The comprehensive landscape of *TTMV::RARA* fusion-driven acute myeloid leukemia: from viral integration mechanisms to clinical outcomes

Shu Sun,[1]* Yongjing Liu,[2]* Qing-Yu Xu,[1]* Jiayu Wang,[3] Li Chen,[4] Zhanrui Cheng,[3] Wei Gao,[5] Huili Wang,[6] Bei Yang,[3] Haiyan Wang,[7] Lijun Wen,[8] Jian Xiao,[9] Jiacheng Lou,[4] Haibo Yu,[1] Na Li,[3] Feng Wang,[10] Yangyang Xie,[11] JinGang Wang,[7] Xianjing Wang,[6] Hongjuan Xue,[5] Kun Chen,[9] Yin Wu,[1] Leping Zhang,[12] Ke Li,[10] Shuhong Shen,[11] Suning Chen,[8] Huan-You Wang,[13] KanKan Wang,[4] Jinyan Huang[2] and Hong-Hu Zhu[1,3]

[1]Department of Hematology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China; [2]Biomedical Big Data Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; [3]Institute for Cancer, Chinese Institutes for Medical Research (CIMR), Beijing, China; [4]Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; [5]Pediatric Hematology and Oncology of Xingtai People's Hospital, Hebei, China; [6]Department of Hematology, The Third People's Hospital of Zhengzhou, Zhengzhou, China; [7]Department of Hematology, The First College of Clinical Medicine Science, China Three Gorges University affiliated Yichang Central People's Hospital, Yichang, Hubei, China; [8]National Clinical Research Center for Hematologic Diseases, The First Affiliated Hospital of Soochow University, Jiangsu Institute of Hematology, Institute of Blood and Marrow Transplantation, Collaborative Innovation Center of Hematology, Soochow University, Suzhou, China; [9]Department of Hematology, Zigong First People's Hospital, Zigong, China; [10]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, NHC Key Laboratory of Biotechnology of Antibiotics, Institute of Medicinal Biotechnology, Chinese Academy of Medical, Beijing, China; [11]Department of Hematology, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; [12]Department of Pediatrics, Peking University People's Hospital, Beijing, China and [13]Department of Pathology, University of California San Diego, La Jolla, CA, USA

SS, YL and QYX contributed equally as first authors.

Supplementary Appendix

Title：The comprehensive landscape of *TTMV::RARA* fusion-driven acute myeloid leukemia: from viral integration mechanisms to clinical outcomes

Shu Sun[1#], Yongjing Liu[2#], Qing-Yu Xu[1#], Jiayu Wang[3], Li Chen[4], Zhanrui Cheng[3], Wei Gao[5], Huili Wang[6], Bei Yang[3], Haiyan Wang[7], Lijun Wen[8], Jian Xiao[9], Jiacheng Lou[4], Haibo Yu[1], Na Li[3], Feng Wang[10], Yangyang Xie[11], JinGang Wang[7], Xianjing Wang[6], Hongjuan Xue[5], Kun Chen[9], Yin Wu[1], Leping Zhang[12], Ke Li[10], Shuhong Shen[11], Suning Chen[8], Huan-You Wang[13], KanKan Wang[4], Jinyan Huang[2]*, Hong-Hu Zhu[1,3]*

Part 1: Supplementary Methods

Part 2: Supplementary Figures and legends

Supplementary Fig. S1. Modular pipeline for the identification and analysis of virus - host integration events used in this study.

Supplementary Fig. S2. Illustration of TTMV genomes and integration breakpoints.

Supplementary Fig. S3. The misalignment of the 3' and 5' breakpoints when TTMV integrates into the host genome.

Supplementary Fig. S4. Multiple sequence alignment of TTMV insertion sequences from the 15 samples.

Supplementary Fig. S5. Gene Set Enrichment Analysis.

Supplementary Fig. S6. Treatment outcomes of ATRA+ATO versus other therapies in AML with *TTMV::RARA*.

Part 3: Supplementary Table

Supplementary Table S1: Sources of the 2,543 Public Database Samples.

Supplementary Table S2: Assembled sequences of *TTMV::RARA* transcripts from 15 patients.

Supplementary Table S3: Clinical and molecular characteristics of AML, APL and *TTMV::RARA* samples used for differential gene expression.

Supplementary Methods

RNA Sequencing and Variant Analysis

RNA sequencing (RNA-seq) data of leukemia samples were collected through public databases, projects or publications. Samples with ambiguous lineage or mixed/obscure phenotypes were classified into AML, B-ALL, and T-ALL subtypes based on unsupervised clustering of gene expression profiles.

Data pre-processing for RNA-seq primarily followed the GATK Best Practices pipeline[1]. Raw RNA-seq sequences were aligned to the reference genome using STAR (v2.7.10a)[2] for bam file generation. For WGS, BWA[3] was used for alignment (v0.7.18) while GATK (v4.6.0)[4] was applied to mark PCR duplicated reads, and perform base quality score recalibration for mutation analysis. Variant calling in leukemia samples was performed using HaplotypeCaller, VarScan (v2.4.4)[5], and SpeedSeq (v0.1.2)[6]. Variants identified by WGS and RNA-seq were combined for the same patient, and non-synonymous variants with a variant allele frequency (VAF) of at least 5% were retained. Gene fusions were identified using STAR-Fusion (v1.11.0), Arriba (v2.4.0)[7], and FusionCatcher (v1.33) from RNA-seq data. For both mutation and fusion, only events detected by at least two tools were retained for further analysis. Arm-level copy number alterations were estimated using RNAseqCNV[8], and FLT3-ITD events were identified using FiLT3r[9] for RNA-seq samples.

TTMV Integration Detection and Annotation

Sequences of 6,806 TTMV genomes were collected from NCBI Virus and were used in this study. The latest release of the GRCh38 genome assembly was used as the host reference in this study. To achieve higher sensitivity, minimap2 (v2.2.6)[10] was employed in single-end read

mode for reads mapping analysis. The generated PAF files were used to perform an initial selection of sequence alignment. Subsequently, viral-human sequence fusion events were identified using an in-house script inspired by the detectIS.pl script from the detectIS package[11]. Potential chimeric reads were selected with following criteria: read alignment had more than one mapping blocks; alignment length was less than 90% of the read length; and the number of matching bases was at least 95% of the alignment length. For each sample, reads not properly aligning with the human genome were selected and further filtered by mapping to the TTMV genomes. The virus and human genome mapping results were integrated, and viral-human sequence integration sites were identified by matching alignment information from the same read, with a 3-base tolerance window. For each detected integration site, the corresponding reads were separated into viral and host regions. Reads with a mean quality lower than 90% of the overall sequencing quality in either single region was discarded. Integration sites identified in only one read were removed. Events with the same integration site were then combined.

For each identified TTMV-host integration event, supporting reads were selected and divided into two parts, resembling the viral sequence and the host sequence, respectively. An in-house function utilizing the dustyScore function in the ShortRead R package[12] was used to remove low-quality sequences for both viral and host parts. A normalized dust score for the sequences from both sides was calculated, and integration events with the score greater than 0.4 on either side were excluded. Highly repetitive sequences present in over 10% of the samples among samples from either host or virus genomes were also excluded. Shared sequence between TTMV genomes and some artificial genomes were found in the preliminary tests, therefore, the reads perfectly matching FOSMID/BAC clones or other artificial vectors were excluded as a

last filter, resulting in reliable genomic insertion events. All integration events were double-checked by IGV[13] visualization of corresponding bam files. Functions from the GViz R package[14] were modified to illustrate mismatches (TTMV sequence) using the bam files.

For each sample with *TTMV::RARA* integration, all reads mapped to the TTMV genomes were assemblied by SPAdes (v4.0.0)[15] using the --rnaviral parameter. Subsequently, the blastn algorithm was applied on the resulting contigs to find the closest match. Initial insertion sites were verified by paired RNA-seq and WGS data. For insertions with micro-homology, single-nucleotide level sites were determined utilizing reads covering same TTMV sequences from initial mRNA and spliced mRNA. According to prior knowledge, longest 3 ORFs were predicted by systemPipeR[16] for the assigned TTMV genomes. The circlize R package[17] was used for visualization of TTMV genomes.

References

1. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43(1110):11.10.1-11.10.33.

2. Dobin A, Gingeras TR. Optimizing RNA-Seq Mapping with STAR. *Methods Mol Biol*. 2016;1415:245-262.

3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

4. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.

5. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576.

6. Chiang C, Layer RM, Faust GG, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods*. 2015;12(10):966-968.

7. Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res*. 2021;31(3):448-460.

8. Bařinka J, Hu Z, Wang L, et al. RNAseqCNV: analysis of large-scale copy number variations from RNA-seq data. *Leukemia*. 2022;36(6):1492-1498.

9. Boudry A, Darmon S, Duployez N, et al. Frugal alignment-free identification of FLT3-internal tandem duplications with FiLT3r. *BMC Bioinformatics*. 2022;23(1):448.

10. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572-4574.

11. Grassi L, Harris C, Zhu J, Hardman C, Hatton D. DetectIS: a pipeline to rapidly detect exogenous DNA integration sites using DNA or RNA paired-end sequencing data. *Bioinformatics*. 2021;37(22):4230-4232.

12. Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*. 2009;25(19):2607-2608.

13. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.

14. Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol*. 2016;1418:335-351.

15. Vasilinetc I, Prjibelski AD, Gurevich A, Korobeynikov A, Pevzner PA. Assembling short

reads from jumping libraries with large insert sizes. *Bioinformatics*. 2015;31(20):3262-3268.

16. H Backman TW, Girke T. systemPipeR: NGS workflow and report generation environment. *BMC Bioinformatics*. 2016;17:388.

17. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811-2812.

Supplementary Figure and legends

Fig. S1
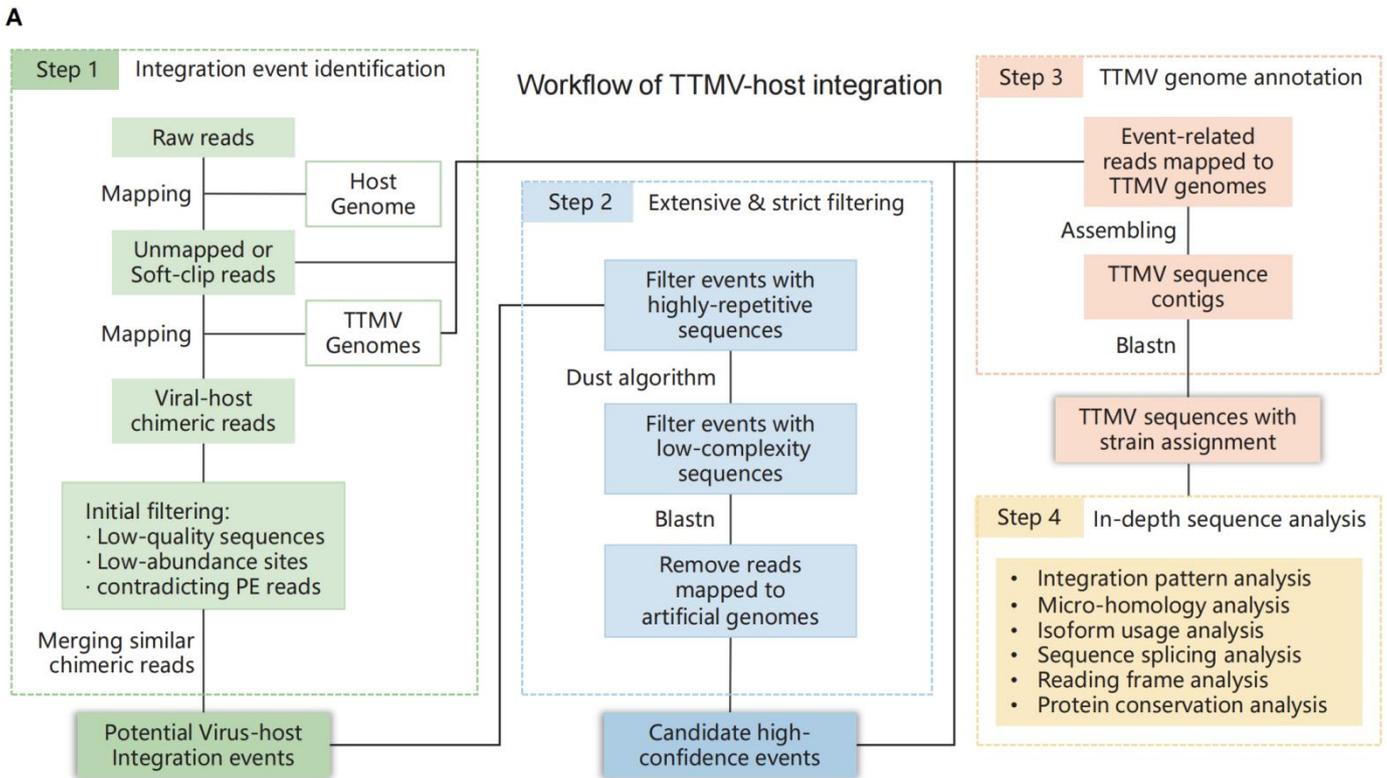
Fig. S1. Modular pipeline for the identification and analysis of virus - host integration events used in this study. A total of 2,553 samples were processed through Step 1 and Step 2 for initial filtering, and the remaining candidate events were subjected to further analysis in Step 3 and Step 4. Detailed methodology is provided in the Methods section.

Fig. S2
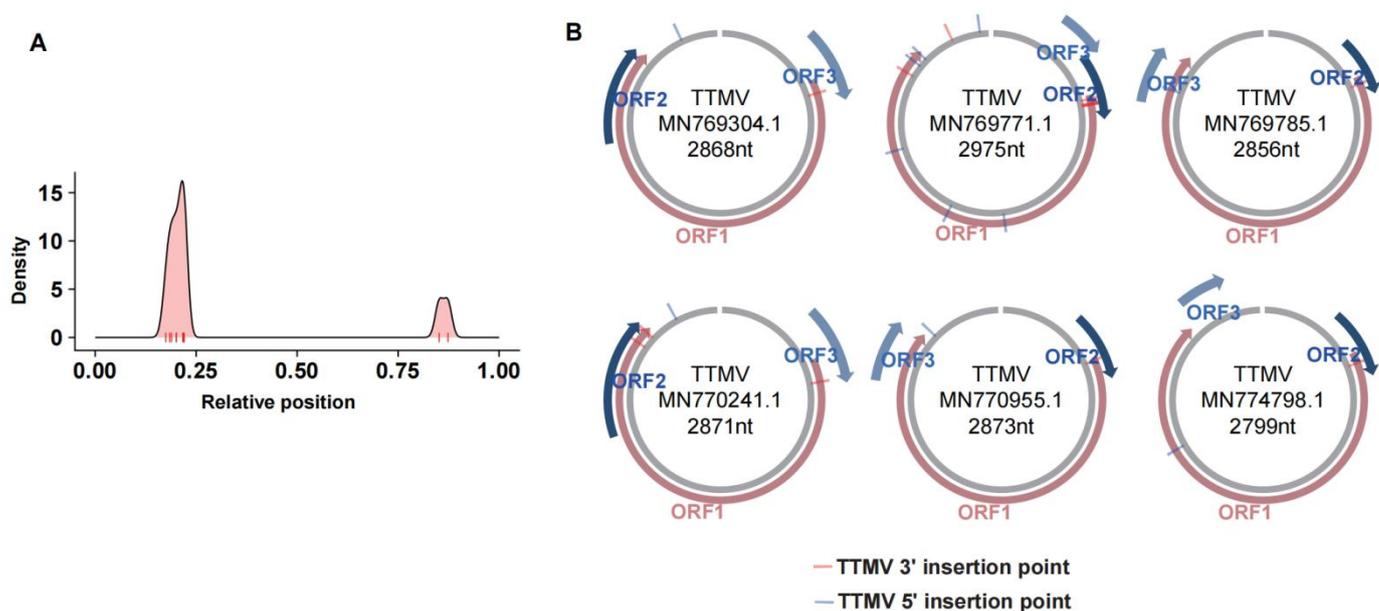
Fig. S2. Illustration of TTMV genomes and integration breakpoints. (A) Distribution of the relative genomic positions of the insertion points at the 3' end of the TTMV genome. The circular genomes are normalized at a [0, 1] interval. (B) The characteristics of the TTMV genomes and the integration sites with the human genome. Arrows indicate the start, end and transcription direction (arrowhead as 3') information for TTMV ORFs. The red lines represent insertion points at the 3' end of the TTMV genome, and the blue lines represent the 5' end insertion points.

Fig. S3

Fig. S3. The misalignment of the 3' and 5' breakpoints when TTMV integrates into the host genome. (A) The 5' breakpoint is positioned in the 3' direction relative to the 3' breakpoint, resulting in the loss of an intronic segment of the host genome. (B) The 5' breakpoint is positioned in the 5' direction relative to the 3' breakpoint, resulting in the duplication of an intronic segment in the host genome.

Fig. S4
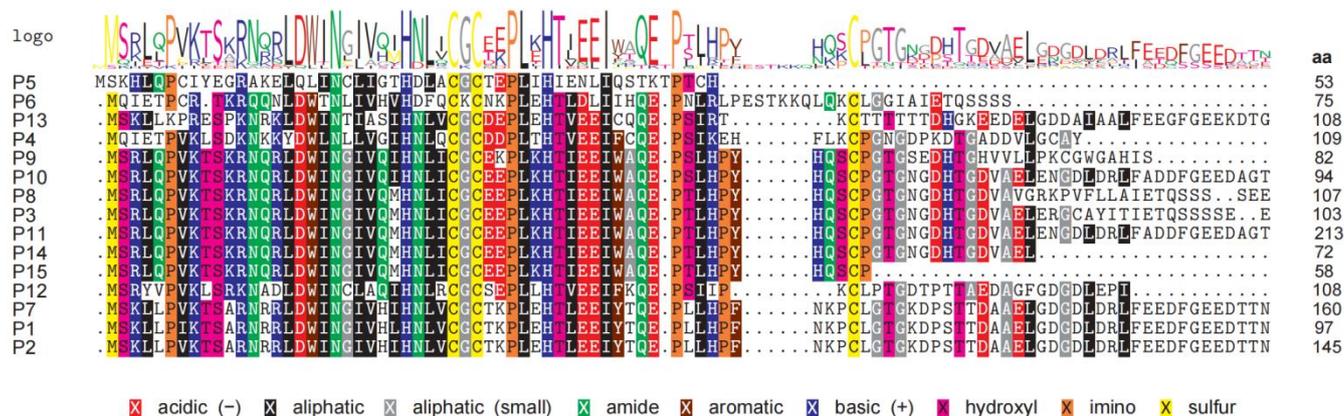
Fig. S4. Multiple sequence alignment of TTMV insertion sequences from the 15 samples. Residues were colored according to chemical properties of their functional groups. The right column indicates the amino acid lengths of TTMV sequence starting from the initial M residue.

Fig. S5
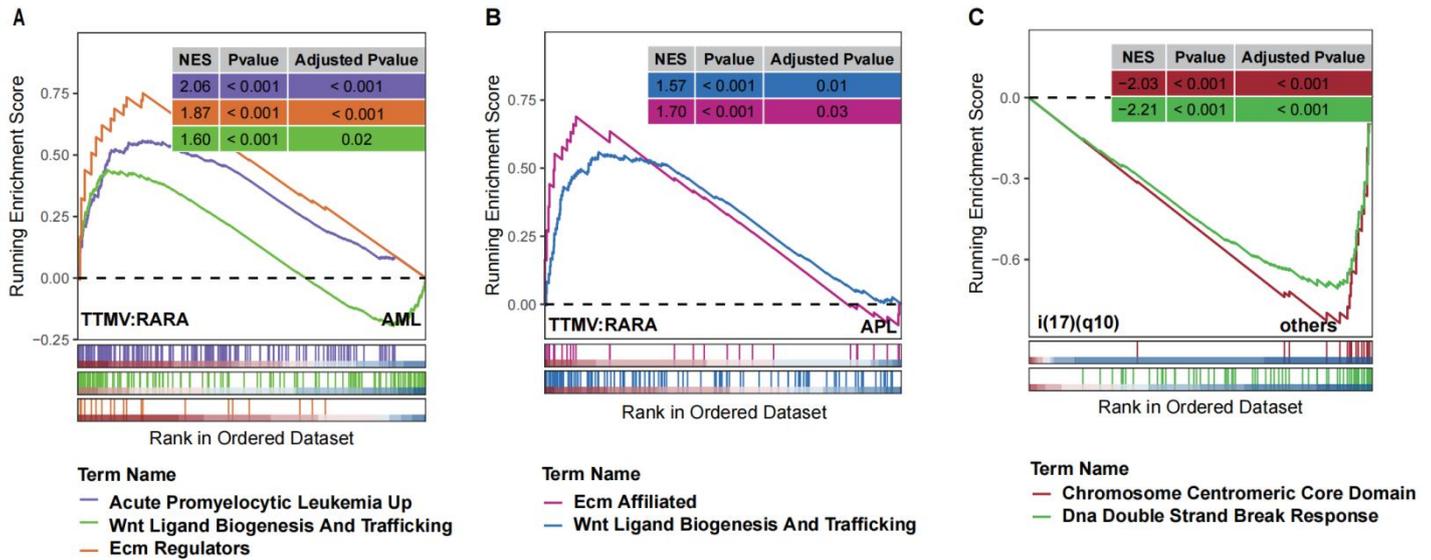
Fig. S5. Gene Set Enrichment Analysis. (A) Functional enrichment of upregulated pathway in AML with *TTMV::RARA* compared to typical AML. (B) Functional enrichment of upregulated pathway in AML with *TTMV::RARA* compared to typical APL. (C) Functional enrichment of downregulated pathway in *TTMV::RARA* patients with i(17)(q10) compared to non i(17)(q10).

Fig. S6
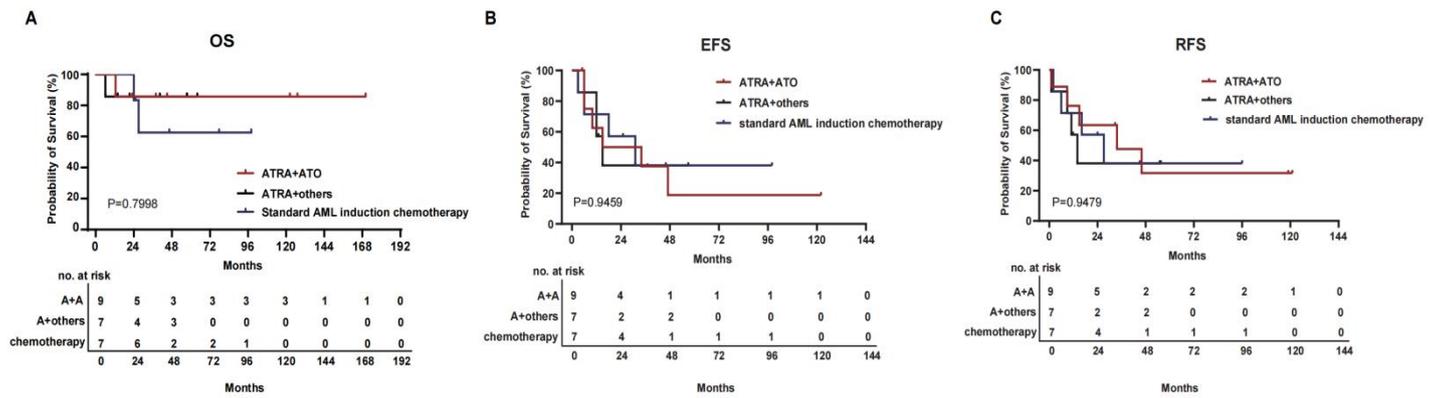
Fig. S6. Treatment outcomes of ATRA+ATO versus other therapies in AML with *TTMV::RARA*. (A) overall survival (OS). (B) Event-free survival (EFS). (C) Relapse-free survival (RFS). ATRA, all-trans retinoic acid; ATO arsenic trioxide.

Supplementary Table

Table S1

Data source	Publication
Beataml	PMID: 30333627
CRA001840	PMID: 32274301
EGAS00001000349	PMID: 27798625
EGAS00001002202	PMID: 29146900
EGAS00001002217	PMID: 30262806
GSE122682	PMID: 33530372
GSE162280	PMID: 33876209
GSE172057	PMID: 33893160
HRA000369	PMID: 33262139
HRA000789	PMID: 35347147
HRA001135	PMID: 34513657
HRA002693	PMID: 36442087
phs000159	PMID: 28760689
the TARGET project	NA (phs000218)
the TCGA project	PMID: 23634996

Table S1: Sources of the 2,543 Public Database Samples. This table details the origin and relevant information of 2,543 samples downloaded from public databases for the study.

Table S2:

Patient ID	Insertion sequence before RARA exon 3	TTMV strain	Splicing	Status
P1	TAATCACCATCATATGACAATTTCCTGGGAGGAGCCACTTACCTATATAACTAAGTGCACTTCCGAATGGCTGAGTTTATGCCGCCGGACGGAGACGCGATAGGAACTATCAGCGGCTTAGCCTGGGCGGGTGCCGAAGATGAGCAAACTACTACCTATTAAAACATCAGCAAGAAACAGAAGACTCGACTGGATAAATGGAATAGTTCACCTACACAACCTTGTCTGCGGCTGCACCAAACCACTGGAACACACCTTAGAAGAAATCTACACTCAAGAACCTCTACTACACCCCTTCAATAAACCATGCCTTGGTACTGGAAAAGACCCCTCTACTACCGACGCCGCCGAGCTTGGAGACGGCGACCTAGACCGCCTTTTCGAAGAAGACTTTGGAGAAGAAGACACCACAAACGTGTGTGCATATATAA	MN770241.1	final transcript	full
P2	TAATCACCATCATATGACTCTTTCCTGGGAGGAGCCACTTACCTATATAACTAAGTGCACTTCCGAATGGCTGAGTTTATGCCGCCGGACGGAGACGCGATAGGAACTATCAGCGGCTTAGCCTGGGCGGGTGCCGAAGATGAGCAAACTACTACCTGTTAAAACATCAGCAAGAAACAGAAGACTAGACTGGATAAATGGAATAGTTCACATCCACAACCTTGTCTGCGGCTGCACTAAACCACTGGAACACACCTTAGAAGAAATCTACACTCAAGAACCTCTACTACATCCCTTCAATAAACCATGCCTTGGTACTGGAAAAGACCCCTCTACTACCGACGCCGCCGAGCTTGGAGACGGAGACCTAGACCGCCTTTTCGAAGAAGACTTTGGAGAAGAAGACACCACAAACGCCGCTTCTACAAGCGGGAATCAACCTCCTCAACCCACAGCCAACACAAGACTCATCGGAGGAGACAGAAACAGAGGAGAAAAGCGAAAAGAGACACTACAGAAGCTCCTCAAGCAGCTCAAACAGCAACAGCACCGATACAGACAGCGGATCACCGAG	MN770241.1	final transcript	full
P5	CTCCCTAAGTCCTTTCTCCCTGGACTTTCAATGCACGAGAGAACAGAGACACACCAAGAATCACCACTCCAAACATCGGAGGAGGAAGAAGAGGAAACATCACTGTTCCACCAGCTCCAGCTCCAGCGAGCCAAGCAGCTCAGAATCAGACAGCGAATAATATCAACATTACAAAAACTTCAACAATTAGAATAGACAGAAGCAAACAAAGTATACTTATTTCCTCCAAAACCTAAACCTTACAATAGGTTTAATCCTCAGGAATTACAAACAGAGATAGAGATAGCCAGCTGGTTAAAAAGACCAGTAAGAACATTTAAAGAAGACCCCCCATACTATCCCTGGCTTCCTCCTACTCCTAAAGTACCCTTCCCAAACTTTAACCTTAATTTTACTGAATAAAGGCCTACAATTTTCACTTAGTGGTGTCTGTTTATATTATTTTCAACTTAAATAAACGTCCACCGCCTCCCAAATACGCAGGCGCAAAAGGGGGCTCCGCCCCCTTAAACCCCCAGGGGGCTCCGCCCCCTAAAACCCCCAAGGGGGCTCCGCCCCCTTACACCCCCTAATTAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAACCGTCAACAAGTTCCTCTTTTTTACATTACTTCCTCATTTCCTCATTATTATTCATGACATTAATTAGTAATCACCGTAATTCCGGGGAGGAGACTTAAACCTATATAACTAAGTACACTTCCGAATGGCTGAGTTTATGCCGCCAGACGGAGACGGGATCACTACAGTGACTCCAGGCTGACCAAGGGCGGGTGCTGAAGATGAGCAAACATCTCCAACCATGCATCTATGAAGGAAGAGCAAAGGAACTACAATTAATTAACTGCCTAATTGGAACCCATGATCTTGCCTGTGGCTGCACGGAACCATTAATTCACATTGAAAACTTAATTCAATCAACTAAAACACCAACATGCCATG	MN769785.1		full
P9	ATGTCAAGACTTCAACCTGTAAAAACTTCCAAAAGAAACCAACGCTTAGACTGGATTAATGGCATCGTCCAGATACACAACTTAATCTGCGGCTGTGAA	MN769771.1	final transcript	full

	AAACCTCTAAAACACACCATTGAAGAAATTTGGGCTCAGGAACCAAGCC TACATCCCTATCACCAATCATGCCCTGGTACTGGAAGCGAAGACCATAC TGGACACGTCGTCCTTCTCCCCAAATGTGGGTGGGGTGCCCACATTTC AG			
P10	ATGTCAAGACTTCAACCTGTAAAAACTTCTAAAAGAAACCAACGCTTAGA CTGGATTAATGGCATCGTCCAGATACACAACTTAATCTGCGGCTGTGAA GAACCTCTAAAACACACCATTGAAGAAATTTGGGCTCAAGAACCAAGCC TACATCCCTATCACCAATCATGCCCTGGTACTGGAAACGGAGACCATAC TGGAGACGTCGCAGAACTAGAAAATGGGGATTTAGACCGTTTGTTCGC CGACGACTTTGGAGAAGAAGACGCAGGCACCAGTACAGG	MN769771.1	final transcript	full
P11	ATGTCAAGACTTCAACCTGTGAAAACTTCTAAAAGAAACCAACGCTTAGA CTGGATTAATGGCATTGTCCAGATGCACAACTTAATCTGCGGCTGTGAA GAACCTCTGAAACACACCATTGAAGAAATTTGGGCTCAAGAACCAACTC TACATCCCTATCACCAATCATGCCCTGGTACTGGAAACGGAGACCATAC TGGAGACGTCGCAGAACTAGAAAATGGGGATTTAGACCGTTTGTTCGC CGACGACTTTGGAGAAGAAGACGCAGGCACCAGTACAGGGAGCAGAC CCCTTTCTCCCAACACCCCAAGAAGCAGCACCGACACAGGACTCATCG GAATCGGAAGAAGAAAAGAAACATTACAGCTCCTCATCCAGCAACACC GAGCAAAGCAGCAAAAGTTCAGGAACCGAATCCTCAGACTATTAACAGA GGAAAGTTCATAAACCTTGGTTGTGTACAAACTGCTCTTTTATTTCCTAA AGATACTTTTAAAAACAGACGCTTTACTACTTCTGAATTCCAACTAGAAC TGGAACTATGTAAAGCTTTTCGCAGACCCCCTAGAACATTCTTTCATGAC ACACCATATTATCCTTGGGTCTCTAACTGCCCCTCCCCTCTTCTCTCTCT AG	MN769771.1		full
P3	CTCCCACACTTAATTATTAAACACTGTAATTTTACACATATCCTGGGAGG AGACTATAAACTATAAGACTAACTACACTTCCGAATGGCTGAGTTTATGC CGCCAGACGGAGACGCGAAAGGAACTTTCAGCGGCTTAGCCTGGGCG GGTGCCGGAGGTGAGTTTACCACCGTAGTCAAGGGGCAATTCGGGCTG GCTAAGTCTGGCGGAACGGGCAAGAAACTTAAATAATATTTTTATTATAG ATGTCAAGACTTCAACCTGTGAAAACCTCTAAAAGAAACCAACGCTTAG ACTGGATTAATGGCATTGTCCAGATGCACAACTTAATCTGCGGCTGTGA AGAACCTCTAAAACACACCATTGAAGAAATTTGGGCTCAAGAACCAACT CTACATCCCTATCACCAATCATGCCCTGGTACTGGAAACGGAGACCATA CTGGAGACGTCGCAGAACTTGAGCGTGGGTGTGCATATATAACCATTG AGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCCCAGCCCTCCCTCGC CACCCCCTCTACCCCGCA	MN769771.1		3prime
P4	AGATGCCAAAATTGCTACCAGTGAAGCTCTCAGACAAAAACAAAAAATA TGACTGGCTAAATTTACTTGTTGGAATCCACAATCTACAATGTGGCTGC GATGATCCCCTTACTCACACTGTAGAAGAAATTTTCTGCCAAGAACCTTC AATTAAAGAGCACTTTCTAAAATGCCCTGGCAATGGAGACCCCAAAGAT ACTGGCGCAGACGACGTACTTGGGTGTGCATATATAACCATTGAGACC CAGAGCAGCAGTTCTGAAGAGATAGTGCCCAGCCCTCCCTCGCCACCC CCTCTACCCCGCATCTACAAGCCTTGCTTTGACTGTCA	MN774798.1		3prime

P7	CTTCCCCTACACGACGCTCTTCCGATCTCGGAGACGCGATAGGAACTAT CAGCGGCTTAGCCTGGGCGGGTGCCGAAGATGAGCAAACTACTACCTG TTAAAACATCAGCAAGAAACAGAAGACTAGACTGGATAAATGGAATAGT TCACATCCACAACCTTGTCTGCGGCTGCACTAAACCACTGGAACACACC TTAGAAGAAATCTACACTCAAGAACCTCTACTGCATCCCTTCAATAAACC ATGCCTTGGTACTGGAAAAGACCCCTCTACTACCGACGCCGCCGAGCT TGGAGACGGAGACCTAGACCGCCTTTTCGAAGAAGACTTTGGAGAAGA AGACACCACAAACGCCGCTTCTACAAGCGGGAATCAACCTCCTCAACC CACAGCCAACACAAGACTCATCGGAGGAGACAGAAACAGAGGAGAAAA GCGAAAAAGAAGGCACGCCATTGAGACCCAGAGCAGCAGTTCTGAAGA GATAGTGCCCAGCCCTCCCTCGCCACCCCCTCTACCCCGCATCTACAA GCCTTGCTTTGTCTGTCAGGACAAG	MN770241.1		3prime
P8	AGATGTCAAGACTTCAACCTGTGAAAACTTCTAAAAGAAACCAACGCTTA GACTGGATTAATGGCATTGTCCAGATGCACAACTTAATTTGCGGCTGTG AAGAACCTCTGAAACACACCATTGAAGAAATTTGGGCTCAAGAACCAAC TCTACATCCCTATCACCAATCATGCCCTGGTACTGGAAACGGAGACCAT ACTGGAGACGTCGCAGTGGGAAGGAAGCCCGTCTTCCTTTTAGCCATT GAGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCCCAGCCCTCCCTCG CCACCCCCTCTACCCCGCATCTACAAGCCTT	MN769771.1		3prime
P6	AGATGCAAATAGAAACACCATGCCGCACGAAAAGACAACAAAATCTTGA CTGGACAAACCTCATTGTTCATGTCCACGACTTCCAGTGCAAATGCAAC AAACCTCTTGAACACACCTTGGATTTAATTATTCACCAAGAACCAAACCT GAGACTTCCAGAATCTACTAAAAAACAACTGCAAAAATGCCTTGGTGGT ATTGCCATTGAGACCCAGAGCAGCAGTTCTG	MN770955.1		3prime
P12	CGAAGGCGCGATAGGAACTATCAGCGTCTGAGCAAGGGCGGGTGCCG AAGATGTCAAGATATGTTCCAGTTAAACTATCAAGAAAAAATGCAGATTT GGACTGGATAAACTGCCTTGCCCAAATTCACAACCTGCGCTGCGGATG CTCGGAACCTCTGCTACACACAGTAGAAGAAATTTTTAAACAAGAACCA TCCATAATTCCAAAATGCCTGCCTACTGGAGATACCCCTACTACCGCAG AAGACGCCGGATTTGGAGACGGAGACCTAGAACCCATTGAGACCCAGA GCAGCAGTTCTGAAGAGATAGTGCCCAGCCCTCCCTCGCCACCCCCTC TACCCCGCATCTACAAGCCTTGCTTTGTCTGTCAG	MN769304.1		3prime
P13	GGTCGTAGACGCGATAGGAACTATCAGCGGCTGAGCTTGGGCGGGTG CCGAAGATGTCAAAGCTGCTAAAACCAAGAGAATCTCCAAAAAACAGAA AATTAGACTGGATAAACACCATTGCCTCCATCCATAACCTTGTCTGCGG CTGTGATGAACCATTAGAACACACTGTAGAAGAAATCTGCCAACAAGAA CCTTCAATTCGCACAAAATGTACTACAACTACAACTACAGACCATGGAAA AGAAGAAGACGAACTTGGAGACGACGCCATCGCCGCCCTTTTCGAAGA AGGATTTGGAGAAGAAAAGATACTGGAAACGACGGCCATTGAGACCC AGAGCAGCAGTTCTGAAGAGATAGTGCCCAGCCCTCCC	unidentified		3prime
P14	ACGCCAGACGGAGACGCGAAAGGAACTTTCAGCGGCTTAGCCTGGGC GGGTGCCGGAGATGTCAAGACTTCAACCTGTGAAAACTTCTAAAAGAAA CCAACGCTTAGACTGGATTAATGGCATTGTCCAGATGCACAACTTAATC TGCGGCTGTGAAGAACCTCTGAAACACACCATTGAAGAAATTTGGGCTC AAGAACCAACTTTACATCCCTATCACCAATCATGCCCTGGTACTGGAAA CGGAGACCATACTGGAGACGTCGCAGAACTAG	MN769771.1		3prime

| P15 | GACGGAGACGCGAAAGGAACTTTCAGCGGCTTAGCCTGGGCGGGTGC CGGAGGTGAGTTTACCACCGTAGTCAAGGGGCAATTCGGGCTGGCTAA GTCTGGCGGAACGGGCAAGAAACTTAAATAATATTTTTATTGTAGATGTC AAGACTTCAACCTGTGAAAACTTCTAAAAGAAACCAACGCTTAGACTGG ATTAATGGCATTGTCCAGATGCACAACTTAATCTGCGGCTGTGAAGAAC CTCTAAAACACACCATTGAAGAAATTTGGGCTCAAGAACCAACTCTACAT CCCTATCACCAATCATGCCCTG | MN769771.1 | | 3prime |

Table S2: Assembled sequences of *TTMV::RARA* transcripts from 15 patients. The table presents speculated inserted TTMV sequences for each sample. Given the possible post-insertion splicing events, both pre- and post-splicing reads may coexist in data. Therefore, it cannot be excluded that read assembly does not necessarily represent the final biological event. The "Splicing" column denotes this phenomenon, with "final transcript" representing high-confidence final sequences. For the nine patients where full-length sequences could not be well-reconstructed, we presented the assembled segment located upstream of the 3' insertion site.

Table S3

Sample ID	Gender	Group	8p gain	i(17)(q10)	NRAS mutation	FLT3-ITD	NPM1 mutation
HRR719193	male	AML					
P12	male	*TTMV::RARA*			exist		
P3	female	*TTMV::RARA*					
HRR719231	male	APL					
P12_2	male	*TTMV::RARA*			exist		
P4	female	*TTMV::RARA*					
P13	female	*TTMV::RARA*					
P2	male	*TTMV::RARA*					
P9	female	*TTMV::RARA*					
P15	female	*TTMV::RARA*					
P5	female	*TTMV::RARA*					
P6	male	*TTMV::RARA*					
HRR719229	male	APL					
HRR719228	male	APL				exist	
HRR719235	female	APL					
HRR719242	female	APL					
HRR719278	female	APL					
HRR719263	female	APL					
HRR719250	female	APL					
HRR719267	female	APL					
HRR719262	male	APL					
HRR719254	male	APL					
HRR719268	female	APL					
HRR719269	female	APL				exist	
HRR719251	female	APL				exist	
HRR719260	male	APL				exist	
HRR719246	female	APL				exist	
HRR719245	female	APL				exist	
HRR719232	male	APL				exist	
HRR719240	male	APL				exist	
HRR719234	male	APL			exist		
P8	female	*TTMV::RARA*	yes				
HRR719249	female	APL					
HRR719512	male	APL					
HRR719243	male	APL					
HRR719258	male	APL					
HRR719237	female	APL					
HRR719253	female	APL					
HRR719226	male	APL			exist		
HRR719255	male	APL					
HRR719266	male	APL					

HRR719256	female	APL					
P1	female	*TTMV::RARA*		yes			
P11	female	*TTMV::RARA*	yes	yes			
P7	male	*TTMV::RARA*		yes			
P10	male	*TTMV::RARA*		yes			
P14	female	*TTMV::RARA*					
HRR719244	female	APL					
HRR719239	female	APL					
HRR719002	male	APL				exist	
HRR718999	female	APL				exist	
HRR719004	male	APL			exist		
HRR719257	male	APL					
HRR719014	male	APL					
HRR719261	female	APL					
HRR719294	female	APL				exist	
HRR719236	male	APL					
HRR719259	female	APL					
HRR719238	male	APL					
HRR719252	male	APL					
HRR719225	female	APL			exist		
HRR719264	male	APL					
HRR719265	female	APL					
HRR719018	male	APL			exist		
HRR719248	male	APL				exist	
HRR719233	female	APL					
HRR719241	male	APL					
HRR719227	female	APL					
HRR719230	male	APL					
HRR719247	female	APL					
HRR718864	male	AML					
HRR718920	male	AML			exist		
HRR718896	male	AML					
HRR719104	male	AML					
HRR719008	male	AML					
HRR719500	male	AML					
HRR719484	male	AML					
HRR718923	female	AML	yes				
HRR719271	male	AML	yes				
HRR719183	male	AML					
HRR719163	female	AML					exist
HRR719027	female	AML					exist
HRR719281	female	AML					
HRR719384	female	AML			exist		exist

HRR719061	female	AML					exist
HRR719419	female	AML					
HRR719340	male	AML					exist
HRR719170	female	AML					exist
HRR719157	male	AML			exist		
HRR719378	female	AML					exist
HRR719402	male	AML					
HRR719383	female	AML					exist
HRR719306	male	AML					exist
HRR719168	male	AML			exist		exist
HRR719224	male	AML					exist
HRR719348	female	AML			exist		exist
HRR719270	male	AML					
HRR719137	male	AML					exist
HRR718997	female	AML					exist
HRR719331	male	AML					exist
HRR718919	male	AML					exist
HRR718889	male	AML					exist
HRR718931	male	AML					exist
HRR718909	male	AML				exist	exist
HRR718946	female	AML			exist		exist
HRR719015	male	AML				exist	exist
HRR719130	female	AML				exist	exist
HRR719375	male	AML				exist	exist
HRR719411	female	AML					exist
HRR719141	male	AML				exist	
HRR719133	female	AML					exist
HRR719088	female	AML					
HRR719308	female	AML					
HRR718875	male	AML					
HRR719188	female	AML					
HRR719486	female	AML					
HRR718883	male	AML					
HRR718871	male	AML			exist		
HRR718860	male	AML					
HRR718938	male	AML					
HRR718927	female	AML					
HRR719144	female	AML				exist	
HRR719109	female	AML				exist	
HRR718988	female	AML				exist	
HRR719189	female	AML				exist	
HRR718878	female	AML					
HRR718873	female	AML				exist	

HRR719017	male	AML				exist	
HRR718908	female	AML				exist	
HRR719062	female	AML	yes				exist
HRR718980	female	AML	yes				
HRR719049	male	AML					
HRR719097	male	AML					
HRR719098	male	AML					
HRR718882	female	AML				exist	exist
HRR718866	male	AML			exist		exist
HRR719346	female	AML					exist
HRR719122	female	AML					exist
HRR719216	male	AML					exist
HRR718975	female	AML				exist	exist
HRR719507	female	AML				exist	exist
HRR719447	male	AML				exist	exist
HRR719090	female	AML				exist	exist
HRR719086	male	AML					exist
HRR719114	female	AML	yes				exist
HRR719085	male	AML				exist	exist
HRR719511	male	AML				exist	exist
HRR719209	male	AML					exist
HRR719182	female	AML					
HRR719293	male	AML				exist	exist
HRR719093	female	AML					exist
HRR719394	female	AML					exist
HRR719323	female	AML				exist	
HRR719175	male	AML				exist	
HRR719079	female	AML				exist	exist
HRR719339	male	AML				exist	exist
HRR719388	female	AML			exist		exist
HRR718963	female	AML				exist	
HRR719149	male	AML				exist	exist
HRR719284	female	AML				exist	exist
HRR719315	female	AML				exist	exist
HRR719147	female	AML				exist	exist
HRR719068	female	AML					exist
HRR719056	female	AML				exist	exist
HRR719353	male	AML					exist
HRR719064	female	AML				exist	
HRR719374	female	AML				exist	exist
HRR719011	male	AML				exist	exist
HRR719360	male	AML				exist	exist
HRR719390	male	AML				exist	exist

HRR719356	male	AML					exist	exist
HRR719201	female	AML					exist	
HRR719210	female	AML						exist
HRR718994	male	AML					exist	exist
HRR719397	female	AML						exist
HRR719118	female	AML					exist	exist
HRR719078	male	AML					exist	
HRR718962	male	AML					exist	exist
HRR719134	female	AML					exist	exist
HRR719075	male	AML					exist	exist
HRR718880	female	AML						exist
HRR718933	female	AML					exist	exist
HRR719044	male	AML						exist
HRR719126	female	AML						exist
HRR719376	female	AML					exist	exist
HRR719211	male	AML					exist	
HRR719220	female	AML						
HRR719091	female	AML					exist	
HRR719310	female	AML						
HRR719071	female	AML						
HRR719334	male	AML			exist			
HRR718894	female	AML						
HRR718940	male	AML						
HRR718917	female	AML					exist	
HRR718881	male	AML						
HRR719187	male	AML			exist		exist	exist
HRR719366	female	AML						
HRR719083	female	AML					exist	exist
HRR719105	female	AML					exist	
HRR719398	female	AML						
HRR719326	female	AML						
HRR719221	female	AML						
HRR719145	female	AML						
HRR719136	male	AML						
HRR719186	female	AML						exist
HRR719362	male	AML						
HRR719080	female	AML						
HRR719121	female	AML						
HRR719177	female	AML					exist	
HRR719092	female	AML			exist		exist	
HRR719320	female	AML					exist	
HRR719302	female	AML	yes					
HRR718977	male	AML						

HRR719051	female	AML					
HRR719171	male	AML					
HRR718926	female	AML					
HRR718967	male	AML					
HRR719371	male	AML					
HRR718989	female	AML					
HRR719026	female	AML			exist		
HRR719276	female	AML				exist	
HRR719053	male	AML				exist	
HRR719165	male	AML					
HRR719096	female	AML					
HRR719167	female	AML				exist	exist
HRR719007	female	AML				exist	exist
HRR718949	female	AML				exist	
HRR719409	female	AML				exist	
HRR719392	male	AML				exist	
HRR719391	male	AML				exist	
HRR719355	female	AML				exist	
HRR718935	male	AML					
HRR718898	female	AML				exist	exist
HRR718862	female	AML					
HRR718906	male	AML				exist	exist
HRR718902	female	AML				exist	exist
HRR719215	male	AML				exist	
HRR719065	male	AML				exist	exist
HRR719337	male	AML					exist
HRR719169	female	AML				exist	exist
HRR719368	male	AML				exist	exist
HRR718981	female	AML				exist	exist
HRR719153	female	AML				exist	exist
HRR719295	female	AML				exist	exist
HRR718982	female	AML				exist	
HRR719077	male	AML				exist	
HRR719162	male	AML				exist	
HRR719289	female	AML				exist	exist
HRR718990	female	AML				exist	exist
HRR719146	male	AML	yes			exist	
HRR719365	female	AML				exist	exist
HRR718903	male	AML					
HRR718939	female	AML					
HRR718948	male	AML					
HRR719487	male	AML					
HRR718893	male	AML					

HRR718904	male	AML					
HRR718876	male	AML					
HRR718913	male	AML					
HRR718929	female	AML					
HRR718891	male	AML					
HRR718943	male	AML					
HRR718885	male	AML					
HRR718900	male	AML					
HRR718870	male	AML					
HRR718911	female	AML				exist	
HRR719425	male	AML					
HRR718861	female	AML					
HRR718877	male	AML					
HRR718942	female	AML					
HRR718925	female	AML					
HRR718916	male	AML			exist		
HRR718907	male	AML			exist		
HRR718944	male	AML			exist		
HRR718947	male	AML					
HRR718865	male	AML			exist		
HRR718921	male	AML			exist		
HRR719273	male	AML					
HRR719344	male	AML					
HRR719286	female	AML					
HRR719283	male	AML				exist	
HRR719151	male	AML					
HRR719135	male	AML				exist	
HRR719021	male	AML					
HRR719336	female	AML					
HRR719297	male	AML					
HRR719410	male	AML					
HRR719496	female	AML					
HRR718924	male	AML				exist	
HRR718951	male	AML					
HRR719112	female	AML					
HRR719125	female	AML					
HRR719101	male	AML					
HRR719072	male	AML					
HRR719089	female	AML					
HRR719046	female	AML					
HRR719329	male	AML				exist	
HRR718971	female	AML					
HRR719041	female	AML					

HRR719407	female	AML					
HRR718970	female	AML					
HRR719066	male	AML					
HRR718992	male	AML					
HRR719127	male	AML					
HRR719342	male	AML					
HRR719381	female	AML					
HRR719058	male	AML					
HRR719386	female	AML					
HRR718987	male	AML					
HRR719285	female	AML	yes				
HRR719031	male	AML					
HRR719074	female	AML					
HRR719370	male	AML					
HRR719330	female	AML					
HRR719207	female	AML					
HRR719099	female	AML					
HRR719178	female	AML					
HRR719203	male	AML					
HRR719174	female	AML					
HRR719012	male	AML					
HRR719369	male	AML					
HRR719372	male	AML					
HRR719035	female	AML			exist		
HRR719213	male	AML					
HRR719350	male	AML					
HRR719119	male	AML					
HRR719205	male	AML					
HRR719009	male	AML					
HRR718966	female	AML					
HRR718958	male	AML					
HRR719354	male	AML			exist		
HRR718915	female	AML					
HRR719290	male	AML				exist	
HRR718969	male	AML				exist	
HRR719019	male	AML			exist		
HRR719087	male	AML					
HRR719404	male	AML			exist		
HRR719338	female	AML			exist		
HRR719318	male	AML					
HRR719202	male	AML					
HRR719272	male	AML					
HRR719084	male	AML					

HRR719212	female	AML					
HRR719138	male	AML					
HRR719191	male	AML					
HRR719322	female	AML					
HRR719349	male	AML					
HRR719110	male	AML					
HRR719341	female	AML					
HRR719314	male	AML					
HRR718995	male	AML					
HRR719036	female	AML					
HRR719113	female	AML					
HRR718914		AML					
HRR719156	female	AML					
HRR719106	female	AML					
HRR719057	female	AML					
HRR719217	female	AML					
HRR719166	male	AML			exist	exist	
HRR718955	male	AML					
HRR718937	male	AML			exist		
HRR719347	male	AML					
HRR719358	male	AML					
HRR719385	male	AML					
HRR718976	male	AML					
HRR719040	male	AML					
HRR719070	female	AML					
HRR719028	female	AML					
HRR718986	female	AML					
HRR719059	male	AML					
HRR719000	male	AML					
HRR719352	male	AML					
HRR719054	female	AML					
HRR719129	male	AML					
HRR719033	female	AML					
HRR719142	female	AML					
HRR719219	female	AML					
HRR718922	female	AML					
HRR718890	female	AML				exist	
HRR718928	male	AML					
HRR718912	male	AML					
HRR719299	male	AML					
HRR719274	male	AML					
HRR719279	male	AML					
HRR719117	male	AML					

HRR719022	female	AML						
HRR719132	female	AML						
HRR718956	female	AML						
HRR719510	male	AML						
HRR719292	female	AML						
HRR718974	male	AML						
HRR719185	female	AML						
HRR718996	female	AML						
HRR719115	male	AML						
HRR718867	male	AML						
HRR718879	male	AML						
HRR718869	female	AML						
HRR718887	male	AML						
HRR718892	male	AML						
HRR718954	male	AML						
HRR718863	female	AML					exist	
HRR718918	male	AML						
HRR718874	female	AML				exist		
HRR718886	male	AML						
HRR719069	female	AML						
HRR719038	female	AML						
HRR719197	female	AML						
HRR719067	male	AML						
HRR719034	male	AML						
HRR719387	male	AML						
HRR719016	female	AML						
HRR719481	male	AML						
HRR719103	female	AML						
HRR718984	male	AML						
HRR719100	female	AML				exist		
HRR718973	female	AML				exist		
HRR718983	male	AML				exist		
HRR719010	male	AML						
HRR719081	female	AML						
HRR719180	male	AML						
HRR719194	male	AML	yes				exist	
HRR719275	female	AML				exist		
HRR718950	male	AML				exist		
HRR718979	male	AML						
HRR719152	male	AML						
HRR719102	female	AML						
HRR719159	female	AML				exist		
HRR718961	male	AML						

HRR718957	female	AML					
HRR718964	male	AML					
HRR719277	female	AML					
HRR718972	male	AML					
HRR719108	female	AML					
HRR719063	male	AML					
HRR719351	male	AML			exist		
HRR719380	male	AML				exist	
HRR718965	male	AML					
HRR719111	female	AML			exist		
HRR719399	male	AML			exist		
HRR719179	female	AML	yes				
HRR719400	female	AML	yes				
HRR718945	female	AML					
HRR719095	female	AML					
HRR719309	female	AML					
HRR719479	female	AML					
HRR718910	male	AML					
HRR719345	male	AML					
HRR719195	female	AML				exist	
HRR719335	male	AML					
HRR719382	female	AML					
HRR719029	male	AML					
HRR719304	female	AML				exist	
HRR719396	female	AML				exist	
HRR718868	female	AML					
HRR719405	female	AML				exist	
HRR719223	male	AML					
HRR719025	male	AML				exist	
HRR719128	female	AML					
HRR719052	female	AML					
HRR719006	male	AML					
HRR719013	female	AML					
HRR719514	male	AML					
HRR719423	female	AML					
HRR718899	female	AML					
HRR718901	male	AML					
HRR718953	male	AML					
HRR718952	female	AML					
HRR719504	male	AML					
HRR719401	male	AML					
HRR719107	female	AML					
HRR719303	male	AML					

HRR719005	female	AML					
HRR718978	female	AML					
HRR719032	male	AML					
HRR719204	male	AML					
HRR719327	female	AML					
HRR719218	male	AML					
HRR719389	female	AML					
HRR718985	male	AML					
HRR719150	male	AML			exist		
HRR719047	male	AML			exist		
HRR719321	male	AML					
HRR719140	male	AML			exist		
HRR719287	male	AML					
HRR719124	male	AML				exist	
HRR719050	male	AML					
HRR719184	female	AML					
HRR719048	male	AML					
HRR719377	male	AML					
HRR719037	female	AML					
HRR719198	female	AML				exist	
HRR719164	female	AML				exist	
HRR719060	male	AML					
HRR719073	male	AML					
HRR719282	male	AML					
HRR719364	male	AML					
HRR719003	male	AML					
HRR718959	male	AML	yes				
HRR719357	male	AML					
HRR718941	male	AML					
HRR719325	male	AML					
HRR719280	female	AML					
HRR719408	female	AML				exist	
HRR719148	female	AML					
HRR719120	male	AML					
HRR719333	male	AML					
HRR719082	female	AML	yes				
HRR719332	male	AML					
HRR719328	male	AML					
HRR719395	female	AML					
HRR719024	female	AML					
HRR719160	male	AML	yes				
HRR719373	female	AML					
HRR719181	male	AML					

HRR719307	male	AML					
HRR719131	male	AML				exist	exist
HRR719367	male	AML					
HRR719343	male	AML				exist	exist
HRR718884	female	AML					
HRR719313	male	AML				exist	
HRR719139	male	AML					
HRR719172	female	AML					exist
HRR719298	female	AML					
HRR718930	female	AML					
HRR719039	male	AML			exist		exist
HRR718905	female	AML				exist	
HRR718897	female	AML					
HRR719317	female	AML				exist	
HRR719192	female	AML					
HRR719495	male	AML					
HRR718932	female	AML			exist		
HRR719158	female	AML			exist		
HRR719485	female	AML					
HRR718934	male	AML					
HRR719305	female	AML		yes		exist	
HRR718895	male	AML					
HRR719406	male	AML	yes				
HRR719393	female	AML					
HRR719222	male	AML					
HRR719312	male	AML					
HRR719359	female	AML					
HRR719214	male	AML					
HRR719206	male	AML	yes				
HRR719190	female	AML					
HRR719045	male	AML	yes				
HRR719154	male	AML					
HRR719023	male	AML					
HRR719196	male	AML					
HRR719361	female	AML					
HRR719001	male	AML					
HRR719475	male	AML					
HRR718888	male	AML					
HRR719116	male	AML					
HRR719055	male	AML					
HRR719379	female	AML					
HRR718998	female	AML					
HRR719316	male	AML					

HRR719030	female	AML					
HRR719291	male	AML					
HRR719311	male	AML					
HRR719363	male	AML				exist	
HRR719143	male	AML			exist		
HRR719043	male	AML	yes				
HRR718991	female	AML					
HRR718993	female	AML					
HRR718936	female	AML	yes				
HRR719319	female	AML			exist		
HRR719493	female	AML			exist		
HRR719296	male	AML					
HRR719288	female	AML			exist		
HRR719324	female	AML					
HRR719076	female	AML			exist		
HRR719301	male	AML				exist	
HRR719300	male	AML					
HRR719020	male	AML	yes				
HRR719208	female	AML					
HRR719173	female	AML					
HRR719176	female	AML					
HRR719155	male	AML			exist		
HRR719199	male	AML					

Table S3: Clinical and Molecular Characteristics of AML, APL and *TTMV::RARA* Samples Used for Differential Gene Expression. The table lists the sample IDs (indicating data sources), gender, group classification, and key genetic alterations for each sample. These samples were used for differential gene expression analysis and unsupervised clustering to explore the molecular landscape of AML with TTMV::RARA and its distinction from classical APL and AML. For mutation events, only hotspot mutations (e.g. W288Cfs*12 for NPM1, G12D or G12V for NRAS) with VAF > 0.1 were presented in this table.