haematologica
Journal of the Ferrata Storti Foundation

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

by Shu Sun, Yongjing Liu, Qing-Yu Xu, Jiayu Wang, Li Chen, Zhanrui Cheng, Wei Gao, Huili Wang, Bei Yang, Haiyan Wang, Lijun Wen, Jian Xiao, Jiacheng Lou, Haibo Yu, Na Li, Feng Wang, Yangyang Xie, JinGang Wang, Xianjing Wang, Hongjuan Xue, Kun Chen, Yin Wu, Leping Zhang, Ke Li, Shuhong Shen, Suning Chen, Huan-You Wang, KanKan Wang, Jinyan Huang and Hong-Hu Zhu

Received: July 16, 2025.
Accepted: November 3, 2025.

Citation: Shu Sun, Yongjing Liu, Qing-Yu Xu, Jiayu Wang, Li Chen, Zhanrui Cheng, Wei Gao, Huili Wang, Bei Yang, Haiyan Wang, Lijun Wen, Jian Xiao, Jiacheng Lou, Haibo Yu, Na Li, Feng Wang, Yangyang Xie, JinGang Wang, Xianjing Wang, Hongjuan Xue, Kun Chen, Yin Wu, Leping Zhang, Ke Li, Shuhong Shen, Suning Chen, Huan-You Wang, KanKan Wang, Jinyan Huang and Hong-Hu Zhu. The comprehensive landscape of TTMV::RARA fusion-driven acute myeloid leukemia: from viral integration mechanisms to clinical outcomes.
Haematologica. 2025 Nov 13. doi: 10.3324/haematol.2025.288721 [Epub ahead of print]

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,*]

[1] Department of Hematology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China;

[2] Biomedical big data center, the First Ailiated 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 & 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 Ailiated 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;

[13] Department of Pathology, University of California San Diego, La Jolla, California, USA;

[#] These authors contributed equally to the research study.

* Corresponding Author:

Hong-Hu Zhu, PhD, MD

Address: Beijing Chao-Yang Hospital, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020; 10 Xitoutiao, Youanmen Wai, Fengtai District, Beijing 100069, China

E-mail: zhuhhdoc@163.com

Jinyan Huang, PhD,

Address: Center for Biomedical Big Data, the First Affiliated Hospital, School of Medicine, Zhejiang University, Zhejiang, China

E-mail: hiekeen@gmail.com

Running title:

The landscape of oncogenic Torque Teno Mini Virus in AML

Keywords

Torque Teno Mini Virus; *TTMV::RARA*; APL; AML;

Authorship Contributions

HHZ designed the study, SS collected clinical data and interpreted the data, YJL and JYH contributed to the processing of the analysis of omics data. HHZ, SS, YJL, and QYX wrote the paper. JYW, ZRC, BY, HBY, NL, FW modified the pictures. LC, WG, HLW, HYW, LJW, JX, JCL,

YYX, JGW, XJW, HJX, KC, YW, LPZ, SHS, SNC, HYW, and KKW collected the clinical samples and compiled the clinical information and omics sequencing data. HHZ, JYH, HYW, and KL critically reviewed the article. All authors read and approved the final manuscript.

Data Availability

For access to the original data, please contact the corresponding author.

Funding

This work was supported by the National Natural Science Foundation of China (82450101, 82370169), the “Dengfeng” Talent Training Program of Beijing Hospitals Authority (DFL20240301), and the funds from the Chinese Institutes for Medical Research, Beijing (CIMR) to Hong-Hu Zhu. CAMS Innovation Fund for Medical Sciences (2024-I2M-TS-017) to Ke Li. National Natural Science Foundation of China (82200130, 82570201), Postdoctoral Fellowship Program of CPSF (GZC20231748), China Postdoctoral Science Foundation (2025T180625, 2024M762179) and Beijing Postdoctoral Research Foundation (2024-ZZ-185) to Shu Sun. National Natural Science Foundation of China (82400147), Beijing Chao-Yang Hospital Golden Seeds Foundation (CYJZ202303), Outstanding Young Talent of the Beijing Overseas Talent Aggregation

Project and the “Phoenix” Project in Chaoyang District, Beijing (Y2024070019) to Qing-Yu Xu.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Abstract

Acute myeloid leukemia (AML) with *TTMV::RARA* fusion represents a novel subtype driven by torque teno mini virus (TTMV) integration into retinoic acid receptor alpha (*RARA*) locus, while current understanding of its molecular features and clinical presentation relies predominantly on isolated case observations. Here, we characterize a large and independent cohort (n=25) through integrative analysis of clinical-omics data, uncovering unique features that distinguish it from classic acute promyelocytic leukemia (APL) and other AML subtypes. Our findings reveal that TTMV integrates exclusively within intron 2 of the *RARA* gene via microhomology-mediated end joining (MMEJ), forming functional *TTMV::RARA* transcripts. Clinically, patients harboring this fusion were predominantly pediatric (72% <18 years) and often presented with extramedullary diseases (24% with myeloid sarcoma, 16% with central nervous system infiltration). Blasts displayed APL-like morphology and immunophenotype but lacked *PML::RARA*, instead harboring *TTMV::RARA* with recurrent i(17)(q10) abnormalities (24%). Unsupervised clustering revealed it as a molecularly distinct subgroup. Transcriptomic profiling identified a Wnt-activated/extracellular matrix-dysregulated signature, driving leukemogenesis via dual mechanisms of clonal expansion and metastatic pathways. Despite achieving a 96% complete remission (CR) rate with induction therapy,

long-term outcomes were significantly inferior, with 2-year event-free survival (EFS) and relapse-free survival (RFS) rates of 53.6% and 53.8%, respectively. Hematopoietic stem cell transplantation (HSCT) achieved durable remission in 9 of 11 patients, particularly those with extramedullary disease or i(17)(q10) abnormalities. Conclusively, this work establishes *TTMV::RARA* as a novel AML subtype, highlighting the need for viral screening in APL-like cases and HSCT prioritization for this subset.

Introduction:

Acute myeloid leukemia (AML) represents a group of molecularly heterogeneous malignancies, characterized by chromosomal rearrangements and genetic aberrations[1-2]. Approximately 40% of AML cases harbor clinically validated pathogenic fusion genes arising from these rearrangements[3]. Key fusion genes, such as *PML::RARA*, have been established as diagnostic markers and therapeutic targets in AML, owing to their distinct clinicopathological features and prognostic relevance[4-6]. The World Health Organization (WHO) has incorporated these molecular alterations into the classification of haematolymphoid tumors, with their clinical application significantly improving patient management and outcomes[7]. However, a subset of AML patients remain without definitive molecular markers, creating a critical gap in targeted therapeutic strategies and underscoring the urgent need to identify novel oncogenic drivers for precision medicine in leukemia.

Recent studies have revealed that Torque Teno Mini Virus (TTMV), conventionally considered as non-pathogenic, can aberrantly integrate into the human genome at the retinoic acid receptor alpha (*RARA*) locus, resulting in the formation of a cross-species "virus-host" fusion gene, *TTMV::RARA*[8-17]. This distinctive genomic rearrangement directly contributes to the development of AML with features resembling acute promyelocytic leukemia (APL), thereby challenging the traditional

paradigm of leukemogenesis mediated by human-human gene fusion and broadening our understanding of AML pathogenesis.

While prior case reports have delineated the basic structural features of this fusion, the precise mechanisms of viral integration remain poorly characterized. Similarly, although clinical observations have documented certain phenotypic traits, an evidence-based treatment consensus and therapeutic guidelines await systematic validation through large-scale studies.

To address these critical knowledge gaps, we have assembled a large multicenter cohort of *TTMV::RARA* cases to date (n=25). This expanded dataset enables us to construct a comprehensive molecular profile of viral integration patterns, establish genotype-phenotype correlations across clinical presentations, and evaluate treatment response patterns to inform evidence-based therapeutic guidelines.

Materials and Methods：

Case identification and study design

This observational, retrospective cohort study (Approval No. 2024-7-31-2) was approved by the Institutional Review Board of Beijing Chaoyang Hospital, Capital Medical University, China, and conducted in compliance with the Declaration of Helsinki.

The study cohort for initial *TTMV::RARA* screening comprised 2,553 AML patients, including 2,543 cases retrieved from published databases (Table S1) and an additional 10 cases recruited from a multicenter study spanning July 2014 to August 2024. These 10 patients exhibited morphological and immunophenotypic features strongly resembling classical APL, but none carried known APL-defining genetic drivers (including classic *PML::RARA* and other *RARA*, *RARG*, or *RARB* fused with human genes), nor the hotspot NPM1 mutations that have been implicated in APL-like presentations. Subsequently, through the application of a stringent filtering strategy that incorporated viral genome data, we ultimately identified four cases harboring the *TTMV::RARA* fusion event. This high frequency emphasizes the potential diagnostic relevance of TTMV in resolving unclassified APL-like cases.

Diagnostic and follow-up information for newly identified *TTMV::RARA* positive patients were submitted by participating centers, while reported case data were extracted from the literature. Collected data included patient demographics, diagnostic/clinical laboratory results, MICM profiling, induction/consolidation chemotherapy regimens, and hematopoietic stem cell transplantation details. Treatment decisions were systematically recorded from medical records. All patients were followed until death or the data cutoff date (August 2024).

Expression-based comparisons

Transcript-level expressions were quantified by Salmon (v1.9.0)[18] and gene-level read counts were aggregated by the tximport R package. To minimize batch effects, an integrated dataset of *TTMV::RARA* samples and a separate dataset[19] (GSA-Human database, accession ID HRA002693) were analyzed for differential expression using DESeq2 (v1.40.2)[20]. The EnhancedVolcano R package was used for visualization of DE results. Variance-stabilizing transformation of gene expression levels was conducted with DESeq2. Unsupervised hierarchical clustering was performed using the Ward.D2 algorithm. Gene Set Enrichment Analysis was performed using the clusterProfiler[21] R package with gene sets from MSigDB and visualized by the GseaVis R package.

Definition of outcomes

Complete remission (CR) and overall response remission (ORR) were defined according to the recommended criteria[22]. Overall survival (OS) was defined as the period from initial diagnosis to the last follow-up with a status of either death or alive assigned. Event-free survival (EFS) was calculated from the time of initial diagnosis to treatment failure (the patient did not have complete remission by month 6), relapse, death or the last follow-up in CR. Relapse-free survival (RFS) was calculated from the time of first CR to the date of first relapse, death or the last follow-up still in CR.

Statistical analyses

Statistical analyses were performed using R software (v4.4.0) and GraphPad Prism (v10.1.2). Differential expression analysis was conducted using Wald's test, survival analysis included Kaplan-Meier curves with log-rank tests for group comparisons and Cox proportional hazards regression for multivariable analysis. $P<0.05$ was considered statistically significant. BH (Benjamini & Hochberg) method was applied for multiple test correction.

Results:

Establishment of the *TTMV::RARA* Clinical and Omics Dataset

To systematically identify TTMV integration at the *RARA* locus, we screened 2,543 publicly available AML RNA-seq datasets and 10 APL-like cases from collaborative cohorts (Figure 1). Following a rigorous multi-step pipeline screening (Supplementary Fig. S1), we identified 10 cases harboring *TTMV::RARA* (Figure 1). By incorporating 15 cases from previous studies[8-17], we established a comprehensive dataset comprising 25 patients in total. This dataset includes clinical profiles for all 25 cases, RNA-seq data for 15 cases, and whole genome sequencing (WGS) data for 2 cases (Figure 1).

General patterns of TTMV integration in AML

An integrative analysis of RNA-seq and WGS data from AML patients with *TTMV::RARA* revealed the presence of seven distinct TTMV strains involved in fusion gene formation. Among these, TTMV strain MN-769771.1 was the most prevalent, accounting for 46.7% (7/15) of all fusion events (Figure 2A-B).

Localization analysis confirmed that all TTMV integration events were exclusively confined to intron 2 of the *RARA* gene, with a pronounced preference for the 3′ breakpoint region (Figure 2A, Supplementary Fig. S2A). This finding aligns with previous observations in the literature[8-17].

Notably, the insertion sites exhibited significant clustering, suggesting the presence of potential hotspot integration loci within this region (p = 0.00178, Figure 2A). Sequence analysis of 15 integration junctions revealed the presence of 2 to 4 base pairs of microhomology in 12 cases (80%) at the integration sites (Figure 2A), indicating that microhomology-mediated end joining (MMEJ)[23] likely drives TTMV-host integration in the majority of patients.

Structural Characteristics of the *TTMV::RARA* Transcripts

The profiling of fusion transcripts revealed that 13 out of 15 cases (86.7%) harbored chimeric *RARA*::*TTMV::RARA* transcripts, with TTMV flanked by 5′ and 3′ *RARA* segments (Supplementary Figs. S2B, S3A-B; Table S2). These findings align with previous reports[14] and provide evidence for the existence of full length *RARA*::*TTMV::RARA* precursor transcripts (Figure 2C).

In depth sequence analysis of the fusion transcripts identified three distinctive structural features. Adjacent to the TTMV start codon, a highly conserved motif was detected in all 13 cases (Supplementary Fig. S4A), while the downstream *RARA* open reading frames remained intact. Additionally, variable retention of *RARA* intron 2 sequences (ranging from 0 to 45 base pairs) was observed (Figure 2D). This tripartite structure suggests a functional hierarchy. The 5′ *RARA* segment of

RARA::*TTMV::RARA* probably functions as a regulatory untranslated region (UTR), enabling the conserved TTMV motif and intact *RARA* ORFs to maintain translational efficiency. The variable retention of intron 2 sequences further implies splicing mediated regulation of this oncogenic fusion transcript, underscoring a complex mechanism by which viral integration dysregulates gene expression to drive leukemogenesis.

Collectively, these findings provide a comprehensive characterization of chimeric *RARA*::*TTMV::RARA* transcript structures that maintain a functional *TTMV::RARA* open reading frame, formed by TTMV viral integration into the human genome. These observations prompt inquiry into the clinical and phenotypic correlations mediated by these fusion genes.

Clinical characteristics of AML patients with *TTMV::RARA*

To elucidate the clinical and molecular characteristics of AML patients with *TTMV::RARA* fusion genes, we comprehensively analyzed 25 cases. The cohort predominantly comprised individuals aged ≤18 years (72%, 18/25), with an approximately equal gender distribution (male:female = 13:12) (Table 1). At the time of diagnosis, fever (36%, 9/25) and bleeding (32%, 8/25) were the most prevalent clinical manifestations (Table 1). Hematological parameters showed median values for white blood cell

(WBC) counts, hemoglobin, and platelet counts of 7.88×10^9/L (range 1 - 41.9×10^9/L), 88 g/L (55 - 113 g/L), and 94×10^9/L (13 - 334×10^9/L), respectively (Table 1). Coagulation studies revealed median prothrombin time and activated partial thromboplastin time of 13.55 and 32.6 seconds, respectively, with fibrinogen and D-dimer levels measured at 196 mg/dL and 10,660 μg/L, respectively (Table 1). Notably, 24% (6/25) of cases exhibited myeloid sarcoma and 16% (4/25) had central nervous system (CNS) infiltration, reflecting an aggressive disease phenotype (Table 1).

The MICM (Morphology, Immunophenotype, Cytogenetics, and Molecular Genetics) profiling of AML patients with *TTMV::RARA* fusion demonstrated both striking parallels and distinct divergences when compared to classic APL. Morphologically, 52% (13/25) of leukemic blasts displayed typical hypergranular APL morphology, while 8% (2/25) showed hypogranular variant APL characteristics. Additionally, Auer rods were observed in 32% (8/25) of the cases (Table 1 and Figure 3A). Flow cytometric analysis revealed high expression levels of CD33 (100%), CD13 (100%), CD117 (79%), MPO (100%), CD99 (100%), and CD38 (87.5%) in the majority of leukemia cells, while CD34 (16.7%), HLA-DR (10%), and CD11b (15%) were rarely expressed (Figure 3B), consistent with APL immunophenotypic patterns.

Conversely, cytogenetic and molecular analyses uncovered distinct characteristics of AML patients with *TTMV::RARA*. Chromosomal abnormalities were detected in 60% (15/25) of patients, among which isochromosome i(17)(q10) (24%, 6/25) and trisomy 8 (8%, 2/25) emerged as the predominant aberrations (Table 1).

Furthermore, systematic screening for recurrent leukemia mutations using multimodal methods including RT-PCR, NGS-targeted sequencing, and bulk RNA-seq analysis identified only one case with FLT3 internal tandem duplication (FLT3-ITD), two cases harboring NRAS p.G12 codon mutations, and two cases with WT1 mutations (Figure 3C). These mutations are lesions commonly observed in APL[24], indicating a distinct mutational landscape for this subtype.

Taken together, these results suggested that while AML patients with *TTMV::RARA* exhibit morphological and immunophenotypic similarities to classic APL, they also displayed significant differences in their cytogenetic, molecular, and clinical profiles.

Transcriptomic Features of AML with *TTMV::RARA*

To determine whether AML patients with *TTMV::RARA* fusion represent a distinct subtype separate from conventional AML and classical APL, we performed a transcriptomic analysis involving 15 *TTMV::RARA* samples.

These were compared against 53 *PML::RARA* (classic APL) samples and 511 non-APL AML samples (Table S3). Unsupervised hierarchical clustering revealed that *TTMV::RARA* samples formed a transcriptionally distinct cluster. While they showed a close relationship with classical APL, a clear separation was observed (Figure 3C), indicating potential molecular differences between these subtypes.

Differential gene expression and functional analysis revealed that AML patients with *TTMV::RARA* exhibited elevated *RARA* expression and robust enrichment of the Wnt signaling pathway (Figure 3D-E and Supplementary Fig. S5A-B), which are associated with cell proliferation and self-renewal[25]. Concurrently, extracellular matrix (ECM) regulators such as *ADAMTS9* and *MMP8* showed marked overexpression (Figure 3D-E and Supplementary Fig. S5A-B), in line with their functions in tissue remodeling and invasive migration[26]. These findings support a dual mechanism model: activation of the Wnt pathway drives clonal expansion, while ECM dysregulation facilitates metastatic dissemination in AML patients with *TTMV::RARA*. The further stratification of the 15 AML samples with *TTMV::RARA* utilizing the same unsupervised clustering approach revealed two distinct subgroups (Figure 3C). These subgroups were differentiated by the presence of isochromosome i(17)(q10), a chromosomal aberration detected in 5 patients (33.3%) through an integrated analysis of G-banding karyotyping and

transcriptomic data. Comparative transcriptomic analysis revealed that the i(17)(q10) positive subgroup (n=5) exhibited significant downregulation of 678 genes (fold-change > 2, adjusted p-value < 0.05) compared to the i(17)(q10) negative subgroup (n=10), with *TP53* showing a 4.2 fold decrease (adjusted p-value = 0.003; Figure 3F). Pathway enrichment analysis (Gene Set Enrichment Analysis, GSEA) demonstrated that the i(17)(q10) positive subgroup was characterized by significant enrichment of pathways related to chromosome centromeric core domains (NES = -2.03, $p<0.001$) and DNA double strand break response (NES = -2.21, $p<0.001$) (Supplementary Fig. S5C). These findings, combined with the known role of *TP53* in DNA repair, suggest a compromised DNA damage repair capacity in the i(17)(q10) positive subgroup. Notably, the prevalence of i(17)(q10) in AML is less than 1%[27], highlighting the unique molecular landscape of AML with *TTMV::RARA*.

These unique features distinguish *TTMV::RARA* patients from those with classic APL and other AML subtypes, warranting further clinical and mechanistic investigation.

Treatments and Outcomes

An analysis of treatment response in AML patients with *TTMV::RARA* revealed an impressive overall complete remission (CR) rate of 96%

following induction therapy (Table 2). However, only 46% of patients achieved CR after the first treatment course. No early deaths occurred within 45 days after induction therapy, indicating a favorable tolerance to the initial therapy. AML patients with *TTMV::RARA* primarily received one of three initial induction regimens: all-trans retinoic acid combined with arsenic trioxide (ATRA+ATO, course duration >14 days, n=9), a short course of ATRA combined with chemotherapy (n=7), and standard AML induction chemotherapy (n=8). As shown in Table 2, the CR rates for the ATRA+ATO, short course ATRA with chemotherapy, and standard chemotherapy were 66.7%, 28.6%, and 50%, respectively, with overall response rates (ORR) of 88.9%, 71.4%, and 62.5%. Although Fisher's exact test did not reveal statistically significant differences in CR rates among the three treatment groups (p = 0.319), the ATRA+ATO group exhibited a relatively higher CR rate and OS compared to the other two groups (Supplementary Fig. S6A-C). The observed trends in treatment response indicate that the ATRA+ATO combination therapy might have a beneficial effect on inducing remission, thereby warranting further investigation with larger sample sizes. Notably, patients with isochromosome i(17)(q10) exhibited significant treatment resistance, with a CR rate of 16.7% (1/6) and ORR of 50% (3/6) (Table 2).

A total of 11 patients underwent hematopoietic stem cell transplantation (HSCT), including 10 allogeneic and 1 autologous HSCT cases. Among

them, 5 presented with myeloid sarcoma or central nervous system involvement, and 1 harbored the i(17q10) chromosomal abnormality, both of which represent high-risk prognostic markers. Clinical outcome analysis revealed that, of the 11 patients, one experienced relapse with multiorgan involvement and died 50 days post-transplantation, while another patient succumbed to disease progression 24 months post-transplantation. The remaining 9 patients were still alive at the last follow-up, with a median follow-up period of 36 months (ranging from 7.7 to 115.5 months). Notably, two of these patients remained in remission for 8 and 9 years post-transplantation, respectively.

Survival analysis revealed 2-year overall survival (OS), event-free survival (EFS), and relapse-free survival (RFS) rates were 84.6% (95% confidence interval [CI]: 59 - 94.8%), 53.6% (95% CI: 30.8 - 71.8%), and 53.8% (95% CI: 31.1 - 72%), respectively (Figure 4A-C). When compared to classical APL patients treated with ATRA+ATO[28-29], AML with *TTMV::RARA* showed significantly lower EFS and RFS despite similar OS rates.

Discussion

This study systematically screened 2,553 cases of AML to provide evidence for the specific integration of TTMV into the human *RARA* gene within a large patient cohort. By integrating clinical and multi-omics data through an international multicenter collaboration, we achieved a thorough characterization of the structural features and clinical manifestations associated with the *TTMV::RARA* fusion gene, thereby providing critical evidence for the establishment of a precision medicine framework for this hematological malignancy

In contrast to the eight previously identified pathogenic viruses[30], TTMV is the first single-stranded DNA virus shown to directly induce oncogenesis by creating a chimeric fusion oncogene. Its carcinogenic mechanism operates independently of classical pathways, such as the expression of viral oncoproteins, genomic instability, or dysregulation of the cell cycle[31], thereby broadening the theoretical framework of viral oncogenesis and offering significant insights into the field of viral oncology.

Our analysis of TTMV integration patterns revealed two key characteristics. Firstly, TTMV can integrate into intron 2 of the *RARA* gene from any region of its own genome and exhibits significant heterogeneity in viral subtypes among different patients. This spatial

randomness and diversity of substrains pose challenges for clinical detection, highlighting the imperative for comprehensive viral integration screening in AML, particularly in cases exhibiting APL-like morphology. Secondly, we observed a marked predilection of TTMV for intron 2 of *RARA*, where we identified a recurrent integration hotspot, designating TTMV MN769771.1 as the predominant pathogenic subtype. MMEJ may be a significant mechanism facilitating this site-specific integration. These findings contribute to a deeper understanding of the interactions between viral and host genomes in leukemogenesis and carry clinical relevance. The identified recurrent integration patterns serve as a specific molecular signature that could aid in the development of diagnostic panels.

While earlier research indicated similarities between AML with *TTMV::RARA* and classical APL[8-17], our investigation has confirmed the distinct clinical and molecular features of this subtype, aligning with the findings of Zhou et al[32]. This subtype predominantly affects pediatric populations and is marked by a notable frequency of extramedullary involvement and recurrent i(17)(q10) abnormalities, which are infrequently observed in classical APL or other AML subtypes[33]. Transcriptomic analysis has further delineated a unique gene expression profile, substantiating its classification as a separate disease entity.

Currently, there is no established induction therapy for this condition. Previous studies have indicated that the TTMV::RARA fusion protein exhibits a dose-dependent response to ATRA[34], however, mutations within the ligand binding domain of *RARA* readily confer treatment resistance[32]. Predictive protein structures of *TTMV::RARA* suggest that C55/C57/C59 residues in TTMV ORF may form arsenic binding sites, thereby conferring sensitivity to ATO[16]. Our research provides direct clinical evidence supporting the efficacy of the ATRA+ATO in these patients, with a discernible trend toward enhanced overall survival, thereby supporting its potential as a first-line therapeutic option.

Utilizing unsupervised clustering analysis, we further categorized AML patients with *TTMV::RARA* into two subgroups exhibiting significant molecular heterogeneity, closely associated with clinical outcomes. The high-risk subgroup, distinguished by a prevalent occurrence of i(17)(q10) abnormalities, exhibited an extremely poor prognosis and a limited response to conventional therapies, including ATRA+ATO. Mechanistic studies revealed that this high-risk cohort displayed downregulated *TP53* expression and activated DNA damage repair pathways, which may contribute to its chemoresistant phenotype. Notably, HSCT demonstrated promising efficacy within this subgroup, presenting a viable strategy to address treatment challenges.

Although the sample size of this study remains limited, it represents one of the largest global cohorts of *TTMV::RARA* cases to date. Our observations regarding clinical manifestations, integration mechanisms, and therapeutic responses align closely with those reported in a separate, substantial cohort of *TTMV::RARA* cases[32], reinforcing the validity of our conclusions and underscoring the importance of multicenter collaboration in studying rare hematological malignancies. To date, nearly 40 cases have been documented across two independent cohorts, indicating that the actual incidence of this malignancy is likely underestimated, primarily due to the absence of routine viral integration screening in current diagnostic workflows.

It should be acknowledged that the limited sample size has constrained the statistical power available for subgroup analyses. The frequency of *FLT3-ITD*, *NRAS/KRAS*, and *WT1* mutations in our cohort was lower than that in previously published cohorts[32]. These variations may reflect underlying population heterogeneity, differences in enrollment criteria, or intrinsic molecular diversity of the disease. Consequently, future initiatives should aim to amalgamate data from multiple centers and increase sample sizes to systematically delineate the clinical-molecular spectrum of this condition and validate the risk stratification model proposed herein. Simultaneously, a thorough investigation of the leukemogenic mechanisms instigated by the TTMV::RARA fusion

protein, the development of highly sensitive diagnostic modalities and targeted therapeutics are paramount objectives for future research.

In conclusion, this study, through a comprehensive examination of the molecular mechanisms and clinical phenotypes associated with the *TTMV::RARA*, establishes it as a novel subtype of AML and proposes an initial risk-stratified diagnostic and therapeutic strategy. Given the unique biological behavior and poor prognosis associated with *TTMV::RARA* positive AML, we advocate for the inclusion of viral genomic testing within next-generation sequencing based clinical diagnostic protocols to enhance detection rates. We are currently conducting further mechanistic studies, with the goal of addressing these critical biological questions and providing more substantive evidence in the future.

References

1. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150(2):264-278.

2. Kihara R, Nagata Y, Kiyoi H, et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia. 2014;28(8):1586-1595.

3. Chen X, Wang F, Zhang Y, et al. Fusion gene map of acute leukemia revealed by transcriptome sequencing of a consecutive cohort of 1000 cases in a single center. Blood Cancer J. 2021;11(6):112.

4. Kayser S, Levis MJ. Updates on targeted therapies for acute myeloid leukaemia. Br J Haematol. 2022;196(2):316-328.

5. Sanz MA, Fenaux P, Tallman MS, et al. Management of acute promyelocytic leukemia: updated recommendations from an expert panel of the European LeukemiaNet. Blood. 2019;133(15):1630-1643.

6. Alcalay M, Zangrilli D, Pandolfi PP, et al. Translocation breakpoint of acute promyelocytic leukemia lies within the retinoic acid receptor alpha locus. Proc Natl Acad Sci U S A. 1991;88(5):1977-1981.

7. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703-1719.

8. Astolfi A, Masetti R, Indio V, et al. Torque teno mini virus as a cause

of childhood acute promyelocytic leukemia lacking PML/RARA fusion. Blood. 2021;138(18):1773-1777.

9. Sala-Torra O, Beppu LW, Abukar FA, Radich JP, Yeung CCS. TTMV-RARA fusion as a recurrent cause of AML with APL characteristics. Blood Adv. 2022;6(12):3590-3592.

10. Chen X, Wang F, Zhou X, et al. Torque teno mini virus driven childhood acute promyelocytic leukemia: The third case report and sequence analysis. Front Oncol. 2022;12:1074913.

11. Chen J, Zhou X, Chen X, et al. Pediatric TTMV::RARA-positive relapsed acute promyelocytie leukemia responsive to venetoclax and achieving long remission after allogenic transplantation. Pediatr Blood Cancer. 2023;70(12):e30665.

12. Chen J, Zhou X, Wang Y, et al. TTMV::RARA-driven myeloid sarcoma in pediatrics with germline SAMD9 mutation and relapsed with refractory acute promyelocytic leukemia. Int J Lab Hematol. 2024;46(1):190-194.

13. Wang L, Chen J, Hou B, et al. Case report of pediatric TTMV-related acute promyelocytic leukemia with central nervous system infiltration and rapid accumulation of RARA-LBD mutations. Heliyon. 2024;10(5):e27107.

14. Tsai HK, Sabbagh MF, Montesion M, et al. Acute Promyelocytic Leukemia With Torque Teno Mini Virus::RARA Fusion: An Approach to

Screening and Diagnosis. Mod Pathol. 2024;37(7):100509.

15. Wang Z, Chen J, Meng J, Zhao M, Liu H, Xiao X. TTMV::RARA-positive acute promyelocytic leukemia with marrow necrosis and central nervous system involvement at disease recurrence. J Clin Exp Hematop. 2024;64(3):237-241.

16. Lou J, Zhang Y, Tao J, et al. Infection-related pathway activation and vulnerability to arsenic trioxide in acute promyelocytic leukemia with *TTMV::RARA*. Leukemia. 2025;39(2):490-494.

17. Xu Q, Peng Y, Sun S, et al. Acute promyelocytic leukemia with *TTMV::RARA* fusion potentially responds to all-trans retinoic acid/arsenic trioxide treatment. Haematologica. 2025;110(6):1426-1431.

18. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-419.

19. Cheng WY, Li JF, Zhu YM, et al. Transcriptome-based molecular subtypes and differentiation hierarchies improve the classification framework of acute myeloid leukemia. Proc Natl Acad Sci U S A. 2022;119(49):e2211429119.

20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

21. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment

tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141.

22. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140(12):1345-1377.

23. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;18(8):495-506.

24. Madan V, Shyamsunder P, Han L, et al. Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. Leukemia. 2016;30(8):1672-1681.

25. Kirstetter P, Anderson K, Porse BT, Jacobsen SE, Nerlov C. Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. Nat Immunol. 2006;7(10):1048-1056.

26. Chastney MR, Kaivola J, Leppänen VM, Ivaska J. The role and regulation of integrins in cell migration and invasion. Nat Rev Mol Cell Biol. 2025;26(2):147-167.

27. Kanagal-Shamanna R, Luthra R, Yin CC, et al. Myeloid neoplasms with isolated isochromosome 17q demonstrate a high frequency of mutations in SETBP1, SRSF2, ASXL1 and NRAS. Oncotarget. 2016;7(12):14251-14258.

28. Kantarjian H, Kadia T, DiNardo C, et al. Acute myeloid leukemia:

current progress and future directions. Blood Cancer J. 2021;11(2):41.

29. de Almeida NAA, Ribeiro CRA, Raposo JV, de Paula VS. Immunotherapy and Gene Therapy for Oncoviruses Infections: A Review. Viruses. 2021;13(5):822.

30. Morales-Sánchez A, Fuentes-Pananà EM. Human viruses and cancer. Viruses. 2014;6(10):4047-4079.

31. Gaglia MM, Munger K. More than just oncogenes: mechanisms of tumorigenesis by human viruses. Curr Opin Virol. 2018;32:48-59.

32. Zhou X, Chen J, Tang YL, et al. Epidemiology, Clinical Features, and Molecular Basis of TTMV::RARA-driven Acute Promyelocytic Leukemia. Blood. 2025 August 14. doi: 10.1182/blood.2025028674 [Epub ahead of print]

33. Zhu HH, Qin YZ, Zhang ZL, et al. A global study for acute myeloid leukemia with RARG rearrangement. Blood Adv. 2023;7(13):2972-2982.

34. Zhou X, Chen X, Chen J, Cao P, Wang F, Zhang Y, et al. The Ingenious Multiple Efficacies of the Integrated Gene Fragment of Torque Teno Mini Virus in Orchestrating Promyelocytic Leukemogenesis Via Hijacking Rara. Blood. 2023;142(Supplement 1):1576.

Table 1. Clinical characteristic of *TTMV::RARA*-AML patients (N=25)	
Characteristic-no (%)	
Age≤18(y)	18 (72)
Male	13 (52)
Clinical presentation-no (%)	
Fever	9 (36)
Bleeding	8 (32)
Blood tests -median (range)	
White blood cell count (×10^9/L)	7.88 (1-41.9)
Hemoglobin (×g/L)	88 (55-113)
Platelet count (×10^9/L)	94 (13-334)
PT (s)	13.55 (11.6-19.6)
APTT (s)	32.6 (23.6-48.2)
Fibrinogen (mg/dL)	196 (52-357)
D-dimer (ug/L)	10660 (1570-38440)
Morphology-no (%)	
APL-like cells -median (range)	77.5 (18.8-99.2)
Hypergranular	13 (52)
Hypogranular	2 (8)
Auer body	8 (32)
Cytogenetics-no (%)	
Normal karyotype	9 (36)
idic(17)(p11.2)	1 (4)
i17(q10)	4 (16)
i17(q10) , +8	2 (8)
Others karyotype	8 (32)
Unkonwn	1 (4)
Myeloid sarcoma-no (%)	**6 (24)**
At diagnosis	4 (67)
At relapse	2 (33)
Central nervous system leukemia-no (%)	**4 (16)**
At diagnosis	3 (75)
At relapse	1 (25)

PT, Prothrombin Time; APTT, Activated Partial Thromboplastin Time.

Table 2. Response to treatment (N=25)	
CR-no. (%)	**24 (96)**
One cycle to CR	11 (46)
Two cycles to CR	11 (46)
>3 cycles to CR	2 (8)
Response to the first induction treatment	**Value**
ATRA+ATO-no. (%)	9
ORR	8 (88.9)
CR	6 (66.7)
PR	2 (22.2)
NR	1 (11. 1)
ATRA[1]+others -no. (%)	7
ORR	5 (71.4)
CR	2 (28.6)
PR	3 (42.8)
NR	2 (28.6)
Standard AML induction chemotherapy-no. (%)	8
ORR	5 (62.5)
CR	4 (50)
PR	1 (12.5)
NR	3 (37.5)
Patients with i(17)(q10)	6
CR -no. (%)	**6 (100)**
One cycle to CR	1(16.7)
Two cycles to CR	3 (50)
>3 cycles to CR	2 (33.3)
Response to the first induction treatment-no. (%)	**Value**
ORR	3 (50)
CR	1 (16.7)
PR	2 (33.3)
NR	3 (50)
Outcomes	
OS rate (2 years,%)	84.6
EFS rate (2 years,%)	53.6
RFS rate (2 years,%)	53.8

CR, Complete Response; ATRA, All-Trans Retinoic Acid; ATO, Arsenic Trioxide; ORR, Objective Response Rate; PR, Partial Response;

NR: No Response; OS, Overall survival; EFS, Event-Free Survival; RFS, Relapse-Free Survival; ATRA[1] indicated that 3 of the 6 patients had a duration of ATRA therapy of less than 8 days, whether continuous or intermittent.

Figure legends

Figure 1. Study design for TTMV screening and characterization of *TTMV::RARA* in AML patients. This schematic illustrates the analytical pipeline for *TTMV::RARA* investigation in AML. Systematic analysis of 25 AML cases with *TTMV::RARA* fusions (10 newly identified from 2,553 screened patients and 15 from published reports) revealed distinct integration patterns of TTMV, along with unique transcriptional profiles and clinical manifestations of AML patients with *TTMV::RARA*.

Figure 2. Characteristics of TTMV virus insertion into the AML genome. (A) TTMV inserts into *RARA* intron 2 at different positions. The red shades indicate sequence homology between viral and host DNA at the insertion site, whereas the hotspot regions denote short genomic segments including insertion sites across multiple samples. **(B)** Types and proportions of TTMV virus substrains inserted into the *RARA* gene among 15 *TTMV::RARA* patients. **(C)** Diagram illustrates the potential alternative splicing pattern of the *RARA*::*TTMV::RARA* fusion gene. **(D)** The length (base pairs) distribution of retained intron 2 sequences before *RARA* exon 3 in the spliced *TTMV::RARA* fusion transcripts. Each column represents a *TTMV::RARA* sample.

Figure 3. The clinical characteristics and transcriptomic landscape of AML with *TTMV::RARA*. (A) The bone marrow aspirate morphology of patients with *TTMV::RARA* showed promyelocytes with mono- to bilobated nuclear contours with dense cytoplasmic purple granules. **(B)** The immunophenotypic features of patients with *TTMV::RARA* showed that the majority of leukemia cells expressed CD38, CD117, CD13, CD33 and MPO. A minority of samples expressed CD34, HLA-DR, CD11b. **(C)** The heatmap displayed the results of unsupervised clustering from the transcriptomic data between 15 *TTMV::RARA* patients, 53 *PML::RARA* patients, and 511 AML patients without *TTMV::RARA* or *PML::RARA*. **(D-E)** Volcano plots showed DEGs between *TTMV::RARA* patients with AML and classic APL, respectively. **(F)** Volcano plots showed DEGs between *TTMV::RARA* patients with i(17)(q10) and those without i(17)(q10).

Figure 4. Clinical outcomes and treatment responses of AML patients with *TTMV::RARA* fusion gene. (A-C) The OS, EFS and RFS of AML patients with *TTMV::RARA*.

Figure 1

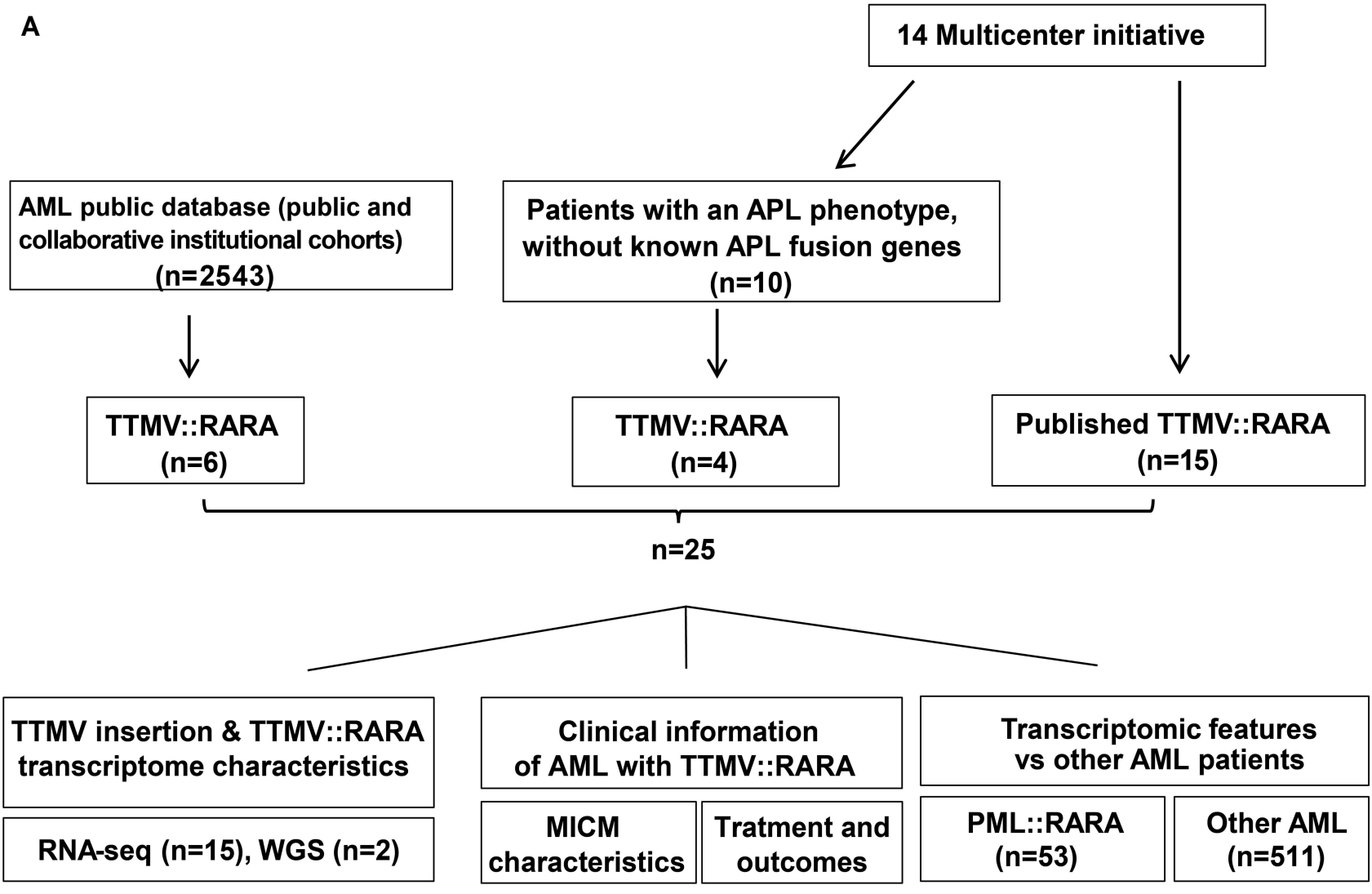

A
14 Multicenter initiative
AML public database (public and collaborative institutional cohorts) (n=2543)
Patients with an APL phenotype, without known APL fusion genes (n=10)
TTMV::RARA (n=6)
TTMV::RARA (n=4)
Published TTMV::RARA (n=15)
n=25
TTMV insertion & TTMV::RARA transcriptome characteristics
RNA-seq (n=15), WGS (n=2)
Clinical information of AML with TTMV::RARA
MICM characteristics
Tratment and outcomes
Transcriptomic features vs other AML patients
PML::RARA (n=53)
Other AML (n=511)

Figure 2

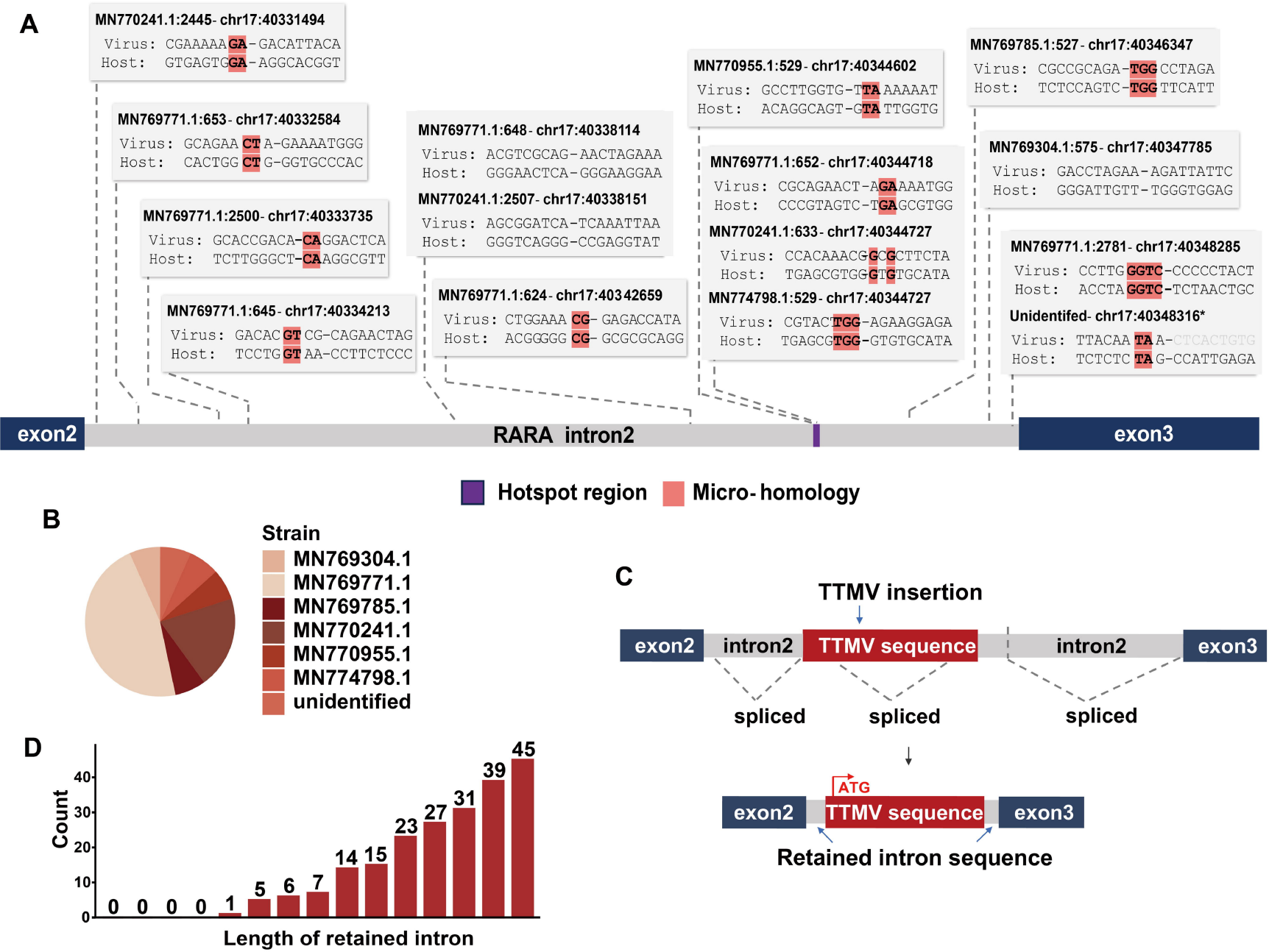

A
MN770241.1:2445- chr17:40331494
Virus: CGAAAAAGA-GACATTACA
Host: GTGAGTGGA-AGGCACGGT
MN769771.1:653- chr17:40332584
Virus: GCAGAACTA-GAAAATGGG
Host: CACTGGCTG-GGTGCCCAC
MN769771.1:2500- chr17:40333735
Virus: GCACCGACA-CAGGACTCA
Host: TCTTGGGCT-CAAGGCGTT
MN769771.1:645- chr17:40334213
Virus: GACACGTCG-CAGAACTAG
Host: TCCTGGTAA-CCTTCTCCC
MN769771.1:648- chr17:40338114
Virus: ACGTCGCAG-AACTAGAAA
Host: GGGAACTCA-GGGAAGGAA
MN770241.1:2507- chr17:40338151
Virus: AGCGGATCA-TCAAATTAA
Host: GGGTCAGGG-CCGAGGTAT
MN769771.1:624- chr17:40342659
Virus: CTGGAAACG-GAGACCATA
Host: ACGGGGGCG-GCGCGCAGG
MN770955.1:529- chr17:40344602
Virus: GCCTTGGTG-TTAAAAAAT
Host: ACAGGCAGT-GTATTGGTG
MN769771.1:652- chr17:40344718
Virus: CGCAGAACT-AGAAAATGG
Host: CCCGTAGTC-TGAGCGTGG
MN770241.1:633- chr17:40344727
Virus: CCACAAACGGCGCTTCTA
Host: TGAGCGTGGGTGTGCATA
MN774798.1:529- chr17:40344727
Virus: CGTACTGGG-AGAAGGAGA
Host: TGAGCGTGG-GTGTGCATA
MN769785.1:527- chr17:40346347
Virus: CGCCGCAGA-TGGCCTAGA
Host: TCTCCAGTC-TGGTTCATT
MN769304.1:575- chr17:40347785
Virus: GACCTAGAA-AGATTATTC
Host: GGGATTGTT-TGGGTGGAG
MN769771.1:2781- chr17:40348285
Virus: CCTTGGGTC-CCCCCTACT
Host: ACCTAGGTC-TCTAACTGC
Unidentifed- chr17:40348316*
Virus: TTACAATAA-CTCACTGTG
Host: TCTCTCTAG-CCATTGAGA
exon2
RARA intron2
exon3
Hotspot region
Micro-homology
B
Strain
MN769304.1
MN769771.1
MN769785.1
MN770241.1
MN770955.1
MN774798.1
unidentified
C
TTMV insertion
exon2
intron2
TTMV sequence
intron2
exon3
spliced
spliced
spliced
ATG
exon2
TTMV sequence
exon3
Retained intron sequence
D
Count
0
10
20
30
40
0 0 0 0 1 5 6 7 14 15 23 27 31 39 45
Length of retained intron

Figure 3

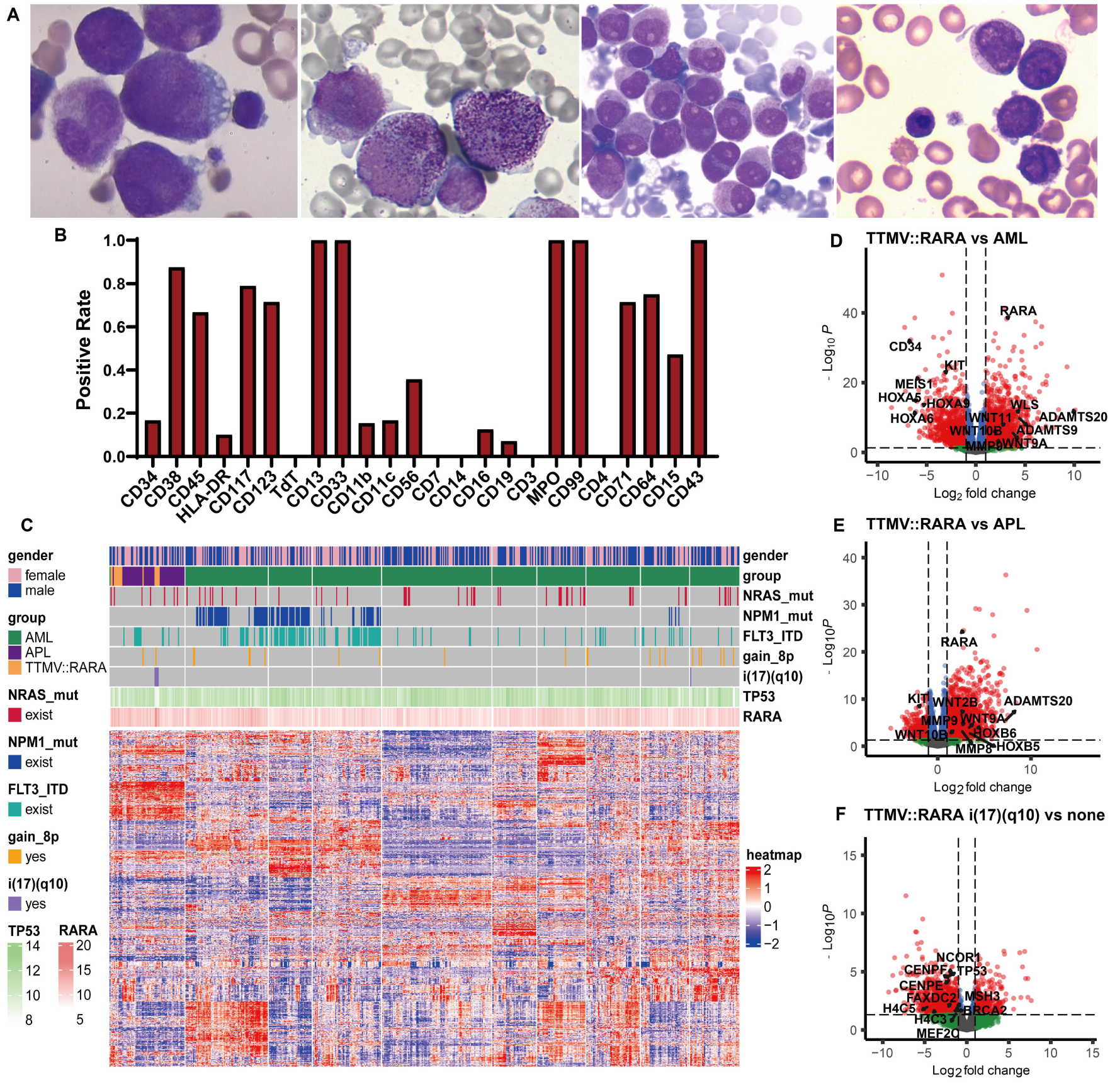
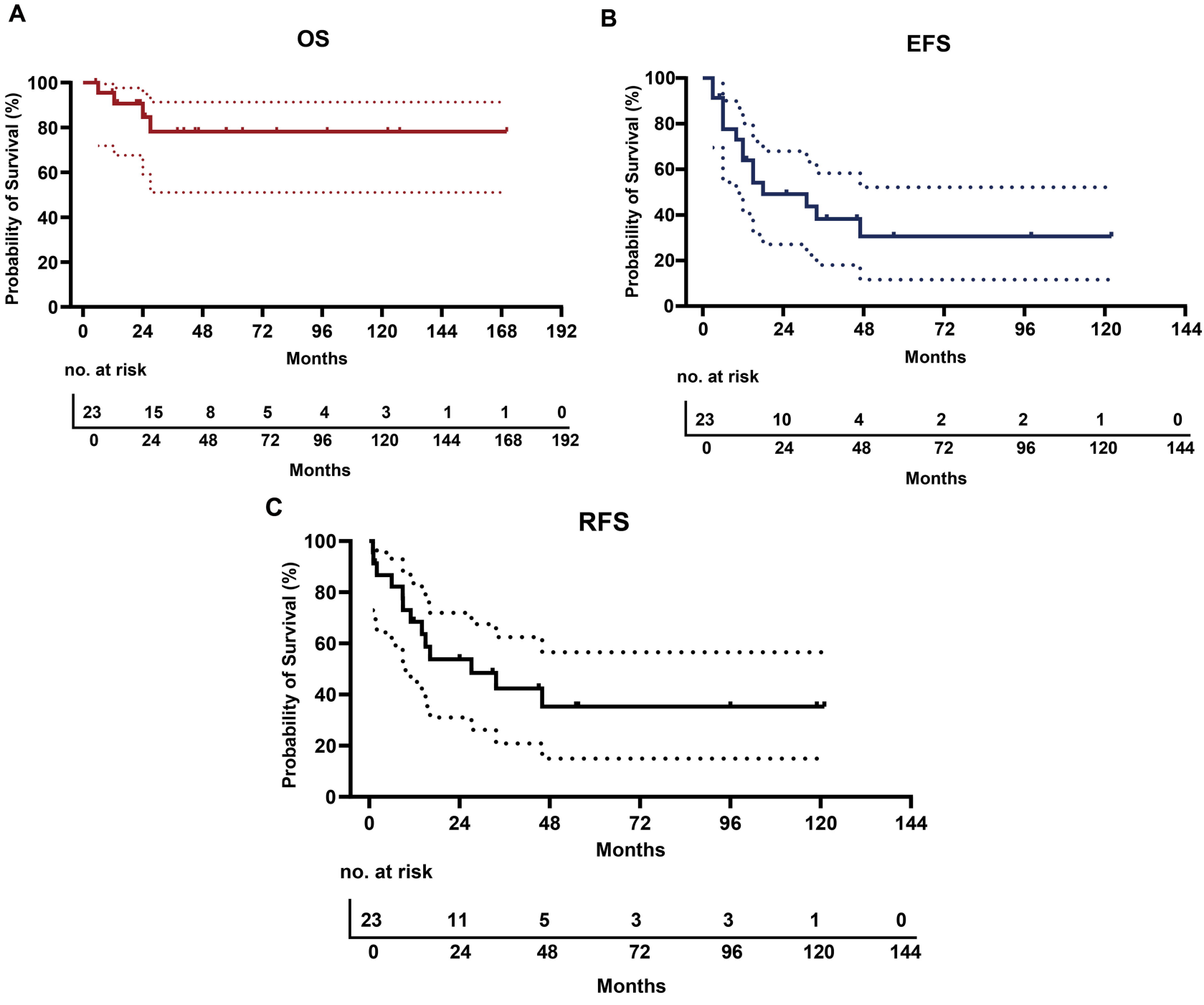

A
B
Positive Rate
1.0
0.8
0.6
0.4
0.2
0.0
CD34
CD38
CD45
HLA-DR
CD117
CD123
TdT
CD13
CD33
CD11b
CD11c
CD56
CD7
CD14
CD16
CD19
CD3
MPO
CD99
CD4
CD71
CD64
CD15
CD43
C
gender
female
male
group
AML
APL
TTMV::RARA
NRAS_mut
exist
NPM1_mut
exist
FLT3_ITD
exist
gain_8p
yes
i(17)(q10)
yes
TP53
14
12
10
8
RARA
20
15
10
5
gender
group
NRAS_mut
NPM1_mut
FLT3_ITD
gain_8p
i(17)(q10)
TP53
RARA
heatmap
2
1
0
−1
−2
D
TTMV::RARA vs AML
- Log10 P
Log2 fold change
RARA
CD34
KIT
MEIS1
HOXA5
HOXA9
HOXA6
WNT11
WLS
ADAMTS20
WNT10B
ADAMTS9
MMP9
WNT9A
E
TTMV::RARA vs APL
- Log10P
Log2 fold change
RARA
KIT
WNT2B
ADAMTS20
MMP9
WNT9A
WNT10B
HOXB6
MMP8
HOXB5
F
TTMV::RARA i(17)(q10) vs none
- Log10P
Log2 fold change
NCOR1
CENPF
TP53
CENPE
FAXDC2
MSH3
H4C5
BRCA2
H4C3
MEF2C

Figure 4


A
OS
Probability of Survival (%)
100
80
60
40
20
0
0
24
48
72
96
120
144
168
192
Months
no. at risk
23
15
8
5
4
3
1
1
0
0
24
48
72
96
120
144
168
192
Months
B
EFS
Probability of Survival (%)
100
80
60
40
20
0
0
24
48
72
96
120
144
Months
no. at risk
23
10
4
2
2
1
0
0
24
48
72
96
120
144
Months
C
RFS
Probability of Survival (%)
100
80
60
40
20
0
0
24
48
72
96
120
144
Months
no. at risk
23
11
5
3
3
1
0
0
24
48
72
96
120
144
Months

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

Part 3: Supplementary Table

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

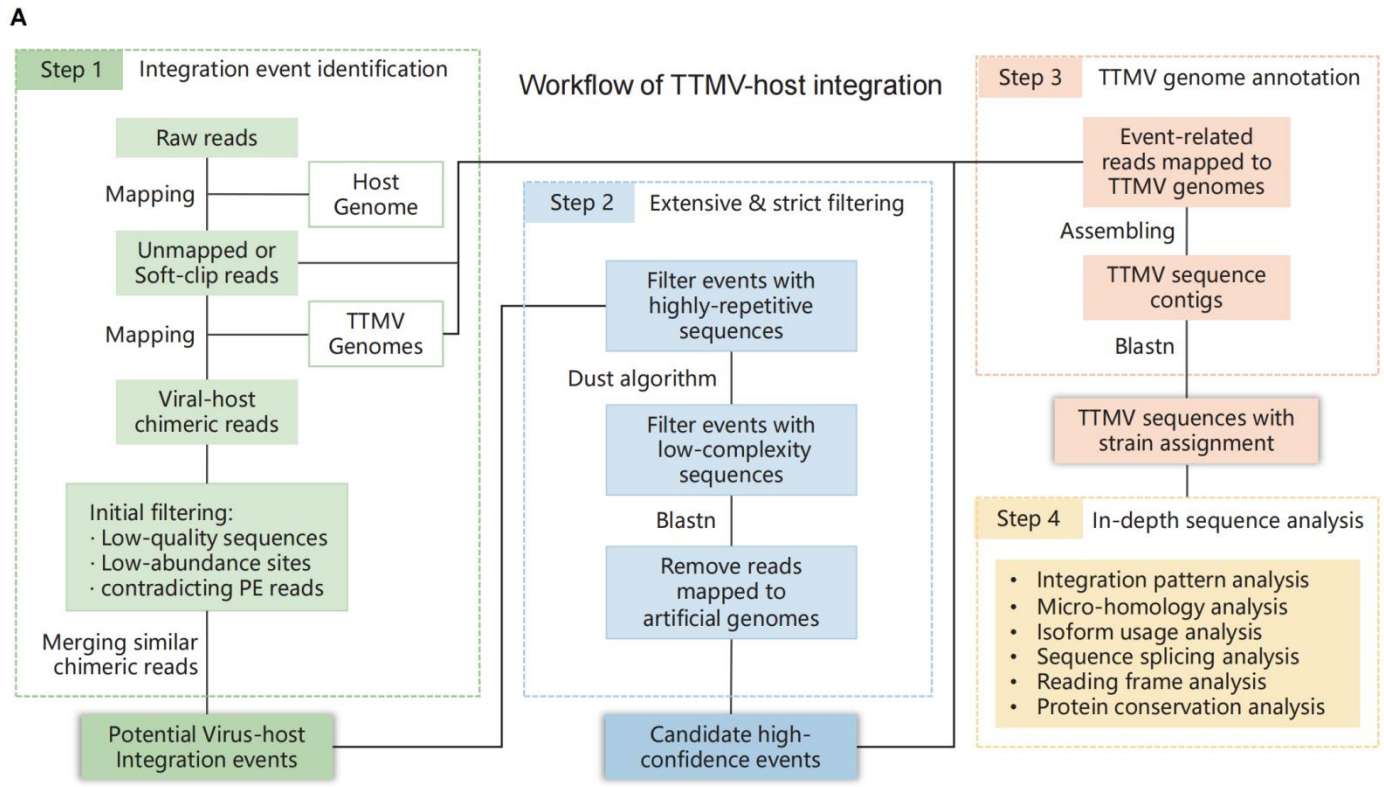


Fig. S1. Modular pipeline for the identification and analysis of virus - host integration events used in this study. A total of 2,553 samples were processed through Step 1 and Step 2 for initial filtering, and the remaining candidate events were subjected to further analysis in Step 3 and Step 4. Detailed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

Fig. S2

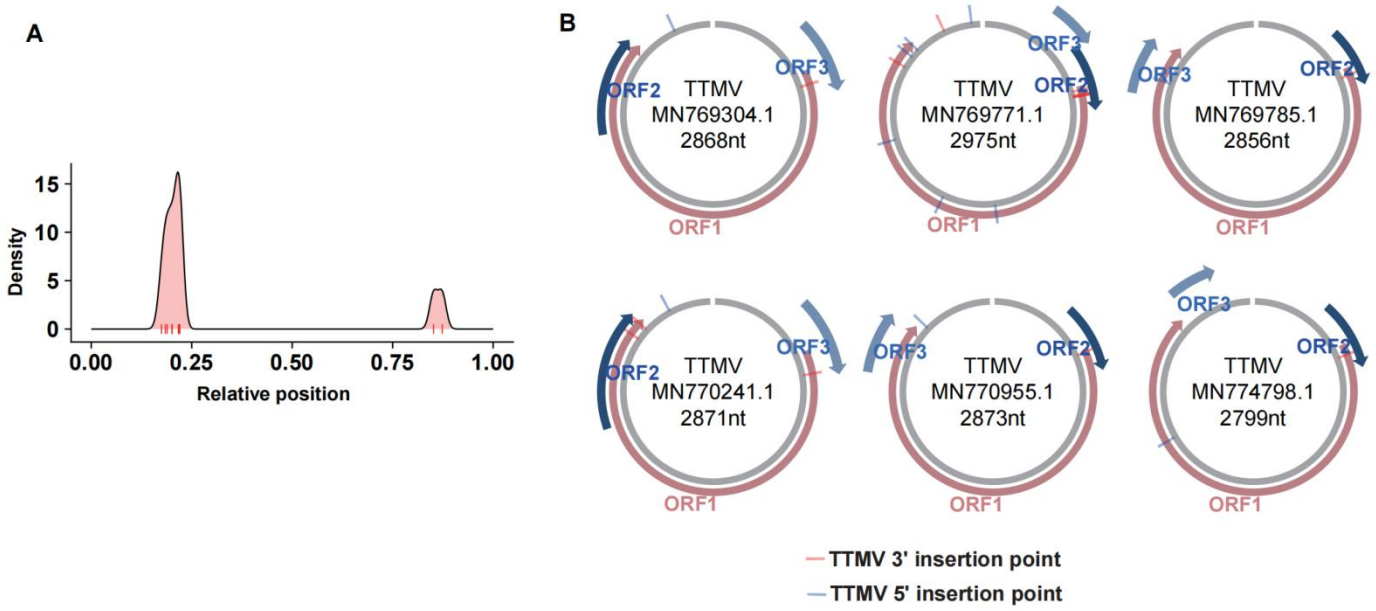


Fig. S2. Illustration of TTMV genomes and integration breakpoints. (A) Distribution of the relative genomic positions of the insertion points at the 3' end of the TTMV genome. The circular genomes are normalized at a [0, 1] interval. (B) The characteristics of the TTMV genomes and the integration sites 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

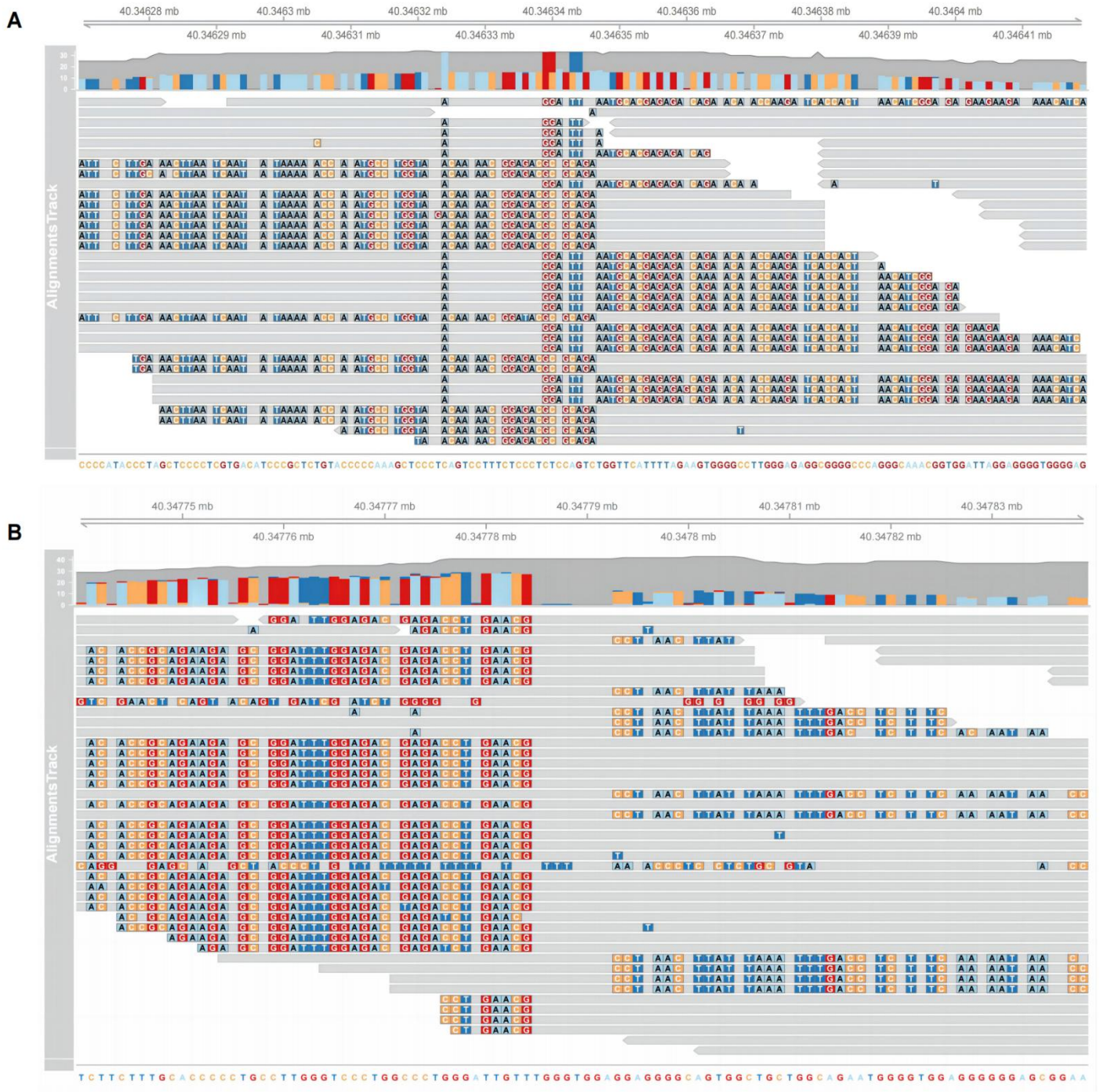


Fig. S3. The misalignment of the 3' and 5' breakpoints when TTMV integrates into the host genome. (A) The 5' breakpoint is positioned in the 3' direction relative to the 3' breakpoint, resulting in the loss of an intronic segment of the host genome. (B) The 5' breakpoint is positioned in the 5' direction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

Fig. S4

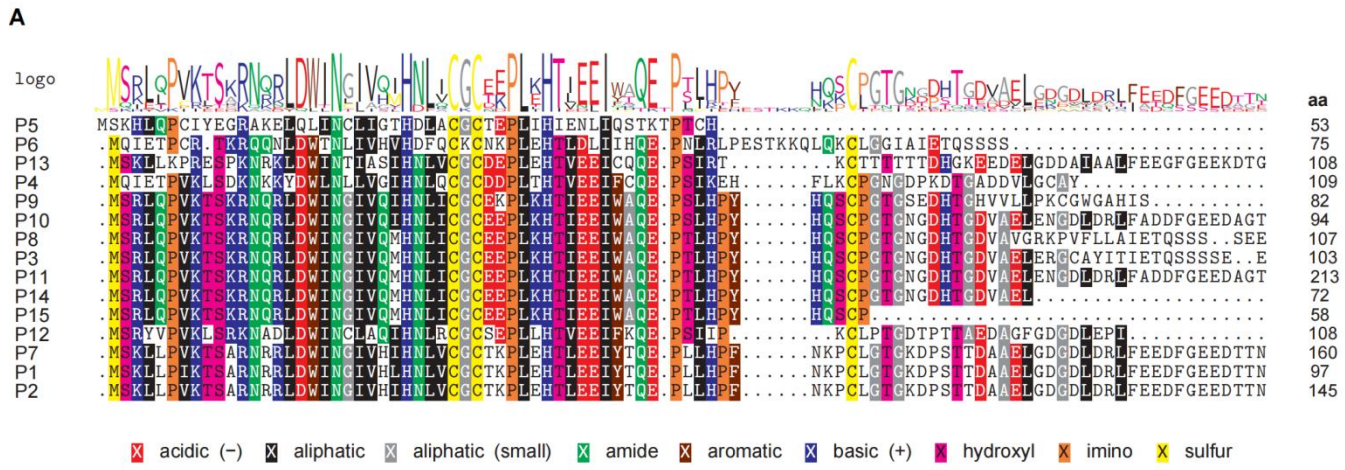


Fig. S4. Multiple sequence alignment of TTMV insertion sequences from the 15 samples. Residues were colored according to chemical properties of their functional groups. The right column indicates the amino acid lengths of TTMV sequence starting from the initial M residue.

Fig. S5

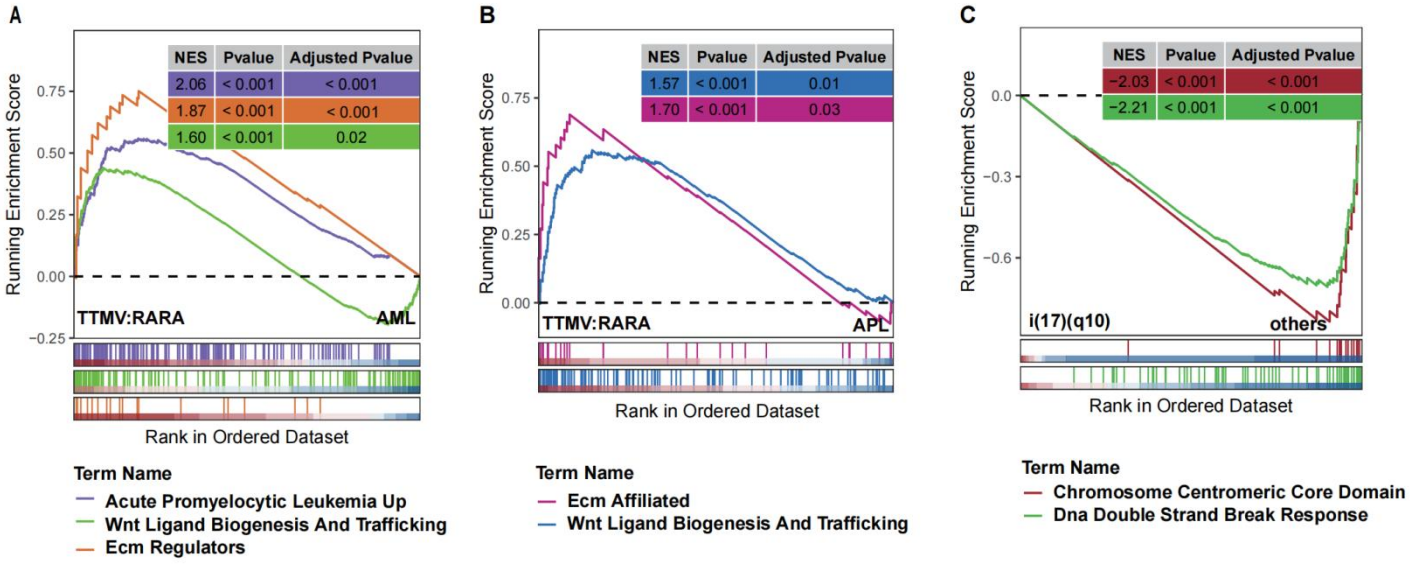


Fig. S5. Gene Set Enrichment Analysis. (A) Functional enrichment of upregulated pathway in AML with *TTMV::RARA* compared to typical AML. (B) Functional enrichment of upregulated pathway in AML with *TTMV::RARA* compared to typical APL. (C) Functional enrichment of downregulated pathway in *TTMV::RARA*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

Fig. S6

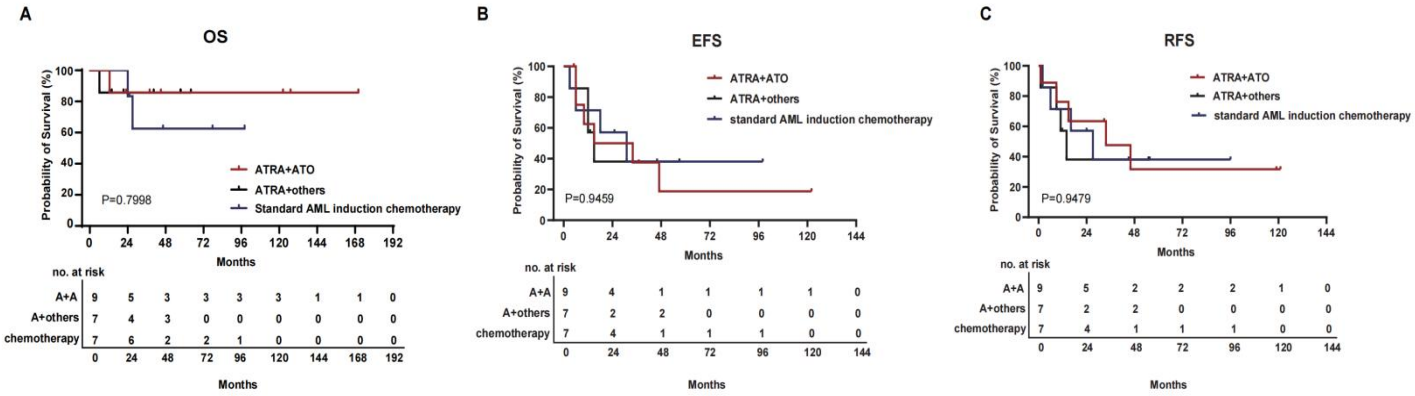


Fig. S6. Treatment outcomes of ATRA+ATO versus other therapies in AML with *TTMV::RARA*. (A) overall survival (OS). (B) Event-free survival (EFS). (C) Relapse-free survival (RFS). ATRA, all-trans retinoic acid; ATO arsenic trioxide.

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	TAATCACCATCATATGACAATTTCCTGGGAGGAGCCACTTACCTATATAA CTAAGTGCACTTCCGAATGGCTGAGTTTATGCCGCCGGACGGAGACGC GATAGGAACTATCAGCGGCTTAGCCTGGGCGGGTGCCGAAGATGAGCA AACTACTACCTATTAAAACATCAGCAAGAAACAGAAGACTCGACTGGAT AAATGGAATAGTTCACCTACACAACCTTGTCTGCGGCTGCACCAAACCA CTGGAACACACCTTAGAAGAAATCTACACTCAAGAACCTCTACTACACC CCTTCAATAAACCATGCCTTGGTACTGGAAAAGACCCCTCTACTACCGA CGCCGCCGAGCTTGGAGACGGCGACCTAGACCGCCTTTTCGAAGAAGA CTTTGGAGAAGAAGACACCACAAACGTGTGTGCATATATAA	MN770241.1	final transcript	full
P2	TAATCACCATCATATGACTCTTTCCTGGGAGGAGCCACTTACCTATATAA CTAAGTGCACTTCCGAATGGCTGAGTTTATGCCGCCGGACGGAGACGC GATAGGAACTATCAGCGGCTTAGCCTGGGCGGGTGCCGAAGATGAGCA AACTACTACCTGTTAAAACATCAGCAAGAAACAGAAGACTAGACTGGAT AAATGGAATAGTTCACATCCACAACCTTGTCTGCGGCTGCACTAAACCA CTGGAACACACCTTAGAAGAAATCTACACTCAAGAACCTCTACTACATC CCTTCAATAAACCATGCCTTGGTACTGGAAAAGACCCCTCTACTACCGA CGCCGCCGAGCTTGGAGACGGAGACCTAGACCGCCTTTTCGAAGAAGA CTTTGGAGAAGAAGACACCACAAACGCCGCTTCTACAAGCGGGAATCA ACCTCCTCAACCCACAGCCAACACAAGACTCATCGGAGGAGACAGAAA CAGAGGAGAAAAGCGAAAAAGAGACACTACAGAAGCTCCTCAAGCAGC TCAAACAGCAACAGCACCGATACAGACAGCGGATCACCGAG	MN770241.1	final transcript	full
P5	CTCCCTAAGTCCTTTCTCCCTGGACTTTCAATGCACGAGAGAACAGAGA CACACCAAGAATCACCACTCCAAACATCGGAGGAGGAAGAAGAGGAAA CATCACTGTTCCACCAGCTCCAGCTCCAGCGAGCCAAGCAGCTCAGAA TCAGACAGCGAATAATATCAACATTACAAAAACTTCAACAATTAGAATAG ACAGAAGCAAAACAAAAGTATACTTATTTCCTCCAAAACCTAAACCTTAC AATAGGTTTAATCCTCAGGAATTACAAACAGAGATAGAGATAGCCAGCT GGTTAAAAAGACCAGTAAGAACATTTAAAGAAGACCCCCCATACTATCC CTGGCTTCCTCCTACTCCTAAAGTACCCTTCCCAAACTTTAACCTTAATT TTACTGAATAAAGGCCTACAATTTTCACTTAGTGGTGTCTGTTTATATTAT TTTCAACTTAAATAAACGTCCACCGCCTCCCAAATACGCAGGCGCAAAA GGGGGCTCCGCCCCCTTAAACCCCCAGGGGGCTCCGCCCCCTAAAAC CCCCAAGGGGGCTCCGCCCCCTTACACCCCCTAATTAATATTCAACAG GAAAACCACCTAATTTAAATTGCCGACCACAAACCGTCAACAAGTTCCT CTTTTTACATTACTTCCTCATTTCCTCATTATTATTCATGACATTAATTAGT AATCACCGTAATTCCGGGGAGGAGACTTAAACCTATATAACTAAGTACA CTTCCGAATGGCTGAGTTTATGCCGCCAGACGGAGACGGGATCACTAC AGTGACTCCAGGCTGACCAAGGGCGGGTGCTGAAGATGAGCAAACATC TCCAACCATGCATCTATGAAGGAAGAGCAAAGGAACTACAATTAATTAA CTGCCTAATTGGAACCCATGATCTTGCCTGTGGCTGCACGGAACCATTA ATTCACATTGAAAACTTAATTCAATCAACTAAAACACCAACATGCCATG	MN769785.1		full
P9	ATGTCAAGACTTCAACCTGTAAAAACTTCCAAAAGAAACCAACGCTTAGA CTGGATTAATGGCATCGTCCAGATACACAACTTAATCTGCGGCTGTGAA	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 GAATCGGAAGAAGAAAAAGAAACATTACAGCTCCTCATCCAGCAACACC 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 AGGATTTGGAGAAGAAAAAGATACTGGAAACGACGGCCATTGAGACCC 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.