

SUMO inhibition reveals temporally distinct intron retention programs shaping the immunopeptidome in acute myeloid leukemia

Pierre Thibault

pierre.thibault@umontreal.ca

University of Montreal <https://orcid.org/0000-0001-5993-0331>

Léa Christophe

University of Montreal

Javier Anleu Alegria

University of Montreal <https://orcid.org/0009-0000-7240-0837>

Mirela Pascariu

Universite de Montreal

Chantal Durette

IRIC, University of Montreal

Éric Bonneil

University of Montreal

Joël Lanoix

Université de Montréal

Mathieu Courcelles

IRIC, Université de Montréal <https://orcid.org/0000-0001-5961-8858>

Marie-Pierre Hardy

IRIC, University of Montreal

Krystel Vincent

IRIC, University of Montreal

Trang Hoang

Institute for research in immunology and cancer

Claude Perreault

Université de Montréal <https://orcid.org/0000-0001-9453-7383>

Article

Keywords:

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11179582/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: **Yes** there is potential Competing Interest. K.V., M.-P.H., P.T. and C.P. are named inventors on patent applications filed by the Université de Montréal and covering Tumor antigens reported in this article (Novel tumor-specific antigens for acute myeloid leukemia and uses thereof No: US20230287070A1 / AU2021256477A1). P.T. and C.P. receive consultant fees from Epitopea. The other authors declare no competing interests.

Abstract

Acute myeloid leukemia (AML) exhibits RNA-processing defects that can expose antigens encoded by non-canonical genomic regions. Here, we combine time-resolved RNA sequencing with discovery and targeted HLA-I immunopeptidomics to determine how SUMOylation controls intron retention and antigen presentation. In three AML cell lines, SUMO inhibition with TAK-981 induced an interferon-associated antigen-processing response and extensively remodeled the immunopeptidome. In MOLM-13 cells, 7,741 dynamically retained introns formed early- and late-rising programs distinguished by intron architecture, splice-site strength and RNA-binding-protein motifs. Among 16,965 HLA-I ligands detected over time, 438 arose from retained introns and 167 mapped to dynamically regulated events, revealing heterogeneous coupling between intron retention and peptide display. TAK-981 further exposed 453 treatment-induced antigens, nearly 60% of which were intron-derived. Targeted mass spectrometry identified 68 sequence-validated intronic ligands from 39 previously unannotated peptideins, with individual retained introns generating multiple, HLA-dependent ligands across AML models. Their source events recurred in untreated primary AML, with subsets enriched relative to healthy hematopoietic samples. These findings establish SUMO-dependent RNA processing as a determinant of non-canonical HLA-I presentation and identify recurrent intron-derived antigens for therapeutic evaluation.

Introduction

Acute myeloid leukemia (AML) is a heterogeneous hematological malignancy characterized by the clonal expansion of differentiation-arrested myeloid progenitors and disruption of normal hematopoiesis¹⁻³. Although intensive chemotherapy, allogeneic hematopoietic stem cell transplantation, and targeted or venetoclax-based therapies have improved outcomes in selected patients, relapse remains frequent and long-term survival is poor, particularly among older adults^{4,5}. This treatment failure is driven in part by therapy-resistant leukemic stem and progenitor cells that survive initial treatment and re-establish the disease. The graft-versus-leukemia effect observed after allogeneic transplantation demonstrates that AML can be eliminated through immune-mediated mechanisms⁶, yet safer immunotherapies remain limited by the scarcity of antigens selectively presented by AML cells.

AML generally carries a low somatic mutation burden, restricting the repertoire of mutation-derived neoantigens available for T-cell recognition⁷. Immune checkpoint inhibitors have shown only modest activity in AML^{8,9}. Their limited efficacy likely reflects heterogeneous antigen presentation, ineffective T-cell priming, and an immunosuppressive bone marrow microenvironment containing regulatory T cells, suppressive myeloid populations, and dysfunctional cytotoxic lymphocytes¹⁰. Checkpoint-ligand expression, including PD-L1, is also heterogeneous and has not consistently predicted response¹¹. Combinations with azacitidine have produced responses in subsets of patients, suggesting that epigenetic remodeling may enhance immune recognition¹². Nevertheless, checkpoint release may be insufficient unless AML cells simultaneously display an adequate repertoire of leukemia-restricted antigens.

Immunopeptidomics provides a direct means of defining this repertoire by identifying peptides displayed by major histocompatibility complex class I (MHC-I) molecules. In a previous study, we combined mass spectrometry (MS)-based immunopeptidomics with proteogenomics to characterize the MHC-I antigen landscape of primary AML specimens¹³. We identified 58 high-confidence tumor-specific antigens (TSAs) whose source transcripts were absent or expressed at very low levels in normal tissues. Notably, 86% originated from genomic regions annotated as non-coding, and nearly half resulted from intron retention. Their source transcripts were shared among patients and expressed in both leukemic blasts and leukemic stem cells. Several TSAs elicited CD8⁺ T-cell responses *in vitro*, while their predicted presentation was associated with T-cell receptor expansion, cytotoxic T-cell activation, immunoediting, and improved survival. These findings established aberrant RNA processing as an important source of immunogenic peptides in AML, but the mechanisms controlling their generation remain poorly understood.

One pathway that may connect AML cell fitness, RNA processing, and immune visibility is protein SUMOylation. This reversible post-translational modification coordinates chromatin organization, transcription, DNA-damage repair, cell-cycle progression, nuclear-body formation, and proteostasis^{14–17}. SUMOylation also contributes to the assembly and function of macromolecular complexes required for pre-mRNA splicing¹⁸. AML cells may depend on this pathway to mitigate transcriptional and proteotoxic stress^{19,20}. Disrupting SUMO conjugation could consequently compromise both leukemic cell survival and RNA-processing fidelity.

TAK-981 (subasumstat) is a first-in-class inhibitor of the SUMO-activating enzyme that rapidly reduces protein SUMOylation²¹. TAK-981 has demonstrated preclinical antileukemic activity by promoting cell-cycle arrest, differentiation, and apoptosis²². It also activates type I interferon signaling and inflammatory programs that can enhance immune recognition. In combination with 5-azacytidine, TAK-981 increased antileukemic activity and natural killer cell-mediated cytotoxicity^{23–25}. Early clinical studies have shown that systemic SUMO inhibition is pharmacologically achievable and induces interferon-associated responses and immune-cell activation²⁶. However, its limited activity as a monotherapy highlights the need to define disease-specific responses that may guide rational combinations.

Because SUMOylation supports the organization of nuclear protein complexes, its inhibition may perturb spliceosome function and promote selective intron retention²⁷. Retained-intron transcripts can undergo nuclear retention or RNA surveillance, reducing canonical protein production, whereas a subset may be translated into short-lived non-canonical proteins or defective ribosomal products that are rapidly degraded and sampled by the MHC-I pathway^{28,29}. We therefore hypothesized that SUMO inhibition could impair AML cell fitness while increasing antigenic visibility by generating non-canonical, intron-derived peptides.

Here, we integrated transcriptomic and quantitative MHC-I immunopeptidomic profiling to define how SUMO inhibition remodels RNA processing and antigen presentation in AML. Across three AML human cell lines, TAK-981 induced an interferon-associated antigen-processing program and redirected HLA-I presentation toward proteins involved in RNA metabolism, translation, chromatin regulation, and proteostasis. Time-resolved analyses in MOLM-13 cells identified 7,741 dynamically retained introns organized into two programs distinguished by their kinetics, intron architecture, splice-site strength, and RNA-binding-protein dependencies. Among 16,965 HLA-I peptides detected over time, 490 originated from alternative splicing, predominantly retained introns, and included peptides generated through multiple transcript-surveillance and translational routes. Targeted immunopeptidomics further identified 68 previously unreported intronic peptides encoded by 39 microproteins, several of which generated multiple HLA ligands across AML models. The recurrence of these intron-retention programs in primary AML samples establishes SUMOylation as a regulator of clinically relevant splicing abnormalities that shape the non-canonical immunopeptidome.

Results

TAK-981 induces an interferon-associated antigen-presentation program in AML cells

We first established conditions that achieved robust inhibition of SUMO conjugation while minimizing nonspecific effects caused by loss of cell viability. Three AML cell lines with distinct cytogenetic, mutational burdens and HLA alleles were examined: MOLM-13, SKM-1 and THP-1 (Supplementary Table 1). Dose- and time-response analyses in THP-1 cells showed progressively greater cytotoxicity with increasing TAK-981 concentrations and exposure times (Supplementary Fig. 1A). Based on these experiments, we selected 20 nM TAK-981 for 48 h as a condition that maintained more than 80% viability while producing measurable pathway inhibition. We subsequently evaluated this condition in all three cell lines. Water-soluble tetrazolium salt 1 (WST-1) measurements for cell counts and viability showed no significant reduction relative to DMSO-treated controls over the 48-h exposure period (Supplementary Fig. 1B). By contrast, immunoblotting revealed a pronounced depletion of SUMO2/3-conjugated proteins in each cell line (Supplementary Fig. 1C).

To examine how SUMO inhibition remodels antigen presentation, we integrated transcriptomics with HLA-I immunopeptidomics and proteogenomic database construction (Fig. 1A). We used paired-end RNA sequencing to quantify gene expression and generate cell line-specific search databases containing canonical proteins and potential treatment-associated translation products. TAK-981-enriched k-mers were assembled into contigs, translated in all reading frames, and combined with personalized canonical proteomes. HLA-I complexes were then isolated with a pan-HLA antibody, and naturally presented MHC-I-associated peptides (MAPs) were identified and quantified by LC-MS/MS. This framework enabled simultaneous assessment of transcriptional responses, global changes in peptide presentation, and MAPs derived from genomic regions not represented in conventional protein databases.

RNA-seq revealed a coordinated immune and inflammatory response to TAK-981. Gene Ontology analysis of genes altered in at least two of the three cell lines identified enrichment for response to type I interferon, antiviral defense, antigen processing and presentation through MHC-I, and positive regulation of T-cell activation (Supplementary Fig. 2A). Protein-protein interaction analysis supported the organization of these genes into connected immune-regulatory modules (Supplementary Fig. 2B). Upregulated nodes included TLR3, AIM2, RGS2 and CD6, consistent with activation of innate sensing and immune-signaling pathways. A smaller downregulated network included CD38 and HMOX1 and was associated with stress-response and apoptosis-related annotations. These data indicate that SUMO inhibition induces broad transcriptional reprogramming rather than activation of a single immune pathway.

The interferon-associated response was accompanied by increased expression of multiple components of the HLA-I antigen-processing machinery. TAP1, TAP2, PSMB8, PSMB9, B2M and HLA class I transcripts were consistently increased after TAK-981 treatment (Supplementary Fig. 2C). Canonical interferon-stimulated genes, including IFIT1, IFITM3, OAS1 and OAS3, were also induced across the three AML cell lines (Supplementary Fig. 2D). The coordinated induction of immunoproteasome subunits, peptide transporters, B2M and HLA-I supports an increased capacity to generate, load and display intracellular MAPs. Together, these findings suggest that TAK-981 shifts cells toward a pro-inflammatory transcriptional state that may enhance immune recognition through increased interferon signaling and MHC-I antigen presentation.

SUMO inhibition remodels the HLA-I immunopeptidome

We next determined whether the transcriptional increase in antigen-processing machinery translated into changes at the level of naturally presented HLA-I peptides. MAPs were isolated from 1×10^8 cells per replicate following 48 h of treatment with TAK-981 or DMSO. In total, 39,424 unique peptide sequences were identified across all cell lines and conditions (Supplementary Table 2). The individual datasets comprised 8,267 quantified MAPs in MOLM-13, 15,463 in SKM-1 and 7,905 in THP-1 (Fig. 1B). Peptide repertoires were highly cell line specific: approximately 85% of the MAPs observed in a given cell line were not detected in the other cell lines consistent with their distinct HLA alleles and their specific binding preferences.

To enable differential-presentation analysis while minimizing bias from incomplete peptide detection and intensity-dependent variance, missing values were imputed by random forest, peptide intensities were $\log_2$ -transformed, and variance-stabilizing normalization was applied (Supplementary Fig. 3). Quantitative comparison of TAK-981- and DMSO-treated cells revealed extensive but cell line-dependent remodeling (Fig. 1C). Peptides were considered differentially presented when their abundance changed by more than two-fold at an adjusted P value of 0.05 or less. In MOLM-13 cells, 425 MAPs increased and 248 decreased after TAK-981 treatment. SKM-1 exhibited the largest response, with 599 increased and 390 decreased MAPs. THP-1 showed a more modest response, comprising 57 increased and 18

decreased peptides. Thus, more MAPs increased than decreased in each model, although the magnitude of remodeling differed between cell lines. These findings are consistent with the shared induction of antigen-processing genes but also indicate that cellular and HLA context determine which source proteins and peptides become preferentially displayed.

To define the biological processes contributing to the remodeled immunopeptidome, we analyzed source proteins of MAPs detected exclusively after TAK-981 treatment or significantly increased in abundance. Both groups were enriched for RNA-centered processes, including pre-mRNA splicing, nuclear RNA export, ribonucleoprotein-complex biogenesis and translation initiation (Fig. 2A). Source proteins of treatment-specific MAPs additionally mapped to chromatin organization, histone modification, transcriptional regulation, DNA repair and telomere organization. By comparison, source proteins of quantitatively increased MAPs were enriched for RNA catabolism, nonsense-mediated mRNA decay, signal recognition particle-dependent co-translational targeting and post-transcriptional gene regulation, as well as chromosome segregation and mitotic processes.

A protein-protein interaction network generated from these source proteins identified highly connected protein hubs (Fig. 2B). These included RNA-splicing and ribonucleoprotein factors such as SNRNP proteins, SF3B1 and HNRNP family members; translation-initiation factors and ribosomal proteins; chaperones and proteasome components; chromatin regulators including KDMs, HDACs and KAT2A; and DNA-damage-response proteins. Related hubs were observed across the three AML models, despite their distinct peptide repertoires. The convergence at the pathway level indicates that SUMO inhibition preferentially redirects HLA-I presentation toward proteins involved in nuclear organization, RNA metabolism, translation, proteostasis and chromatin control. These observations are compatible with altered synthesis, assembly, or turnover of SUMO-regulated protein complexes and provided a rationale to examine whether treatment also exposed peptides from non-canonical translation products.

SUMO inhibition extends a pre-existing intronic antigen repertoire in AML

To define the baseline AML antigen repertoire and benchmark our pipeline against our previous analysis of untreated primary AML samples¹³, we analyzed MAPs detected in time-matched DMSO controls separately from those emerging or increasing after TAK-981 exposure. BamQuery was used to reverse translate each peptide, mapped compatible coding sequences to the genome and prioritized the most likely MAP-coding sequence using RNA-seq support. Expression in the corresponding AML cell line was then compared across untreated AML cohorts, GTEx tissues, myeloid precursor cells (MPCs), blood, bone marrow (BM), and medullary thymic epithelial cells (mTECs), enabling constitutively presented MAPs to be classified as tumor-associated antigens (TAAs) or tumor-specific antigens (TSAs) using established criteria (see Methods)^{30,31}.

Because non-canonical genomic assignments are particularly sensitive to database size and spectral ambiguity, candidate antigens underwent additional validation. Experimental spectra were compared

with Prosit predicted fragmentation patterns, retention times and sequence coverage. Candidates were excluded when an isomeric leucine/isoleucine variant mapped to a canonical protein or when assignment involved highly polymorphic HLA, immunoglobulin or T-cell-receptor loci. Approximately one-third of the initial candidates did not meet these criteria and were removed. To validate the identity of all candidates, we compared their endogenous MS/MS spectra with their corresponding synthetic counterparts. All retained candidates were subsequently validated using synthetic peptides, based on concordance of fragment-ion series and quantitative spectral-similarity metrics (Supplementary File 1).

Prior to TAK-981 treatment, we identified 204 TAAs and 73 TSAs across the three AML cell lines, with SKM-1 contributing the largest number (Fig. 3A). Among the TAAs, 113 were unaffected while 70 increased and 21 decreased significantly following TAK-981 treatment (Supplementary Table 3). More than 95% of TAAs originated from in-frame coding regions (Fig. 3B), although a smaller subset mapped to non-coding exons or other non-canonical regions. TAAs included peptides derived from established cancer-associated proteins, including PRAME and FLT3, further supporting the biological relevance of the workflow.

TSAs displayed a broader genomic distribution, with 57% arising from putatively non-coding regions, predominantly introns, but also exon-intron junctions, intergenic regions and non-coding RNAs (Fig. 3B). The detection of retained-intron-derived TSAs before treatment independently recapitulated our previous observation in primary AML that aberrant intronic translation is an important source of leukemia-restricted antigens¹³.

We assessed the predicted immunogenicity of the TSAs using PRIME 2.0, with viral epitopes and experimentally defined non-immunogenic peptides as reference groups. TSA scores had a median of 0.12 and were significantly higher than those of non-immunogenic peptides (Fig. 3C). More than 30% of the TSAs, including 14 intron-derived peptides, scored at or above the lower bound of the viral antigens (0.15–0.27), and approximately 5% exceeded this range. Although these predictions do not demonstrate T-cell recognition, they prioritize candidates with features associated with immunogenicity. Supporting the biological relevance of this antigen class, we previously demonstrated T-cell immunogenicity for three intron-derived MAPs initially identified in primary AML cells (ALPVALPSL, SLLSGLLRA, and DPLLLRI)¹³.

Most TSA-coding sequences were undetectable or expressed at very low levels in normal tissues, mTECs and MPCs but were detected in AML cell lines and patient samples from two independent AML cohorts (Fig. 3D, 3E and Supplementary Table 3). This restricted expression supports their classification as TSAs and a reduced likelihood of central tolerance or off-tumor recognition. Some TSA-coding regions increased transcriptionally after TAK-981 exposure, whereas others changed in presentation without a corresponding increase in source RNA, implicating additional regulation through RNA processing, translation, protein turnover or HLA loading.

Many TSA-coding sequences were recurrently expressed in the Leucegene and TCGA-AML cohorts, although their abundance varied considerably among patients (Fig. 3E). This heterogeneity makes reliance on a single universal antigen unlikely and instead supports multi-antigen strategies combining recurrent TSAs with targets selected according to each tumor's transcriptome and HLA genotype.

An additional 453 MAPs emerged or increased following TAK-981 treatment and arose from coding sequences rarely detected in untreated normal tissues or AML samples. Because GTEx, Leucegene and TCGA comprise untreated samples and cannot establish whether these sequences are similarly induced by TAK-981 in normal cells, we designated them as TAK-981-induced antigens rather than TSAs. Notably, nearly 60% originated from retained introns (Supplementary Table 3). Thus, SUMO inhibition both remodels the pre-existing intronic TSA repertoire and exposes additional retained-intron-derived antigens. These observations prompted us to determine whether peptide-presentation kinetics track retained-intron accumulation directly or are further shaped by transcript fate, translation and antigen processing.

Time-resolved transcriptomics identifies two distinct programs of intron retention

To determine how SUMO inhibition alters splicing over time, we treated MOLM-13 with 20 nM TAK-981 and performed RNA-seq in triplicate at 0, 12, 24, 36 and 48 h (Fig. 4A). Alternative splicing events were quantified with SpliceWiz³² using the splice-over Percent Spliced-In (PSI). Intron-retention events quantified at four or more of the five time points were fitted with polynomial models, and significant temporal profiles were grouped by fuzzy c-means clustering, retaining events with membership scores > 0.7. Quality-control metrics were consistent, replicates clustered tightly and principal component analysis separated samples by treatment duration for both transcriptomic and splicing analyses (Supplementary Fig. 4A,B). Dynamic IR events were identified from profiles quantified at four or more of the five time points using polynomial fits benchmarked against randomized profiles at a 5% FDR (Supplementary Fig. 4C). Significant profiles were grouped by fuzzy c-means clustering; silhouette analysis supported a two-cluster solution, and only events with membership scores > 0.7 were retained.

Among the alternative-splicing classes examined, intron retention showed the largest absolute temporal change (Fig. 4B). Of approximately 50,000 quantified events, 7,741 met the temporal significance and cluster-membership criteria. Ordering these events by cluster and estimated half-rise time revealed two programs (Fig. 4C). Cluster 1 (late-rising) comprised 5,250 introns that accumulated progressively, with the largest changes at 36–48 h. Cluster 2 (early-rising) contained 2,491 introns that increased earlier but more modestly. Thus, TAK-981 did not uniformly reduce splicing efficiency; distinct intron classes became susceptible at different stages of the response.

To identify features associated with TAK-981 sensitivity, we compared both clusters with efficiently spliced and stably retained introns (Fig. 4D). Responsive introns were generally longer than those in the reference groups. Cluster 1 contained the longest and most AU-rich events and was enriched for internal

introns. Cluster 2 had weaker 5' donor and 3' acceptor scores and was enriched for last and other 3'-proximal introns. The early response therefore preferentially involved introns that were relatively shorter and GC-richer than those in Cluster 1 but had weaker splice sites features associated with susceptibility to intron retention^{33,34}. These results suggest that introns dependent on core splice-site recognition respond rapidly to SUMO inhibition, whereas long, AU-rich internal introns accumulate gradually as auxiliary splicing-factor availability or recycling becomes limiting. The intermediate properties of stably retained introns further support a graded spectrum of susceptibility to SUMO inhibition.

RNA-binding protein (RBP) motif analysis further supported distinct trans-factor requirements between the two clusters (Fig. 4E). Sequences flanking the splice sites were enriched for motifs recognized by SR proteins, HNRNPs, U2AF-associated factors and DDX helicases. Enrichment was stronger in Cluster 1, consistent with its long, AU-rich introns and their reliance on auxiliary splice-site recognition, whereas the weaker, more 5'-biased enrichment in Cluster 2 suggested a greater contribution of suboptimal core splice sites. Several implicated RBPs are known or predicted SUMO substrates^{35–37} or nuclear-speckle components³⁸, providing a plausible link between SUMO inhibition and altered splicing, although the contributions of individual factors remain to be established.

Functional enrichment further distinguished the two programs (Supplementary Fig. 5A). Cluster 1 genes were enriched for ribonucleoprotein-complex biogenesis, RNA processing and nucleocytoplasmic transport, whereas Cluster 2 was associated more strongly with ubiquitin-dependent protein catabolism and proteostasis. Disruption of these processes could influence RNA maturation, translation and antigen processing, thereby contributing to downstream changes in transcript fate and HLA-I presentation.

Both clusters were distributed across multiple chromosomes, indicating a genome-wide response (Supplementary Fig. 5B). Network analysis identified connected modules associated with RNA regulation, transcription and epigenetic control, and antigen processing, accompanied by coordinated expression changes over time (Supplementary Fig. 5C). Together, these data define a large, late-accumulating program involving long, AU-rich internal introns with greater auxiliary-factor dependence and a smaller, earlier program enriched for 5'-proximal introns with weaker splice sites. These distinct properties provide a framework for examining how each program contributes to non-canonical peptide generation.

Distinct intron-retention programs are associated with temporally regulated presentation of intronic HLA-I peptides

To relate RNA-processing changes to antigen presentation, we profiled the MOLM-13 HLA-I immunopeptidome at 0, 12, 24, 36, 48 and 60 h after TAK-981 treatment, using time-matched vehicle controls and three biological replicates per condition. We identified 16,965 unique MAPs across the time course with 12,366 – 14,187 quantified at each time point (Fig. 5A). BamQuery assigned 490 putative

MAPs to alternative-splicing events: 438 to retained introns and 52 to alternative 3'- or 5'-splice-site usage (Fig. 5B, Supplementary Table 4).

Of the 438 retained-intron-derived MAPs, 167 (38.0%) mapped to dynamically regulated IR events: 122 to the late-accumulating Cluster 1 and 45 to the earlier Cluster 2. These MAPs arose from 160 distinct IR events, indicating that some retained introns generated multiple HLA-I ligands. Cluster 1-associated MAPs were approximately 2.9-fold more numerous than those associated with Cluster 2, partly reflecting the larger size and progressive accumulation of Cluster 1. The remaining 271 MAPs mapped to introns without significant temporal regulation, potentially reflecting constitutive presentation, limited quantitative sensitivity, or restricted access of nuclear-retained transcripts to translation.

We next classified the 160 peptide-producing and dynamically related IR events according to their predicted effects on translation and RNA surveillance (Fig. 5C). Four transcript-fate categories were identified: 5'UTR, in which retention occurred upstream of the annotated initiation site; PTC-free, in which retention altered the translated sequence without introducing a premature termination codon (PTC); NMD escape, in which the PTC was positioned to escape nonsense-mediated decay (NMD); and NMD-sensitive, in which the PTC was predicted to trigger NMD. Most events were NMD-sensitive, with smaller subsets predicted to avoid NMD through PTC position, absence of a PTC, or localization within the 5' untranslated region.

These categories suggest several routes by which retained introns can contribute to HLA-I presentation. Because NMD requires translation³⁹, pioneer translation of these transcripts is expected to generate truncated, short-lived polypeptides that are rapidly degraded and may provide efficient substrates for proteasomal processing and MHC-I presentation⁴⁰. Transcripts lacking a PTC or carrying one in an NMD-insensitive position may instead produce extended or alternative isoforms subject to constitutive or quality-control-mediated degradation. Retention within the 5' untranslated region may alter translation initiation or expose intronic initiation sites that generate short upstream open reading frames. Intronic antigen presentation may therefore arise through multiple translation and RNA-surveillance pathways.

Representative examples of each predicted fate are shown in Fig. 5D. For each locus, the transcript structure and genomic position of the retained intron were integrated with the predicted intronic open reading frame, alternative initiation codons, termination codons, and the position of the identified MAP. The examples include an intron retained within the last exon that introduces a PTC but is predicted to escape NMD, an intron-retaining transcript without a PTC that encodes an alternative protein isoform, a retained intron located upstream of the annotated translation-initiation site, and a PTC-generating retained intron predicted to trigger NMD. In each case, the identified MAP mapped within the translated intronic sequence and was supported by its experimental MS/MS spectrum. These examples demonstrate that all four predicted transcript fates can generate detectable HLA-I ligands.

We next examined the expression of these MAP-coding regions across AML and normal tissues (Supplementary File 2). Figure 5E focuses on the examples in Fig. 5D together with additional peptides

derived from intron 1 of the pseudogene *AC018630.2*. Most coding regions were undetectable or expressed at very low levels in GTEx tissues, myeloid precursor cells and mTECs, whereas several recurred in the LeuceGene and TCGA-AML cohorts. Expression varied among AML subtypes and patients, indicating that individual antigens are unlikely to provide universal coverage. When detected in normal blood, bone marrow or other tissues, expression remained below 8.55 RPHM making MAP generation unlikely¹³. Furthermore, low expression in mTECs decreases the risk of central immune tolerance induction⁴¹.

Because data-dependent acquisition (DDA) can miss low-abundance HLA-I ligands, we asked whether individual IR events generated additional MAPs not detected in the discovery dataset. From the 160 peptide-associated IR events, we selected 39 predicted microproteins based on PSI, predicted HLA-binding capacity and recurrence across MOLM-13, SKM-1, and THP-1 cells. These comprised nine events increased by TAK-981 in all three cell lines, five additional high-ranking events from each of THP-1 and SKM-1, and the 20 highest-ranking remaining MOLM-13 events. Predicted strong binders were analyzed by targeted MS in cells treated with vehicle or TAK-981 for 12 or 60 h, using three biological replicates of 5×10^7 cells per condition.

Targeted analysis detected 68 distinct intronic MAPs from 39 predicted microproteins ranging from 15 to 146 amino acids, with 19 IR events generating two or more peptides (Supplementary Table 5). All 68 MAPs were sequence-confirmed using synthetic peptide standards (Supplementary File 1). Retention of intron 5 in *G3BP2*, for example, generated a predicted 63-amino-acid microprotein that yielded four partially overlapping MAPs (FFNVIICL, SILGNMYFY, NYLIDLPLM and YLIDLPLMY) in MOLM-13 cells (Fig. 6A). Targeted analysis confirmed the original identification by data-dependent acquisition (DDA) and revealed three additional ligands. All four increased after 12 h of TAK-981 exposure, although their baseline detection and degree of induction differed.

IR events initially identified in MOLM-13 also generated MAPs in SKM-1 and THP-1 cells. Eleven events produced detectable peptides in at least two cell lines, although not every candidate was targeted in all three models. Retention of intron 9 in *RC3H2*, a late-rising NMD-sensitive event, encoded a predicted 54-amino-acid microprotein that yielded five MAPs (KTPDNINSV, LISNQLHTL, ISNQLHTLL, LFKQTGIIQR and KQTGIIQRI) across the three cell lines (Fig. 6B). Most increased after 60 h of TAK-981 treatment. Thus, a common intron-derived microprotein can generate different HLA-I ligands according to cellular and HLA context.

All 68 MAP sequences were absent from both our previous AML antigen dataset¹³ and the HLA Ligand Atlas, supporting their classification as previously unreported ligands. Their predicted source microproteins also lacked entries in imPi, the Human Microprotein Atlas or OpenProt databases (Supplementary Table 5), supporting their designation as previously unannotated microproteins. These findings show that DDA captures only part of the antigenic output of retained introns. SUMO inhibition may expand the intronic immunopeptidome by increasing the availability of selected retained-intron transcripts and exposing multiple HLA-compatible peptides from each translated microprotein.

Finally, we examined whether the peptide-generating IR events identified in MOLM-13 were also present in primary AML, and how their retention compared with healthy hematopoietic samples. PSI values from the TAK-981-time course were compared with those from selected untreated Leucegene AML samples representing FAB M1-M5 subtypes, including splice-factor-wild-type, SRSF2-mutant and U2AF1-mutant cases, as well as healthy myeloid precursor cells (MPCs), bone marrow and blood (Fig. 7). Both early- and late-rising IR events recurred in primary AML, although their magnitude varied considerably among patients. Their presence in untreated AML indicates that TAK-981 amplifies pre-existing disease-associated splicing abnormalities rather than creating an entirely artificial state.

Comparison with healthy samples revealed two broad patterns. A subset of IR events recurred across primary AML samples but showed little or no retention in healthy hematopoietic cells, consistent with AML-enriched splicing abnormalities; this pattern was particularly apparent among late-rising events. Other events displayed comparable PSI values in AML and healthy samples, suggesting that they represent physiological or broadly shared IR programs rather than leukemia-restricted events. Importantly, detection of a source IR event in healthy cells does not establish its translation or HLA I presentation, which may remain selectively enhanced in AML through differences in transcript fate, translation, antigen processing or HLA genotype. Mean expression values for the corresponding MAP-coding regions across GTEx tissues, Leucegene and TCGA AML cohorts, pooled mTECs, and DMSO- or TAK-981-treated AML cell lines are provided in Supplementary File 2.

The IR events occurred across FAB M1-M5 subtypes, with no single subtype accounting for the complete signature. Some showed higher relative PSI values in particular subtypes or in *SRSF2*- or *U2AF1*-mutant AML, consistent with the effects of these mutations on splice-site recognition. However, the same events were also detected in splice-factor-wild-type samples, indicating that canonical spliceosome mutations are not required. Other contributing mechanisms may include epigenetic state, transcriptional dynamics, RBP availability and post-translational regulation of spliceosome function⁴².

Together with the targeted immunopeptidomic results in Fig. 6, these findings show that recurrent IR events can generate multiple, cell-context-dependent HLA I ligands across AML cell lines and that their source events recur in primary AML. Although peptide presentation was not assessed directly in the Leucegene or healthy cohorts, the relative enrichment of these IR events in AML, together with the previous identification of intron-derived MAPs in primary AML samples¹³, support the biological and translational relevance of the splicing programs exposed by SUMO inhibition.

Discussion

This study establishes SUMOylation as a regulatory interface between RNA processing and HLA-I antigen presentation in AML. TAK-981 induced an interferon-associated antigen-processing program, remodeled the canonical immunopeptidome, and exposed peptides from non-canonical genomic regions. Time-resolved analyses identified two kinetically and structurally distinct intron-retention programs, subsets of which encoded short microproteins that generated multiple HLA-I-associated

peptides. The corresponding IR events recurred across AML cell lines, FAB subtypes and primary LeuceGene transcriptomes, including untreated samples, indicating that SUMO inhibition amplifies rather than solely creates RNA-processing abnormalities intrinsic to AML. Comparison with healthy hematopoietic samples further distinguished AML-enriched events from IR programs shared with normal cells, revealing that TAK-981 acts on both disease-associated and physiological layers of intron retention.

Intronic antigen presentation is not unique to AML. Cancer-associated intron retention has been reported across multiple tumor types, including in the absence of mutations in core splicing factors⁴³, and retained-intron neoepitopes have been detected by MS in melanoma, lymphoma and myeloid cell lines⁴⁴. Non-canonical HLA-I ligands have also been identified in ovarian, colorectal, lung and other cancers^{30,45–47}. AML may nevertheless be particularly enriched for such antigens: in primary AML, 48% of 58 high-confidence TSAs originated from retained-intron translation¹³. This apparent enrichment should be interpreted cautiously because published studies differ in sample input, sequencing depth, database construction, normal-tissue subtraction, and peptide-validation criteria.

Several features of AML provide a plausible basis for extensive intron-derived antigen presentation, including differentiation arrest, epigenetic dysregulation, high transcriptional activity, and frequent perturbation of RNA-processing pathways¹³. Accordingly, TAK-981-responsive introns were detected in *SRSF2*- or *U2AF1*-mutant AML. These patterns suggest that lineage state and splice-factor genotype modulate the magnitude of individual events. Their recurrence in splice-factor-wild-type AML nevertheless demonstrates that canonical spliceosome mutations are not required. Instead, these programs probably reflect interactions between intron architecture and trans-acting constraints such as RBP availability, transcription kinetics and post-translational regulation of spliceosomal complexes. SUMOylation of PRPF3 promotes interactions required for U4/U6·U5 tri-snRNP assembly or recruitment, whereas disruption of the SUMO cycle impairs splicing²⁷. Our findings suggest that SUMO inhibition may disrupt both core splice-site recognition and auxiliary RBP-dependent processing.

The predicted fates of peptide-producing transcripts provide insight into how retained introns enter the HLA-I pathway. Most were classified as NMD-sensitive, supporting a model in which pioneer translation of premature-termination-codon-containing transcripts generates short-lived products suitable for proteasomal processing. However, intronic MAPs also arose from transcripts lacking premature termination codons, carrying last-exon termination codons predicted to escape NMD or retaining introns upstream of the canonical initiation site. Intronic antigen production is therefore not confined to a single NMD-defective-ribosomal-product route. Nuclear export, alternative initiation, translation efficiency, protein turnover, proteasomal cleavage and HLA affinity collectively determine antigen presentation. Intron retention is necessary for these peptides to exist but is insufficient to predict their MHC I presentation.

Targeted MS substantially expanded the antigenic repertoire initially captured by DDA. IR events represented by a single discovery peptide yielded additional sequence-validated MAPs in the same AML

model and distinct HLA-compatible ligands in other cell lines. Overall, this approach identified 68 intronic MAPs derived from 39 predicted microproteins, none of which was represented in the Human Microprotein Atlas or OpenProt. The recovery of multiple MAPs from individual IR events supports the productive translation of antigen-rich polypeptides rather than isolated epitopes and demonstrates how HLA genotype diversifies their presentation. Pending functional characterization, these products are best classified as peptideins – microproteins with indeterminate functional potential ⁴⁸. Peptideins yielding MHC I-ligands exhibit greater evolutionary constraint than those with detectable ligands ⁴⁸. Together with the recurrence of the corresponding IR events across AML cell lines and primary AML transcriptomes, this observation supports their biological relevance but does not establish function. Nevertheless, some peptideins may have roles beyond serving as antigen precursors, as non-canonical ORFs can encode functional proteins required for cancer-cell fitness ⁴⁹, including regulators of membrane and metabolic processes ⁵⁰. Illustrative IR events occurred within *G3BP2* and *RC3H2*, genes involved in RNA fate and ribonucleoprotein-granule biology. *G3BP2* is a stress-granule scaffold whose SUMOylation regulates granule assembly and disassembly ⁵¹, whereas *RC3H2* encodes Roquin-2, a post-transcriptional repressor of immune-regulatory transcripts ⁵². These loci connect intronic antigen production to RNA surveillance, stress adaptation and immune regulation.

These findings suggest that recurrent IR events triggered by TAK-981 may provide broader therapeutic coverage than individual peptides. Individual peptideins generated several HLA-binding MAPs, while the same IR event yielded different ligands across AML models according to HLA context. Such events could support multivalent vaccines or TCR-based therapies tailored to patient HLA genotype. Prioritization will nevertheless require confirmation of peptide presentation on primary blasts and leukemic stem cells, T-cell recognition, clonal conservation and absence from essential normal tissues upon TAK-981 administration.

Finally, the results broaden the therapeutic rationale for SUMO inhibition. TAK-981 has direct antileukemic activity ²², synergizes with 5-azacytidine in AML ²⁴ and activates type-I-interferon-dependent antitumor immunity in preclinical models ²⁶. First-in-human studies have established pharmacological target engagement and immune activation, although monotherapy efficacy was limited in unselected tumors ⁵³. Our data suggest that SUMO inhibition can couple leukemic stress and innate immune activation to expansion of the non-canonical immunopeptidome. Testing TAK-981 with vaccines based on drug-induced intronic MAPs, engineered T cells or checkpoint blockade will determine whether expansion of intronic MAP diversity can be converted into durable antileukemic immunity.

Methods

Reagents

TAK-981, an inhibitor of the SUMO-activating enzyme, was purchased from MedChemExpress (HY-11178). It was dissolved in dimethyl sulfoxide (DMSO; Sigma-Aldrich), aliquoted, and stored at – 80°C for

no longer than six months.

Cell culture

The human acute myeloid leukemia (AML) cell lines MOLM-13 (DSMZ ACC 554), SKM-1 (DSMZ ACC 547), and THP-1 (DSMZ ACC 16) were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ); their characteristics are summarized in Supplementary Table I. After thawing, cells were allowed to recover for seven days and were confirmed to be mycoplasma-free using an InvivoGen detection kit. Cells were maintained at 37°C and 5% CO₂ in humidified incubators in RPMI-1640 medium (ATCC modification; Gibco) supplemented with 10% heat-inactivated fetal bovine serum (Wisent) and 1% penicillin–streptomycin (Gibco).

For each experiment, cultures were divided between DMSO vehicle-control and TAK-981 treatment groups. Unless otherwise stated in the figures or legends, cells were exposed to 20 nM TAK-981 or the corresponding DMSO vehicle for 48 h. Cells were then harvested by centrifugation at 1,100 rpm for 10 min, washed in ice-cold phosphate-buffered saline (PBS), counted, and pelleted. Aliquots of 5×10^6 cells were stored at –80°C for western blotting and MS-based proteomic profiling, whereas aliquots of 1×10^8 cells were stored for downstream immunopeptidomic analyses. Western blotting was used to confirm treatment efficacy. Four biological replicates were collected per condition for the 48-h experiments and three per condition for the time-course experiments.

Western blotting

Cell pellets were lysed in 400 µL of 8 M urea, 20 mM imidazole, 100 mM NaH₂PO₄, 5 mM 2-mercaptoethanol (β-ME), and 10 mM Tris-HCl (pH 6.3). Lysates were vortexed for 40 s, sonicated for 10 s, and centrifuged at 3,700 rpm for 5 min. Protein concentrations were determined by Bradford assay (Bio-Rad). On ice, lysates were mixed with water and 5× Laemmli buffer containing 10% SDS, 30% glycerol, 5% β-ME, 0.02% bromophenol blue, and 250 mM Tris-HCl (pH 6.8), and heated at 95°C for 10 min. Proteins were separated on 4–12% precast Bis-Tris gels (Bio-Rad), transferred to nitrocellulose membranes (VWR), and assessed for transfer quality by Ponceau S staining.

Membranes were blocked for 1 h at room temperature in TBST containing 2.5% bovine serum albumin and 2.5% non-fat milk. Primary antibodies were applied overnight at 4°C and for an additional 1 h at room temperature, followed by secondary antibodies for 1.5 h at room temperature. Membranes were washed three times for 15 min in TBST and developed using ECL Plus (Cytiva); signals were recorded on Fujifilm. Primary antibodies were rabbit anti-HLA-A (Abcam, ab52922; 1:1,000) and mouse anti-SUMO2/3 (Abcam, ab81371; 1:1,000). Secondary antibodies were HRP-conjugated goat anti-rabbit IgG (Cell Signaling Technology, 7074S; 1:10,000) and horse anti-mouse IgG (Cell Signaling Technology, 7076S; 1:10,000), all diluted in blocking buffer.

Cell proliferation assay

DMSO- or TAK-981-treated AML cells were seeded in 96-well plates at 1.5×10^4 cells per well in 100 µL of culture medium. Three biological replicates, each comprising eight technical replicates, were analyzed

per condition. At 3 and 48 h after seeding, 10 μ L of WST-1 reagent (Roche) was added, and cells were incubated for 3 h at 37°C and 5% CO₂. Plates were shaken for 1 min, and absorbance was measured at 450 nm, with a 620-nm reference, using an Infinite 200 Pro reader (Tecan).

RNA extraction and sequencing

Total RNA was isolated from 2×10^6 treated cells using TRIzol (Invitrogen) and the RNeasy Mini kit (Qiagen) according to the manufacturers' recommendations. RNA quality was assessed using an Agilent 2100 Bioanalyzer; all samples had an RNA integrity number of 10. Poly(A)-selected libraries were prepared using the KAPA mRNA HyperPrep kit (Roche), quantified by Qubit, and assessed for size using the Bioanalyzer DNA 1000 assay. Libraries were diluted to 10 nM, normalized by qPCR with the KAPA Library Quantification Kit, pooled at equimolar concentrations, and sequenced as 100-bp paired-end reads on an Illumina NextSeq2000, yielding a mean of 130 million read pairs per sample. Library preparation and sequencing were performed at the IRIC Genomics Platform.

Transcriptomic analyses

Bioinformatic processing. Reads were trimmed with Trimmomatic v0.35⁵⁴ and aligned to GRCh38 using STAR v2.7.5.1a⁵⁵, with GENCODE v33 annotations based on Ensembl 99. Gene-level counts were obtained from STAR, and normalized gene- and transcript-level expression, reported as transcripts per million (TPM), was estimated with RSEM⁵⁶ for stranded libraries. Analyses were performed at the IRIC Bioinformatics Platform.

HLA genotyping. HLA alleles were inferred from next-generation sequencing data using OptiType⁵⁷ and are reported in Supplementary Table 1.

Differential expression. Gene counts were obtained from the aligned FASTQs using RSubread⁵⁸ and normalized to log2 counts-per-million with voom⁵⁹ from the limma package⁶⁰. Within each cell line, fold changes were calculated from log counts-per-million values for TAK-981 relative to DMSO. Genes with $|\log_2 \text{ fold change (FC)}| > 1$ in at least two AML cell lines were considered commonly differentially expressed and subjected to Gene Ontology (GO) enrichment analysis with clusterProfiler⁶¹, using all genes quantified by RNA-seq in each cell line as the background. Commonly dysregulated genes were queried against STRING v11.5⁶² to generate protein–protein interaction networks, which were visualized in Cytoscape v3.9.1⁶³. Subnetworks were grouped using PANTHER⁶⁴ according to UniProt annotations (35) and assigned GO terms.

Database generation

Global cancer databases were generated as previously described^{46,65} by combining canonical and treatment-specific proteomes. For the canonical component, sample-specific exomes were created by incorporating single-nucleotide variants with quality scores > 20 , identified using FreeBayes v1.0.2⁶⁶, into pyGeno v2.0.0⁶⁷. Quantified annotated open reading frames (ORFs; TPM > 0) were translated in silico and appended in FASTA format to the canonical proteome.

For the treatment-specific component, DMSO samples were used to remove reads not induced by TAK-981. RNA-seq reads were divided into 33-nucleotide k-mers, and k-mers occurring fewer than twice in DMSO samples were retained. Overlapping k-mers were assembled into contigs and translated in all three reading frames using an in-house Python script. The resulting peptides were concatenated into sequences of approximately 10,000 amino acids, separated by “JJ” linkers to reduce database entries. For each cell line, the canonical and TAK-981-specific proteomes were then combined into a personalized global cancer database.

Isolation of MHC-associated peptides

W6/32 antibodies (Bio X Cell) were coupled to Protein A Sepharose 4B beads (Invitrogen) and stored at 4°C in PBS (pH 7.2) containing 0.02% NaN₃. Cell pellets (1 × 10⁸ cells per replicate) were resuspended with 0.5 mL PBS and mixed with 1 mL of ice-cold lysis buffer containing 1% CHAPS and protease inhibitor cocktail. After tumbling for 45 min at 4°C, lysates were centrifuged at 16,600g for 20 min at 4°C. Supernatants were transferred to tubes containing 0.5 mL (500 µg) of W6/32–Sepharose beads, and MHC-associated peptides (MAPs) were immunoprecipitated as described previously⁶⁸. Eluted MAPs were dried by vacuum centrifugation and stored at – 20°C until MS analysis.

Mass spectrometry

Dried peptide extracts were resuspended in 4% formic acid (FA), loaded onto a 25-cm IonOpticks Aurora C18 column, and separated on a Vanquish Neo nanoLC system at 300 nL/min using a 2-h gradient from 0% to 38% acetonitrile in 0.2% FA. Peptides were analyzed in positive-ion mode on an Orbitrap Ascend Tribrid mass spectrometer equipped with a Nanospray Flex source operated at 3 kV. Full MS spectra were acquired at 120,000 resolution, followed by data-dependent HCD MS/MS of the most abundant multiply charged precursors over *m/z* 350–760. MS/MS spectra were acquired at 30,000 resolution with a normalized automatic gain control target of 100%, a 59-ms injection time, and a normalized collision energy of 27%. Instrument control used Xcalibur v4.6 (Thermo Fisher Scientific).

MAP identification and quantification

Identification. Database searches were performed in PEAKS XPro v10.6 (Bioinformatics Solutions Inc.)⁶⁹, with precursor and fragment mass tolerances of 10 ppm and 0.01 Da, respectively. Methionine oxidation and asparagine/glutamine deamidation were specified as variable modifications. Results were imported into MAPDP⁷⁰ and filtered to retain unique 8–11-mer peptides with a NetMHCpan-4.1b eluted-ligand rank ≤ 2%⁷¹ at a 5% false discovery rate (FDR). FDR was estimated from PEAKS decoy hits generated by the decoy-fusion strategy.

Quantification. Label-free quantification was performed in PEAKS XPro. After MAPDP filtering, only peptides detected in at least two of three replicates in at least one condition were retained. Missing values were imputed by random forest, and intensities were variance-stabilizing normalized using DEP2⁷².

Differential presentation after 48 h. TAK-981 and DMSO conditions were compared using the moderated *t*-test implemented by the DEP2 test_diff() function, with Benjamini–Hochberg correction. Peptides detected in both conditions were considered differentially presented when $|\log_2FC| > 1$ and adjusted $P \leq 0.05$. Peptides detected in at least two DMSO replicates but no TAK-981 replicate were classified as repressed; those detected in at least two TAK-981 replicates, but no DMSO replicate were classified as induced.

Kinetic analysis. To define temporal MAP profiles, only peptides with abundance ratios available at four or more of six time points were retained. Each profile was fitted with first- to fourth-order polynomials and compared with repeatedly randomized data to estimate curve-fitting FDRs and coefficients of determination (R^2). For each polynomial order, profiles passing a 5% FDR threshold were retained. Fuzzy c-means clustering was performed with Mfuzz⁷³; cluster number and fuzzifier were optimized as recommended⁷⁴, and only MAPs with membership scores > 0.7 were assigned.

Source-gene analysis. GO enrichment was performed with clusterProfiler⁶¹ using source genes of differentially presented or clustered MAPs; the background comprised all source genes represented in each sample's immunopeptidome. Genes shared by at least two cell lines were queried in STRING v11.5 (33), and interaction networks were visualized in Cytoscape v3.9.1⁶³ (34). Subnetworks were grouped using GO terms and UniProt functional annotations⁶⁴.

MAP coding-sequence assignment and quantification

BamQuery³¹ was used to assign a MAP-coding sequence (MCS) and genomic source to each MAP from RNA-seq data. Each peptide was reverse-translated into all possible nucleotide sequences, which were aligned to the human genome. When several loci encoded the same peptide, in-frame protein-coding loci were prioritized. If ambiguity remained and no canonical protein encoded the MAP, the locus with the highest RNA-seq expression was selected. MCS abundance was calculated by summing reads spanning the sequence at its genomic position. Assigned MCSs were quantified across normal tissues (GTEx, medullary thymic epithelial cells, blood, bone marrow, and myeloid precursor cells, MPCs) and AML cohorts (TCGA-AML and Leucegene).

Tumor-antigen candidate selection

Tumor-antigen candidates were classified from MCS expression in cancer and normal tissues as described previously¹³. Tumor-specific antigens (TSAs) required (i) expression < 8.55 reads per hundred million (RPHM) in $> 90\%$ of normal samples, including GTEx tissues except testis, MPCs, medullary thymic epithelial cells, blood, and bone marrow, and (ii) mean expression at least twofold higher in TCGA-AML or Leucegene-AML than in MPCs. Tumor-associated antigens (TAAs) required (i) expression > 8.55 RPHM in $> 10\%$ of samples from any benign tissue and (ii) mean expression at least twofold higher in cancer than in MPCs and other normal samples.

Candidate spectra were compared with spectra predicted by the Prosit 2020 non-tryptic HCD model ⁷⁵, and retention times were evaluated using the Prosit 2019 iRT model and DeepLC ⁷⁶, calibrated against experimental values. Candidates were retained if they had (i) a Prosit spectral angle > 0.60, (ii) an absolute predicted-to-observed retention-time difference ≤ 6 min, and (iii) ≥ 50% sequence coverage by MS/MS fragments. Because leucine and isoleucine cannot be distinguished by standard MS, candidates containing either residue were assessed with MS-Homology in Protein Prospector (<https://prospector.ucsf.edu/mshomology>); candidates with an I/L variant mapping to a canonical protein were excluded.

Immunogenicity prediction

TSA immunogenicity was predicted using PRIME 2.1 ⁷⁷, with known viral and non-immunogenic peptides as positive and negative controls, respectively. Scores were compared using the Kruskal–Wallis test.

Validation with synthetic peptides

All putative TSAs and all intronic MAPs included in the final reported datasets, including peptides detected by global DDA and targeted analyses, were synthesized by GenScript or GL Biochem. Synthetic peptides were dissolved in DMSO at 1 nmol/μL and diluted to 200–500 fmol/μL in 4% formic acid. Peptides were separated on a 25-cm IonOpticks Aurora C18 column over 70 min using an Evosep One coupled to an Orbitrap Ascend Tribrid mass spectrometer. Targeted acquisition was performed using peptide-specific inclusion lists. MS1 scans were acquired at 120,000 resolution, with a normalized AGC target of 100% and a 251-ms maximum injection time. HCD MS/MS scans were acquired at 45,000 resolution, with a normalized AGC target of 5,000%, a 91-ms maximum injection time and a normalized collision energy of 27%. Data were analyzed in PEAKS 12, and synthetic and endogenous spectra were compared using mirror plots generated by an in-house script. Because the endogenous discovery samples and synthetic standards were analyzed using different chromatographic configurations, retention-time co-elution was not used as a criterion for synthetic-standard confirmation. Only peptides satisfying a spectral angle of at least 0.6 were retained in the final putative TSA and intronic MAP datasets.

Splicing analysis and identification of aberrant-splicing-derived neoantigens

Detection and characterization of aberrant splicing. Alternative splicing was analyzed with SpliceWiz ³² using RNA-seq from MOLM-13 cells treated with TAK-981 for 12, 24, 36, or 48 h or with DMSO at 0 h (three biological replicates per condition). The reference was built from Ensembl GRCh38 release 99 using SpliceWiz buildFullRef() function. FASTQ files generated as described above were realigned with STAR through the SpliceWiz STAR_alignExperiment() and processBAM() functions, and the resulting files were used to quantify splice-junction reads, IR, and novel splicing events. Analyses of potential cryptic-antigen sources focused on alternative 3' splice sites, alternative 5' splice sites, and IR events. For temporal analysis, only IR events with PSI values available at four or more of the five RNA-seq time

points were retained. Each profile was fitted with first- to fourth-order polynomial models and compared with randomized profiles to derive order-specific R^2 thresholds corresponding to a 5% FDR. Significant profiles were grouped by fuzzy c-means clustering; the cluster number was selected by silhouette analysis, and only events with membership scores > 0.7 were retained.

Potential trans-acting regulators of clustered IR events were examined by testing RNA-binding-protein motif enrichment within ± 200 nucleotides of both splice sites using RBPmap (<http://rbpmap.technion.ac.il/>). Cis features included 3'- and 5'-splice-site strength, calculated with the maximum-entropy model ⁷⁸, GC content, intron length, and intron position. Genomic distributions were visualized with karyoploteR ⁷⁹.

Analysis of patient samples. Alternative splicing was also evaluated in Leucegene RNA-seq samples randomly selected from FAB M1–M5 AML and from wild-type ($n = 18$), SRSF2-mutant ($n = 5$), or U2AF1-mutant ($n = 5$). A subset of 160 IR events was quantified across these groups using the SpliceWiz pipeline described above. DESeq2 ⁸⁰ was used to compare spliceosome-mutant cases with wild-type controls, with Benjamini–Hochberg correction. Events with $|\log_2FC| > 1$ and adjusted $P \leq 0.05$ were considered differentially expressed.

Analysis of healthy hematopoietic samples. Similarly to patient samples, alternative splicing quantification was performed in myeloid precursor cells (MPCs, $n = 11$), purified blood ($n = 20$) and bone marrow (BM, $n = 11$).

Database construction and intronic-peptide identification. Selected aberrantly spliced isoforms were translated in silico into all possible 8–11-mers of the 3-frames and appended to the personalized database for each AML cell line. Candidates were identified as described above. Intronic peptides were assigned with BamQuery, and their genomic coordinates were intersected with SpliceWiz IR events. Putative ORFs were inferred by considering canonical and near-cognate start codons and canonical stop codons in all three translated intronic reading frames. Because most predicted ORFs encoded fewer than 100 amino acids, they were designated intron-derived microproteins.

Targeted detection of MAPs from intron-derived peptideins. The 160 IR events modulated over the MOLM-13 TAK-981 time course and yielding at least one intronic peptide were evaluated in all three AML cell lines. Within each cell line, events were ranked by mean percent-spliced-in value and by the numbers of predicted weak and strong HLA binders encoded by the putative intron-derived microprotein. The final targeted panel contained 39 peptideins: all nine events consistently upregulated by TAK-981 in the three cell lines, the five highest-ranked additional events in THP-1, the five highest-ranked additional events in SKM-1, and the 20 highest-ranked remaining events in MOLM-13. Cell-line-specific inclusion lists comprised predicted strong binders from common and corresponding cell-line-specific microproteins (MOLM-13: 455, SKM-1: 114, THP-1: 77 peptides).

MOLM-13, SKM-1, and THP-1 cells were treated with DMSO or TAK-981 and harvested at 12 and 60 h, the time points associated with maximal presentation of peptides from the selected IR events. Three

biological replicates of 5×10^7 cells were collected per condition. MAPs were isolated by immunoprecipitation as described above and analyzed by parallel reaction monitoring using cell-line-specific inclusion lists and the acquisition parameters used for synthetic peptides.

Declarations

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium⁸¹ via the PRIDE partner repository⁸² at <http://www.ebi.ac.uk/pride> with the dataset identifier PDX084611. The transcriptomic data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus⁸³ and are accessible through GEO website: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE348672> (Token: olkvocugplgdzix).

Reviewer access details

For mass spectrometry data log in to the PRIDE website using the following details:

PRIDE website: <http://www.ebi.ac.uk/pride>

Project accession: PDX084611

Token: yZGsUPLabFpo

For transcriptomic data log in to the GEO site using the following details:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE348672>

GEO accession: GSE348672

Token: olkvocugplgdzix

Competing interests:

K.V., M.-P.H., P.T. and C.P. are named inventors on patent applications filed by the Université de Montréal and covering Tumor antigens reported in this article (Novel tumor-specific antigens for acute myeloid leukemia and uses thereof No: US20230287070A1 / AU2021256477A1). P.T. and C.P. receive consultant fees from Epitopea. The other authors declare no competing interests.

Additional information

The online version contains supplementary material available at XXX

Author contributions:

Conceptualization: L.C., P.T. and C.P. Methodology: L.C., M.P., J.L., C.D., E.B., J.A.A., M.-H.P., M.C. and P.T. Formal analysis: L.C., J.A.A., M.-P.H., M.C. and P.T. Investigation: L.C., M.P., J.L., C.D. K.V. and E.B. Resources: P.T. and C.P. Writing original draft: L.C. and P.T. Writing review and editing: all authors. Visualization: L.C., J.A.A., M.-P.H., M.C. and P.T. Project administration: P.T. and C.P. Supervision: P.T. and C.P.

Acknowledgements

This study was supported by grants from the Canadian Cancer Society (705604) to C.P. and P.T. and the National Science and Engineering Research Council (NSERC 311598) to P.T. We thank Raphaëlle Lambert from the IRIC genomics facility and the McGill Genome Center for RNA-seq. We are grateful to Robin Minati and Joey Paradis for assistance with preliminary experiments on THP-1 cells at the FACS. We also acknowledge the assistance of Gregory Ehx for valuable discussions and Jean-Philippe Laverdure for essential bioinformatics analyses. We thank the IRIC proteomics facility for MS analyses conducted in this study and the Genotype-Tissue Expression (GTEx) Project for providing RNA-seq data. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. We thank the TCGA Research Network for the data generation upon which the results shown here are based: <https://www.cancer.gov/tcga>.

References

1. Dohner H et al (2017) Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 129:424–447
2. Fennell KA, Bell CC, Dawson MA (2019) Epigenetic therapies in acute myeloid leukemia: where to from here? *Blood* 134, 1891–1901
3. Kantarjian H et al (2024) Current status and research directions in acute myeloid leukemia. *Blood Cancer J* 14:163
4. Leotta S et al (2022) Prevention and Treatment of Acute Myeloid Leukemia Relapse after Hematopoietic Stem Cell Transplantation: The State of the Art and Future Perspectives. *J Clin Med* 11
5. Nazha A, Ravandi F (2014) Acute myeloid leukemia in the elderly: do we know who should be treated and how? *Leuk Lymphoma* 55:979–987
6. Socie G (2025) Challenges in GVHD and GVL after hematopoietic stem cell transplantation for myeloid malignancies. *Blood* 146:926–937
7. Saini SK et al (2022) Neoantigen reactive T cells correlate with the low mutational burden in hematological malignancies. *Leukemia* 36:2734–2738
8. Bewersdorf JP, Stahl M, Zeidan AM (2019) Immune checkpoint-based therapy in myeloid malignancies: a promise yet to be fulfilled. *Expert Rev Anticancer Ther* 19:393–404

9. Gomez-Llobell M, Peleteiro Raindo A, Climent Medina J, Gomez Centurion I (2022) Mosquera Orgueira, A. Immune Checkpoint Inhibitors in Acute Myeloid Leukemia: A Meta-Analysis. *Front Oncol* 12:882531
10. Reinhardt C, Ochsenbein AF (2025) Immune checkpoints regulate acute myeloid leukemia stem cells. *Leukemia*
11. Sun C, Mezzadra R, Schumacher TN (2018) Regulation and Function of the PD-L1 Checkpoint. *Immunity* 48:434–452
12. Daver N et al (2018) Hypomethylating agents in combination with immune checkpoint inhibitors in acute myeloid leukemia and myelodysplastic syndromes. *Leukemia* 32:1094–1105
13. Ehx G et al (2021) Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* 54:737–752e710
14. Boulanger M, Paolillo R, Piechaczyk M, Bossis G (2019) The SUMO Pathway in Hematomalignancies and Their Response to Therapies. *Int J Mol Sci* 20:3895
15. Han ZJ, Feng YH, Gu BH, Li YM, Chen H (2018) The post-translational modification, SUMOylation, and cancer (Review). *Int J Oncol* 52:1081–1094
16. Tharuka MDN, Courelli AS, Chen Y (2025) Immune regulation by the SUMO family. *Nat Rev Immunol*
17. Vertegaal ACO (2022) Signalling mechanisms and cellular functions of SUMO. *Nat Rev Mol Cell Biol* 23:715–731
18. Pozzi B, Mammi P, Bragado L, Giono LE, Srebrow A (2018) When SUMO met splicing. *RNA Biol* 15:689–695
19. Bossis G et al (2014) The ROS/SUMO axis contributes to the response of acute myeloid leukemia cells to chemotherapeutic drugs. *Cell Rep* 7:1815–1823
20. Chen X et al (2021) Hyper-SUMOylation of ERG Is Essential for the Progression of Acute Myeloid Leukemia. *Front Mol Biosci* 8:652284
21. Langston SP et al (2021) Discovery of TAK-981, a First-in-Class Inhibitor of SUMO-Activating Enzyme for the Treatment of Cancer. *J Med Chem* 64:2501–2520
22. Kim HS et al (2023) TAK-981, a SUMOylation inhibitor, suppresses AML growth immune-independently. *Blood Adv* 7:3155–3168
23. Chen Y (2023) A new immuno-oncology target - SUMOylation. *Trends Cancer* 9:606–608
24. Gabellier L et al (2024) SUMOylation inhibitor TAK-981 (subasumstat) synergizes with 5-azacytidine in preclinical models of acute myeloid leukemia. *Haematologica* 109:98–114
25. Lam V et al (2023) T Cell-intrinsic Immunomodulatory Effects of TAK-981 (Subasumstat), a SUMO-activating Enzyme Inhibitor, in Chronic Lymphocytic Leukemia. *Mol Cancer Ther* 22:1040–1051
26. Lightcap ES et al (2021) A small-molecule SUMOylation inhibitor activates antitumor immune responses and potentiates immune therapies in preclinical models. *Sci Transl Med* 13:eaba7791
27. Pozzi B et al (2017) SUMO conjugation to spliceosomal proteins is required for efficient pre-mRNA splicing. *Nucleic Acids Res* 45:6729–6745

28. Vendramin R et al (2026) Nonsense-mediated mRNA decay inhibition reshapes the cancer immunopeptidome. *Immunity* 59:1398–1421e1320
29. Weller C et al (2025) Translation dysregulation in cancer as a source for targetable antigens. *Cancer Cell* 43:823–840e818
30. Apavaloaei A et al (2025) Tumor antigens preferentially derive from unmutated genomic sequences in melanoma and non-small cell lung cancer. *Nat Cancer* 6:1419–1437
31. Cuevas MVR et al (2023) BamQuery: a proteogenomic tool to explore the immunopeptidome and prioritize actionable tumor antigens. *Genome Biol* 24:188
32. Wong ACH, Wong JJ, Rasko JEJ, Schmitz U (2023) SpliceWiz: interactive analysis and visualization of alternative splicing in R. *Brief Bioinform* 25
33. Monteuuis G, Wong JJJ, Bailey CG, Schmitz U, Rasko JE (2019) J. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res* 47:11497–11513
34. Naro C et al (2024) Splicing targeting drugs highlight intron retention as an actionable vulnerability in advanced prostate cancer. *J Exp Clin Cancer Res* 43:58
35. Hendriks IA et al (2018) Site-specific characterization of endogenous SUMOylation across species and organs. *Nat Commun* 9:2456
36. Lamoliatte F et al (2014) Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat Commun* 5:5409
37. Li C et al (2020) Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility. *Nat Commun* 11:834
38. Wu J et al (2024) Dynamics of RNA localization to nuclear speckles are connected to splicing efficiency. *bioRxiv*
39. Maquat LE, Tarn WY, Isken O (2010) The pioneer round of translation: features and functions. *Cell* 142:368–374
40. Apcher S, Daskalogianni C, Fahraeus R (2015) Pioneer translation products as an alternative source for MHC-I antigenic peptides. *Mol Immunol* 68:68–71
41. Dong H, Zhou S, Chen X, Deng X, Fang A (2024) Pan-cancer analysis of the prognostic significance of ACKR2 expression and the related genetic/epigenetic dysregulations. *Expert Rev Clin Immunol* 20:225–236
42. Porras-Tobias AL, Caldera A, Castro-Piedras I (2025) Intron Retention: A Reemerging Paradigm in RNA Biology and Post-Transcriptional Gene Regulation. *Genes (Basel)* 16
43. Dvinge H, Bradley RK (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* 7:45
44. Smart AC et al (2018) Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 36:1056–1058
45. Bedran G et al (2023) The Immunopeptidome from a Genomic Perspective: Establishing the Noncanonical Landscape of MHC Class I-Associated Peptides. *Cancer Immunol Res* 11:747–762

46. Courcelles M et al (2026) Noncanonical Transcription and Splicing Shape the Colorectal Cancer Immuno-peptidome in MSI and MSS Tumors. *Mol Cell Proteom* 25:101581
47. Raja R et al (2025) Immunogenic cryptic peptides dominate the antigenic landscape of ovarian cancer. *Sci Adv* 11:eads7405
48. Deutsch EW et al (2026) Expanding the human proteome with microproteins and peptideins. *Nature* 654:813–825
49. Prensner JR et al (2021) Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat Biotechnol* 39:697–704
50. Polycarpou-Schwarz M et al (2018) The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. *Oncogene* 37:4750–4768
51. Yuan Y et al (2025) TRIM28-mediated SUMOylation of G3BP1/2 regulates stress granule dynamics. *Cell Chem Biol* 32:1336–1352e1336
52. Vogel KU et al (2013) Roquin paralogs 1 and 2 redundantly repress the Icos and Ox40 costimulator mRNAs and control follicular helper T cell differentiation. *Immunity* 38:655–668
53. Juric D et al (2025) A First-In-Human Study of the SUMOylation Inhibitor Subasumstat in Patients with Advanced/Metastatic Solid Tumors or Relapsed/Refractory Hematologic Malignancies. *Cancer Res Commun* 5:2025–2038
54. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
55. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
56. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527
57. Szolek A et al (2014) OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 30:3310–3316
58. Liao Y, Smyth GK, Shi W (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 47:e47
59. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15:R29
60. Ritchie ME et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47
61. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284–287
62. Szklarczyk D et al (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 51:D638–D646

63. Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504
64. Thomas PD et al (2022) Making genome-scale phylogenetics accessible to all. *Protein Sci* 31:8–22
65. Laumont CM et al (2018) Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* 10
66. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907
67. Daouda T, Perreault C, Lemieux S, pyGeno (2016) A Python package for precision medicine and proteogenomics. *F1000Res* 5, 381
68. Lanoix J et al (2018) Comparison of the MHC I Immunoepitidome Repertoire of B-Cell Lymphoblasts Using Two Isolation Methods. *Proteomics* 18:e1700251
69. Zhang J et al (2012), *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics* 11, M111 010587
70. Courcelles M et al (2020) A Cloud-Based Computational Platform for Immunoepitidomics Analyses. *J Proteome Res* 19:1873–1881
71. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 48:W449–W454
72. Feng Z, Fang P, Zheng H, Zhang X (2023) DEP2: an upgraded comprehensive analysis toolkit for quantitative proteomics data. *Bioinformatics* 39
73. Kumar L, M EF (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2:5–7
74. Schwammle V, Jensen ON (2010) A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* 26:2841–2848
75. Gessulat S et al (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* 16:509–518
76. Bouwmeester R, Gabriels R, Hulstaert N, Martens L, Degroeve S (2021) DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *Nat Methods* 18:1363–1369
77. Gfeller D et al (2023) Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8(+) T-cell epitopes. *Cell Syst* 14:72–83e75
78. Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377–394
79. Gel B, Serra E (2017) karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33:3088–3090
80. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550

81. Deutsch EW et al (2020) The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res* 48:D1145–D1152

82. Perez-Riverol Y et al (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47:D442–D450

83. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210

Figures

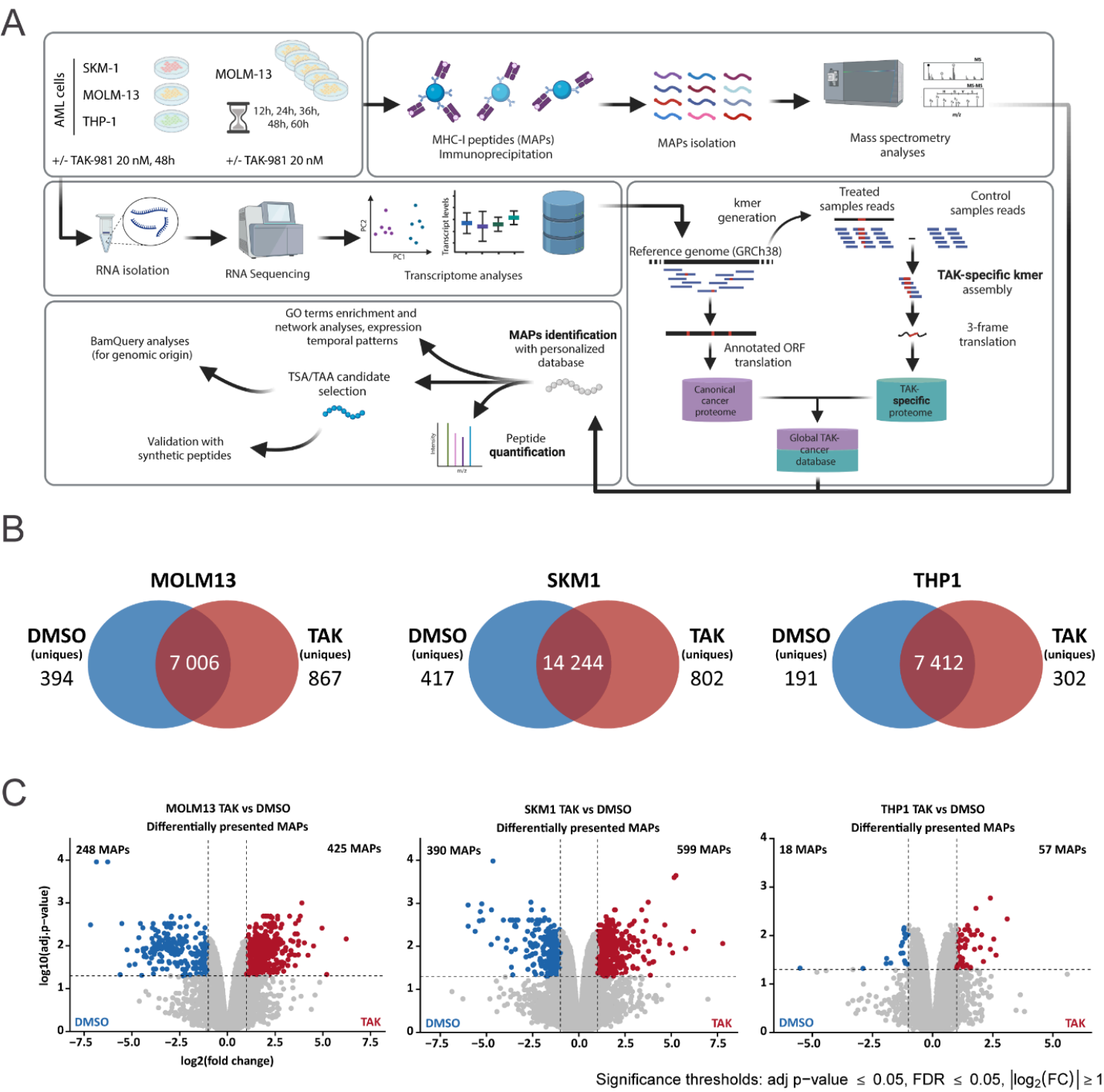


Figure 1

Application of a proteogenomic workflow reveals that TAK-981 induced reshaping of the AML immunopeptidome and enhanced MHC class I peptide presentation. A) AML cells (SKM-1, MOLM-13, THP-1) were treated with 20 nM TAK-981 or DMSO for 48h. A time-course experiment was also performed in MOLM-13 cells treated with 20 nM TAK-981 for 0 to 60 hours. RNA sequencing was conducted to enable transcriptomic profiling and to generate cell-specific customized proteome databases. MAPs isolated via immunoprecipitation were analyzed by LC-MS/MS, and spectra were searched against the corresponding customized databases using PEAKS XPro. Identified peptide candidates were sequence-validated, and their genomic origins were annotated using BamQuery. Final filtering steps were applied to distinguish tumor-associated antigens (TAAs) and tumor-specific antigens (TSAs) among the MAPs. Figure created with Biorender. B) Venn diagrams showing, for each AML cell line, the number of MHC-associated peptides (MAPs) uniquely quantified in DMSO- or TAK-981–treated conditions, as well as those shared between both. C) Volcano plots illustrating the impact of TAK-981 on the immunopeptidomic landscape. Each dot represents a quantified MAP; significantly upregulated peptides ($\log_2$ fold change > 1, adjusted $p < 0.05$) are shown in red, while downregulated peptides ($\log_2$ fold change < -1, adjusted $p < 0.05$) are shown in blue. Fold change was calculated as the ratio of peptide abundance in TAK-981–treated versus DMSO-treated cells. Adjusted p-values were computed using the Benjamini-Hochberg (B-H) method with a false discovery rate (FDR) threshold of 5%. AML, acute myeloid leukemia; GO, gene ontology; MAP, MHC-I associated peptide; TA, tumor antigen; TAA, tumor-associated antigen; TSA, tumor-specific antigen.

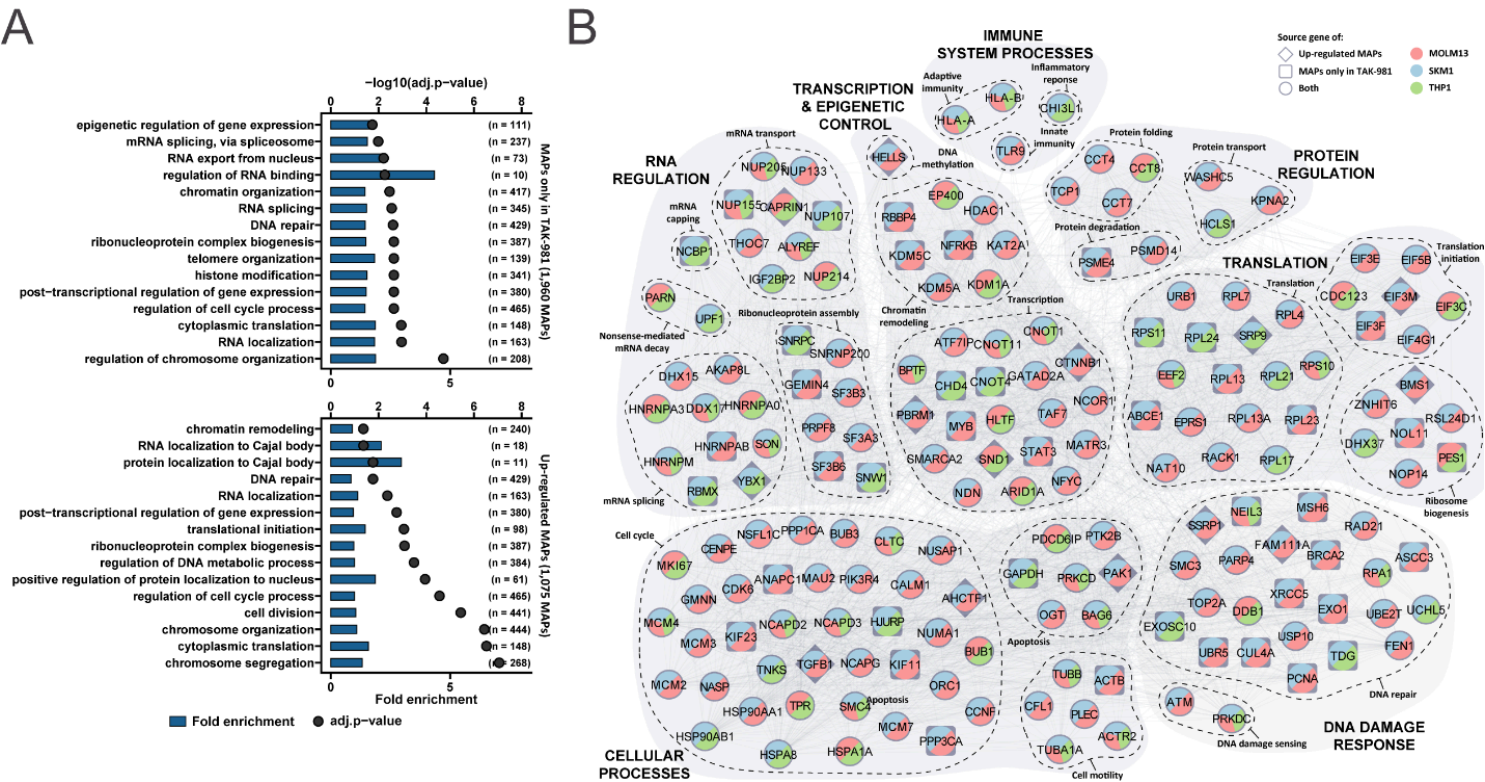


Figure 2

Immunopeptidomics reveals enrichment of epigenetic and RNA processing pathways in AML cells following SUMOylation inhibition. A) Gene Ontology (GO) enrichment analysis of source genes encoding MAPs either uniquely detected in TAK-981–treated cells (top) or significantly upregulated following treatment (bottom), aggregated across all AML cell lines. Only GO terms related to biological processes are shown. Adjusted *p*-values were calculated using the Benjamini-Hochberg method. B) Protein–protein interaction network of source genes encoding MAPs exclusively identified and/or significantly upregulated in TAK-981–treated cells. Subnetworks of genes are organized according to associated GO terms. Node shape depicts the type of MAPs generated, while colors indicate in which cell line the source gene is detected.

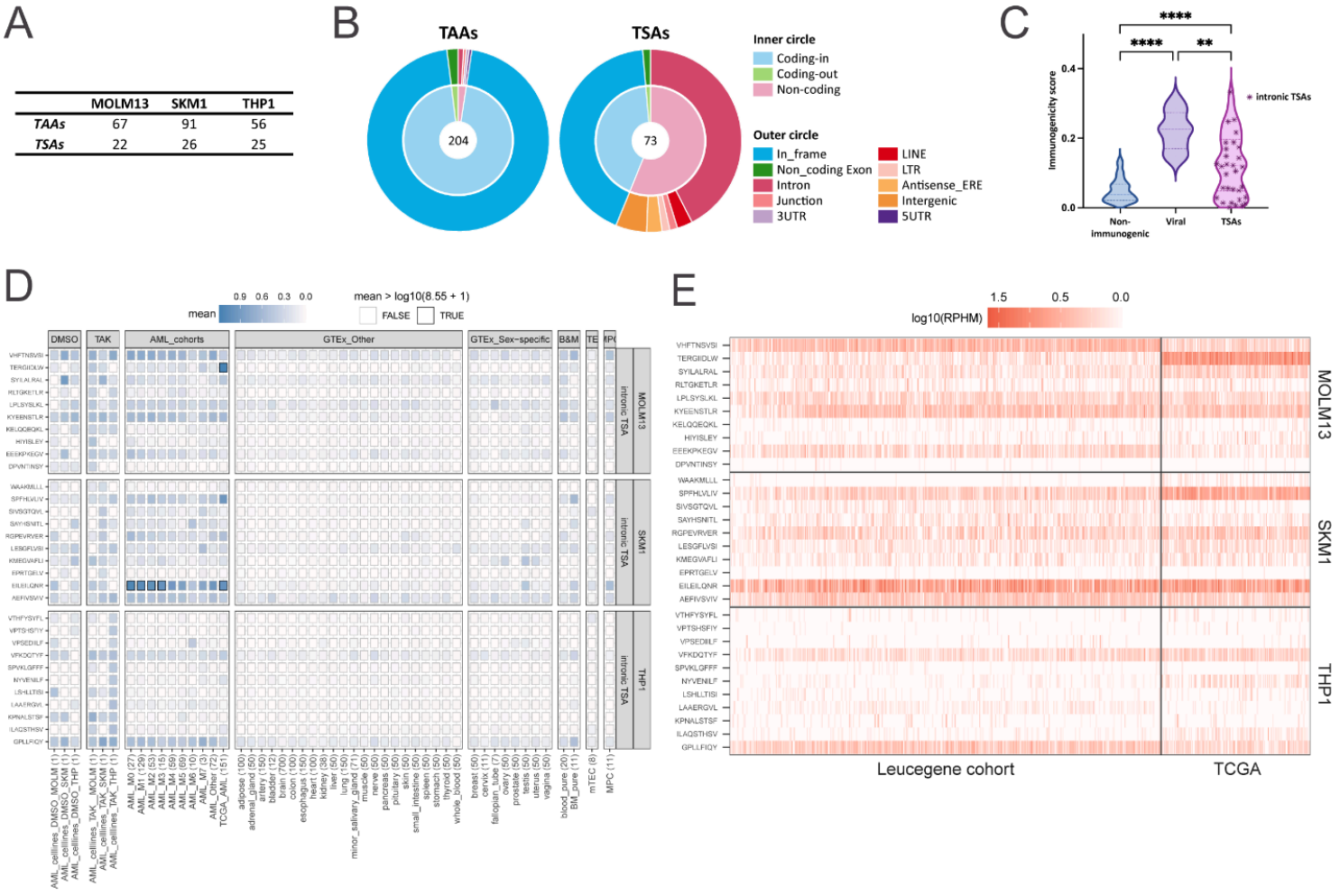


Figure 3

Identified Tumor-specific antigens (TSAs) originate from non-coding regions, exhibit low expression in normal tissues, and can be modulated by SUMOylation inhibition. A) Summary table showing the number of tumor-associated antigens (TAAs) and tumor-specific antigens (TSAs) identified in each AML cell line. B) Stacked pie charts illustrating the genomic origin of TAAs and TSAs. C) Violin plots comparing predicted immunogenicity scores of TSAs, assessed using PRIME 2.0, against positive controls (viral peptides) and non-immunogenic MAPs. Intronic TSAs are highlighted with purple asterisks. Statistical significance was determined using the Kruskal–Wallis test (***p* < 0.01; *****p* <

0.0001). D) Heatmap showing the mean RNA expression [$\log(\text{RPHM} + 1)$] of intronic TSA-coding regions across normal tissues (GTEx), AML patient samples (Leucegene and TCGA cohorts), pooled medullary thymic epithelial cells (mTECs), and AML cell lines treated with DMSO or TAK-981. Black frames highlight regions with mean expression >8.55 RPHM. E) Heatmap of mean RNA expression [$\log(\text{RPHM} + 1)$] of intronic TSA-coding sequences across two AML patient cohorts: Leucegene ($n=437$) and TCGA ($n=151$).

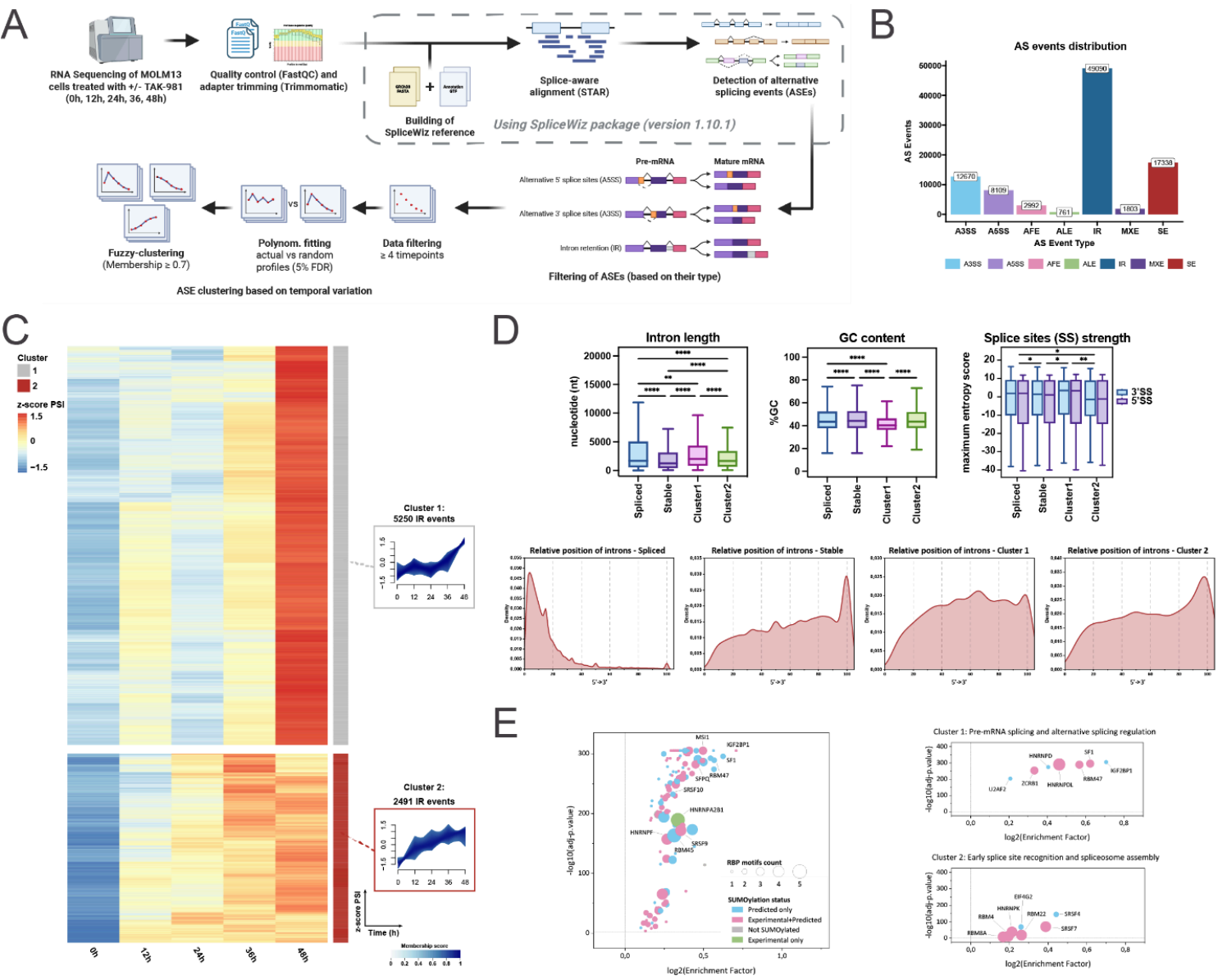


Figure 4

Loss of SUMOylation impairs splicing efficiency in MOLM-13 cells. A) Schematic overview of the experimental design for time-resolved alternative splicing analysis of MOLM-13 cells following TAK-981 treatment. B) Distribution of identified alternative splicing (AS) events. C) A total of 7,741 IR events displayed time-dependent changes and were clustered into two distinct groups based on temporal expression profiles, with a membership score > 0.7 . D) Boxplots showing comparison between retained introns modulated by TAK-981 (Cluster 1 and 2), retained introns not regulated by TAK-981 (Stable), and introns not regulated by TAK-981 (Spliced).

properly spliced introns (Spliced) for the corresponding feature. The curve graphs show the distribution of the introns (red line) along their full-length corresponding transcript (in the 5' → 3' direction). E) Motif enrichment analysis of RNA-binding proteins at the splice sites of IR events. Enrichment was calculated by comparing clustered versus non-clustered IR events and averaged across all motifs associated with each RNA-binding protein (left). The left panel shows the global enrichment analysis, while the right panels highlight the top enriched RNA-binding proteins for Cluster 1 (top) and Cluster 2 (bottom). Statistical significance was determined using the Kruskal–Wallis test (* $p \leq 0.05$; ** $p \leq 0.01$; **** $p \leq 0.0001$; ns non-significant).

For each example, the transcript structure, retained intron, and intronic ORF encoding the peptide are shown. Nucleotide and translated amino acid sequences are displayed, with alternative start codons highlighted in green and stop codons in red. Identified intronic peptides are highlighted in red. Representative MS/MS spectra of each peptide are also shown. E) Heatmap showing the mean RNA expression [$\log(\text{RPHM} + 1)$] of MAP from intron-derived microprotein coding regions across normal tissues (GTEx), AML patient samples (Leucegene and TCGA cohorts), pooled medullary thymic epithelial cells (mTECs), and AML cell lines treated with DMSO or TAK-981. Black frames highlight regions with mean expression >8.55 RPHM.

Examples of microproteins from alternative ORFs

Peptide identified in:

■ MOLM-13
■ SKM-1
■ THP-1
■ DDA original study

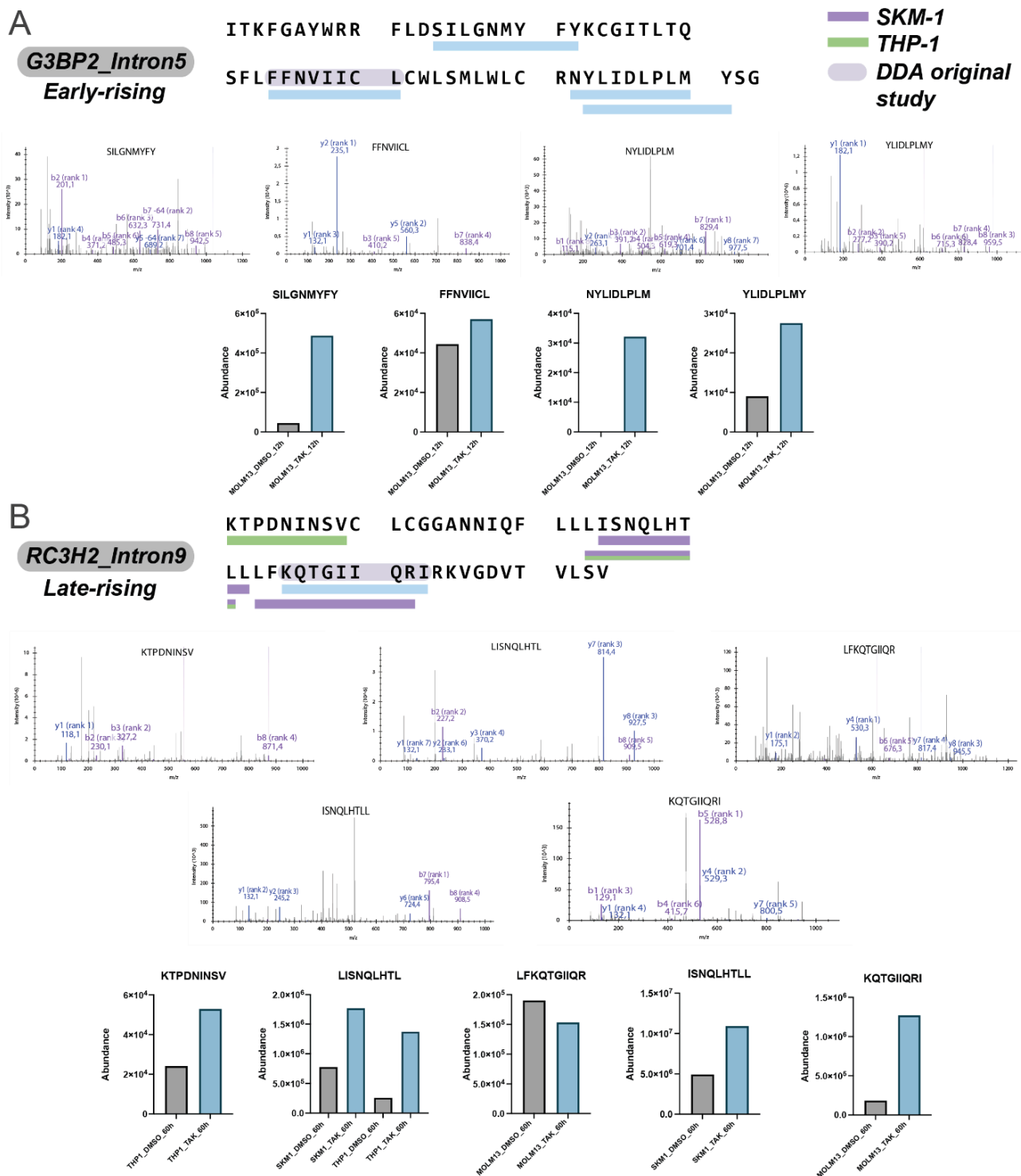


Figure 6

TAK-981 modulates the presentation of HLA-I peptides encoded by intronic alternative open reading frames. A) Amino-acid sequence of an alternative ORF encoded within *G3BP2* intron 5. Horizontal bars map the identified peptides onto the predicted microprotein sequence. Representative annotated MS/MS spectra are shown for NYLIDLPLM, SILGNMYFY, FFNVIICL and YLIDLPLMY. Bar plots compare their measured abundances in MOLM-13 cells treated with DMSO or TAK-981 for 12 h. **B)** Corresponding

Figure 7

Distribution of TAK-981-responsive intron-retention events across primary AML and healthy hematopoietic samples. Heatmap showing row-standardized z scores of log10-transformed Percent Spliced-In (PSI) values for selected peptide-generating IR events. Columns represent triplicate MOLM-13 samples collected at 0, 12, 24, 36 and 48 h after treatment with TAK-981 (20 nM; n = 15); untreated Leucegene AML samples with wild-type splicing factors representing FAB M1-M5 subtypes (n = 18), SRSF2-mutant AML (n = 5) or U2AF1-mutant AML (n = 5); and healthy MPC (n = 11), bone-marrow (n = 11) and blood (n = 20) samples. Rows are grouped according to their early- or late-rising response to TAK-981 in MOLM-13 cells. Column annotations indicate sample type, AML subtype and splice-factor genotype. Because values were standardized within each IR event, colors represent relative PSI patterns across samples rather than absolute differences among events.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile2.pdf](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable1.pdf](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryFigures.pdf](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryFile1.pdf](#)