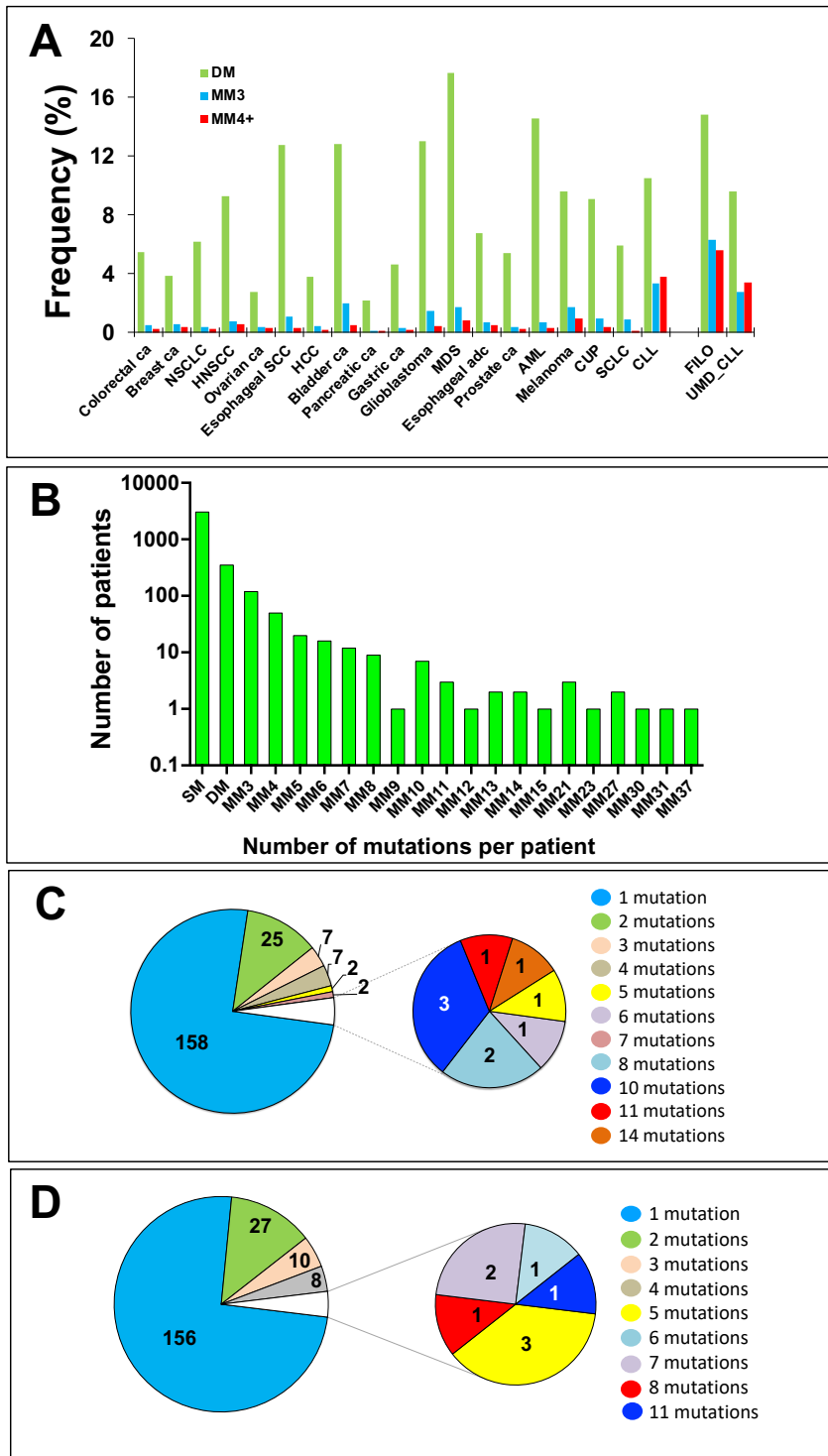
A: Cancer types were classified according to organ sites as defined in OncoTree. LYMPH: class including all CLL patients (5795) as well as other lymphoid cases (5392). **B:** Cancer types were classified according to tumor types.

Number of variants



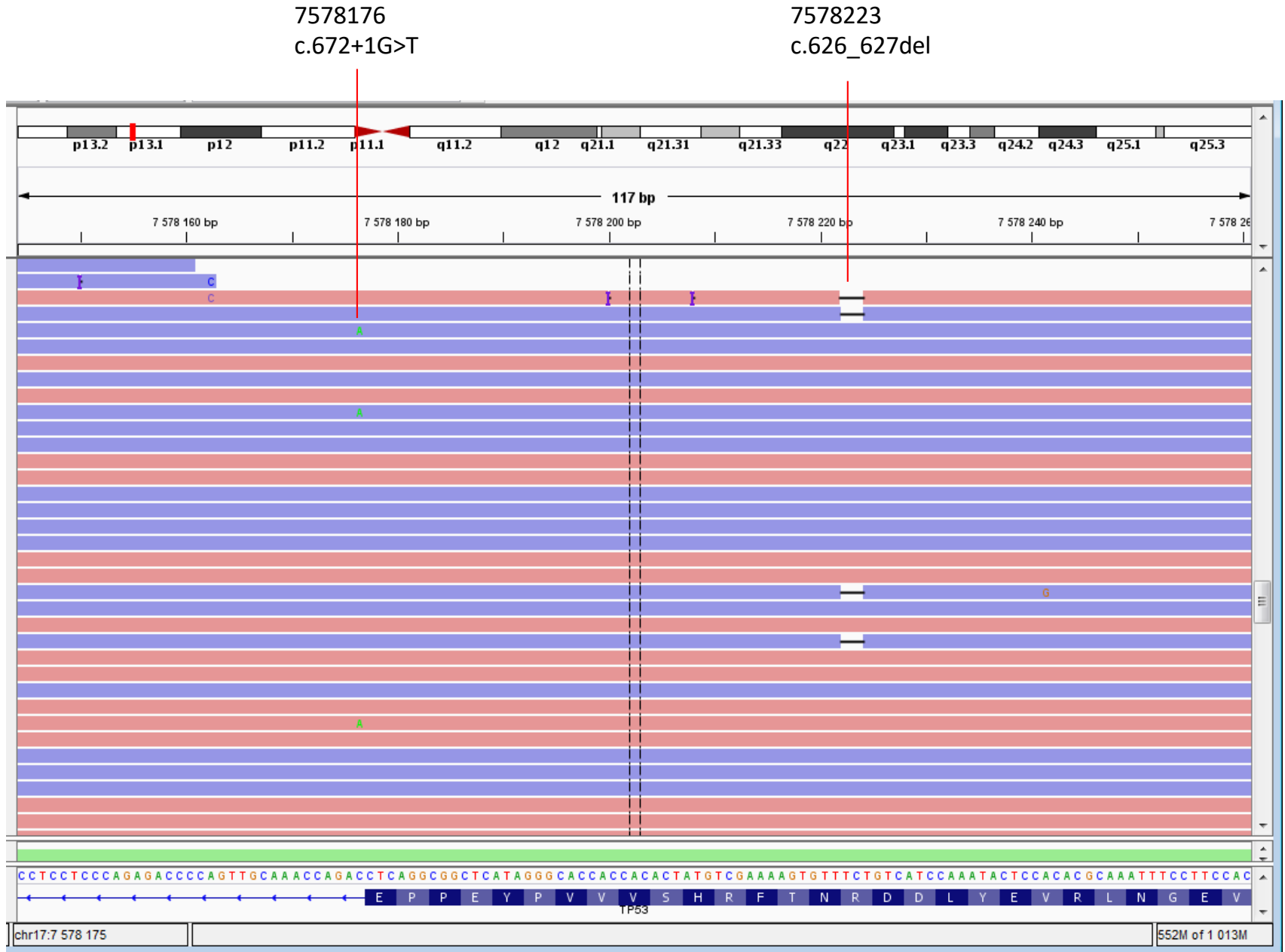
Supplementary Figure S8: CLL hotspot mutation in the splice acceptor signal of intron 6.

Distribution of mutations in the 20 splice signals of the 10 introns of the *TP53* gene for various types of cancer. Data are taken from the most recent release of UMD_TP53.

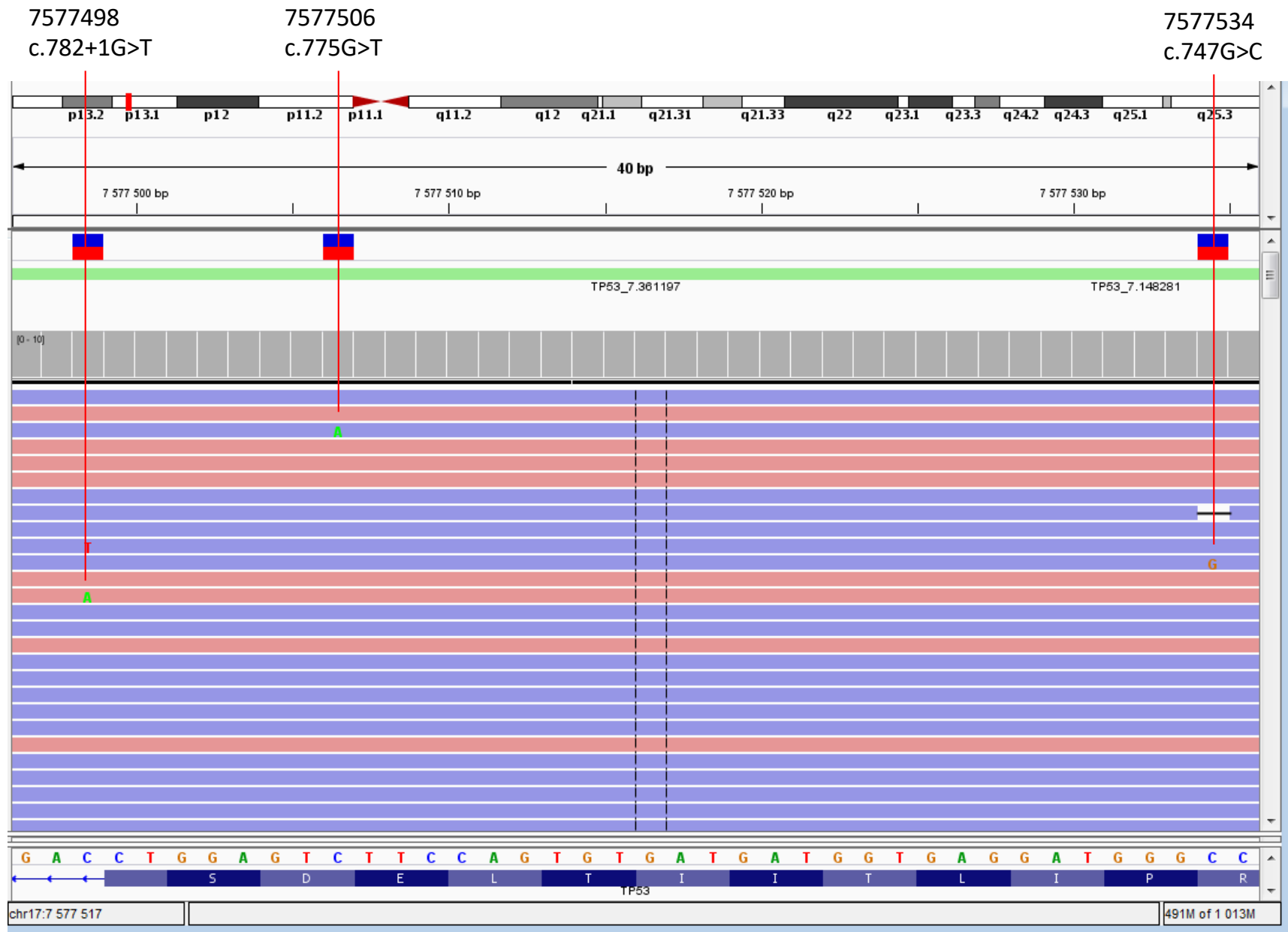


Supplementary Figure S9: Wide variant distribution in the number of *TP53* variants in CLL patients.

A: *TP53* cancer types were analyzed for tumors carrying two (DM), three (MM3) or more than three (MM4+) *TP53* variants. CLL tumors were split into two subgroups including either FILO or UMD_CLL patients. **B:** Number of mutations per patient. SM: patients with a single mutation; DM: patients with two *TP53* variants; MMn: multi-mutated patients with n mutations. **C and D:** Distribution of the number of mutations in tumors from the studies of Catherwood *et al.* and Bomben *et al.* {Bomben *et al.*, 2021, #169498 ; Catherwood *et al.*, 2022, #111726}.



AVC_62

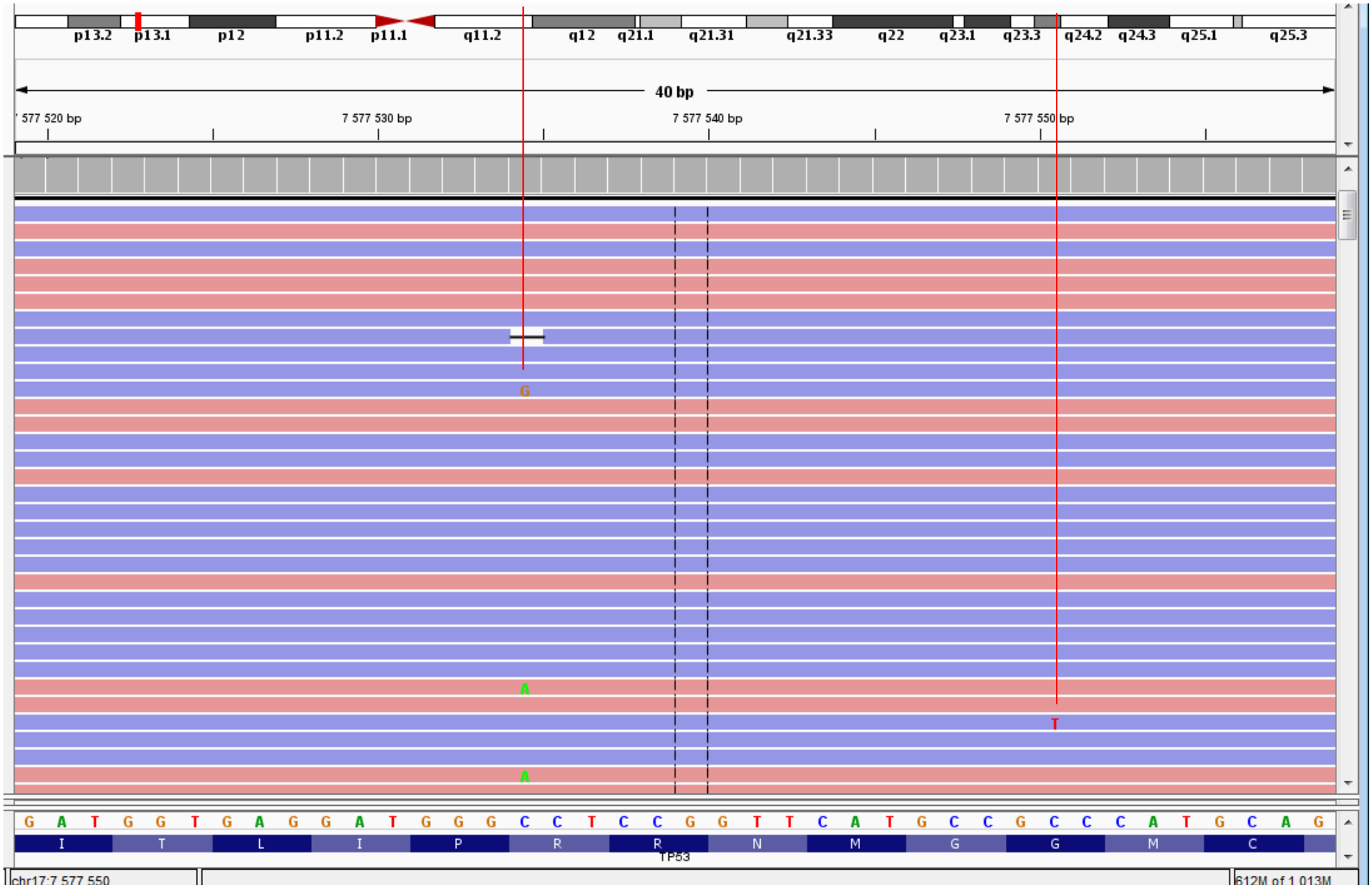


Supplementary Figure S10b

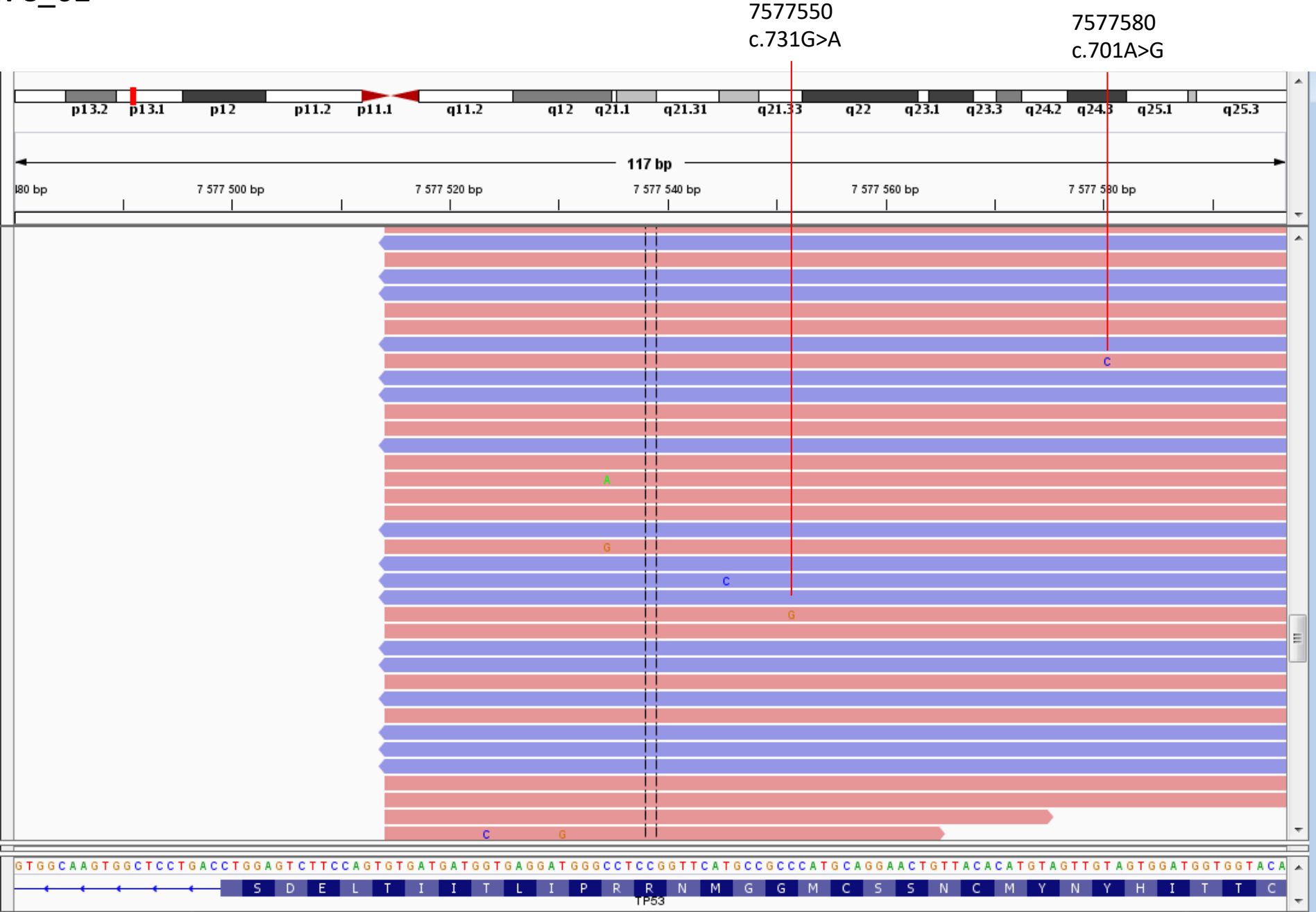
AVC_62

7577534
c.747G>C

7577550
c.731G>A



Supplementary Figure S10c

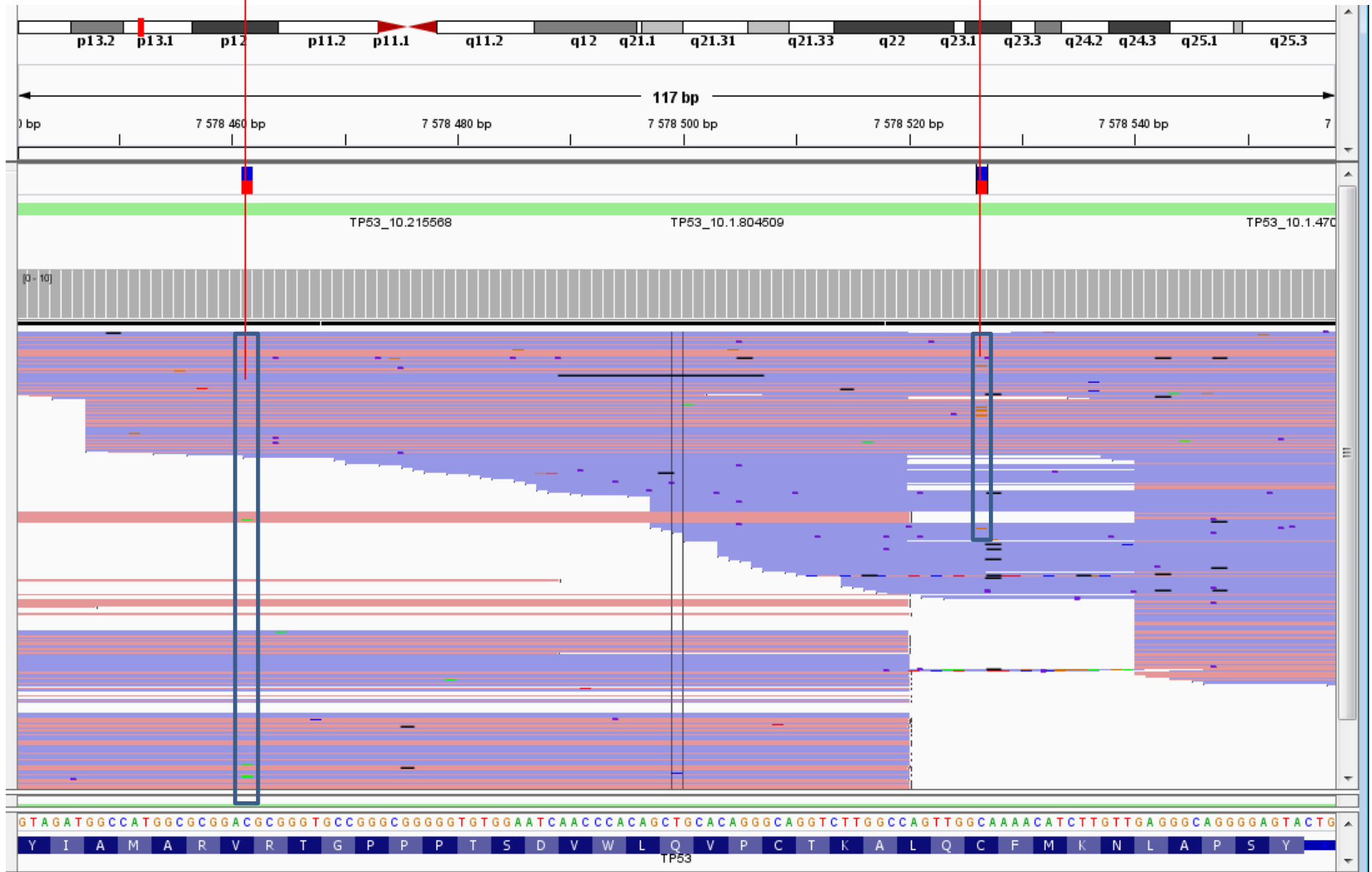


Supplementary Figure S10d

AVC_62

7578461
c.469G>T

7578526
c.404G>C

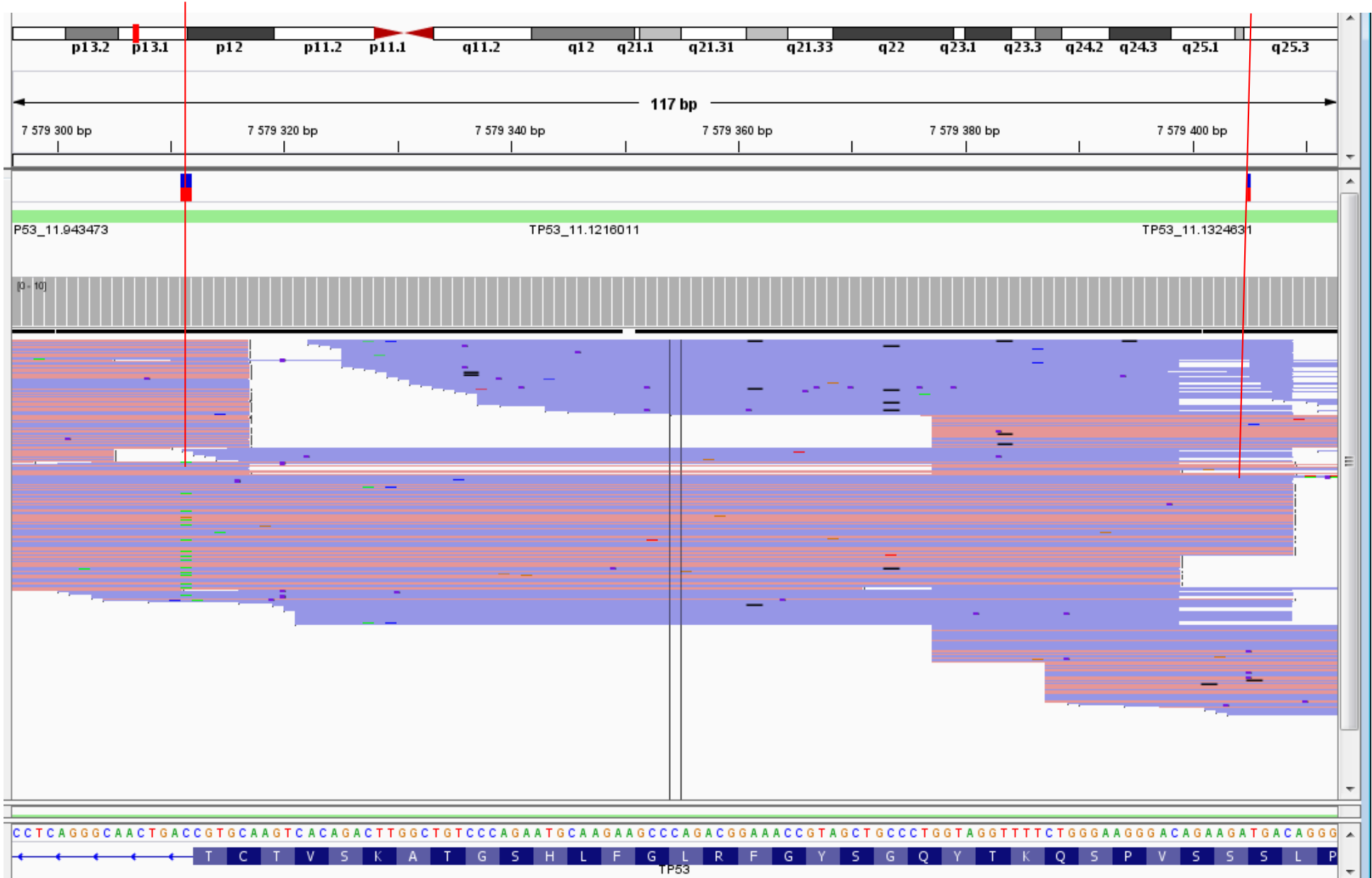


Supplementary Figure S10e

AVC_62

7579311
c.375+1G>T

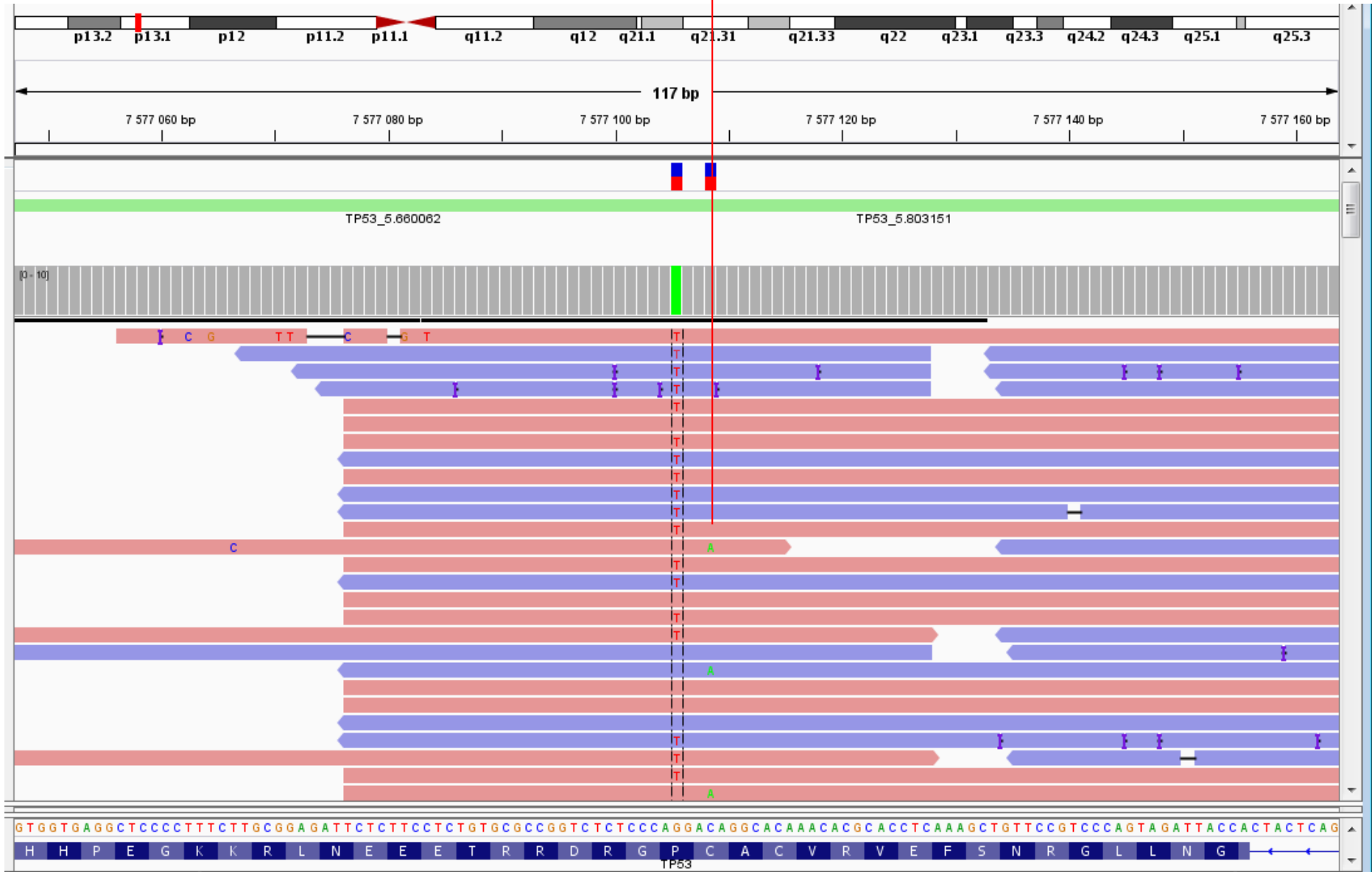
7579404
c.282_283ins



Supplementary Figure S10f

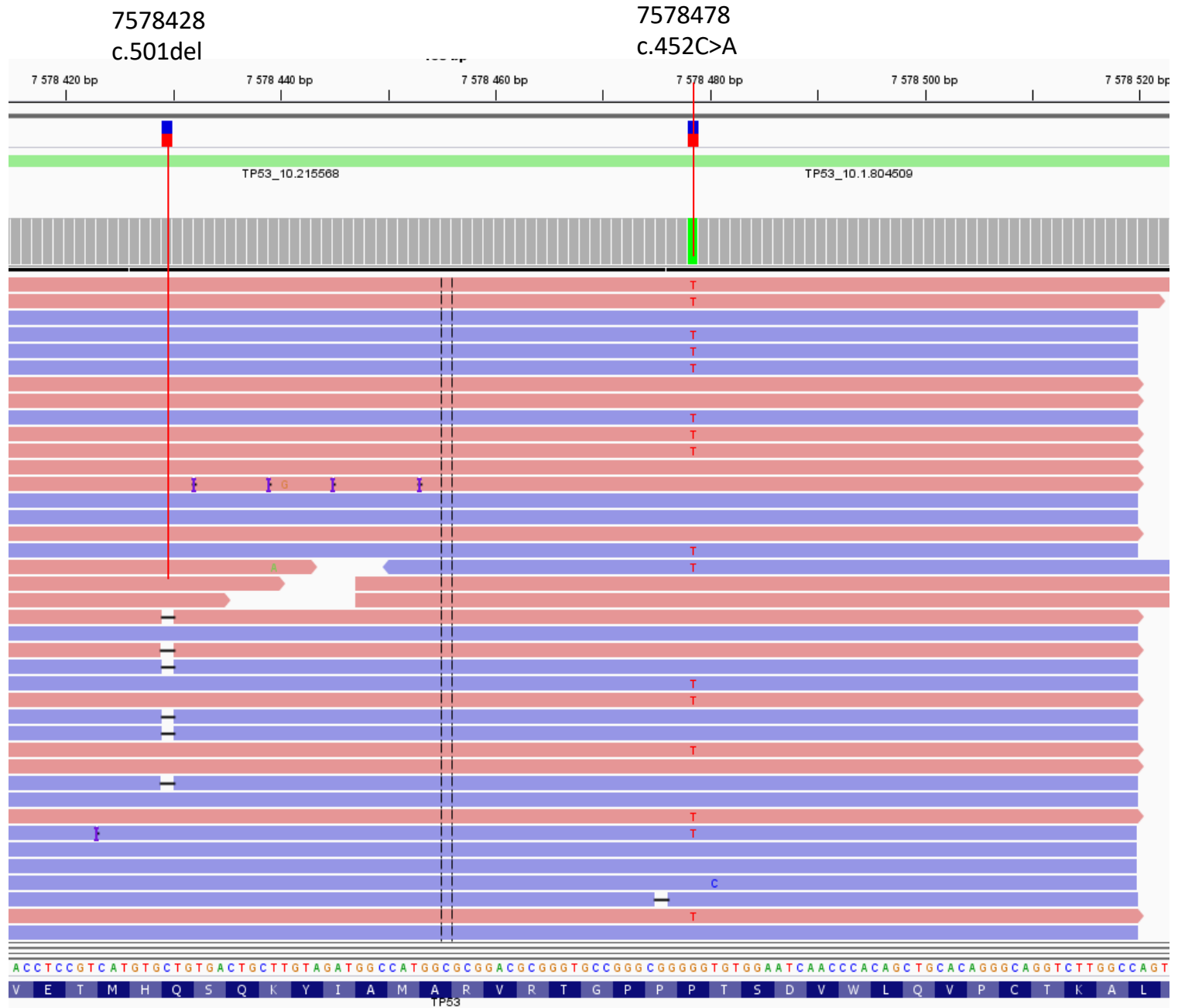
AVC_20

7577105 7577108
c.833C>A c.830G>T



Supplementary Figure S10g

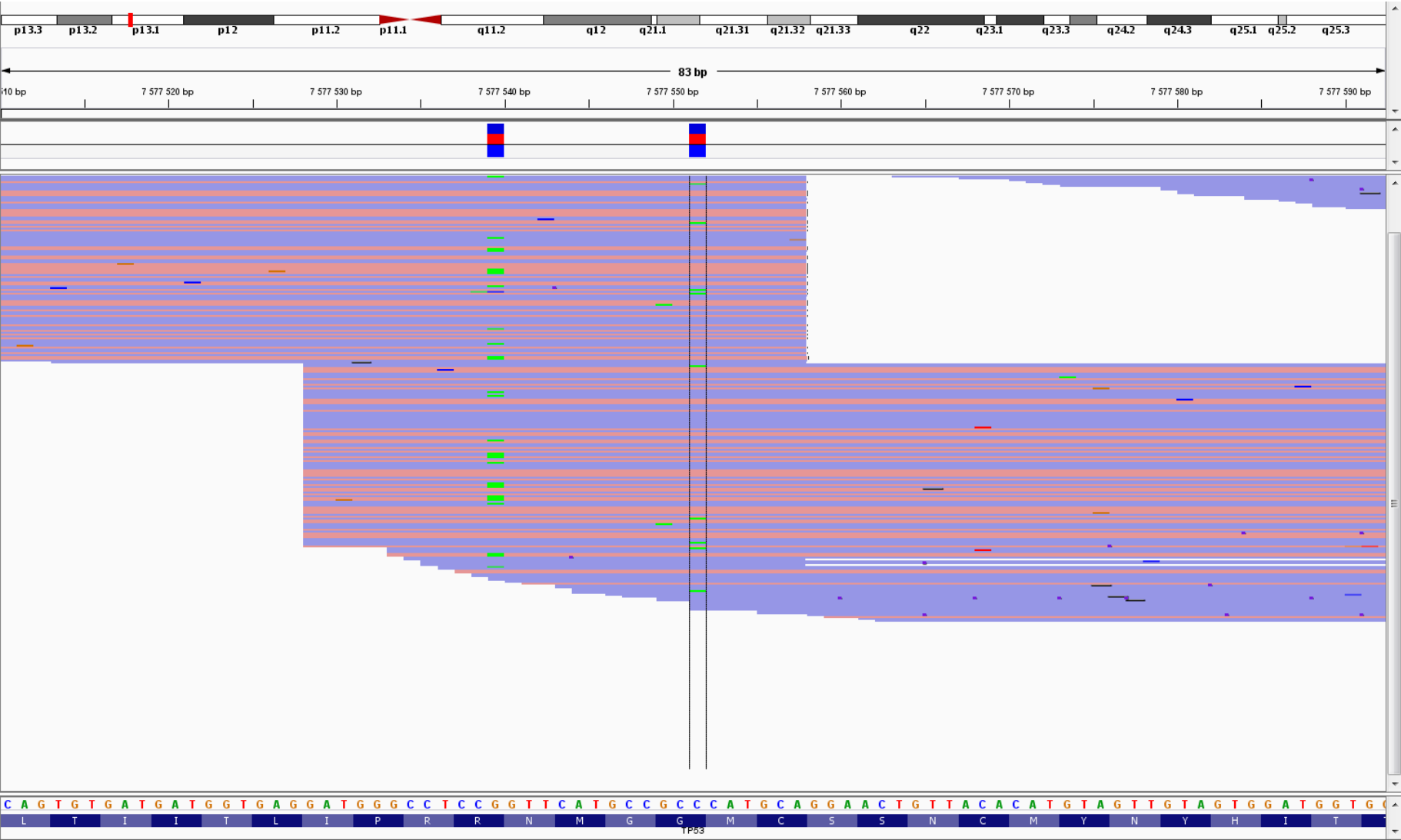
AVC_69



Supplementary Figure S10h

7577539
c.742C>T

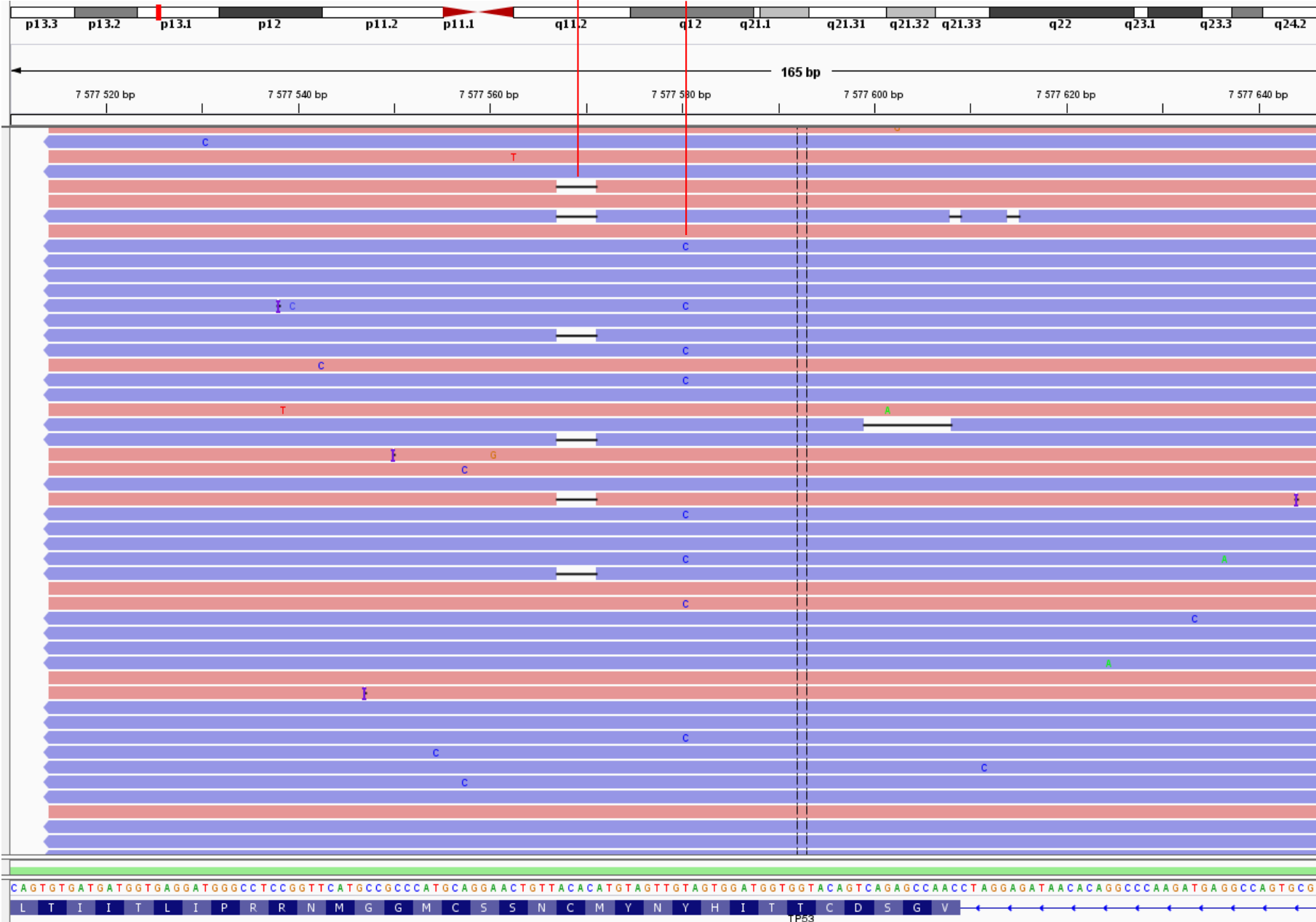
7577551
c.730G>T



Supplementary Figure S10i

AVC_47

7577570
c.711_714del 7577580
c.701A>G



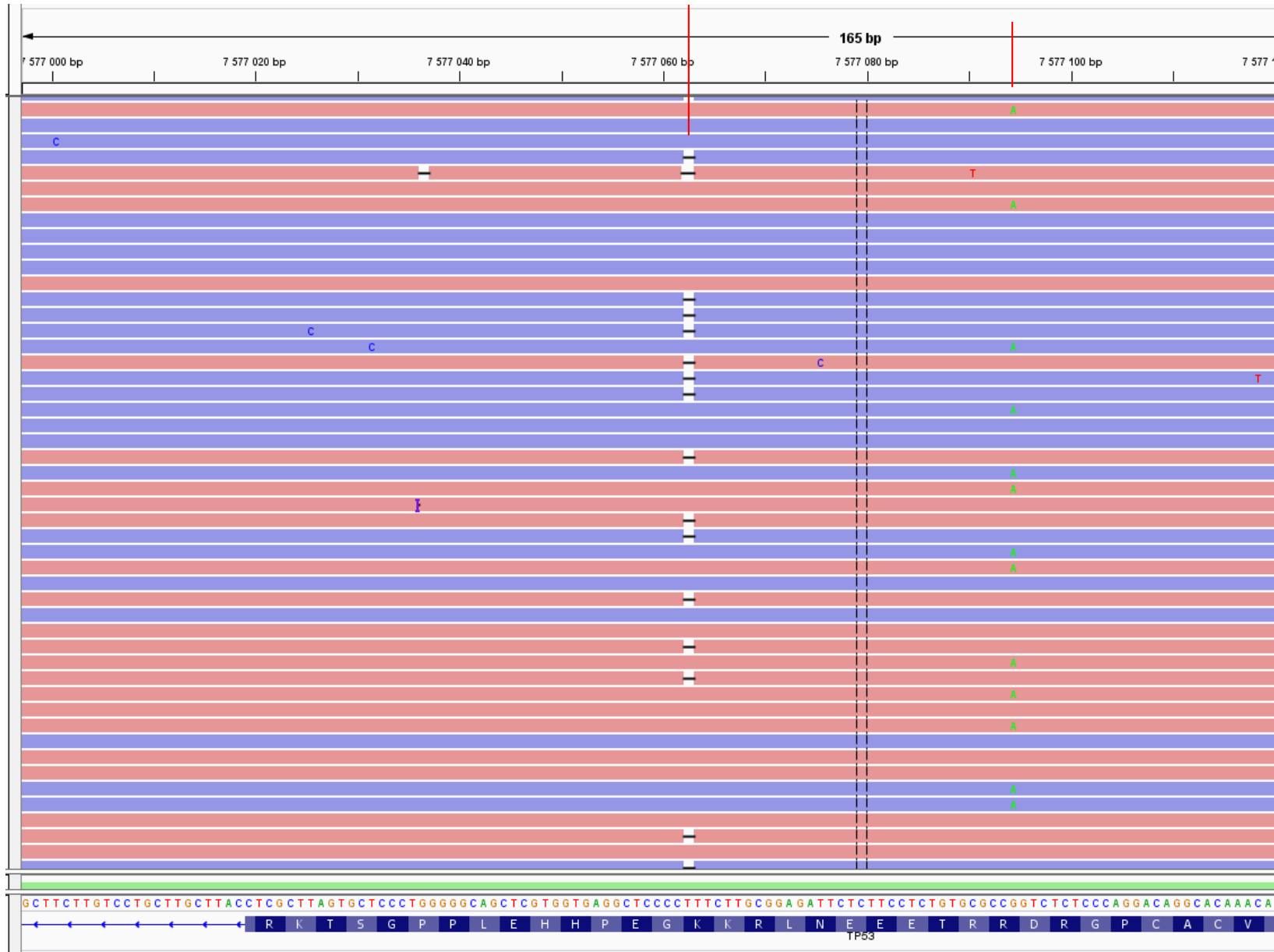
Supplementary Figure S10j

ANG_9

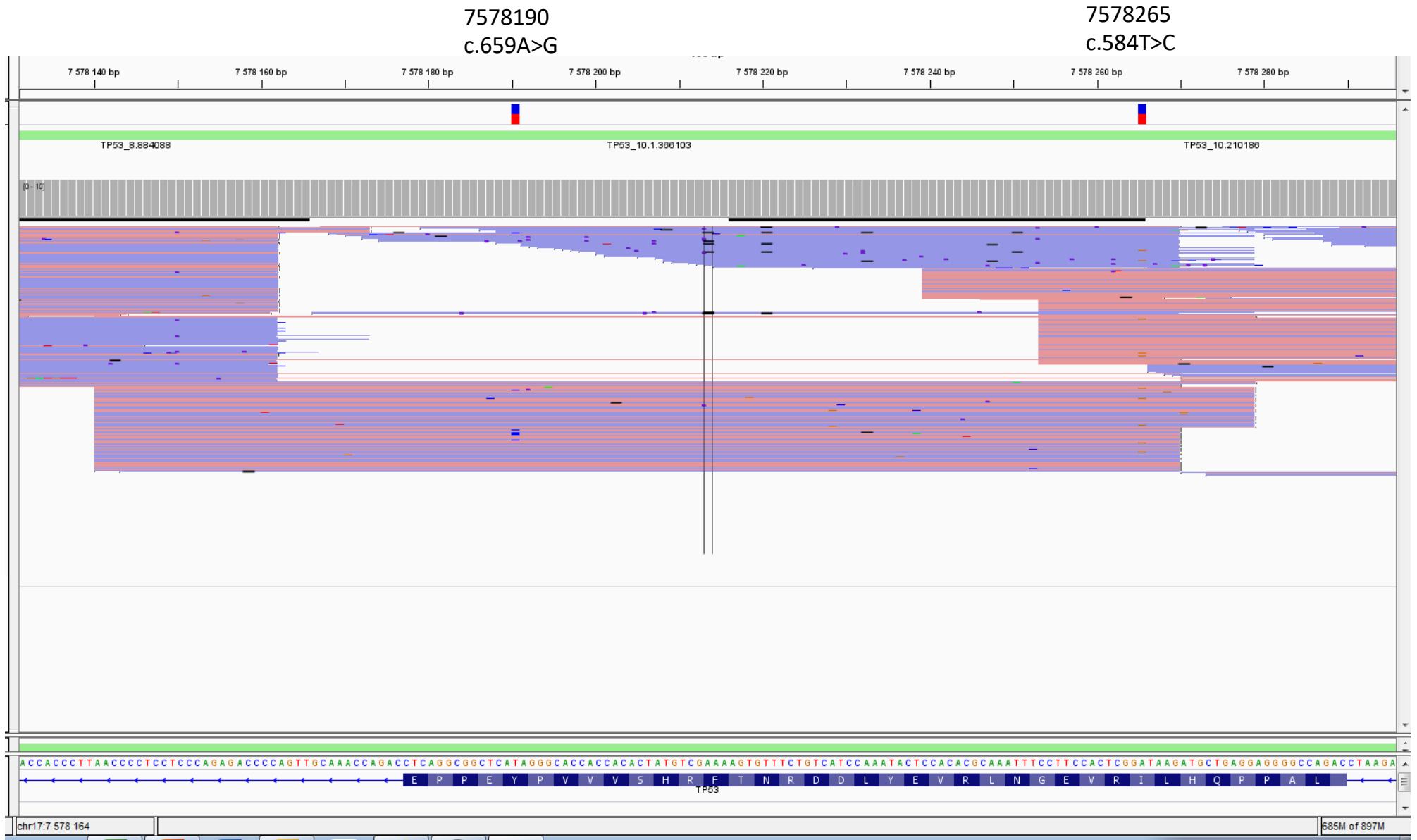
GUI NIC

7577057
c.876delA

7577094
c.844C>T

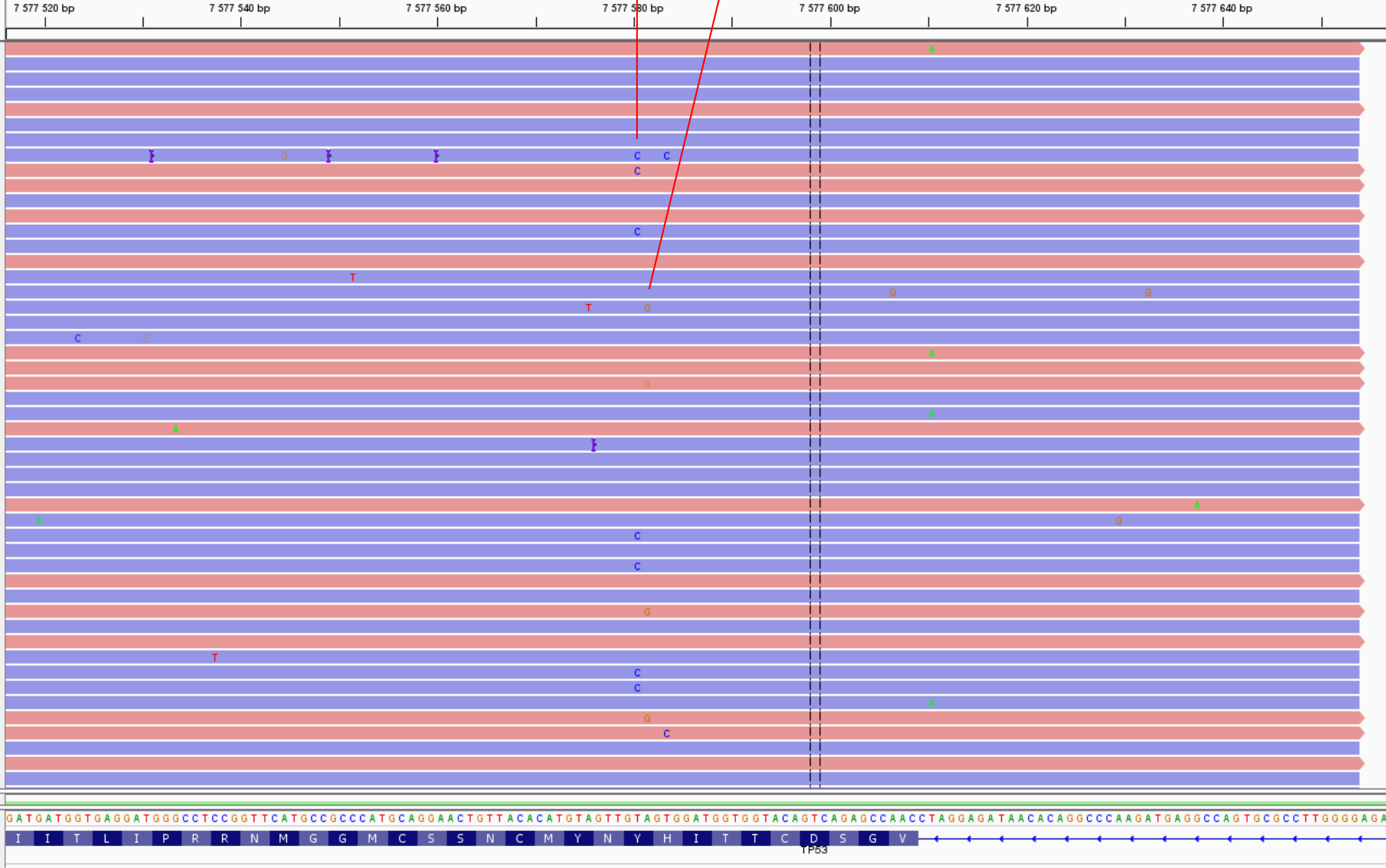


Supplementary Figure S10k



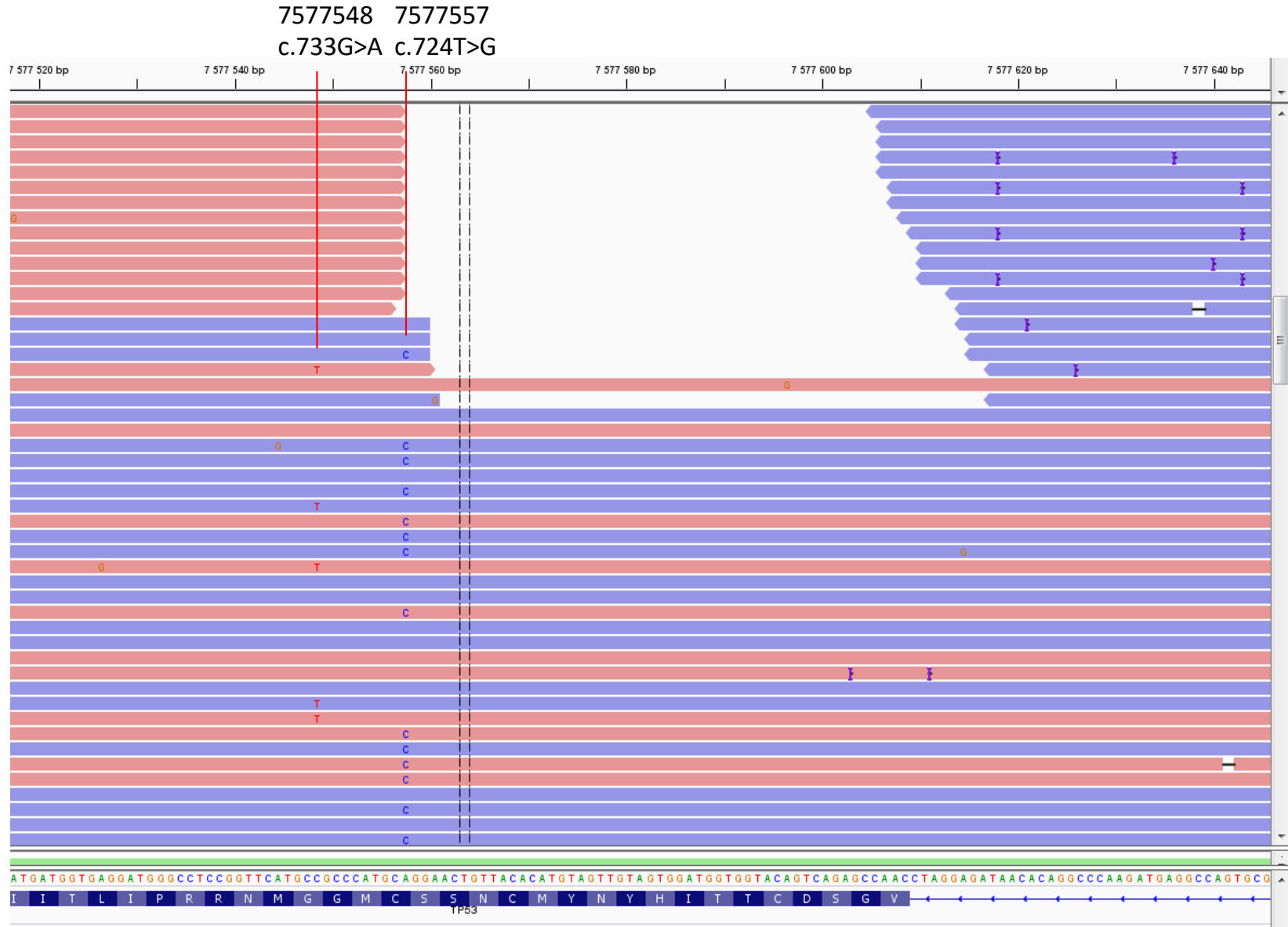
Supplementary Figure S10I

7577580 7577581
c.701A>G c.700T>C



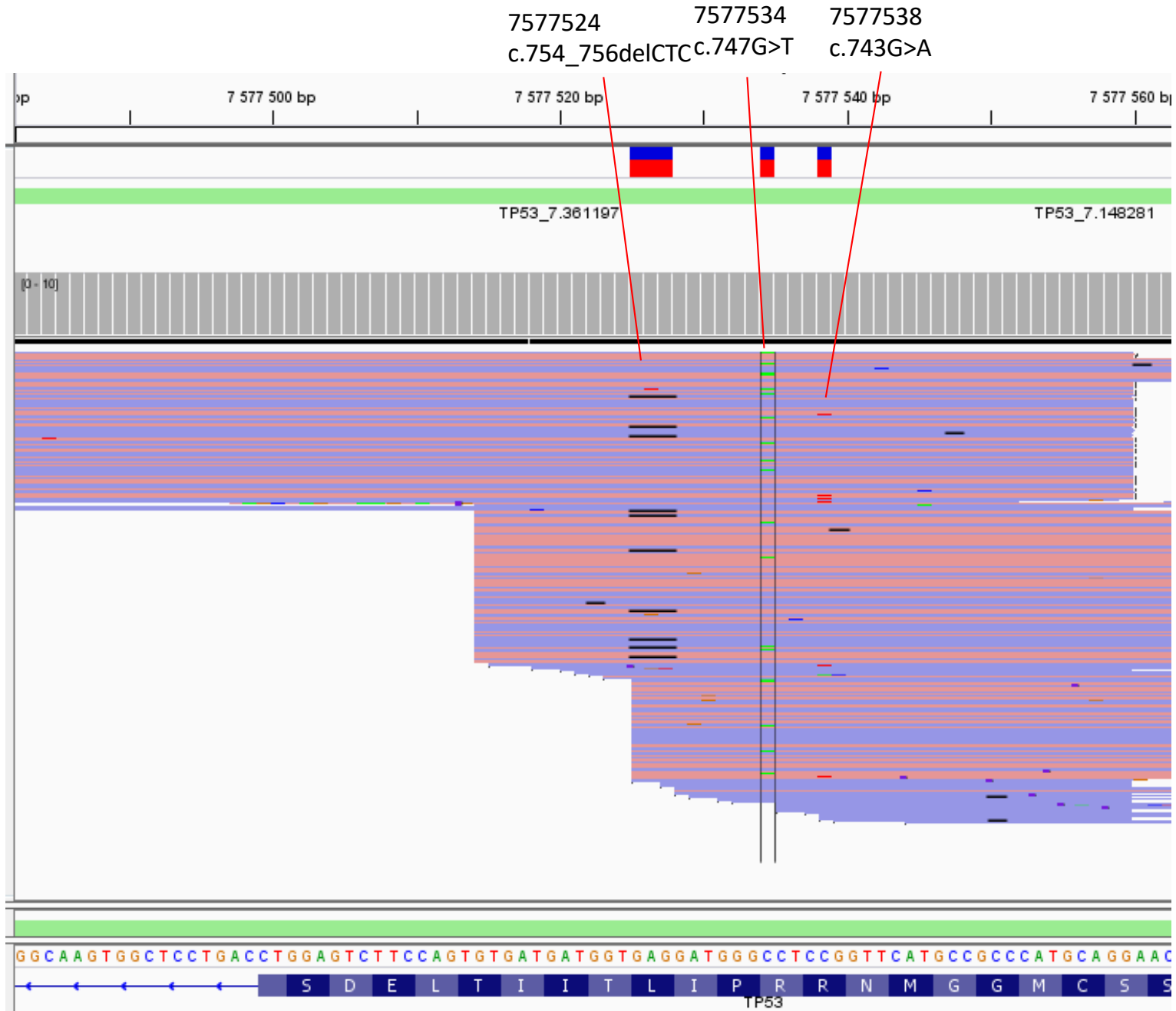
Supplementary Figure S10r

ANG_17



Supplementary Figure S10n

ANG_12



Supplementary Figure S10o

Supplementary Figure S10 to 10p: *TP53* mutations in FILO-cohort polymutated patients are in different alleles.

Each IGV plot depicts the phasing of patients with multiple *TP53* variants. Individual DNA sequencing reads harboring mutations within the same exon and no more than 50 nucleotides apart were analyzed.

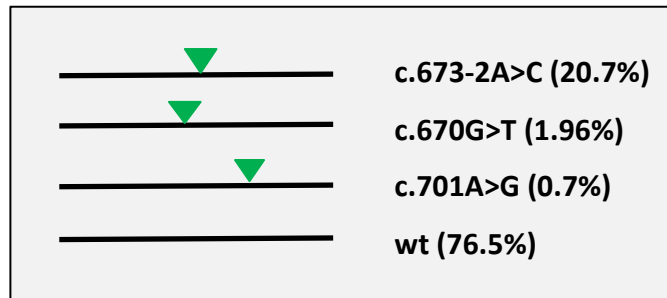
Patient Sw1

Sequencing

Sanger/NGS (short reads)

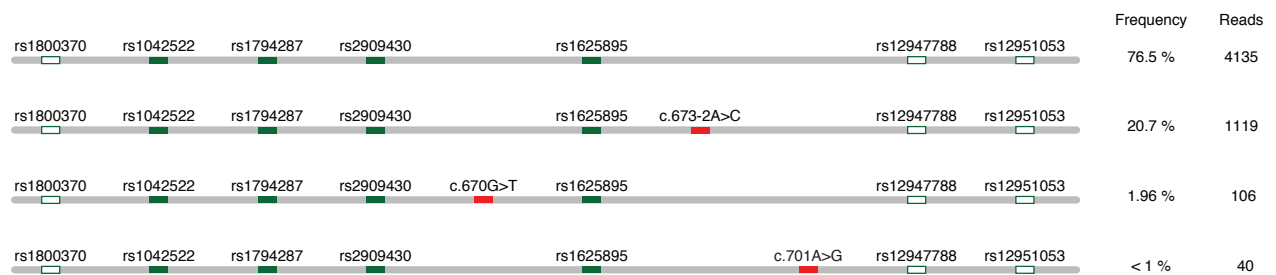
- ▼ c.919+1G>C (32.2%)
- ▽ c.673-2A>C (22.6%)
- ▽ c.701A>G (13%)
- ▾ c.626_627del2 (5.7%)
- ▼ c.670G>T (0.5%)

SMRT sequencing (long reads)



- ▼ Mutation outside the amplicon used for the long-range sequencing
- ▼ Mutation detected after manual examination but below the cut-off used for clinical validation (5%)
- ▽ Mutation detected only by the long-range sequencing
- ▾ Mutation no detected by long range sequencing

Haplotype



Supplementary Figure S11A

Patient Sw3

Sequencing

Sanger/NGS (short reads)

- ▼ c.626_627del (77.5%)
- ▼ c.379T>C (15.9%)

SMRT sequencing (long reads)

- ▼ c.626_627del (84.1%)
- ▼ c.379T>C (6.55%)
- wt (9.27%)

- ▼ Mutation outside the amplicon used for the long-range sequencing
- ▼ Mutation detected after manual examination but below the cut-off used for clinical validation (5%)
- ▽ Mutation detected only by the long-range sequencing

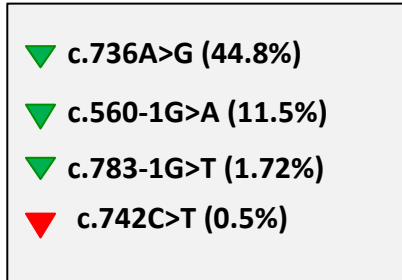
Haplotype

	rs1800370	rs1042522	rs1794287	rs2909430	c.626_627del2	rs1625895	rs12947788	rs12951053	Frequency	Reads
	□	■	■	■	■	■	□	□	84.1 %	9390
	□	■	■	■		■	□	□	9.27 %	1035
	□	■	■	■	■	■	□	□	6.55 %	731

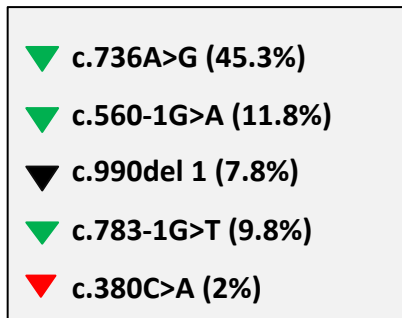
Supplementary Figure S11B

Sequencing

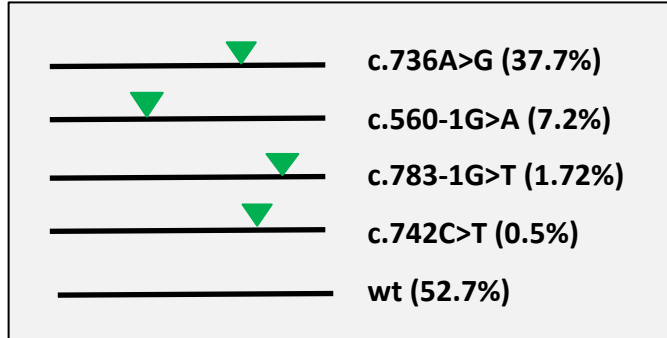
Sanger/NGS (short reads)



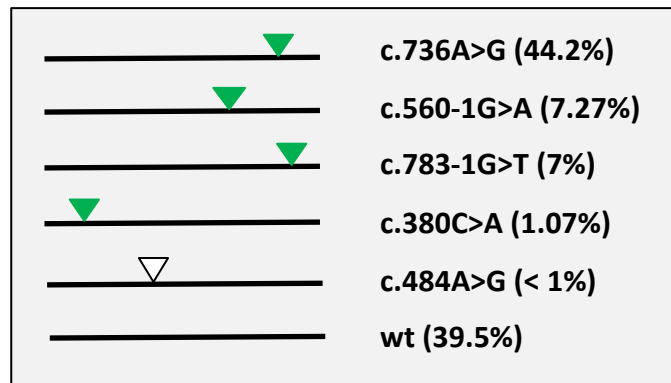
3 years



SMRT sequencing (long reads)



3 years

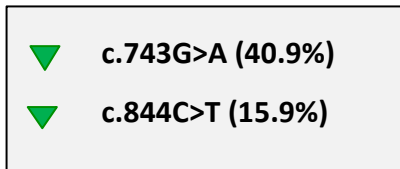


- ▼ Mutation outside the amplicon used for the long-range sequencing
- ▼ Mutation detected after manual examination but below the cut-off used for clinical validation (5%)
- ▽ Mutation detected only by the long-range sequencing

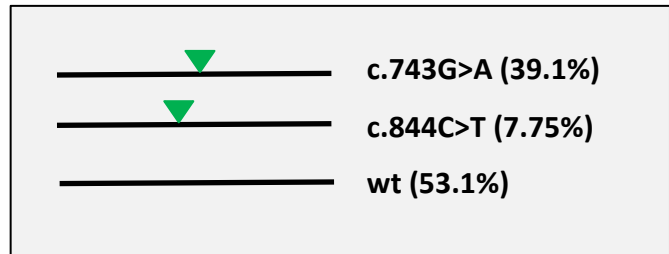
Patient Sw4

Sequencing

Sanger/NGS (short reads)

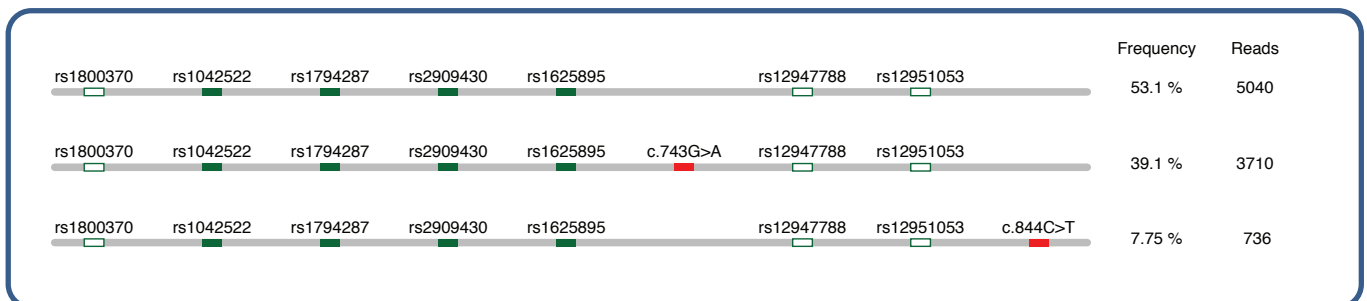


SMRT sequencing (long reads)



- ▼ Mutation outside the amplicon used for the long-range sequencing
- ▼ Mutation detected after manual examination but below the cut-off used for clinical validation (5%)
- ▽ Mutation detected only by the long-range sequencing

Haplotype

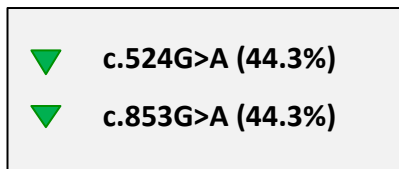


Supplementary Figure S11D

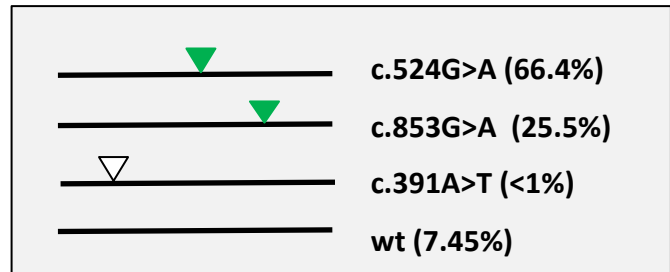
Patient Sw5

Sequencing

Sanger/NGS (short reads)



SMRT sequencing (long reads)



- ▼ Mutation outside the amplicon used for the long-range sequencing
- ▼ Mutation detected after manual examination but below the cut-off used for clinical validation (5%)
- ▽ Mutation detected only by the long-range sequencing

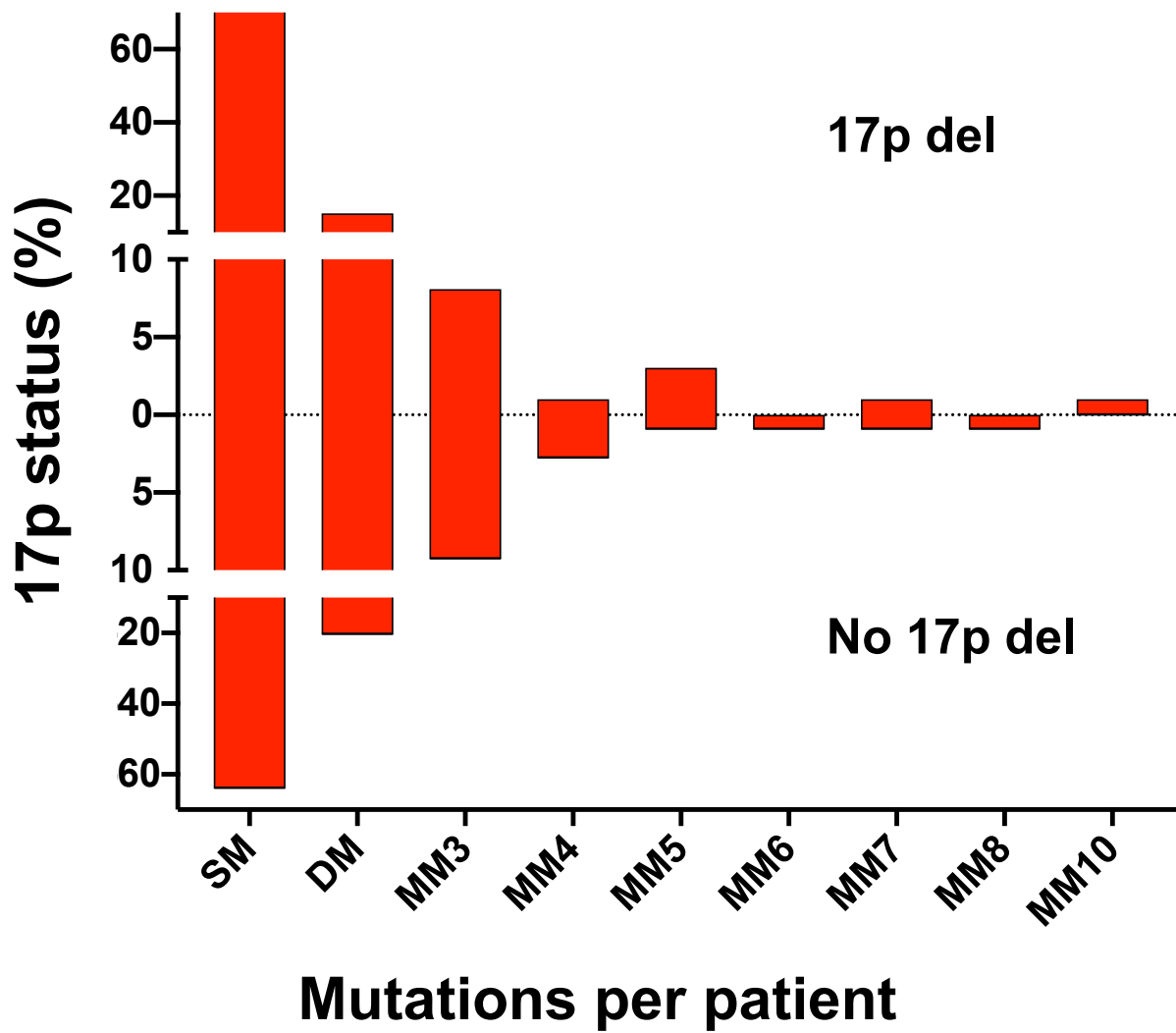
Haplotype

rs1800370	rs1042522	rs1794287	rs2909430	c.524G>A	rs1625895	rs12947788	rs12951053	Frequency	Reads	
rs1800370	rs1042522	rs1794287	rs2909430	c.524G>A	rs1625895	rs12947788	rs12951053	66.4 %	1088	
rs1800370	rs1042522	rs1794287	rs2909430		rs1625895	rs12947788	rs12951053	c.853G>A	25.5 %	419
rs1800370	rs1042522	rs1794287	rs2909430		rs1625895	rs12947788	rs12951053	7.45 %	122	
rs1800370	rs1042522	rs1794287	rs2909430	c.391A>T	rs1625895	rs12947788	rs12951053	< 1 %	8	

Supplementary Figure S11E

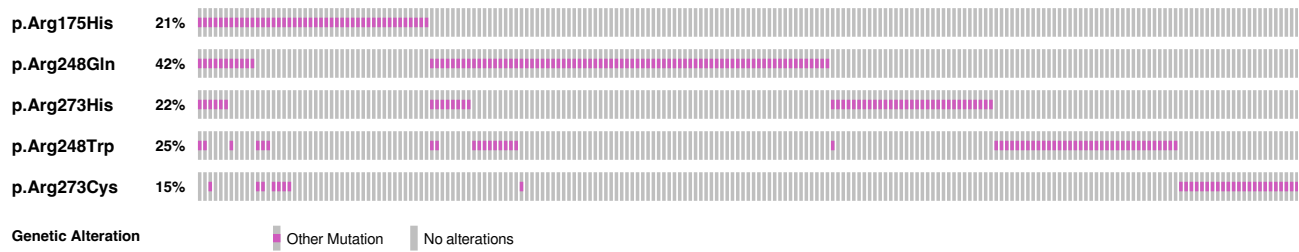
Supplementary figure S11 A to E: Long-range sequencing of CLL patients shows that *TP53* mutations in multi-mutated patients are in different alleles.

A to E: Detailed analysis of the 5 patients included in this study. For each patient, 2 sections are available i.e. sequencing and haplotype. Sequencing: conventional NGS analysis is shown on the left. No allelic distribution can be inferred from this type of analysis. SMRT sequencing is shown on the right. It provides an accurate picture of the allelic distribution of each *TP53* variant, as well as the remaining wt allele {Lodé et al., 2018, #147689}. The frequencies of the different alleles are shown in parentheses. 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 (SNPs) are shown in white (allele 1) and green (allele 2) to make a distinction for heterozygote cases.



Supplementary Figure S12: Multiple *TP53* variants are found at the same frequency in patients with or without 17p deletion.

SM: patients with a single *TP53* mutation; DM: patients with two *TP53* mutations; MM3 to MM10: patients with three to ten *TP53* mutations.



Supplementary Figure S13: Distribution of *TP53* hotspot variants in CLL patients.

CLL patients included in both UMD_CLL and the FILO cohort were used for this analysis. Data were analyzed by OncoPrinter (cBioPortal Version 1.14.0) {Gao et al., 2013, #80688}.

Supplementary Table S1: Patient characteristics.

Number of patients	n=683
Median age (years), [min;max]	70 [23;98]
Cytogenetics, n (%)	
Trisomy 12	22/181 (12%)
Deletion 17p	127/257 (49%)
Deletion 11q	24/175 (14%)
Complex karyotype	62/169 (36%)
IGHV status, n (%)	
Mutated	204/464 (44%)
Unmutated	260/464 (56%)

Supplementary Table S2: Mutations summary.

	Number of variants (%)		
	Conventional sequencing	NGS	Total
Missense	136 (69.4)	632 (73.5)	768 (72.7)
Nonsense	17 (8.7)	52 (6)	69 (6.5)
Synonymous	1 (0.5)	0 (0)	1 (0.1)
Splice site	13 (6.6)	69 (8)	82 (7.8)
Intronic	4 (2)	0 (0)	4 (0.4)
Frameshift	19 (9.7)	93 (10.8)	112 (10.6)
Inframe	6 (3.1)	14 (1.6)	20 (1.9)
Total	196 (100)	860 (100)	1056 (100)
	Number of patients (%)		
Variants per patients	Conventional sequencing	NGS	Total
1 variant	150 (87.2)	351 (68.7)	501 (73.4)
2 variants	20 (11.6)	81 (15.9)	101 (14.8)
3 variants	2 (1.2)	41 (8)	43 (6.3)
4 and + variants	0 (0)	38 (7.4)	38 (5.6)
Total	172 (100)	511 (100)	683 (100)

Supplementary Table S3: Mutation description.

(Provided as an Excel file)

Table S4: Benign polymorphisms identified in CLL patients included in UMD.

HG19_Variant	Variant_Type	Event_type	Mutation_effect	Protein P1 TP53_alpha NP_000537.3	LRG_321t1	occurrence in UMD_CLL	ACMG_class
chr17:g.7577091G>A	SNV	G>A	Missense_variant	p.Arg283Cys	c.847C>T	16	Benign
chr17:g.7577069C>T	SNV	C>T	Missense_variant	p.Arg290His	c.869G>A	9	Benign
chr17:g.7577577T>C	SNV	T>C	Missense_variant	p.Asn235Ser	c.704A>G	8	Benign
chr17:g.7578464G>A	SNV	G>A	Missense_variant	p.Arg156Cys	c.466C>T	3	Benign
chr17:g.7579705C>T	SNV	C>T	Missense_variant	p.Val31Ile	c.91G>A	3	Benign
chr17:g.7579358C>T	SNV	C>T	Missense_variant	p.Arg110His	c.329G>A	2	Benign
chr17:g.7579470C>T	SNV	C>T	Missense_variant	p.Val73Met	c.217G>A	1	Benign
chr17:g.7579439G>A	SNV	G>A	Missense_variant	p.Ala83Val	c.248C>T	1	Benign

All these variants were defined as rare single nucleotide polymorphisms (SNPs) in the human population; they do not display any loss of function {Doffe et al., 2021, #89023; Soussi, 2022, #253203}. These variants were removed from UMD_CLL for all analyses described in the present study.