Pan-cancer analysis reveals context-dependent roles of LINE-1 ORF1p in immune regulation and copy number alterations

Efiyenia Ismini Kaparos[1,2,3], Wenjing Zhang[1,2], Ryusei Miyanaga[1,3], Wilson McKerrow[1,2], Jef Boeke[1,2], Teresa Davoli[1,2], Paolo Mita[1,2]*, Kelly V. Ruggles[1,3]*, David Fenyö*[1,2]

[1]Institute for Systems Genetics, New York University Grossman School of Medicine, New York, NY, USA
[2]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY, USA
[3]Division of Precision Medicine, Department of Medicine, New York University Grossman School of Medicine, New York, NY USA

Corresponding Authors:

David Fenyö, PhD

Institute for Systems Genetics

Department of Biochemistry and Molecular Pharmacology

New York University Grossman School of Medicine

Tel: 212-263-2216

Email: David@fenyolab.org

Kelly V. Ruggles, PhD

Division of Precision Medicine, Department of Medicine

New York University School of Medicine

227 East 30th St #823, New York, NY 10016

Tel: 347-510-6960

Email: Kelly.Ruggles@nyulangone.org

Paolo Mita, PhD

Institute for Systems Genetics

Department of Biochemistry and Molecular Pharmacology

New York University Grossman School of Medicine

Tel: 212-263-6934

Email: Paolo.Mita@nyulangone.org

* Denotes co-corresponding and co-senior authors

ABSTRACT

L1 comprises 17% of the human genome, with 50-150 full-length sequences capable of retrotransposition. Although largely inactive in normal somatic tissues, tumorigenesis leads to L1 derepression and overexpression of its RNA binding chaperone protein, ORF1p. A potential cancer biomarker, ORF1p expression is a hallmark of multiple cancers and an early event in precursor lesions. Our study provides a comprehensive pan-cancer analysis of ORF1p using CPTAC proteogenomic data, revealing a dichotomous role in modulating immune responses. Integrated analyses and supervised learning models divide ORF1p-high tumors into two groups: (1) high ORF1p associates with immunosuppression, reduced interferon signaling and diminished immune cell infiltration (HNSCC and LSCC), and (2) ORF1p-high tumors associate with immune activation (UCEC). Linear regression models reveal that cancer-specific aneuploidies may underlie this immune dichotomy, highlighting the prognostic significance of ORF1p and informing new strategies to leverage detection of plasma-circulating ORF1p to enhance immunotherapeutic efficacy in different tumor contexts.

INTRODUCTION

Transposable elements comprise about half of the human genome, with retrotransposons accounting for just over 40%.[1,2] Among them, the long interspersed nuclear element-1 (LINE-1 or L1) makes up 17% of these sequences and is the only active autonomous element.[1] Although most of the ~500,000 L1 copies in the human genome are truncated or mutated, about 50–150 copies remain full-length with intact open reading frames (ORF).[3–5] These elements retrotranspose through a “copy-and-

paste" mechanism, in which L1 RNA is reverse transcribed and integrated into new genomic sites.[6,7] This process requires the two L1 encoded proteins: ORF1p, an RNA-binding protein, and ORF2p, which provides endonuclease and reverse transcriptase functions and also capable to drive the mobilization of non-autonomous repeats such as SINEs.[8–11] Together with L1 RNA, ORF1p and ORF2p form ribonucleoprotein particles (RNPs) that mediate nuclear entry and integration.

Due to its potentially mutagenic nature, L1 is suppressed in most normal somatic tissues.[12,13] Host cells have evolved multiple defense mechanisms to regulate and restrict L1 activity, but in many forms of cancer L1 repression mechanisms are lost and repeat elements are often derepressed.[14–20] Pattern recognition receptors (PRR) recognize derepressed L1 byproducts triggering type I interferon (IFN) responses similar to the one activated by exogenous retrovirus invasions. [21–23]

Holistically understanding the functional role of L1 derepression in cancer requires an examination of L1 activity, defined here as its ability to re-insert into new random genomic loci, as well as the consequences of L1 expression across its lifecycle, including its RNA regulation, ORF1p/ORF2p expression and L1 RNP interaction with the tumorigenic cellular context.[24] Current *in vivo* and *in vitro* studies investigating L1 in cancer have relied on engineered L1 overexpression systems and reporter assays that miss physiologically relevant contexts such as the complete tumor ecosystem, comprised of cancer-associated fibroblasts and immune system interactions.[25–31] Despite the extensive genomic characterization of L1 activity in cancer[32–35], comprehensive pan-cancer proteomic analyses of L1 ORF1p aimed at understanding the consequences of its overexpression in neoplasia are still limited.[36,37] Our study aims to fill this gap by

characterizing ORF1 protein expression, both because its expression is more easily detectable by proteomic approaches compared to ORF2p[38] and because recent studies have highlighted a complex role for ORF1p in tumorigenesis.[25,39–42] In addition to the establishment of ORF1p as a potential diagnostic marker for multiple cancers, a study in pancreatic cancer identified an unexpected feed-back loop in which ORF1p, through its mRNA binding activity, may decrease the IFN-I response triggered by its own L1 mRNA.[42]

To better understand the impact of L1 ORF1p derepression in *in vivo* tumor contexts, we used the NCI-supported Clinical Proteomic Tumor Analysis Consortium's (CPTAC) harmonized pan-cancer dataset (n = 1,072) spanning 10 cancer types[43], and integrated L1 quantifications with proteogenomic data with emphasis on L1 ORF1p. In our analysis, we grouped patients based on their ORF1p protein expression levels and applied supervised learning to map downstream effects on gene, protein, and phosphoprotein expression, as well as immune-related changes and copy number alterations. We identified tissue specific consequences of L1 expression in the cellular adaptation to high ORF1p expression and in the immune response triggered by L1 derepression and describe molecular vulnerabilities unique to persistent ORF1p overexpression with prognostic potential.

RESULTS

Pan-cancer characterization of L1 ORF1p

We quantified L1 RNA and ORF1p in all 10 CPTAC tumor types and identified novel L1 somatic insertions where whole genome sequencing (WGS) was available (Figures 1A and S1A) as previously described.[44] Of the 1,043 samples with

transcriptomic, proteomic or WGS data available, we identified 929 tumors with at least one L1 measurement detected (Figure S1B). This included BRCA, CCRCC, COAD, OV, and UCEC tumors.[44,45] Out of the 929 tumors in our analysis, 808 included ORF1 protein quantification (Figures 1B and S1C), with the highest median ORF1p expression identified in UCEC and the lowest in BRCA (Figure S1D). 776 samples in our dataset had L1 RNA quantification (Figures 1B and S1C), with the highest median L1 RNA expression in HNSCC and the lowest in GBM (Figure S1D). In GBM tumors, we detected ORF1p abundance in only 8% of samples (Figure S1E), consistent with previous findings showing that L1 retrotransposition and ORF1p expression are infrequent in aggressive brain tumors.[40,46,47] Thus, we have excluded GBM from all proteomic-based analyses in this study.

Except for CCRCC, L1 RNA and ORF1p correlated strongly (Spearman $R > 0.4$) and significantly (P value < 0.001) across cancer types, suggesting that L1 ORF1p level is mainly dependent on transcription and less on post-transcriptional regulations (Figures 1B and S2A). Incidence of novel somatic L1 insertions correlated most strongly with L1 RNA expression (Spearman $R > 0.6$, P value $< 10^{-12}$) in HNSCC and LSCC tumors (Figure S2B), which also had the greatest volume of insertions ranging from 11-60 (Figure 1B). These findings suggest that HNSCC and LSCC are highly L1 active tumors, as previously shown[32,33], and may indicate an enhanced capacity of these tumors to sustain elevated L1 expression and activity. This observation could be at least partially explained by the fact that the vast majority of HNSCC and LSCC tumors in these cohorts harbor p53 mutations (Figures 1B and S2C), a known pre-requisite for L1 expression.[20,44] On the other hand, our measurements also show similar p53 mutation frequencies in PDAC and

OV compared to HNSCC and LSCC tumors, suggesting that additional regulatory mechanisms beyond p53 influence L1 expression. Overall, these results confirm our ability to quantify L1 at various points of its lifecycle (RNA, ORF1 protein and novel insertions) (Figure 1A) and demonstrate that L1 is derepressed across all CPTAC tumor types (Figures 1B).

To characterize L1 ORF1p expression in cancer, we first grouped samples as ORF1p-high (top quartile) or ORF1p-low (bottom quartile) and assessed the association of this grouping with multimodal molecular features, cancer types, subtypes, and clinical outcomes (Figure 1C). Past studies have shown an association between L1 RNA expression and methylation status and poor outcomes across several cancer types.[48,49] To assess the relationship between L1 protein expression and survival in CPTAC, we stratified our samples by ORF1p and L1 RNA expression, grouping samples into L1 high and L1 low (Methods and Figure 1C). Using pan-cancer Kaplan–Meier analysis, we found significantly lower survival for ORF1p-high patients compared to ORF1p-low across CPTAC cohorts (log-rank P = 0.021, Figures 1D and S3A). Parallel L1 RNA analysis revealed no significant difference in overall survival between L1 RNA-high and L1 RNA-low patients across cancer types, suggesting that L1 ORF1p overexpression has a functional role that affects cancer outcomes independent of L1 RNA expression (Figure S3B). Survival analyses within individual cancer types did not yield any significant differences in overall survival based on ORF1p expression, possibly due to small sample sizes (Figure S3C).

Clinical associations with ORF1p expression

We evaluated the association of ORF1p expression with clinical and demographic variables (Figure 2A). Interestingly, L1 ORF1p expression was significantly increased in current smokers (P = 0.006) and past smokers (P = 0.015) when compared to non-smokers in HNSCC, with no significant differences based on smoking identified in other cancer types. LUAD patients with East Asian inferred ancestry had significantly higher (P value = 0.047) ORF1p levels compared to those with inferred European and African ancestry.

Tumor mutational burden (TMB), defined as the number of somatic mutations per megabase of a genomic sequence, is believed to generate immunogenic neoantigens that enhance patient response to immune checkpoint inhibitors (ICI).[50] To determine associations of ORF1p with TMB in our cohorts, we categorized TMB into groups using quartiles from the TMB distribution per cohort (Q1, Q2, Q3 and Q4). ORF1p expression was found to be significantly higher in tumors with higher TMB levels in HNSCC (P = 0.026 for TMB Q2, 0.00013 for TMB Q3 and 0.027 for TMB Q4) and LSCC (P = 0.012 for TMB Q4). HNSCC and LSCC are tumor types characterized by high TMB[50], suggesting that L1 ORF1p expression is positively associated with an increasing TMB in squamous cell carcinomas (Figure 2A). Though L1 RNA and ORF1p expression are significantly higher in p53 mutant tumors compared to p53 WT tumors at the pan-cancer level (Figure S4A-S4B), significant differences between ORF1p expression and p53 mutation status at the cohort-specific level were detected only in BRCA, LUAD and UCEC (Figure 2A), while LSCC was the only cohort to have a significant difference in L1 RNA expression and p53 mutation status (Figure S4C). Though lack of significance at the cohort level is largely due to sample size, there may be additional p53 independent mechanisms of regulation

for the other tumor types. No significant differences in L1 ORF1p expression were found for stage, sex, MSI-high (MSI-H) status, race, and ethnicity (Figure 2A). Parallel analyses with L1 RNA resulted in similar significant trends seen in the ORF1p analysis, with higher L1 RNA expression correlating with increased TMB levels in HNSCC and LSCC (Figure S4C). Current smokers of HNSCC had significantly higher L1 RNA levels than non-smokers (P = 0.0074), while current smokers of UCEC had significantly lower L1 RNA expression than non-smokers (P = 0.036), a trend also seen at the L1 protein level as well (Figures 2A and S4C).

ORF1p expression in tumor subtypes

We then assessed ORF1p expression by CPTAC genomic molecular subtypes to determine whether ORF1p expression can be utilized as an additional patient stratification metric (Figures 2B-J and S5A-I).[51–59] HNSCC, LSCC and UCEC showed significant differences in ORF1p expression across molecular subtypes, with the highest ORF1p expression observed in the classical subtypes in HNSCC and LSCC, respectively, and the CNV-H subtype in UCEC compared to all other subtypes. The lowest expression of ORF1p was found in mesenchymal and secretory subtypes of HNSCC and LSCC, respectively (Figures 2B, 2E, 2H, and S5, $P < 0.001$). Similarly, the chromosomal instability (CIN) genomic subtype of COAD had significantly higher ORF1p expression than the mesenchymal subtype (Figure S5C).

High ORF1p-expressing classical subtype tumors in HNSCC and LSCC are known to have similar molecular profiles, including *TP53* mutation, strong smoking history, low immune activation, and high CIN.[55,58,60] We found that tumors in these subtypes also had

significantly lower cytotoxic immune scores compared to other molecular subtypes (Figures 2C and 2F) and significantly higher CIN (Figures 2D and 2G). Low ORF1p-expressing mesenchymal and secretory subtypes of HNSCC and LSCC were characterized by upregulation of immune-related pathways based on cytotoxic immune score (Figures 2C and 2F). Similar to HNSCC and LSCC, ORF1p high-expressing CIN-subtype tumors of COAD had significantly lower cytotoxic immune score levels compared to the mesenchymal subtype ($P < 0.01$), along with higher CIN ($P < 0.01$) as expected and significantly higher aneuploidy scores compared to the MSI subtype ($P < 0.0001$) (Figure S5C).

High ORF1p UCEC CNV-H tumors are characterized by frequent copy number alterations and increased expression of cell cycle and antiviral response proteins.[54,61] We found that these tumors had the highest levels of CIN compared to all other UCEC subtypes, similar to the classical subtype of HNSCC and LSCC (Figure 2J; $P < 0.001$). Although there were no significant differences between UCEC subtypes for cytotoxic immune score (Figure S5I), CNV-H tumors had significantly higher CGAS-STING scores, representing antiviral signaling activation, than CNV-L tumors (Figure 2I) and the highest aneuploidy scores when compared to all other molecular subtypes of UCEC, as expected (Figure S5I).

Collectively, these data demonstrate that ORF1p expression can distinguish between more and less aggressive molecular subtypes within each cancer type, with high ORF1p expression consistently associated with high CIN and reduced immune response in HSCC, LSCC and COAD and, in contrast, associated with increased immune response in UCEC.

ORF1p modulates the expression and post-translational modification of proteins involved in IFN response

Previous work in pancreatic cancer showed that repetitive elements, including L1, are able to trigger IFN-I activation through "viral mimicry" processes.[25,29,42] *In vitro* and *in vivo* studies identified cytoplasmic dsRNA, RNA:DNA hybrids or dsDNA as potential molecules produced by retrotransposons expression and activity and able to trigger and IFN-I response.[23,62–65] Moreover, the L1 ORF1p protein itself was shown to "mask" dsRNA L1 adducts decreasing the activation of IFN-I mediated responses, inducing an opposite effect from the one triggered by L1 mRNA expression.[42]

To gain a pan-cancer view of the L1 involvement in viral mimicry pathways during tumorigenesis and gain insights on the molecular details at play in *in vivo* contexts, we first conducted differential proteomic expression analysis accounting for tumor purity between ORF1p-high and ORF1p-low samples for each cancer type (Methods), identifying tumor specific and pan-cancer associations (Table S1, Figure S6). After filtering for proteins with an adjusted $P < 0.2$ in at least four of the cohorts, we identified a common signature of 17 proteins differentially expressed in ORF1p-high samples (Figure 3A). Of the 17 proteins identified 10 are directly connected to RNA metabolism and/or ribonucleoprotein metabolism (Figure 3A, bold), suggesting that RNA-linked processes may be key for the tissue and tumor response to ORF1p expression. This is perhaps not surprising considering the well-known function of ORF1p as RNA binding protein.[66,67] Specifically, the ORF1p proteomic signature includes reduced expression of selenium binding protein 1 (SELENBP1) and vacuolar protein sorting VPS13C and

increased expression of transcription factor TFB2M, nuclear export receptor XPO5, replication factor RFC4 and protein phosphatase Mg2+/Mn2+ dependent 1G (PPM1G) (Figure 3A). Several of these proteins are established repressors of IFN signaling, inflammation or viral defense processes[68–70], suggesting that tumors may actively attempt to limit the L1-triggered IFN response (Table 1).

In particular PPM1G is a known regulator of the innate immune adapter protein STING1 (STimulator of INterferon response cGAMP interactor 1) and MAVS (Mitochondrial AntiViral Signaling) protein.[69] To determine whether increased expression of PPM1G in ORF1p-high tumors also results in changes of phosphorylation of its downstream targets STING and MAVS, we leveraged CPTAC's phosphoproteomic data and assessed differences between ORF1p-high and ORF1p-low groups for all STING and MAVS phosphosites (Figure 3B and Table S2). Due to overexpression of PPM1G in most cancer types, we expected to find decreased phosphorylation of STING and MAVS in ORF1p-high tumors. Indeed, ORF1p-low samples in HNSCC had significantly higher phosphorylation of MAVS-S253 (P = 0.03) and STING1-S345 (P = 0.039) compared to ORF1p-high samples. No other cancer types exhibited significant differences in phosphorylation of PPM1G targets in this direction (Table S2). To determine if dephosphorylation of STING and MAVS affects downstream activation of the IFN program[71,72], we evaluated the phosphorylation status of inhibitor of nuclear factor kappa B kinase subunit beta (IKBKB) across cancer types and found reduced phosphorylation of IKBKB in HNSCC ORF1p-high samples at phosphosites Ser675 and Ser697 (Figure 3C and Table S2). Phosphorylation of IKBKB triggers phosphorylation of the inhibitor that silences the NF-κB complex[73,74], and consistent with this relationship, expression of

downstream IFN-stimulated genes was found to be significantly negatively correlated with PPM1G protein expression in HNSCC and LSCC (adjusted *P* value < 0.05). Tumor types with the highest levels of ORF1p, HNSCC and LSCC, clustered together in their IFN expression signature (Figure 3D). OV and COAD tumors also harbored few significant negative correlations between IFN-stimulated gene expression and PPM1G protein expression (Figure 3D).

Together, these findings indicate that ORF1p-high HNSCC and LSCC tumors upregulate proteins to reduce and control IFN-I signaling. One such protein is PPM1G, a phosphatase whose upregulation correlates with dephosphorylation of MAVS, STING and IKBKB proteins, negatively modulating the IFN-I and NF-κB innate immune response (Figure 3E). This mechanism was significantly observed in HNSCC tumors (Figures 3B-C), likely facilitating tolerance to L1 overexpression.

ORF1p expression regulates immune pathways and immune cell infiltration in a tumor-specific manner

Despite the established role of IFN mediated inflammation in cancer development, studies focused on the consequences of the L1-mediated IFN activation on immune cell recruitment and activity are still limited.[75,76] To more systematically investigate the role of ORF1 protein expression in immune regulation, we performed gene set enrichment analysis (GSEA) on differentially expressed proteins in ORF1p-high tumors to assess ORF1p-associated pathway dysregulation for each cancer type (Figures S7-9). Previous studies have reported enrichment in pathways of DNA repair and cell cycle regulation in ORF1p-high tumors.[44] Consistent with this, we found that in all tumor types except for

CCRCC, pathways of oxidative phosphorylation, glycolysis, and MYC and MTORC1 signaling were significantly enriched in tumors with high ORF1p-expressing tumors (Figure 4A). Further, we found that pathways involving innate immune and inflammatory signaling, hypoxia, epithelial mesenchymal transition (EMT), apoptosis and myogenesis were downregulated in ORF1p-high LSCC, PDAC, COAD and HNSCC tumors (Figure 4A). These findings align with previous studies that have demonstrated a negative association of immune response with L1 retrotransposition.[29,35,42] Interestingly and more specifically, we also noticed that several IFN Hallmark pathways were upregulated in ORF1p-high for CCRCC, OV, UCEC and BRCA that, in this context, seem to behave in an opposite manner compared to LSCC, PDAC, COAD and HNSCC as for our previous results (Figure 4A, bold).

These data suggest that, in specific cancer types, tumors with higher expression of ORF1p downregulate inflammatory pathways with potential implications in tumor immune response. To explore this hypothesis we further categorized tumors into immune subtypes (cold, cool, warm and hot) based on a previously published CPTAC pan-cancer study linking oncogenic drivers to functional states.[77] When comparing ORF1p expression by immune subtype at the pan-cancer level, we observed significantly lower ORF1p expression in immune-hot tumors compared to those in the immune-warm and immune-cool subtypes (Figure 4B). Similar significant trends were observed at the individual cancer level for COAD, HNSCC, LSCC and LUAD (Figure S10), validating pathway-based clustering (Figure 4A). Assessing L1 RNA expression by immune subtypes, we also observed significantly lower L1 RNA in the immune-hot subtype at the pan-cancer level and individually for HNSCC, LSCC and LUAD (Figure S11). Together,

these findings suggest a negative relationship between L1 expression and host immune response in the squamous cell carcinomas (Figure 4C).

To further evaluate this relationship, we performed regression analyses to predict the effect of ORF1p expression on immune cell type. RNA-derived xCell[78] immune cell type estimations were used as model input, correcting for WES-based tumor purity for bulk sample correction (Figure 4C). The only statistically significant finding at the pan-cancer level was a negative correlation between ORF1p expression and activated myeloid dendritic cells (Figure S12). Independent, cancer-specific regression revealed that the associations between ORF1p expression and immune cell infiltration are unique to each tumor type (Figures 4D and S13). Significant associations (adjusted $P < 0.05$) between ORF1p expression and immune cell activation were only identified in HNSCC, LSCC and LUAD. The top significantly associated cell types with ORF1p expression in HNSCC and LSCC tumors, were, respectively, plasmacytoid dendritic cells (pDCs) and activated myeloid DCs in the innate arm and CD4+ effector memory T-cells and class-switched memory B-cells for the adaptive arm (Figures 4D and S13E). The involvement of pDCs is particularly notable given their role as key regulators of antiviral immunity via inflammatory cytokine and their ability to differentiate into myeloid DCs[79], that we also found negatively correlated with ORF1p expression in HNSCC (adjusted $P < 0.05$) (Figure 4D).

This data corroborate our hypothesis that ORF1p expression coincides with reduced innate immune response, potentially compromising tumor immunosurveillance.[80,81] This findings are also in line with our pathway-level expression analysis of the interferon alpha and gamma response and the MAPK/NF-κB axis (TNF-a

signaling via NF-κB) that shows depletion of these pathways in ORF1p-high tumors of both HNSCC and LSCC (Figure 4A and S8-9). These pathways are indeed known to drive pDC maturation into activated myeloid DCs[82,83] and their downregulation suggests that high L1 ORF1p expression is associated with a microenvironment depleted of proinflammatory signaling resulting in fewer functional DCs. In contrast, molecular and pathway analysis in UCEC show opposite trends compared to HNSCC and LSCC (Fig. 2B-J, 3D and 4A). Notably in UCEC, ORF1p expression has a positive correlation with immune activation (adjusted P value < 0.1), with the strongest association observed for activated myeloid DCs (Figure S13H). These results show that ORF1p overexpression has tumor-type specific effects on immune infiltration, dividing CPTAC cancers into two distinct categories: those where ORF1p expression correlates with immune suppression (e.g. HNSCC and LSCC) and those where it correlates with activated immune response (e.g., UCEC). This dichotomy is consistent with previously conflicting reports of both positive and negative immune correlations with L1 activation.[22,29,30,42]

To gain insight into the molecular mechanisms that drive the functional interaction between L1 ORF1p expression and the immune response, we derived a series of RNA- and protein-based immune scores representing cytosolic dsRNA triggered responses by RIG-I-like receptors (RLR score) and cytosolic dsDNA or DNA:RNA hybrids triggered pathways (cGAS/STING scores). We also included IFN signaling, cytolytic activity of the adaptive immune response, and tumor inflammation scores based on previous published signatures (Methods). We built cancer-specific linear regression models to predict each immune score by ORF1p expression. Corroborating our previous analyses, we found that HNSCC and LSCC exhibited similar response to ORF1p expression, with clear negative

correlations between ORF1p expression and nearly all RNA- and protein-based immune scores (Figure 4E, adjusted *P* value < 0.01). Notably, and again consistent with our previous findings, UCEC showed an opposite association between ORF1p and immune response, with positive associations identified between ORF1p expression and RLR and IFN signature scores (Figure 4E). These findings also suggest that in UCEC repeat element-derived molecular patterns likely trigger mostly an antiviral response through dsRNA sensing mediated by RIG-I and/or MDA5 rather than cytosolic dsDNA or DNA:RNA hybrid detection triggered by the cGAS/STING pathway.

ORF1p expression predicts cancer-specific aneuploidies

To investigate the underlying reasons why the immunogenicity of ORF1p overexpression varies across cancer types, we evaluated the relationship between somatic copy number alterations (SCNAs) and ORF1p. Aneuploidy, the state of having gained or lost chromosomes, is a fundamental hallmark of cancer.[84] Aneuploid states in specific cancer types have been linked to immune evasion and altered immune cell infiltration.[85] We argue that if tumors attempt to decrease inflammation and IFN signaling for their growth and development, a potential solution to achieve this goal would be the loss or gain of genomic regions rich in genes involved in immune regulation (e.g. the IFN cluster on chromosome 9p).[86] We therefore assessed the correlation of chromosome arm-level copy number (CN) ratios with ORF1p expression and found cancer-specific significant associations (adjusted *P* value < 0.05) (Figure S14). Interestingly, as with our previous analyses, HNSCC and LSCC followed similar trends in the arm-level aneuploidies correlating with ORF1p expression, with chromosomes 1q gain, 2p gain, 3q

gain, and 4p loss significantly correlating with ORF1p expression in both tumor types (Figure S14).

We next binarized cytoband-level copy number ratios as gains (CN > 0.3) or losses (CN < -0.3) and performed individual generalized regressions to determine whether ORF1p expression can predict the gain or loss of specific cytogenic bands across cancer types (Methods). We found that ORF1p significantly (adjusted $P < 0.05$) predicts 1q gain in both HNSCC and LSCC (Figure 5A and S15A). Chromosome 1 is a chromosome known to contain several immune regulating genes.[87] When assessing binary chromosome 1q gain status in HNSCC and LSCC, tumors with 1q gain in both cancer types had significantly higher ORF1p expression (Figures 5B and S15B). In UCEC, ORF1p expression was correlated with different significant gains and losses (Figure 5C), chief among them gain of chromosome 2p. Indeed, UCEC tumors with 2p gain have significantly higher ORF1p expression than those with no amplification (Figure 5D)

Given the strong prediction of ORF1p for unique CN alterations in HNSCC, LSCC and UCEC, we investigated whether specific CN alterations in tumors might enable high ORF1p expression through the amplification or loss of specific pathways. We categorized CN alterations into 4 groups: those (1) positively associated with ORF1p and CN gain (Pos-Gain); (2) positively associated with ORF1p and CN loss (Pos-Loss); (3) negatively associated with ORF1p and CN gain (Neg-Gain); and (4) negatively associated with ORF1p and CN loss (Neg-Loss) and filtered for an adjusted $P < 0.2$ (Figure S15C). We then performed over-representation analysis (ORA) on each of these gene sets per cancer type, determining which biological pathways were enriched for ORF1p-predicted aneuploidies grouped into each category (Table S3). Next, we performed similar ORA for

genes on specific chromosome arms significantly predicted by ORF1p in HNSCC, LSCC and UCEC to determine pathways unique to each arm (Tables S4-6).

As expected, we detected a significant enrichment for negative regulation of humoral immune response and negative regulation of immune effector processes, for genes on chromosome 1q in HNSCC and LSCC (Figure 5E and S16). Conversely, we identified pathways involved in immunoglobulin production and natural killer cell-mediated cytotoxicity for genes enriched on chromosome 2p in UCEC (Figure 5F). Together, these findings suggest that ORF1p expression is associated with aneuploidies enriched in genes involved in immune response and that aneuploidies found in HNSCC and LSCC versus UCEC can explain the ORF1p immune dichotomy.

To better understand the protein-level impacts of ORF1p-associated aneuploidies in HNSCC and UCEC, we isolated genes located on specific 1q and 2p cytoband regions significantly (adjusted $P < 0.05$) predicted by ORF1p expression using overall 1q and 2p CN ratios as covariates, respectively, to identify expression relationships independent from total copy number dosage (Figures 5G-H). In both HNSCC and UCEC, we identified proteins negatively correlated with ORF1p despite overall amplification of their respective chromosomes, indicating that chromosome copy number does not by itself explain proteomic associations with ORF1p for these genes and that additional ORF1p-dependent mechanisms may contribute to their deregulation.

Altogether, these findings suggest that ORF1p expression can direct the formation of specific aneuploidies that modulate innate immune and immune cell activation pathways. Further experimental validation is needed to determine whether derepression of TEs in early tumorigenesis leads to L1/IFN-permissive CN alterations or if ORF1p

expression and specific CN alteration develop independently through a common causative event.

DISCUSSION

To our knowledge, this work provides the first large-scale proteogenomic pan-cancer study of L1 with patient-level L1 insertion, RNA and ORF1p quantifications and systematic integration with multi-omic data types. By leveraging CPTAC pan-cancer proteogenomic data, we demonstrate that ORF1p-high tumors are characterized by unique proteogenomic and immune changes that are tumor-specific and related to copy number alterations, potentially resulting in a permissive molecular environment for L1 expression (Figure 6).

Our comprehensive analysis reveals that ORF1p expression and its molecular consequences subdivide tumors into two distinct groups (Figure 6). The first is an immunosuppressed group, including HNSCC and LSCC, which exhibited high ORF1p expression, especially in classical subtypes, and negative associations between ORF1p expression and immune scores and cell types. Consistent with previous studies showing particularly high L1 permissiveness in squamous cell carcinomas, heightened ORF1p expression in HNSCC and LSCC may be linked to the high prevalence of p53 mutations in these tumors, as p53 is a direct repressor of L1.[33,34] Within these cancer types the highest ORF1p levels were observed in the classical subtypes, both of which have been associated with high chromosomal instability, strong smoking evidence and p53 mutation in each respective tumor type.[60] Given that current and past smokers with HNSCC, irrespective of molecular subtype, had significantly higher ORF1p expression than non-

smokers, our findings corroborate existing evidence that hypomethylation of L1 is associated with smoke-exposed epithelia and that this hypomethylation is associated with an increased risk of HNSCC.[88] These squamous cell carcinomas also show negative correlations between ORF1p expression and immune infiltration, particularly plasmacytoid and activated myeloid dendritic cells. The association between ORF1p expression and 1q gain in HNSCC underscores the potential role of ORF1p overexpression in reshaping immune-regulatory circuits through genomic alterations. Not only are the observed over representation of genes on 1q linked to the negative regulation of immune processes[87], but ORF1p is negatively correlated with proteins located on 1q independent of 1q copy number dosage. The top 3 most significantly anticorrelated 1q proteins with ORF1p included RCSD1, a cytoskeletal regulatory molecule that has been shown to associate with immune infiltrating cells[89], CD247, the gene encoding the essential CD3 zeta (ζ) chain component of the T cell receptor (TCR) complex[90], and PTPRC (or CD45), a hematopoietic-specific transmembrane protein tyrosine phosphatase that regulates Src kinases required for T- and B-cell antigen receptor signal transduction.[91] Though further *in vitro* and *in vivo* experiments are needed to determine whether ORF1p is a cause or consequence of SCNAs, our findings suggest that copy number alone does not explain the association of immunosuppressive characteristics with ORF1p expression in these tumor types.

Conversely, the second group of tumor types is an immune-active group, including UCEC tumors, that show the opposite trend, with positive associations between ORF1p expression and immune cell infiltration, particularly activated myeloid DCs and upregulation of antiviral response pathways. UCEC tumors with the highest ORF1p

expression were identified in the CNV-H subtype known for high copy number variations, genomic instability and elevated expression of cell cycle and antiviral response proteins.[61] In this tumor type, one of the most significant copy number alterations predicted by ORF1p expression were gains in chromosome 2p. Unlike 1q, over-representation analysis in chromosomal region 2p is linked to pathways that foster immune activation, such as immunoglobulin production and natural killer cell-mediated cytotoxicity, suggesting a distinct mechanism of immune modulation in UCEC compared to the squamous cell carcinomas. Independent of chromosome 2p amplification, 2p proteins were still significantly positively correlated with ORF1p, including antiviral dsRNA sensor EIF2AK2 (PKR)[92] and USP39, a deubiquitinase that stabilizes STAT1 to sustain type I IFN immunity.[93] This disparity raises intriguing questions about the role of ORF1p overexpression on shaping cancer-specific aneuploidies and immune responses, which could be reflective of the unique evolutionary pressures or tissue-specific contexts inherent to each cancer[94], for example squamous cell carcinomas (HNSCC and LSCC) versus adenocarcinomas (UCEC) in our study.

This dichotomy suggests cancer specific adaptations to L1 derepression and our data indicate that ORF1p expression may compromise tumor immunosurveillance in squamous cell carcinomas by reducing type I IFN activation and subsequent pDC infiltration. Further, our data suggest that reduced infiltration of key immune cells, including plasmacytoid and activated myeloid dendritic cells in HNSCC and LSCC, potentially compromises tumor immunosurveillance by decreasing type I interferon secretion and adaptive immune response recognition in these tumor types. Further experimental validation is needed to connect ORF1p's role in masking dsRNA from

repetitive elements and precise mechanisms by which ORF1p may modulate immune response in the context of immune cell type recruitment.[29,42]

Our study also offers valuable insights into the role of ORF1p expression in modulating global protein and phosphoprotein expression. We identified PPM1G phosphatase upregulation in ORF1p high tumors as particularly compelling, as ORF1p-high HNSCC tumors upregulate PPM1G to negatively modulate immune activation by regulating post-translational modification of innate immune adapter proteins and subsequent type I IFN gene expression programs, similar to the immune evasion effect caused by hijacking of PPM1G during Kaposi's sarcoma-associated herpesvirus (KSHV) infection.[69] The observed reduction in STING1 and MAVS phosphorylation in ORF1p high HNSCC tumors suggests that PPM1G overexpression dampens antiviral responses, facilitating a tolerogenic environment that supports L1 overexpression. Additionally, proteins upregulated in ORF1p-high tumors include those in RNA and RNP metabolism, as ORF1p-high tumors may upregulate these proteins to stabilize L1 RNA expression and support active L1 translocation.[9,67,95]

This study improves our understanding of ORF1p overexpression in cancer, but several questions remain. First, though we identify significant associations of ORF1p with proteins, immune pathways and aneuploidies, precise molecular mechanisms remain unclear and need more basic experimental follow-up. Second, sample sizes for specific demographic variables and somatic mutation statuses, such as p53, varied per cancer type, warranting validation analyses with independent proteogenomic datasets. In the age of personalized medicine, our findings advance ORF1p as a candidate biomarker by introducing novel tumor-specific associations of ORF1p with immune infiltration and copy

number alterations that can enhance clinical utility.[75,96–98] Emerging development of ultrasensitive immunoassays for detection of plasma-circulating ORF1p can be leveraged to predict immune checkpoint inhibitor (ICI) resistance, especially in ORF1p-high tumors, and to longitudinally monitor immune adaptation to intercept emerging resistance during ICI therapy.[99–101]

Our comprehensive pan-cancer characterization of L1 ORF1p provides valuable insights into the molecular underpinnings of L1 activity in cancer. These findings pave the way for future research to explore the therapeutic potential of targeting ORF1p and its associated molecular consequences in specific cancer treatments.

METHODS

CPTAC Pan-Cancer Data

All data used in this study can be downloaded from the NCI's Proteomic Data Commons (PDC) at https://proteomic.datacommons.cancer.gov or https://proteomic.datacommons.cancer.gov/pdc/cptac-pancancer. Details regarding sample acquisition, harmonization of the pan-cancer data processing, demographic composition and methods can be found in the CPTAC pan-cancer resource paper.[43]

Human Subjects

In this study, we used proteogenomic and clinical data obtained by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) from over 1,000 eligible participants. This cohort (n = 1,072) spans 10 cancer types: breast adenocarcinoma (BRCA, n = 122)[51], clear cell renal cell carcinoma (CCRCC, n = 110)[56], colon adenocarcinoma (COAD, n = 96)[52],

glioblastoma multiforme (GBM, n = 99)[102], head and neck squamous cell carcinoma (HNSCC, n = 110)[55], lung squamous cell carcinoma (LSCC, n = 108)[58], lung adenocarcinoma (LUAD, n = 110)[57], pancreatic ductal adenocarcinoma (PDAC, n = 140)[59], high-grade serous ovarian cancer (OV, n = 82)[53] and uterine corpus endometrial carcinoma (UCEC, n = 95).[54] Protocols and consent documentation were reviewed by the institutional review boards at the tissue source sites, ensuring compliance with CPTAC guidelines. All specimens were collected from patients during routine tumor resections, with informed consent obtained from the patients, in accordance with CPTAC standards and guidelines.

Clinical Data

Clinical information was downloaded from CPTAC Data Coordinating Center (DCC). Clinical variable names and units were standardized and unified. Clinical and molecular data can be accessed and downloaded from the NCI's Proteomic Data Commons (PDC) at https://proteomic.datacommons.cancer.gov.

Specimen Acquisition

CPTAC collected tumor and normal adjacent tissue (NAT) and whole blood samples derived from patients diagnosed with BRCA, CCRCC, COAD, GBM, HNSCC, LSCC, LUAD, OV, PDAC and UCEC. All cases in the study met the histology acceptance criteria, and detailed information regarding sample processing can be found in each respective flagship publication and the CPTAC pan-cancer resource paper.[43]

Somatic Copy Number Variation (CNV) Data

CNV ratios were derived from whole genome (WGS) and whole exome sequencing (WES) for each tumor type as described in Li et al., 2023.[43]

Somatic Mutation Data, Mutational Burden and Signatures

Somatic mutation calls from WES data for matched tumor/normal samples were analyzed using a hg38 WES characterization pipeline and harmonized as described in Li et al., 2023.[43] Estimated tumor mutational burden and mutational signatures were derived using the Genomic Identification of Significant Targets in Cancer (GISTIC2.0)[103] algorithm and MutSig2V[104] as described in Li et al., 2023.[43]

RNA Expression Data

Global RNA expression data was obtained from gene-level read counts (Fragments Per Kilobase of transcript per Million mapped reads, FPKM) and processed as described in Li et al., 2023.[43] Data can be downloaded from the Genomic Data Commons (GDC) at https://portal.gdc.cancer.gov/ or from https://proteomic.datacommons.cancer.gov/pdc/cptac-pancancer.

Cell Type Enrichment Deconvolution (RNA)

Cell type abundance was inferred by the xCell web tool[78], which performed the cell type enrichment analysis from gene expression data for 64 immune and stromal cell types (default xCell signature) as described in Li et al., 2023.[43] The xCell immune score per sample that integrates the enrichment scores B cells, CD4+ T-cells, CD8+ T-cells, DC,

eosinophils, macrophages, monocytes, mast cells, neutrophils, and NK cells. The xCell microenvironment score is the sum of the immune score and stroma score.

Protein Expression Data

Global protein expression data and peptide-spectrum match (PSM) counts were processed as described in Li et al., 2023[43] and can be downloaded from the PDC.

Phosphoprotein Expression Data

Global, site-level phosphoprotein expression data from phosphopeptide-enriched proteomic data were processed as described in the methods of Li et al., 2023.[43] Data can be downloaded from the PDC.

Adjustment of Phosphoprotein Abundance by Protein Expression

To analyze phosphorylation changes that happen specifically at the phosphoprotein level, we performed a linear regression between phosphosite abundance (BCM pipeline) and protein abundance (BCM pipeline) of each phosphosite using the base R package stats (lm function). We then computed the response residuals for each phosphosite, which normalizes the difference between the observed protein abundance and expected value based on protein expression (residual) by their standard deviation.

Adjustment of Somatic Copy Number by Tumor Purity and Ploidy

Gene-level copy number and chromosome arm-level copy number tables were adjusted for tumor purity and ploidy. The log2 copy number is first linearized. The normal cell

contribution to the copy number signal is calculated by multiplying the fraction of normal cells in the sample by the deviation of the sample from diploid (normal cell is assumed to be diploid) and by the integer 2. The normal cell contribution is then divided by the tumor purity multiplied by the ploidy to determine the correction which is then added back to the linearized copy number input and log2 transformed. This adjustment corrects the tumor copy number by unmixing the tumor signal from normal cell contamination and adjusts for a non-diploid tumor baseline.

Calculation of Chromosome Arm and Cytoband-Level Copy Number

WES log2 gene-level CNV matrices from the BCM pipeline were used to derive chromosome arm-level and cytoband-level copy number matrices per cancer type by calculating the mean copy number of all genes for that chromosome arm or cytoband.

Calculation of Aneuploidy Score

To estimate the extent of tumor aneuploidy, we calculated the sum of the absolute value of the copy number for each chromosome arm per cancer type.

L1 Quantifications

Intact L1 RNA

L1 RNA was quantified as described in McKerrow et al., PNAS, 2022[44] by passing available RNA sequencing BAM files from the GDC to L1EM[105] on the Cancer Genomics Cloud. L1EM utilizes expectation maximization to estimate locus-specific L1 expression and to separate proper L1 expression from passive co-transcription that includes L1 RNA

but does not support retrotransposition. For intact L1 RNA quantifications, full length L1 loci with no stop codon in either ORF1 or ORF2 were considered expressed if at least two read pairs per million (FPM) were assigned to that locus and less than 10% of the RNA assigned to that locus was estimated to be passive co- transcription. Total intact L1 RNA expression was estimated by adding together the FPM values for each such locus. This number is an estimate of the total amount of active L1 RNA expression from intact loci. By “active” we mean an expression that is driven by the L1 promoter (i.e. the TSS is within 25 bp of the start of L1). Samples were excluded if no locus was detected at 2 FPM as this may be due to data quality rather than a lack of L1 RNA.

L1 ORF1p

L1 ORF1p was quantified as described in McKerrow et al., PNAS, 2022.[44] ORF1p was quantified from tandem mass tag (TMT) labeled mass spectrometry (MS/MS) proteomics data. X!Tandem[106] was used to search mass spectra against a combined database that included both the standard Ensembl human proteome and in silico translated L1 ORF1 sequences from hg38 to the human Ensembl proteome. Oxidation of methionine (+15.994915@M) was included as a potential modification. Carbamidomethylation of cysteine (+57.022@C) was set as a fixed modification in addition to modifications appropriate to the particular isobaric labeling used (+144.102063@[, +144.102063@K for iTRAQ4 and +229. 162932@[, +229.162932@K for TMT10). Peptides were quantified by calculating the log ratio between the reporter intensity for each sample and the reporter intensity of an internal control. Peptide/peptide spearman correlations were calculated and peptides that had a correlation of at least 0.6 with two other peptides were retained.

We then identified matches to 20 ORF1p peptides: DFVTTRPALK, EALNMER, EWGPIFNILK, LIGVPESDGENGTK, LIGVPESDVENGTK, LSFISEGEIK, LTADLSAETLQAR, NLEECITR, NVQIQEIQR, QANVQIQEIQR, REWGPIFNILK, RNEQSLQEIWDYVK, SNYSELR, SNYSELREDIQTK, VSAMEDEMNEMK, YQPLQNHAK, DFVTTRPALQELLK, LENTLQDIIQENFPNLAR, VSAMEDEMNEMKR, and NEQSLQEIWDYVK. If a peptide were detected multiple times in the same sample, the median across peptide spectral matches was used. Then to get a protein quantification, the median was taken across all peptides in the preceding list that were identified in that sample. Finally, the quantification was translated by the median log ratio for all human proteins to account for variation in the size of the input sample. This pipeline was implemented and executed as a workflow on the Cancer Genomics Cloud.

High Confidence Somatic Insertions

Somatic insertions were identified as described in McKerrow et al., PNAS, 2022[44] by using the Mobile Element Locator Tool (MELT)[107] v2.1.15 on the Cancer Genomics Cloud. MELT was run on each pair of matched cancer/normal WGS datasets (no WGS for BRCA, OV, and COAD). Only insertions with the greatest evidence (ASSESS = 5) were considered. To be considered a high-confidence somatic insertion, the insertion had to be called heterozygous in cancer and homozygous absent in normal, with the log likelihood of both of these genotypes being at least 10 greater than the log likelihood of the next most probable genotype.

ORF1p-High versus ORF1p-Low Analyses

Derivation of ORF1p-High and ORF1p-Low Groups

ORF1p-high and OR1p-low groups were derived using quartiles from the distribution of ORF1p expression per individual cancer type. Samples whose ORF1p expression was less than or equal to the first quartile were considered “low”, and samples whose ORF1p expression was greater than or equal to the third quartile were considered “high”. Samples not labeled as “high” or “low” were excluded from differential expression analyses.

Differential Proteomic Expression Analysis

The DEqMS R package[108] was used to perform differential protein expression analysis using proteomic data for each individual cancer type comparing ORF1p-high versus ORF1p-low samples. DEqMS is a robust method for both labelled and label-free MS data that considers peptide spectrum matches (PSMs) per protein to adjust for the inherent dependence of protein variance on the number of peptides used for quantification. The consideration for the actual MS-proteomics data structure improves the accuracy of differentially expressed protein (DEP) detection. Individual observed and normalized log2 reference intensity protein abundance tables from the BCM pipeline were used as input for each cancer type, filtered for ORF1p-high and ORF1p-low samples, along with matched PSMs obtained from the UMich-GENCODE34 proteomic tables. ORF1p expression was not included in these matrices, and rows with NAs were removed. WES tumor purity was used as a covariate in the model. For samples with missing tumor purity values, we used the mean tumor purity of all samples in the analysis.

Gene Set Enrichment Analysis for Differential Protein Expression

The clusterProfiler R package[109] was used to perform GSEA for human MSigDB hallmark gene sets. Proteins ordered by decreasing logFC from DEqMS were used as the gene list input. The minGSSize was set to 30, the maxGSSize was set to 500, and the pvalueCutoff was set to 1 to capture all results. The pAdjustMethod was set to “BH”.

RNA- and Protein Expression-Based Scores

Calculation of Tumor Inflammation Score

The tumor inflammation signature is a published set of 18 genes, including CD276, HLA-DQA1, CD274, IDO1, HLA-DRB1, HLA-E, CMKLR1, PDCD1LG2, PSMB10, LAG3, CXCL9, STAT1, CD8A, CCL5, NKG7, TIGIT, CD27, and CXCR6, that enrich for clinical benefit to immune checkpoint blockade by measuring the level of tumor microenvironment inflammation.[110] The score was calculated as referenced by ranking the gene expression values within each sample and taking the weighted average of the gene ranks within each sample. RNA and protein expression was used to infer tumor inflammation score, however RNA-based scores had less gene missingness.

Calculation of Cytotoxic Immune Score

The cytotoxic immune score was calculated as referenced in Davoli et al., Science, 2017[85] by ranking the expression of each gene per sample and then taking the sum of the ranks per sample. The gene set captures the expression of genes characteristic of the adaptive immune response and immune-mediated cytolytic and cytotoxic activities and includes

CD247, CD2, CD3E, GZMH, NKG7, PRF1, and GZMK. RNA and protein expression was used to infer cytotoxic immune score.

Calculation of CGAS-STING, RLR, and Interferon Signature Scores

These scores were inferred by single sample scoring using the singscore method from the singscore R package.[111] This method does not depend on background samples resulting in scores that are stable regardless of the composition and number of samples being scored. The CGAS-STING activation score (cytosolic dsDNA sensing score) includes CGAS, STING1, CXCL10, IRF3, TBK1, and STAT1.329,357 The RLR score (cytosolic dsRNA sensing score) includes DDX58 (RIG-I), IFIH1 (MDA5), MAVS, EIF2AK2 (PKR), and IRF3.[42,92,112] The curated list of 140 IFN and ISGs from Sun et al., Immunity, 2024[29] was used as the IFN signature score. RNA and protein expression was used to infer all scores.

Survival Analysis

The survival[113] and survminer[114] R packages were used to perform survival analysis. The Kaplan-Meier curve of overall survival was used to compare the pan-cancer prognosis among ORF1p status. Log-rank test (survminer) was used to test the differential survival outcomes between ORF1p-high and ORF1p-low groups.

Correlations and Linear Regressions

Correlations were run using the cor.test() function in R. The method argument was set to “spearman” and the alternative argument set to “two.sided”. Linear regressions were run

to account for covariates, such as tumor purity or CN ratio, using the R functions lm() for predicting continuous variables and glm() to predict binary (0, 1) outcomes, such as the prediction of chromosome gain (0, 1) or loss (0, 1).

Over-Representation Analysis (ORA) of Chromosomal Aneuploidies

ORA for All Gained or Lost Chromosomes Associated with ORF1p Expression

Individual linear models were run per cancer type to predict the binary (0,1) gain (CN > 0.3) or loss (CN < -0.3), as previously described, of cytogenic bands by ORF1p expression using tumor purity as a covariate:

1. *Gain (1) or No Gain (0) ~ ORF1p + WES Tumor Purity*
2. *Loss (1) or No Loss (0) ~ ORF1p + WES Tumor Purity*

Results were filtered for ORF1p associations with adjusted P values < 0.2. Cytogenic bands were stratified based on the direction of ORF1p association (positive or negative) and type of copy number event (gain or loss). Four gene lists, Pos-Gain, Pos-Loss, Neg-Gain and Neg-Loss were constructed from genes within all cytogenic bands that fit each respective category and used as input for ORA to determine overall pathways genomically gained or lost.

ORA for Individual Gained or Lost Chromosomes Associated with ORF1p Expression

To determine pathways unique to specific chromosomal aneuploidies associated with ORF1p, we performed similar ORAs as described above for individual chromosome arm aneuploidies. This time, results were more strictly filtered for significance (adjusted P <

0.05), and genes located on individual chromosome arms (p or q) were aggregated into gene lists. ORA was run for each individual significant aneuploidy to determine which pathways enriched on unique chromosome arms are associate with ORF1p expression.

Gene Set Collections for ORA

Gene set collections were obtained from the Molecular Signatures Database (MSigDB) using the msigdbr R package.[115] ORA was performed using multiple complementary approaches to capture diverse biological processes and pathways:

Gene Ontology Biological Process (GO:BP) ORA was assessed using the enrichGO function from the clusterProfiler R package.[109] Analysis parameters included a minimum gene set size of 10, maximum gene set size of 500, and BH multiple testing correction. The background universe consisted of all genes present in the regression analysis. Hallmark Pathway ORA was performed using the enricher function with MSigDB Hallmark gene sets as the reference database. The same statistical parameters and background universe were applied as for GO:BP analysis. BioCarta Pathway ORA was evaluated using BioCarta gene sets with identical statistical parameters to ensure consistency across analyses. For KEGG Pathway ORA, gene symbols were converted to Entrez IDs using the bitr function from the org.Hs.eg.db annotation package.[116] Enrichment was performed using enrichKEGG function with organism set to "hsa" (Homo sapiens). Reactome Pathway ORA was conducted similar to KEGG analysis using the enrichPathway function with Entrez IDs and organism set to "human". Gene lists containing fewer than 5 valid Entrez IDs were excluded from KEGG and Reactome analyses to ensure statistical robustness. Results were visualized using dot plots

generated with the dotplot function of the clusterProfiler package, showing the top 10 enriched terms per analysis.

Statistical Tests

Wilcoxon rank sum tests with BH multiple hypothesis correction were used for boxplot visualizations with more than 2 groups. A Student's T-test was used for comparisons between 2 groups with assumed normalized distribution (expression data). Wilcoxon rank sum tests between 2 groups were performed when the data was not assumed to be normally distributed. The threshold for significance for all boxplot visualizations is alpha < 0.05. For correlation and linear regression analyses with more than one variable per cancer type, BH multiple hypothesis correction was used to control for false discovery rate.

Plot Visualizations

All R-based results were visualized using ggplot[117], ggpubr[118] and ComplexHeatmap[119] R packages. All biological and methods schematics were created with BioRender.com.

ACKNOWLEDGEMENTS

This work utilized the computing resources at the NYU Grossman School of Medicine High Performance Computing (HPC) Facility and was supported by NIH P01AG051449 (D.F) and NIH/NCI U24CA210972 (D.F). T.D. and P.M. were supported by grants from NIH (R37 R37CA248631, R01 R01HG012590, and R01 R01DK135089), the Pershing Square Sohn Cancer Prize and an NFCR grant.

FIGURES

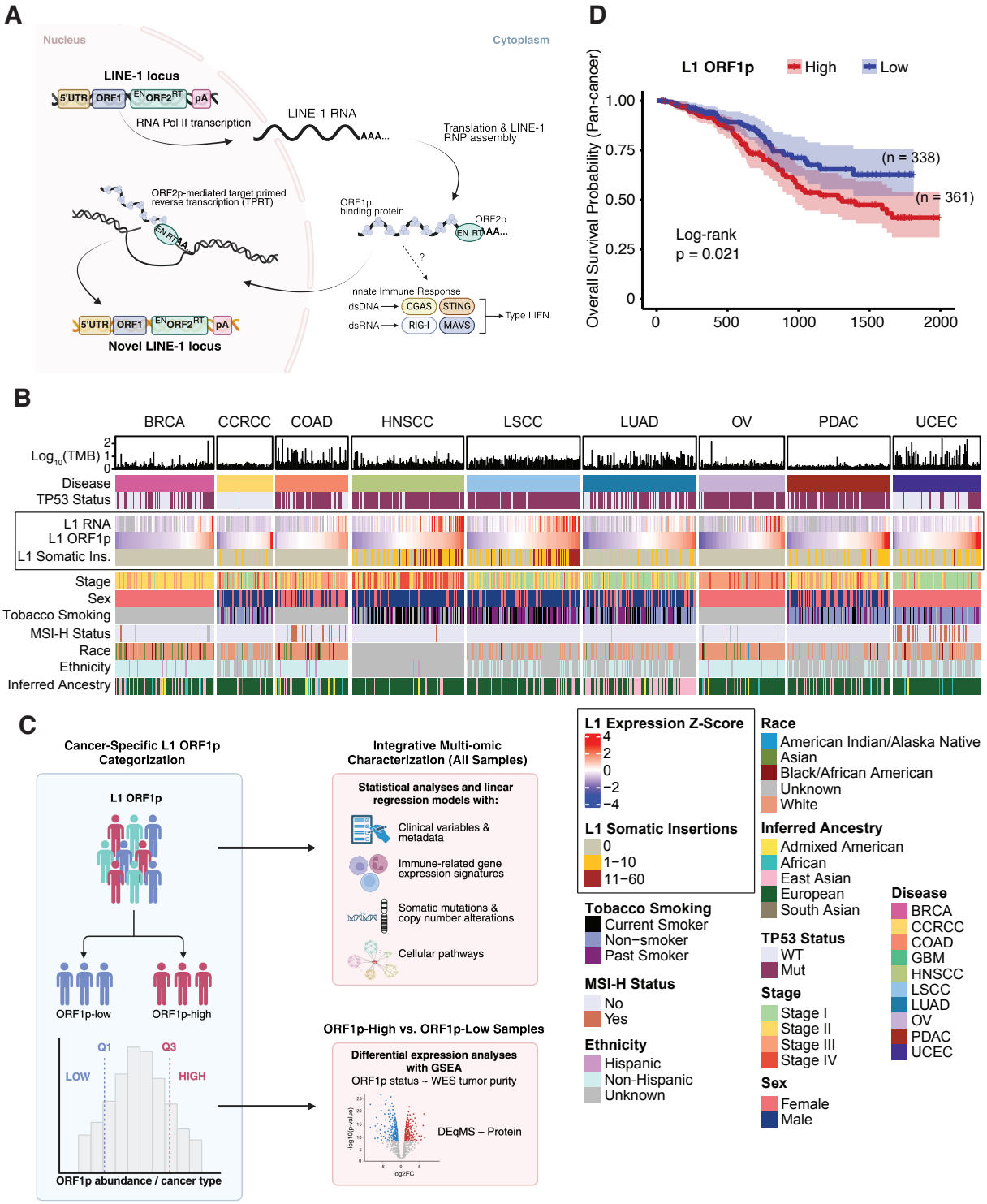


Figure 1. Pan-cancer overview of L1 activity.

(A) Schematic of L1 life cycle. L1 consists of two open reading frames, ORF1 and ORF2, and is transcribed by the cell's machinery. L1 mRNA translocates to the cytoplasm for ORF1/2 translation and ribonucleoprotein assembly whereby it re-enters the nucleus for ORF2p-mediated target primed reverse transcription (TPRT) into a new genomic locus. Formation of L1 dsRNA from L1 mRNA and DNA:RNA hybrids and L1 cDNA from reverse transcription can be sensed by cytosolic DNA and RNA sensors to trigger an innate immune response and type I IFN induction. (B) Heatmap ordered by increasing ORF1p expression stratified by CPTAC cancer type profiling clinical variables. L1 quantifications are indicated by the black box on the heatmap and legend. (C) Schematic outlining ORF1p-high and ORF1p-low sample categorization and downstream proteogenomic analysis. (D) Pan-cancer Kaplan-Meier plot comparing OS for patients stratified by ORF1p status, log rank test. Numbers in parentheses represent the pan-cancer sample size for each group.

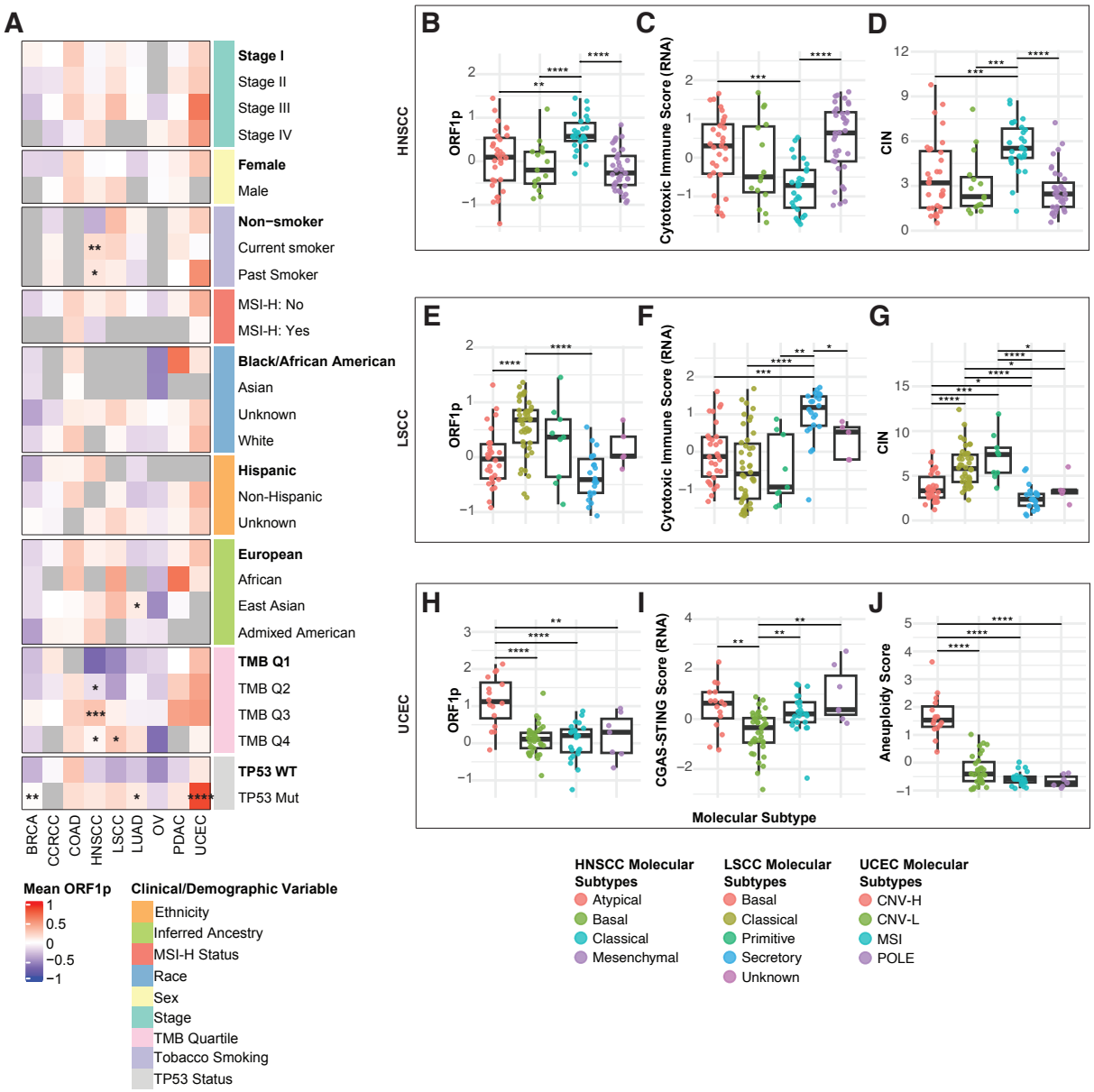


Figure 2. Clinical variable associations with L1 ORF1p.

(A) Mean ORF1p expression stratified by clinical variable sub-categories. Rows are sub-categories grouped by clinical variables. Columns are cancer types. Bolded sub-categories indicate the reference group used for Wilcoxon rank sum test of ORF1p distribution with BH p-value correction for variables with more than 2 sub-categories for

each clinical variable. Significant comparisons indicated with asterisk (*). (B-G) Boxplots of (B,E) ORF1p, (C,F) cytotoxic immune score and (D,G) CIN stratified by molecular subtypes of HNSCC and LSCC, respectively. (H-J) Boxplots of (H) ORF1p, (I) CGAS-STING score and (J) aneuploidy score stratified by molecular subtypes of UCEC. $p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.

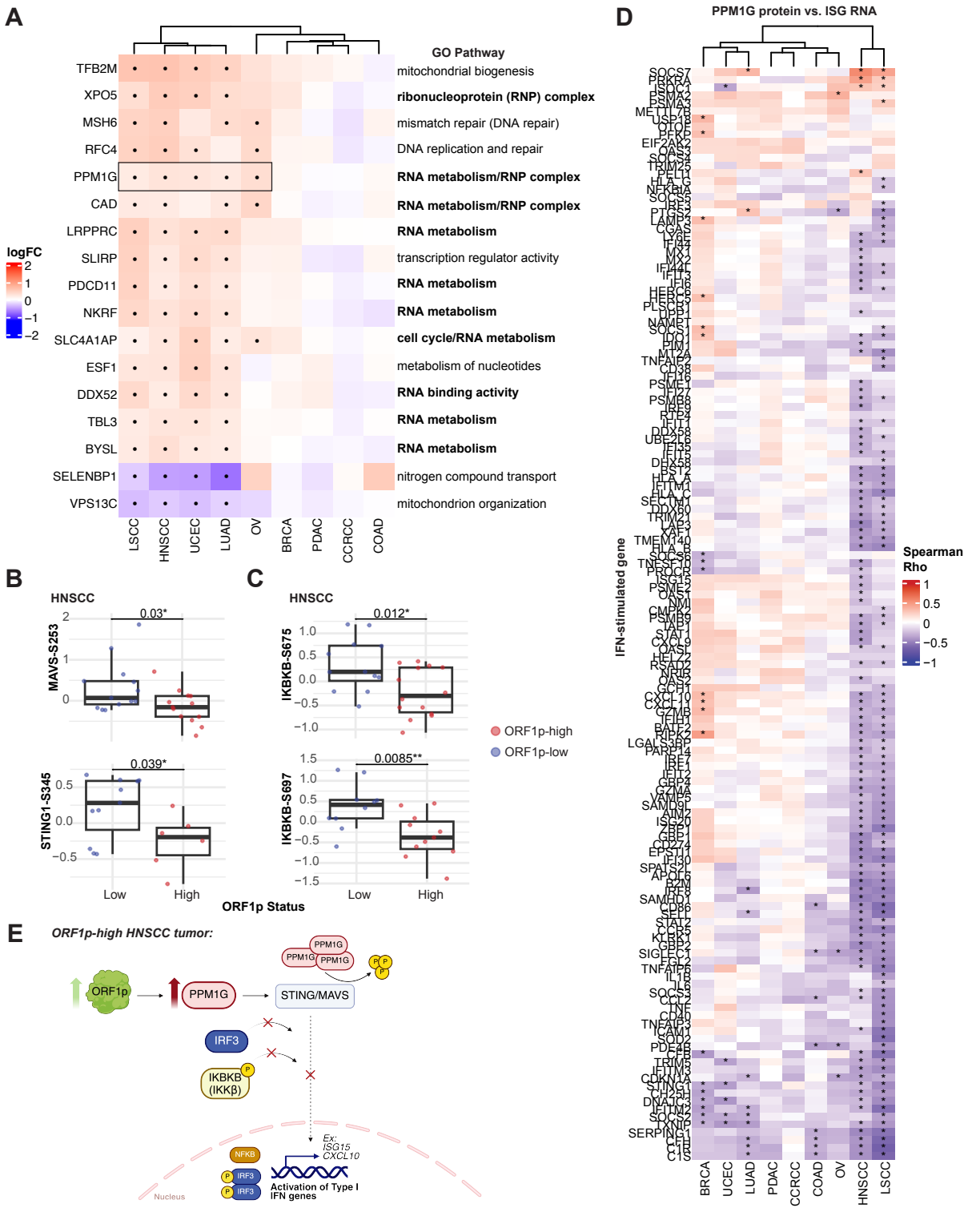


Figure 3. ORF1p modulates the expression and post-translational modification of proteins involved in IFN response

(A) Heatmap of logFC for differentially expressed proteins in ORF1p-high tumors filtered for P adjusted < 0.2 in at least four cohorts indicated by black dot (•). Associated Gene Ontology (GO) pathway to the right of heatmap with RNA/RNP metabolism pathways bolded. PPM1G is boxed to denote further investigation in panels B-D. (B-C) Protein-normalized phosphosite abundance of immune genes downstream of PPM1G significant in ORF1p-high and ORF1p-low groups of HNSCC. Two-tailed T-test. (D) Spearman correlation of RNA abundance of type I IFN genes from IFN signature gene set (Methods) with PPM1G protein in HNSCC. (E) Proposed model summarizing findings from panels A-D explaining post-translational control of immune suppression in ORF1p-high tumors of HNSCC. For T-tests: $p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.

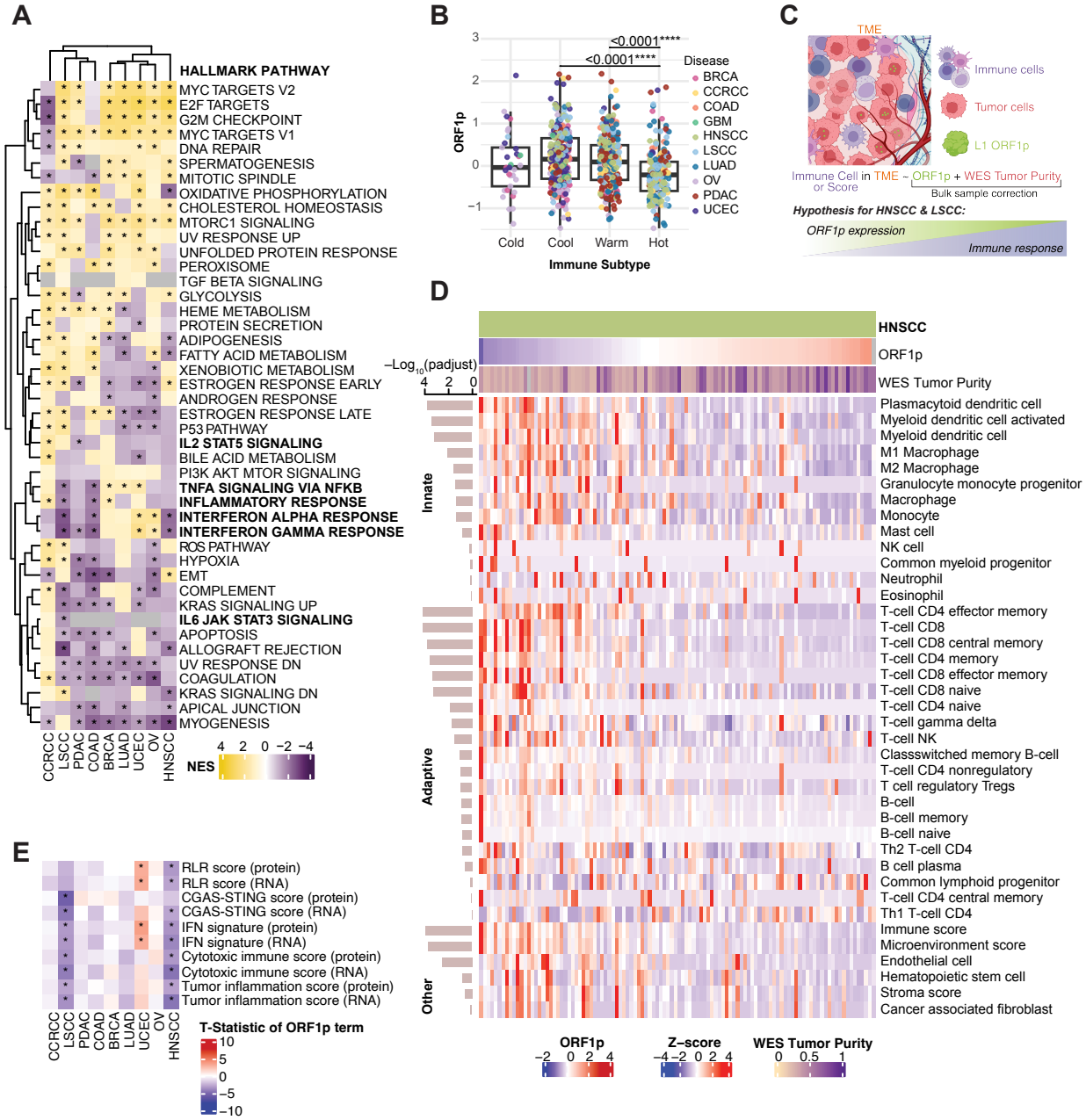


Figure 4. ORF1p expression regulates immune pathways and immune cell infiltration in a tumor-specific manner.

(A) Heatmap of normalized enrichment scores (NES) from GSEA of HALLMARK pathways. Input is a gene list ranked by decreasing logFC from differential proteomic

expression analysis (per cancer type). Rows are pathways and columns are cancer types. Asterisk (*) indicates significant NES (adjusted P < 0.01). Immune-related pathways are bolded. (B) ORF1p abundance by CPTAC immune subtypes from Li et al., 2023327: Cold, Cool, Warm and Hot. Each point is a sample colored by cancer type. Two-sided Wilcoxon rank sum test (BH-adjusted P values). (C) Schematic of tumor microenvironment (TME) explaining linear regression model used in panels D-E and decreased immune response in ORF1p-high squamous cell carcinoma tumors. (d) Heatmap of xCell immune cell type deconvolution scores ordered by increasing ORF1p expression in HNSCC. WES tumor purity covariate also annotated. Rows are immune cell types grouped by branches of the immune system. Rows within each group are ordered by decreasing significance of the P adjusted value of the ORF1p term from the linear regression model. (E) Heatmap of T-statistic from linear regression predicting the immune score (row) by ORF1p expression and tumor purity. Scores are calculated based on RNA and protein expression. Asterisk (*) indicates significant association with ORF1p expression (P adjusted < 0.01). Positive T-statistic indicates positive correlation, while negative T-statistic indicates negative correlation. For box plots: p ≤ 0.05, **p≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.

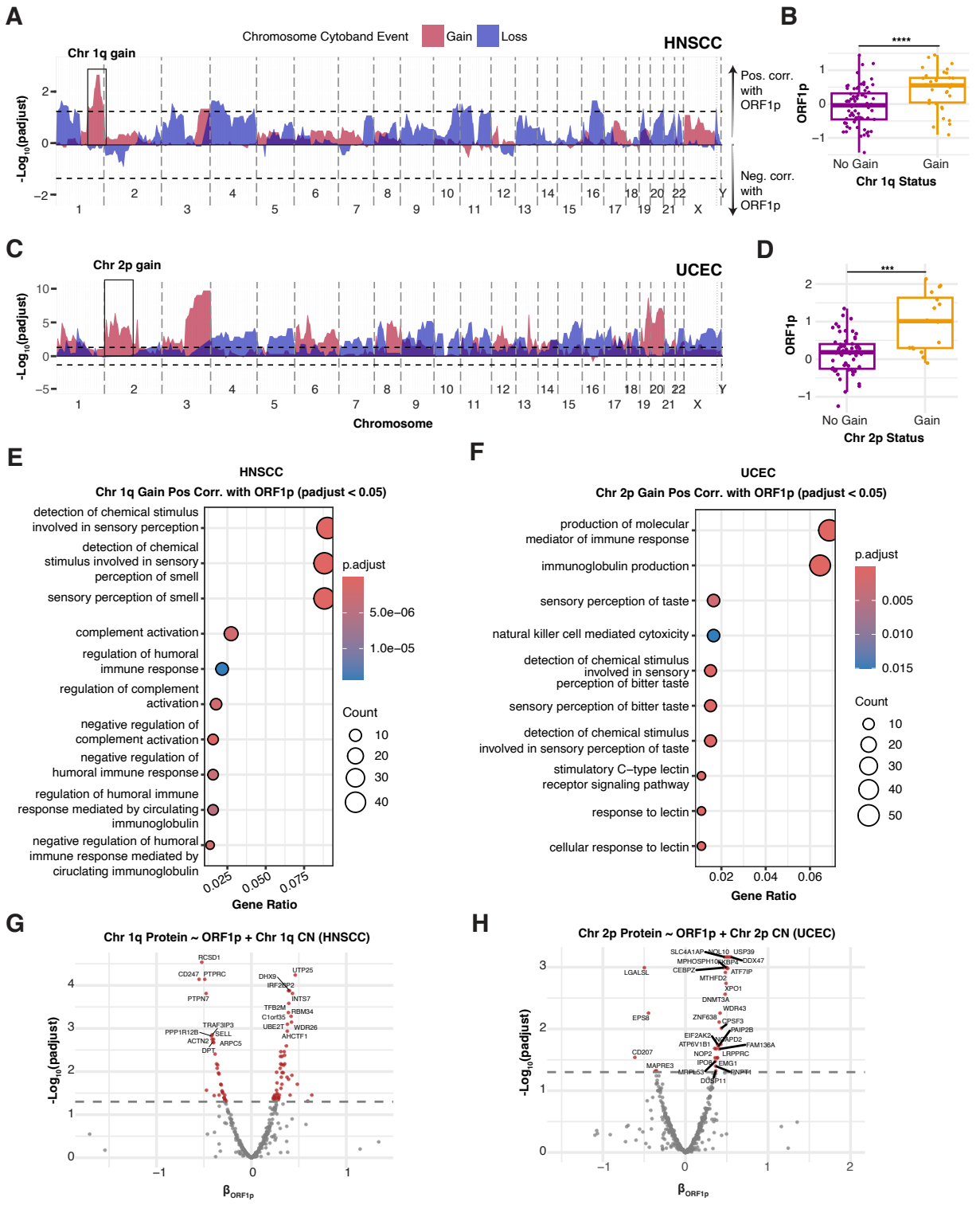
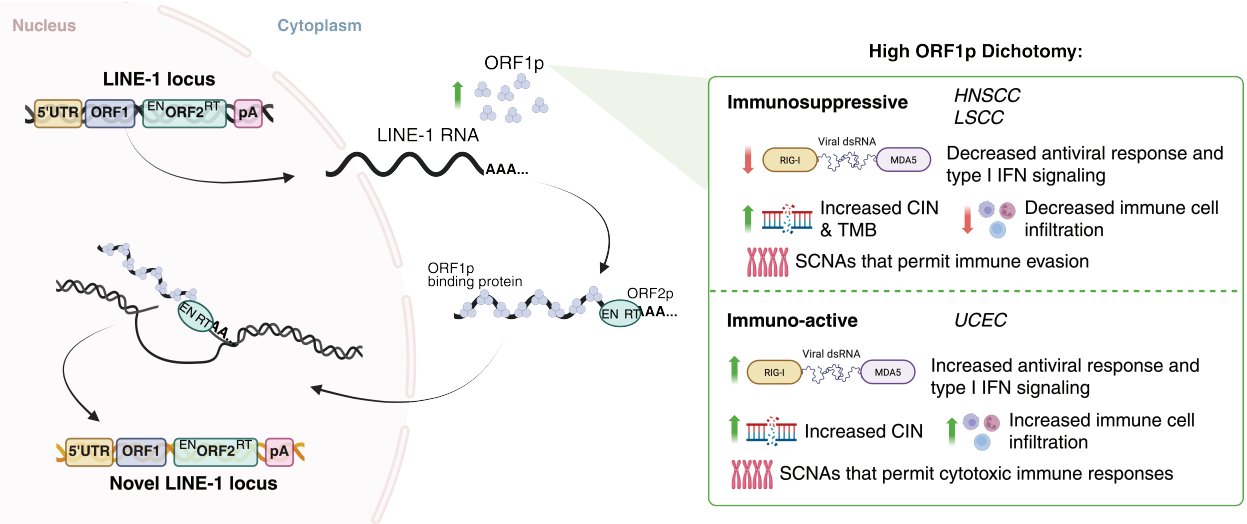


Figure 5. ORF1p expression predicts cancer-specific aneuploidies.

(A) Area plot represents the association between chromosome cytoband gains (red) and losses (blue) and ORF1p expression in HNSCC in a multivariable logistic regression model (see Methods for formula). Each chromosome is split by a grey vertical dotted line. The black horizontal dotted line represents adjusted P value = 0.05. Positive -Log10(adjustedP) indicates the cytoband gain or loss is positively associated with ORF1p expression, while a negative -Log10(adjustedP) indicates the cytoband gain or loss is negatively associated with ORF1p expression. Chromosome 1q gain is boxed to emphasize further analysis in B, E, and G. (B) ORF1p expression in HNSCC is stratified by gain or no gain of 1q. (C) Area plot represents the association between chromosome cytoband gains (red) and losses (blue) and ORF1p expression in UCEC. Chromosome 2p gain is boxed to emphasize further analysis in D, F and H. (D) ORF1p expression in UCEC is stratified by gain or no gain of 2p. (E) Dot plot showing top 10 GOBP pathways from ORA for ORF1p-predicted (adjusted $P < 0.05$) gained 1q cytobands that are positively associated with ORF1p in HNSCC. (F) Dot plot showing top 10 GOBP pathways from ORA for ORF1p-predicted (adjusted $P < 0.05$) gained 2p cytobands that are positively associated with ORF1p in UCEC. (G) Volcano plot -Log10(adjustedP) vs. β-coefficient of ORF1p from linear model predicting protein expression of genes located on significantly predicted (adjusted $P < 0.05$) gained 1q cytobands in HNSCC. Arm-level CN ratio (continuous) is used as a covariate to determine associations independent of CN. (H) Volcano plot -Log10(adjustedP) vs. β-coefficient of ORF1p from linear model predicting protein expression of genes located on significantly predicted (adjusted $P < 0.05$) gained 2p cytobands in UCEC.

Figure 6. Summary schematic of ORF1p-high dichotomy.

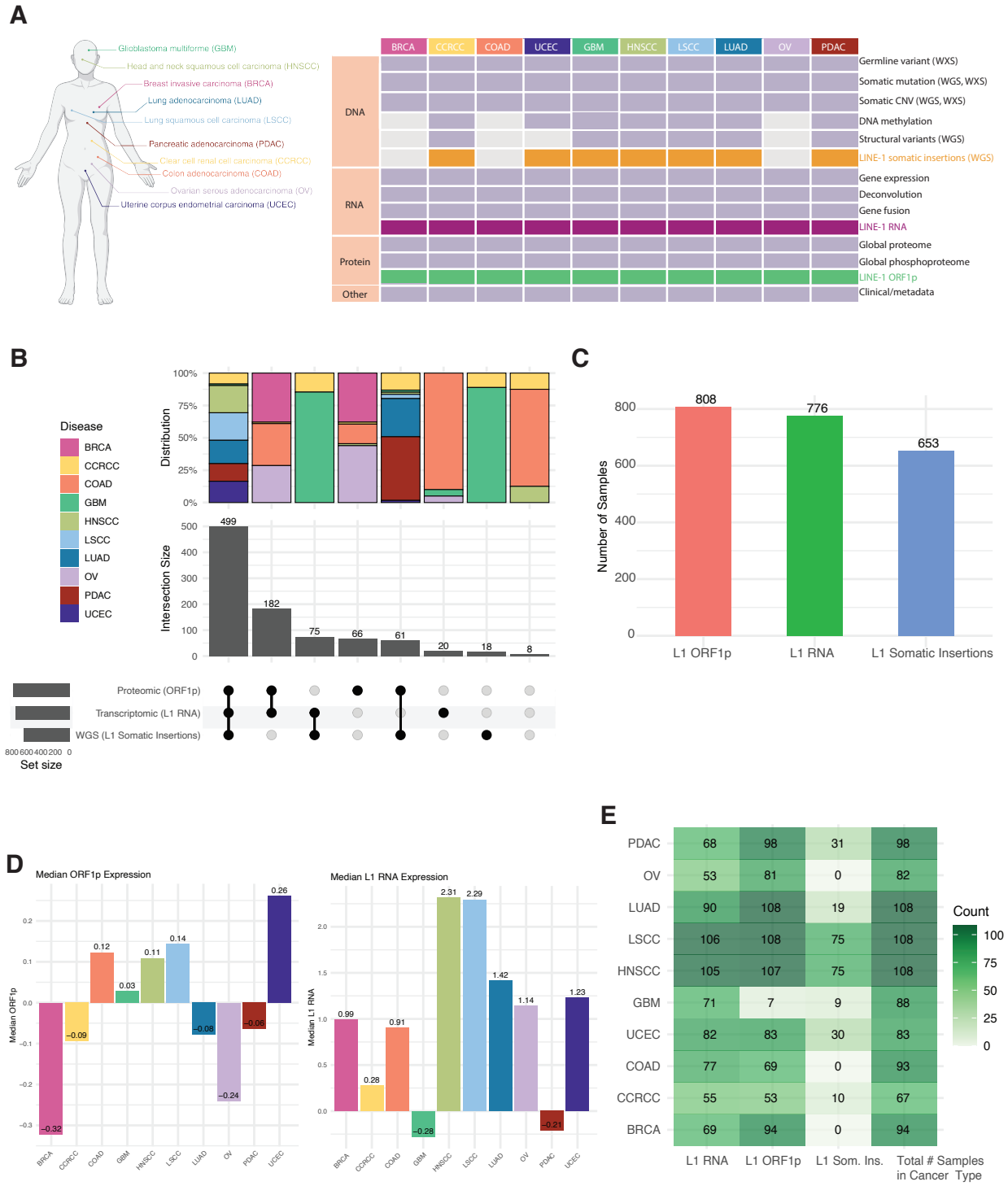


ORF1p-high tumors can be divided into two groups: an (1) immunosuppressive group characterized by negative correlations between ORF1p expression and antiviral responses and innate immune signaling, increased CIN and TMB, and decreased immune cell infiltration in the background of copy number alterations that permit immune evasion, and an (2) immuno-active group characterized positive correlations between ORF1p expression and antiviral/innate immune responses, increased CIN and immune cell infiltration in the background of copy number alterations that permit cytotoxic immune functions.

Table 1. ORF1p-high proteomic signature from differential protein expression.

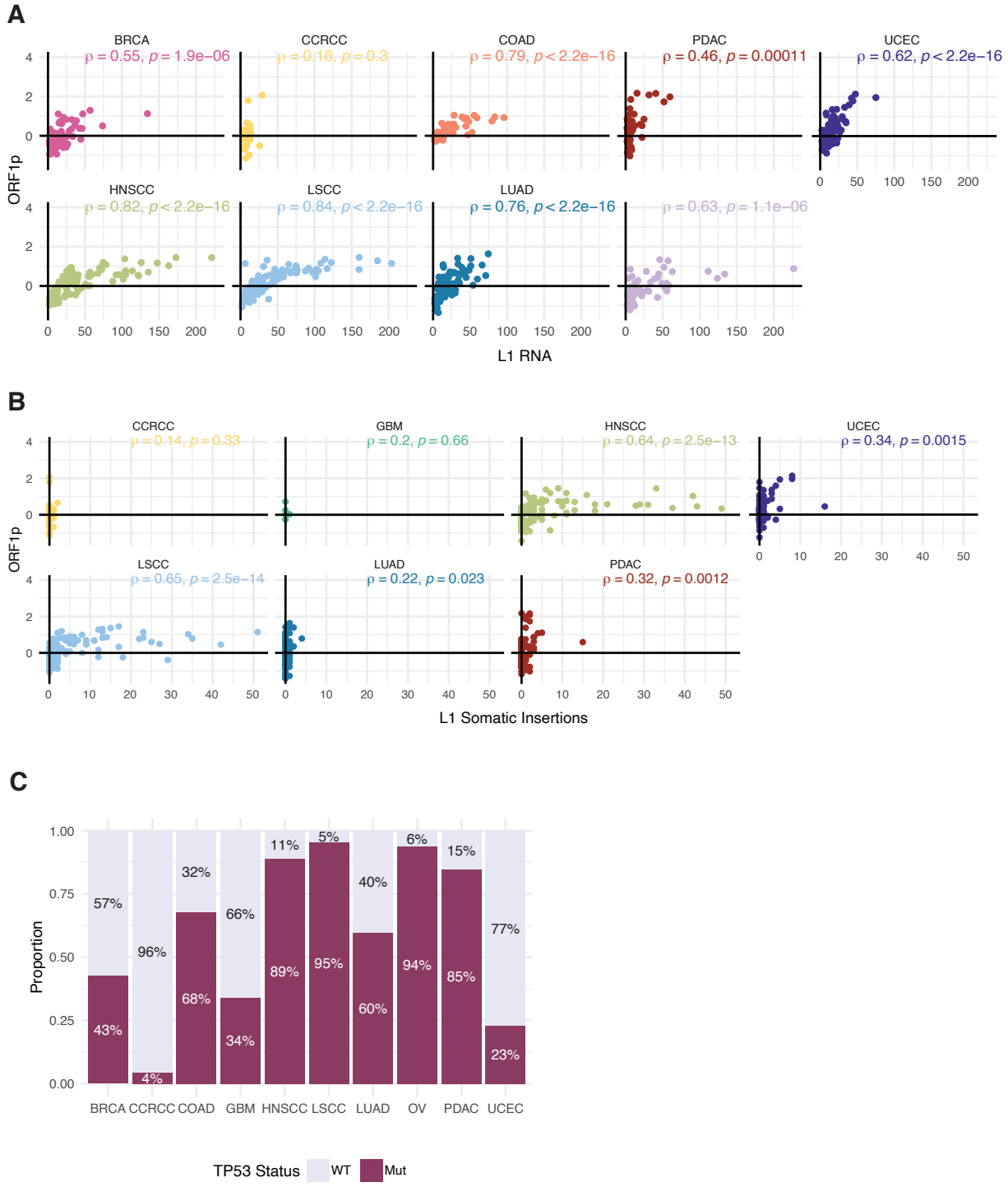
Protein	Known function	Direct involvement in IFN response/innate immunity:	Indirect involvement in IFN response/innate immunity:	Activator or repressor of IFN response or innate immunity?	Notes	References (PMID)
TFB2M	Mitochondrial transcription factor B2; essential for mitochondrial DNA transcription initiation		YES	Activator?	In ovarian cancer, TFB2M overexpression increases cytosolic mtDNA. This activates the TLR9/NF-κB/IL-6 pathway	PMID: 35577351
XPO5	Exportin-5; mediates nuclear export of pre-miRNAs and some proteins		YES	Repressor	XPO5 can regulate interferon response by regulating the microRNA (miRNA) pathway, which is a key component of the innate immune system. Some viruses, such as SARS-CoV-2, can target and degrade XPO5 to inhibit the maturation of miRNAs. XPO5 also binds dsRNA and stimulates RBPs and ADAR1export from the nucleus	PMID: 19124606 PMID: 14631048
MSH6	DNA mismatch repair protein; also functions in antibody somatic hypermutation and class switch recombination		YES	Repressor?	Reduced expression of the DNA mismatch repair (MMR) protein MSH6 is associated with a more active immune response and a stronger interferon signature in several cancers, particularly endometrial cancer. This inverse relationship suggests that MSH6's normal function may suppress immune signaling, and its loss can trigger an immune response that is more sensitive to immunotherapy	PMID: 38390329
RFC4	Replication factor C subunit; involved in DNA replication and repair	NO	NO			
PPM1G	Protein phosphatase; regulates pre-mRNA splicing and chromatin structure	YES		Repressor	Known regulator of the innate immune adapter proteins stimulator of interferon response cGAMP interactor 1 (STING1) and mitochondrial antiviral signaling protein (MAVS).	PMID: 33219031
CAD	Multifunctional enzyme in de novo pyrimidine biosynthesis (carbamoyl-phosphate synthase 2, aspartate transcarbamylase, dihydroorotase)		YES	Repressor	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD) influences the interferon (IFN) response through its role in pyrimidine synthesis and through complexing with innate immune proteins. CAD also deamidates the NF-κB pathway transcriptional factor RelA (also known as p65)	PMID: 38743960
LRPPRC	Leucine-rich pentatricopeptide repeat-containing protein; regulates mitochondrial gene expression and mRNA stability	YES		Repressor	LRPPRC interacts with and inhibiting the mitochondrial antiviral signaling protein MAVS, a key adaptor protein in antiviral immunity	PMID: 30070380
SLIRP	Stem-loop interacting RNA binding protein; mitochondrial RNA stability and regulation	YES		Activator	SLIRP amplifies the interferon (IFN) response by stabilizing and promoting the release of mitochondrial double-stranded RNAs (mt-dsRNAs), which then activate MDA5, a pattern recognition receptor	PMID: 40253699
PDCD11	Programmed cell death protein 11; nucleolar protein involved in rRNA processing and NF-κB pathway regulation		YES	Repressor	PDCD11 does not have a well-documented role in directly regulating the interferon (IFN) response. However, there is evidence of an indirect, regulatory connection through its effect on nuclear factor-kappa B (NF-κB); PDCD11 interacts with ADAR1	PMID: 39673305 PMID: 32709934
NKRF	NF-κB repressing factor; transcriptional repressor of NF-κB target genes including IFN-β and cytokines	YES		Repressor	NKRF plays a key role in the interferon (IFN) response by inhibiting the transcription of certain genes, including the IFN-β gene. NKRF's function is to prevent an excessive or prolonged immune reaction that could harm the host	PMID: 37084822
SLC4A1AP	Solute carrier family 4A1 adaptor protein; involved in centrosome duplication and mitotic spindle formation	NO	NO			
ESF1	ESF1 homolog; nucleolar protein required for pre-rRNA processing	NO	NO			
DDX52	DEAD-box helicase 52; functions in ribosome biogenesis and RNA processing			Repressor?	DDX52 appears to act as an antiviral protein against some viruses, but there is no direct evidence that DDX52 is part of a canonical interferon-response pathway.	PMID: 29146961
TBL3	Transducin beta-like 3; WD40 repeat protein involved in RNA metabolism	YES		Repressor	TBL3 is an interferon-stimulated gene (ISG) whose expression increases during an interferon antiviral response and that inhibits IRF1 (interferon (IFN) regulatory factor-1)	PMID: 25051498
BYSL	Bystin-like protein; ribosome biogenesis and early embryonic cell adhesion	NO	NO			
SELENBP1	Selenium-binding protein 1; implicated in redox regulation, oxidative stress response, and cancer biology		YES?	Repressor?/Activator?	SELENBP1 is involved in regulating the inflammatory and oxidative stress pathways that are modulated by the interferon response. Its exact role is complex and may vary depending on the context, such as in autoimmune diseases versus cancer	PMID: 36253721 PMID: 38172565
VPS13C	Vacuolar protein sorting 13 homolog C; lipid transfer protein at organelle contact sites, mutations linked to Parkinson's disease	YES		Repressor	Loss of the lipid-transport protein VPS13C activates the innate immune interferon response through the cGAS-STING pathway. This occurs because dysfunctional lysosomes, caused by a lack of VPS13C, become leaky, leading to the escape of mitochondrial DNA (mtDNA) into the cytosol	PMID: 35657605

SUPPLEMENTAL FIGURES



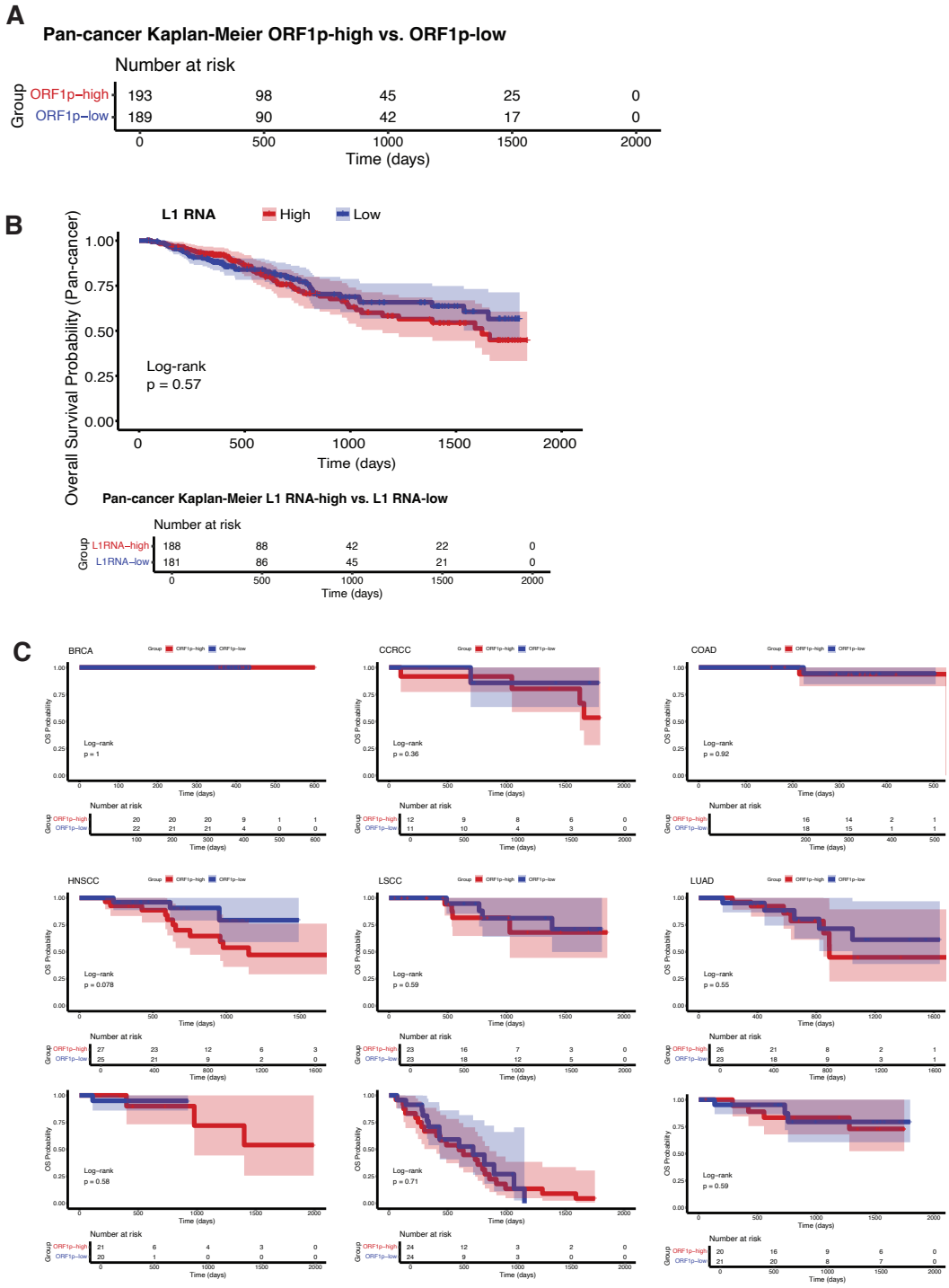
Supplemental Figure 1. Breakdown of L1 quantifications across CPTAC samples.

(A) Left: Schematic of CPTAC cancer tissue origin. Right: Heatmap of data type availability across CPTAC cancer types. (B) Upset plot counting number of samples containing L1 ORF1p, RNA, and insertion data across cancer types. Proportion bar plot visualizes distribution of cancer types for each set. 499 samples had quantifications in all three L1 measurements. (C) Bar plot summarizing the number of samples with each L1 measurement (n denoted above bar). (D) Bar plot summarizing the median L1 RNA and ORF1p expression level for each cancer type (median denoted in or above bar). (E) Heatmap summarizing the number of samples with each LI1 measurement for each cancer type and the total number of samples analyzed for L1 for each cancer type.



Supplemental Figure 2. L1 measurement correlations and p53 status breakdown.

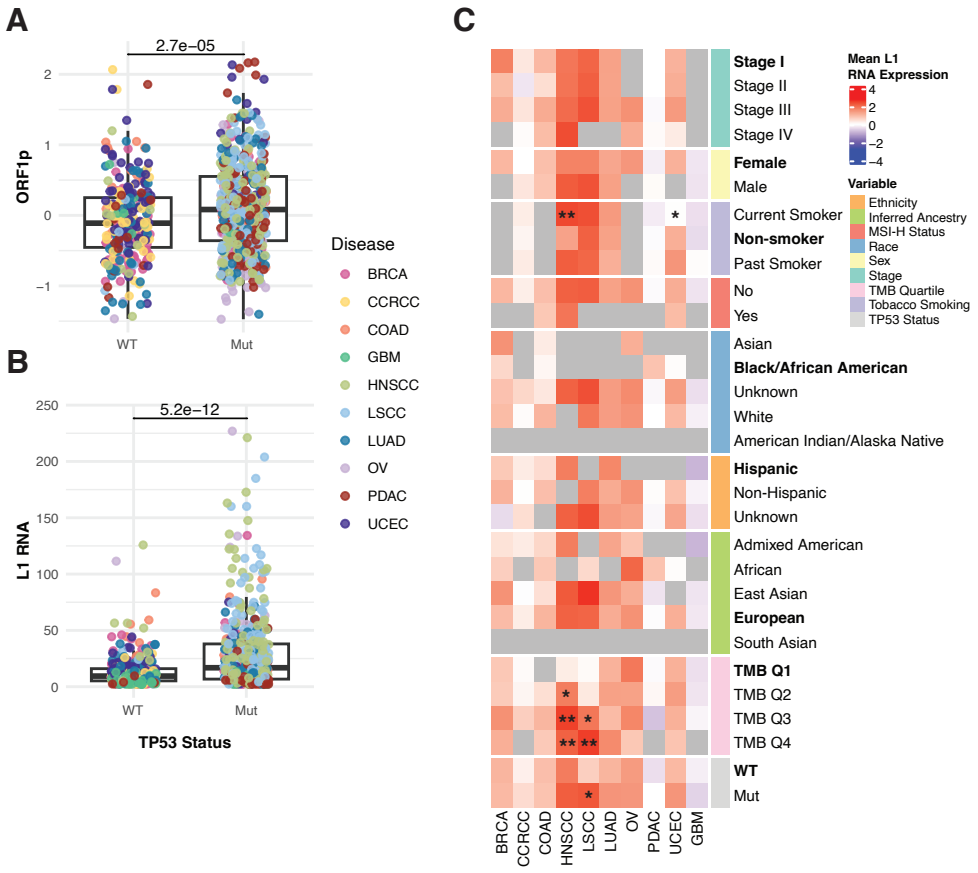
(A) Spearman correlations of L1 ORF1p (Log2 ratio) vs. L1 RNA (FPKM) per cancer type. (B) Spearman correlations of L1 ORF1p vs. L1 somatic insertion counts per cancer type. ρ = Spearman rho coefficient, p = p-value (C) Proportion bar plot summarizing TP53 mutant or WT status of samples per cancer type. Proportion denoted as percentage.



Supplemental Figure 3. Kaplan-Meier survival analyses.

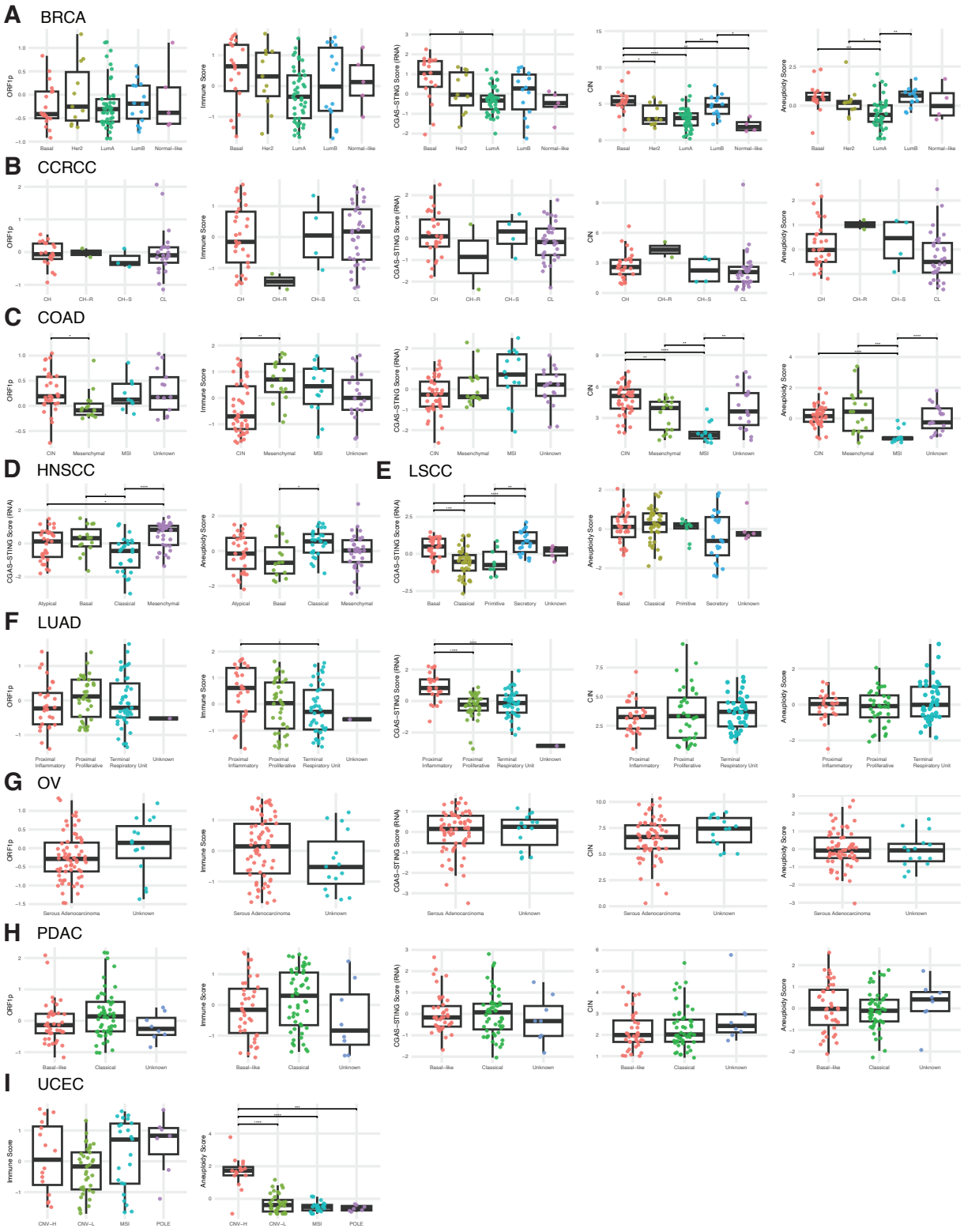
(A) Pan-cancer sample numbers at risk for ORF1p-high and ORF1p-low groups. (B) Pan-cancer Kaplan-Meier plot comparing OS for patients stratified by L1 RNA status, log rank

test. Numbers in parentheses represent the pan-cancer sample size for each group. (C) Individual Kaplan-Meier analyses for ORF1p-high vs. ORF1p-low samples per cancer type.



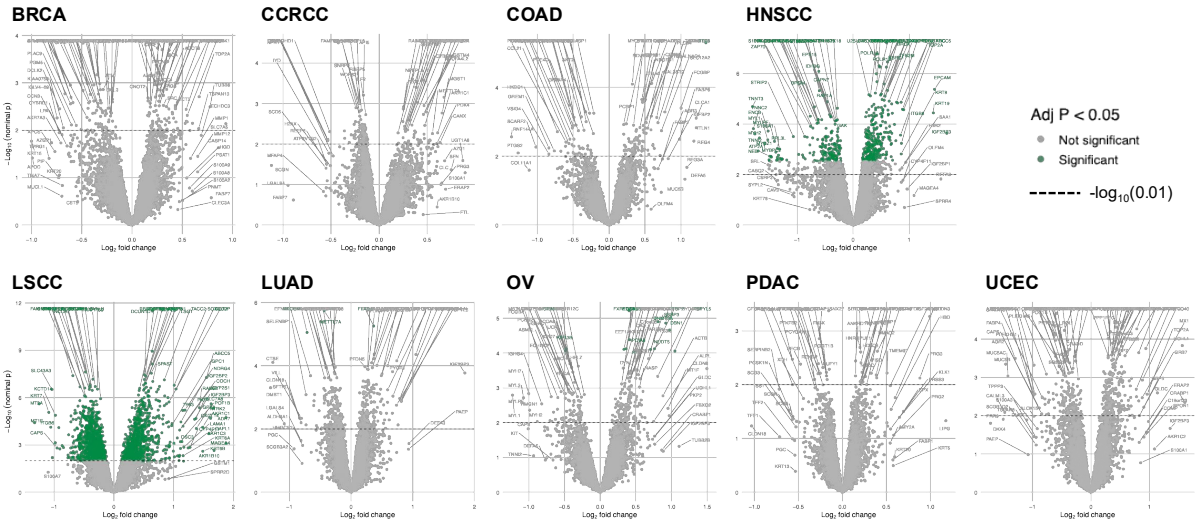
Supplemental Figure 4. Clinical variable associations with L1, continued.

(A-B) Pan-cancer ORF1p (A) and L1 RNA (B) levels by p53 status. Student's T-test for significance. Cancer type colors for both box plots indicated on the right of (B). (C) Mean L1 RNA expression stratified by clinical variable sub-categories. Rows are sub-categories grouped by clinical variables. Columns are cancer types. Bolded sub-categories indicate the reference group used for Wilcoxon rank sum comparison of L1 RNA distribution with BH p-value correction for variables with more than 2 sub-categories for each clinical variable. Significant comparisons indicated with asterisk (*).



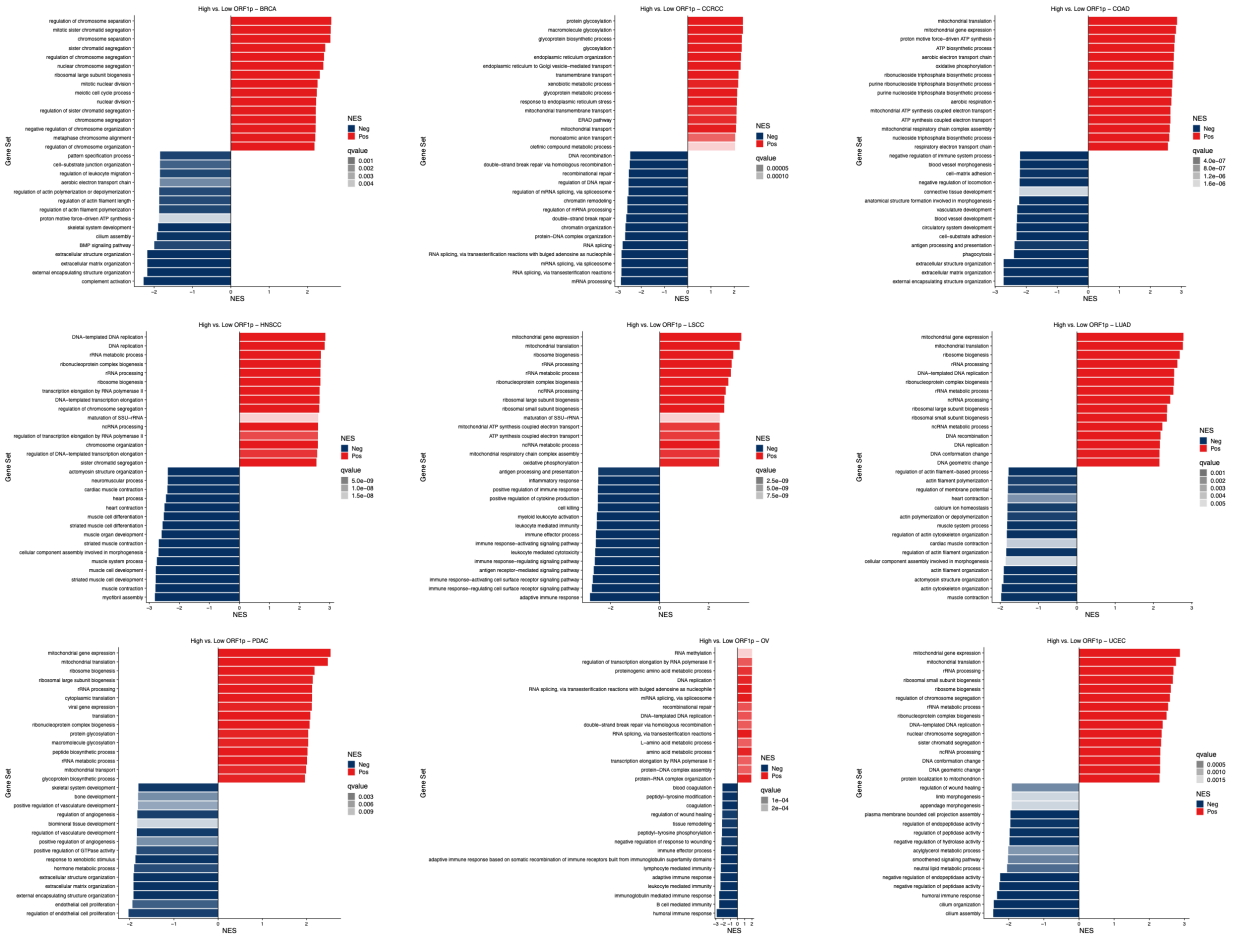
Supplementary Figure 5. ORF1p expression, immune scores, CIN and aneuploidy scores stratified by tumor subtypes.

(A-C, F-G) ORF1p expression, cytotoxic immune score (RNA-based), CGAS-STING score (RNA-based), CIN and aneuploidy score stratified by genomic subtypes per cancer type. (D-E, G) CGAS-STING scores (D-E) and cytotoxic immune score (G) and aneuploidy scores (D-E,G) for HNSCC, LSCC and UCEC, respectively. Pairwise Wilcoxon rank sum tests (BH p-value correction) were performed to test for significant comparisons. $p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.



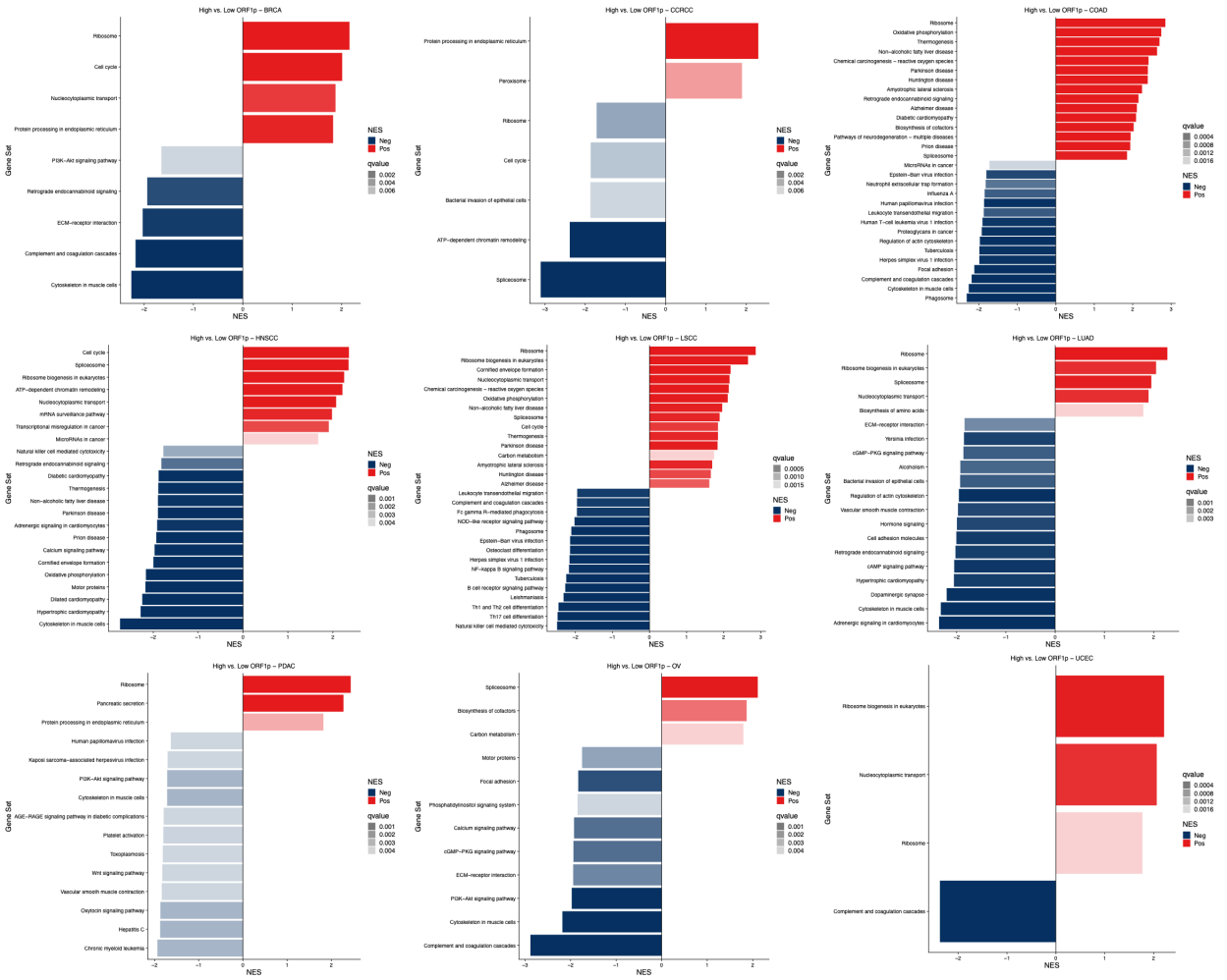
Supplementary Figure 6. Volcano plots from DEqMS differential protein expression.

Representative volcano plots form differential proteomic expression analysis plotting the -Log10 of the nominal p-value versus the Log2 fold change (logFC) of expression. Horizontal dotted line indicates a nominal p-value of < 0.01. Significant logFCs based on the BH-adjusted p-value threshold < 0.05 are colored in green.



Supplementary Figure 7. Gene Ontology Biological Pathway (GOBP) enrichment results.

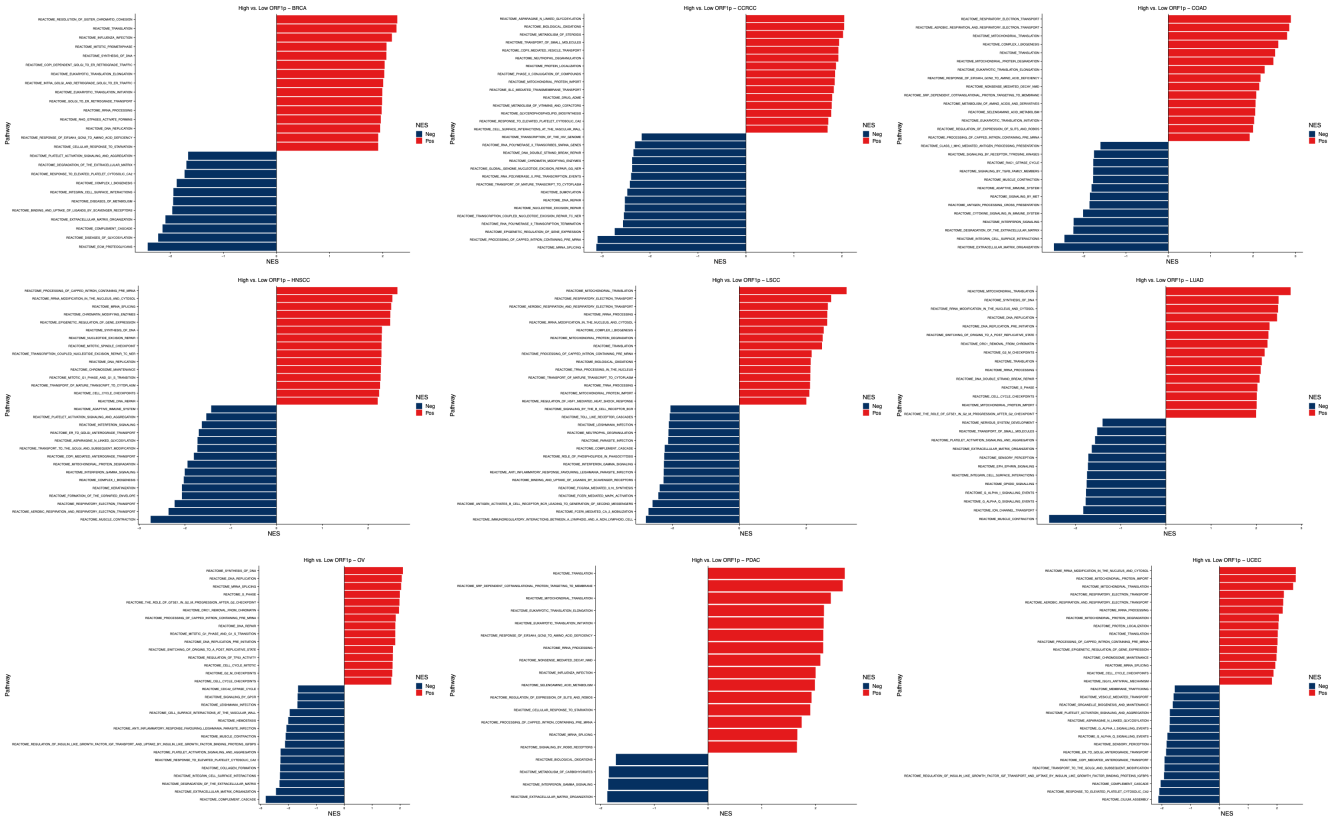
Input is a gene list ranked by decreasing logFC from differential proteomic expression analysis (per cancer type). GOBP GSEA results were filtered to prioritize pathways with strong enrichment signals. Pathways with an adjusted p-value < 0.2 were retained and then arranged by increasing false discovery rate, followed by adjusted p-value, and then by the magnitude of the normalized enrichment score (NES). Gene sets were classified as positively (Pos) or negatively (Neg) enriched based on the sign of the NES. Within each enrichment direction, the top 15 gene sets with the highest absolute NES values were selected. The final results were sorted by descending NES.



Supplementary Figure 8. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment results.

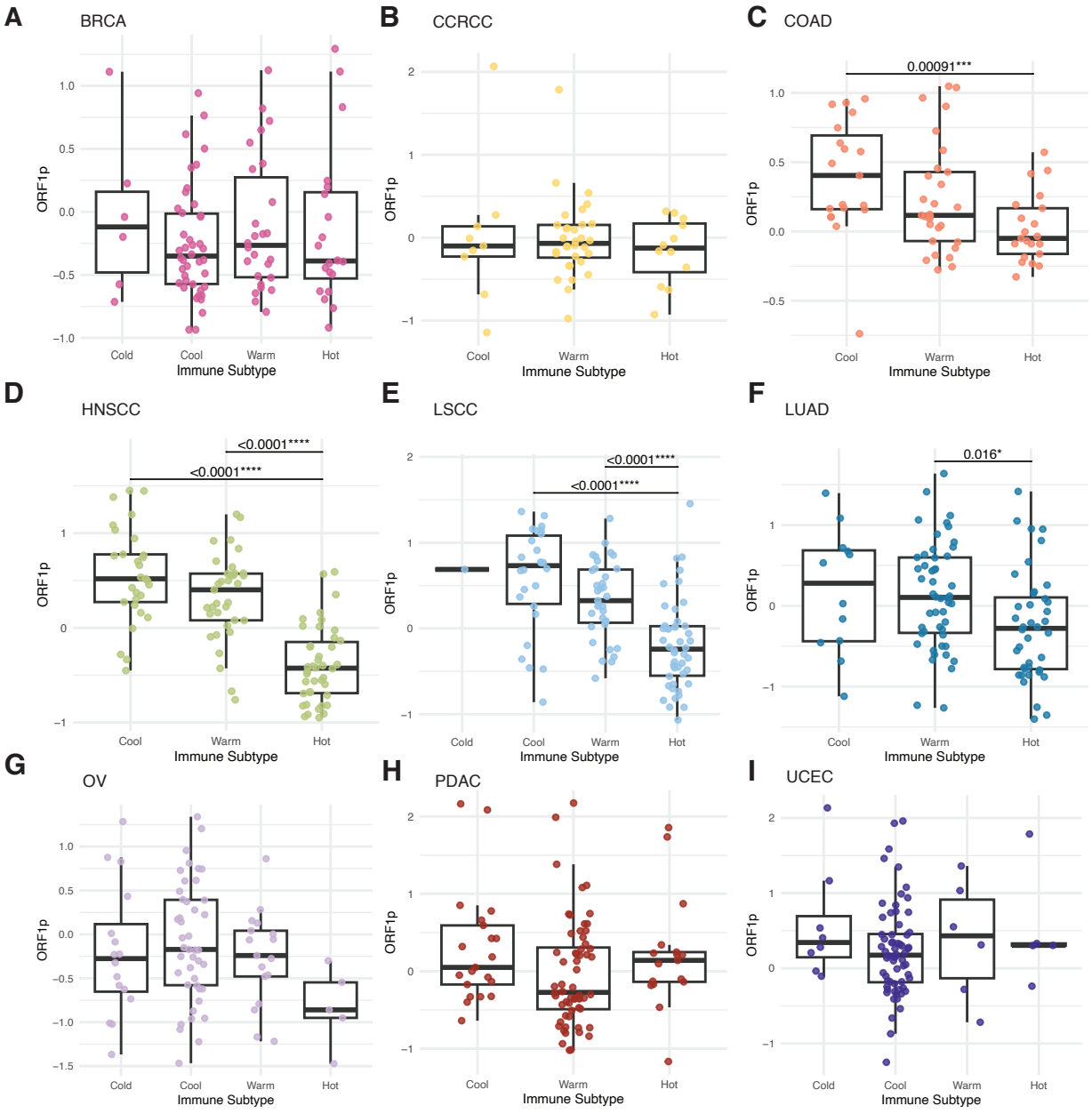
Input is a gene list ranked by decreasing logFC from differential proteomic expression analysis (per cancer type). KEGG GSEA results were filtered to prioritize pathways with strong enrichment signals. Pathways with an adjusted p-value < 0.01were retained and then arranged by increasing false discovery rate, followed by adjusted p-value, and then by the magnitude of the normalized enrichment score (NES). Gene sets were classified as positively (Pos) or negatively (Neg) enriched based on the sign of the NES. Within

each enrichment direction, the top 15 gene sets with the highest absolute NES values were selected. The final results were sorted by descending NES.



Supplementary Figure 9. Reactome enrichment results.

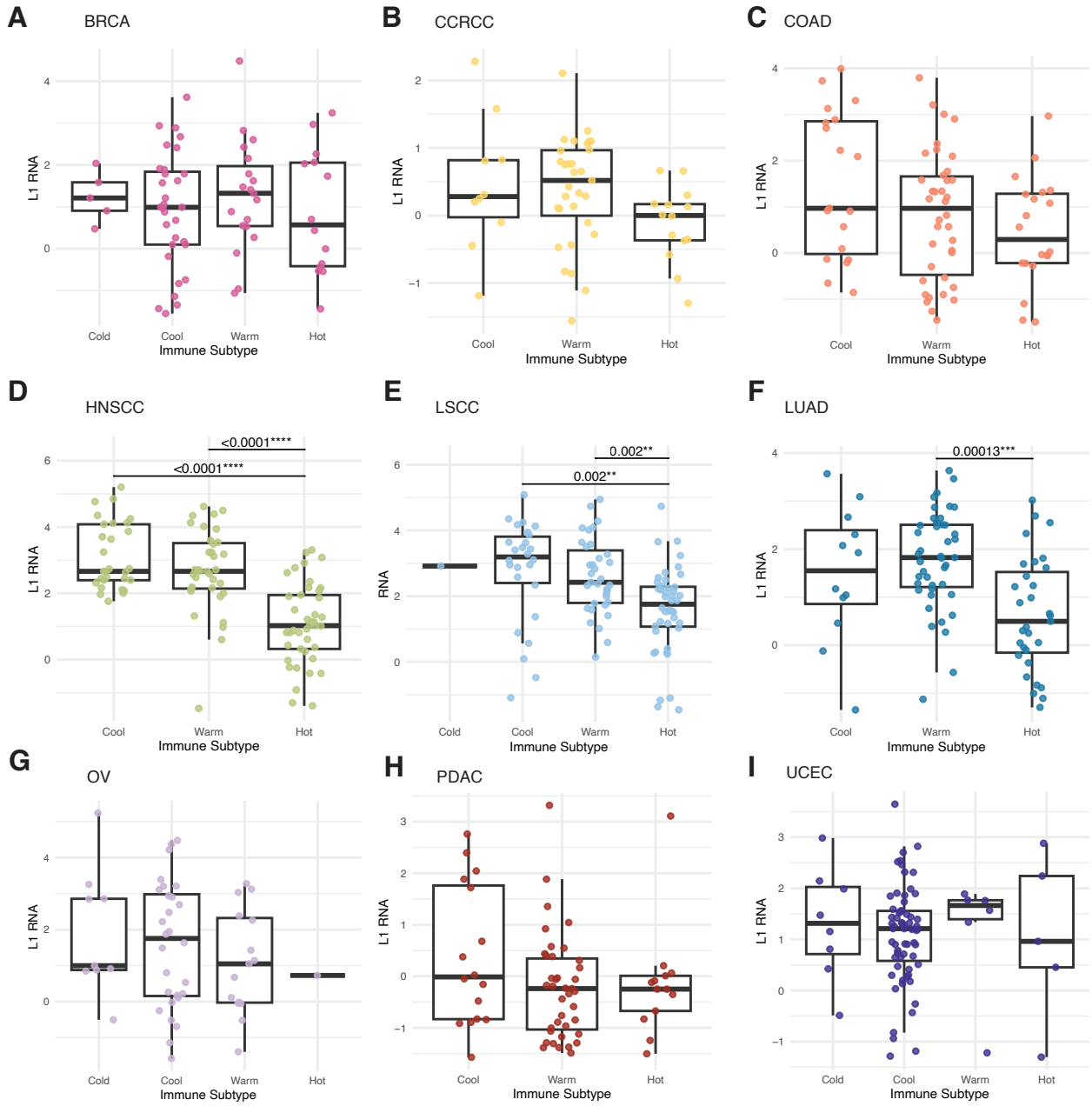
Input is a gene list ranked by decreasing logFC from differential proteomic expression analysis (per cancer type). REACTOME GSEA results were filtered to prioritize pathways with strong enrichment signals. Pathways with an adjusted p-value < 0.01were retained and then arranged by adjusted p-value, followed by the magnitude of the normalized enrichment score (NES). Gene sets were classified as positively (Pos) or negatively (Neg) enriched based on the sign of the NES. Within each enrichment direction, the top 15 gene sets with the highest absolute NES values were selected. The final results were sorted by descending NES.



Supplementary Figure 10. ORF1p by CPTAC pan-cancer immune subtypes per cancer type.

(A-I) Box plots showing ORF1p expression stratified by CPTAC pan-cancer immune subtypes within individual cohorts. Two-sided Wilcoxon rank sum tests (BH-adjusted P

values) are performed to test for significant comparisons. $p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.



Supplementary Figure 11. L1 RNA by CPTAC pan-cancer immune subtypes per cancer type.

(A-I) Box plots showing L1 RNA expression stratified by CPTAC pan-cancer immune subtypes within individual cohorts. Two-sided Wilcoxon rank sum tests (BH-adjusted P

values) are performed to test for significant comparisons. $p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$.

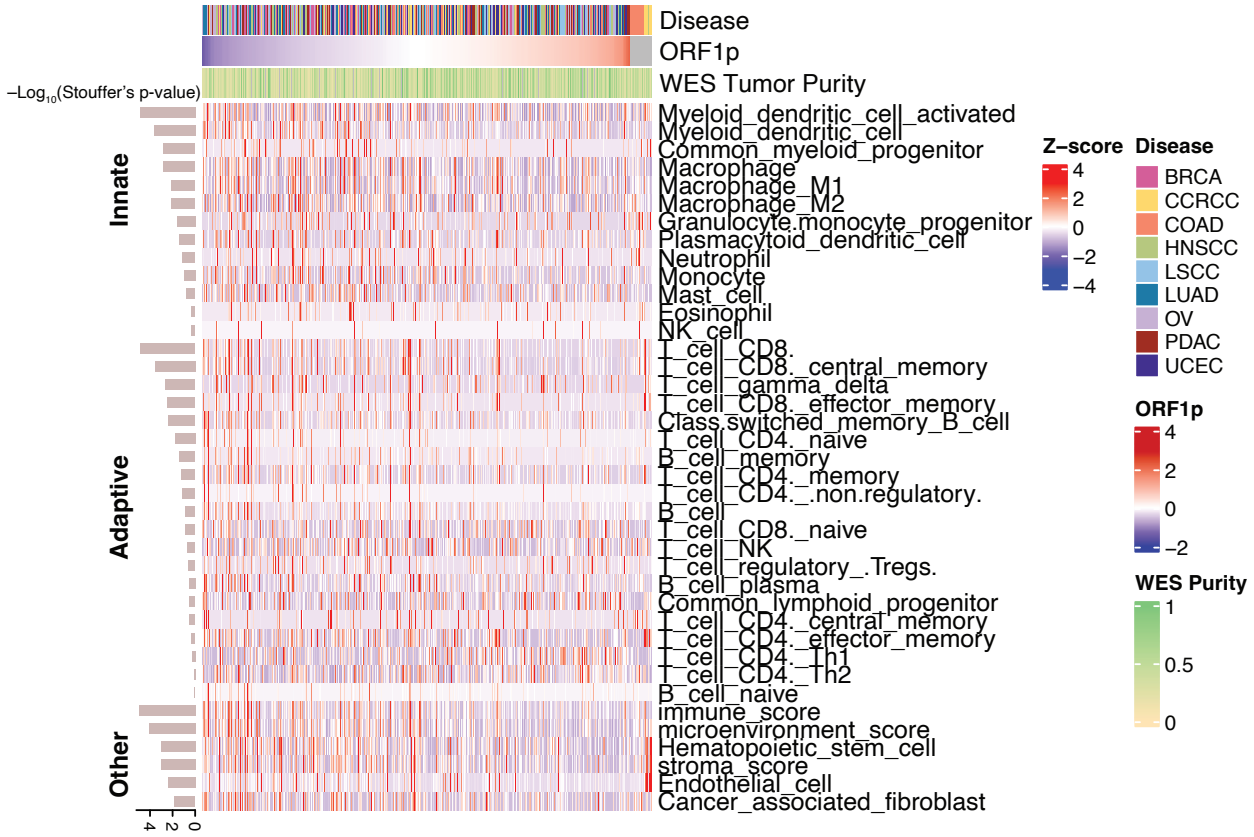
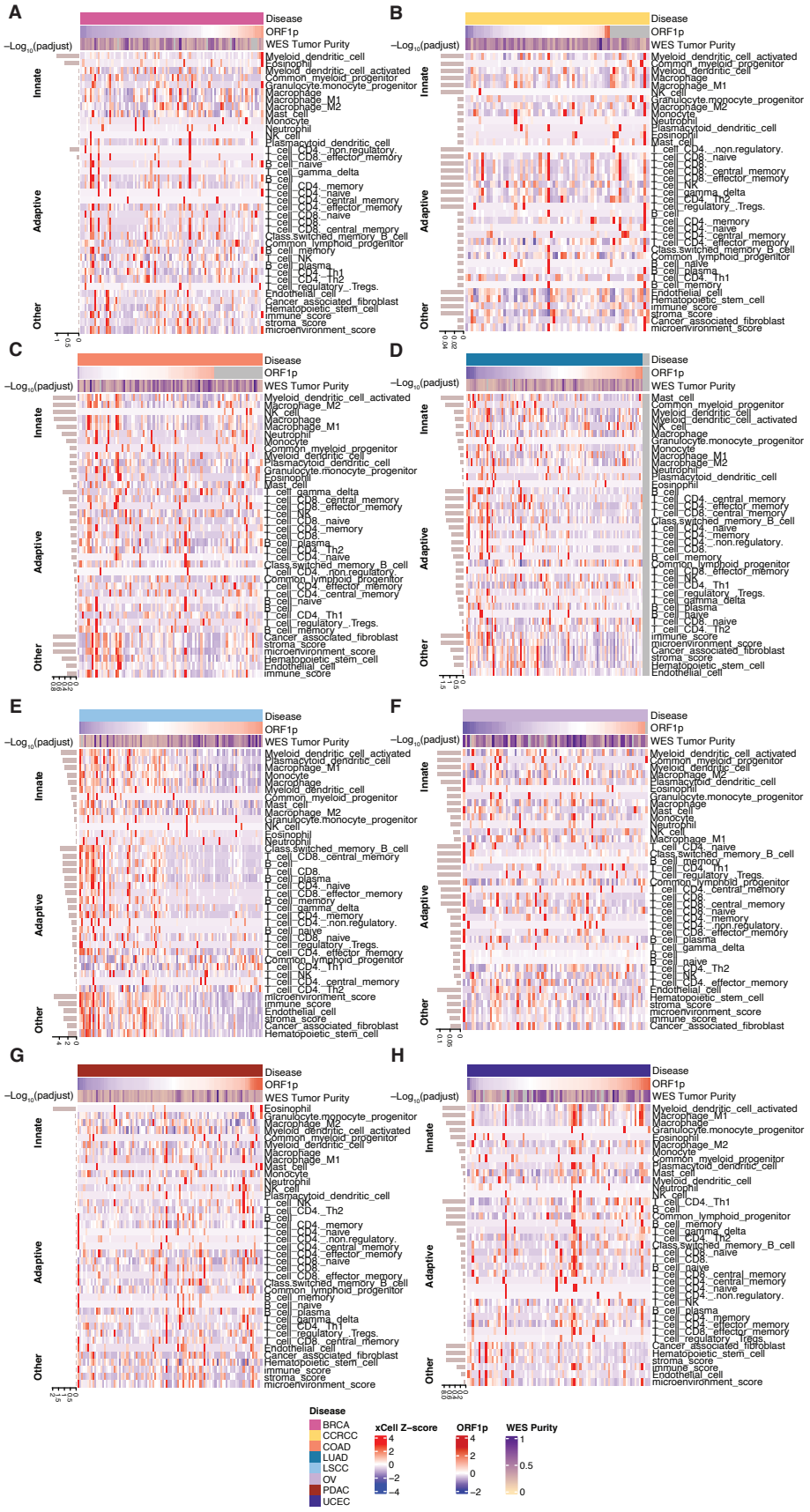


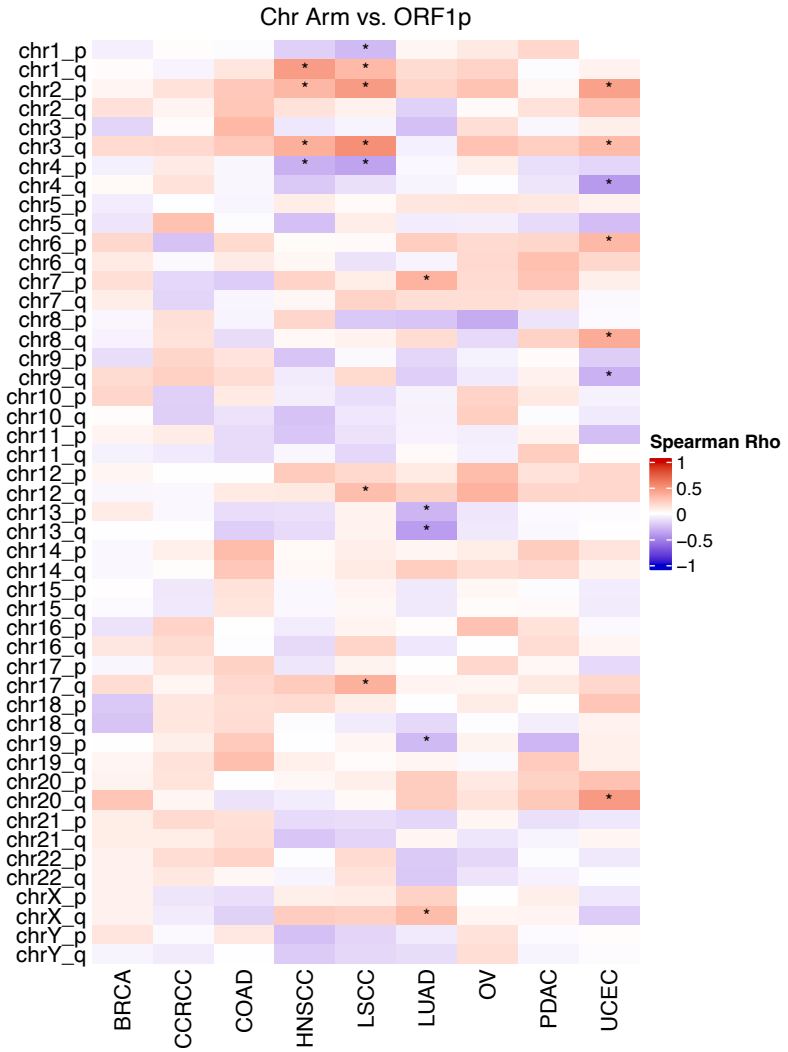
Figure 12. Pan-cancer associations of immune cell types with ORF1p.

To assess pan-cancer associations between L1 ORF1p expression and xCell immune cell type scores, we performed meta-analyses of correlation p-values across cancer types for xCell cell type. Correlation results between ORF1p and xCell scores were compiled for each cancer type from previously computed linear regressions. For each immune cell type, we combined the per-cancer p-values across all remaining cancer types using Stouffer's method, a z-score-based p-value combination approach. Rows are immune cell types grouped by branches of the immune system. Rows within each group are ordered by decreasing significance of the combined Stouffer's P value of the ORF1p term.


A
B
C
D
E
F
G
H
Disease
ORF1p
WES Tumor Purity
−Log10(padjust)
Innate
Adaptive
Other
Disease
BRCA
CCRCC
COAD
LUAD
LSCC
OV
PDAC
UCEC
xCell Z−score
ORF1p
WES Purity

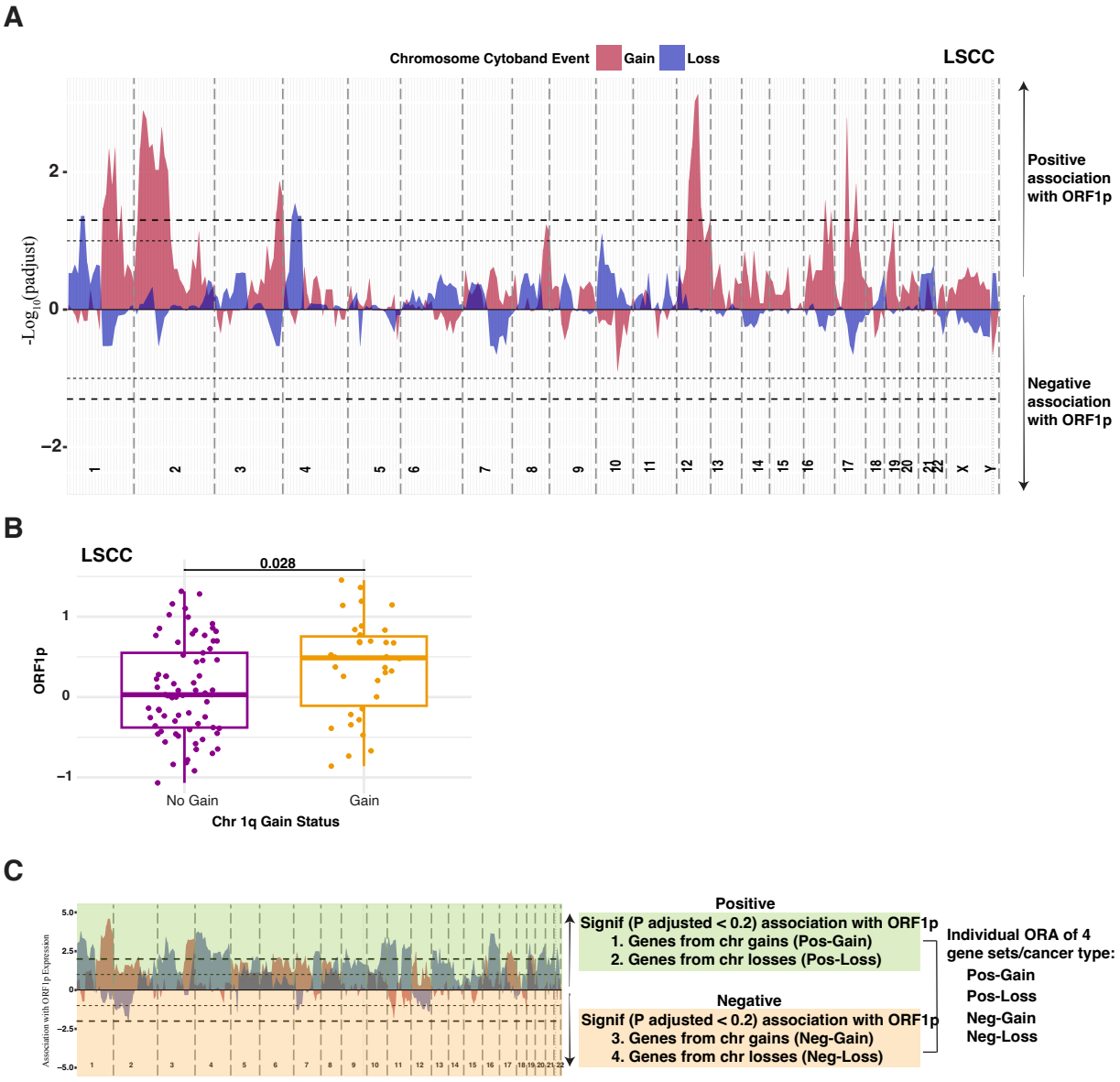
Supplemental Figure 13. Association of ORF1p with xCell immune cell type scores per cancer type.

Heatmap of xCell immune cell type deconvolution scores ordered by increasing ORF1p expression. WES tumor purity covariate also annotated. Rows are immune cell types grouped by branches of the immune system. Rows within each group are ordered by decreasing significance of the P adjusted value of the ORF1p term from the linear regression model.



Supplemental Figure 14. Spearman correlations of chromosome arm-level CN ratios with ORF1p.

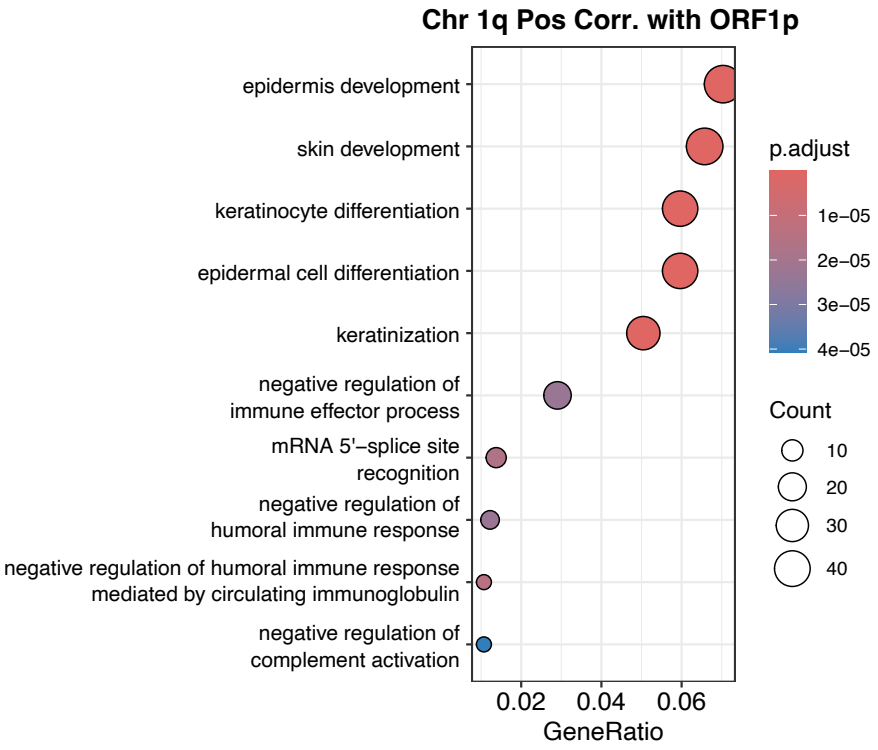
For each chromosomal arm, we computed Spearman's rank correlation between its copy number ratio and ORF1p expression. Spearman correlation coefficient (rho) and multiple testing correction was performed using the BH method to obtain adjusted p-values per cancer type. Significance (adjusted P value < 0.05) is annotated (*).



Supplementary Figure 15. ORF1p predicts aneuploidies in LSCC.

(A) Area plot represents the association between chromosome cytoband gains (red) and losses (blue) and ORF1p expression in LSCC as described in Figure 23 legend. Chr 1q is boxed. (B) ORF1p expression in LSCC is stratified by gain or no gain of 1q. Student's T-test for significance. (C) Schematic outlining ORA workflow for gained (red) or lost

(blue) genes significantly associated with ORF1p expression in the positive (green) or negative (orange) direction. ORA was performed for each gene set per cancer type.



Supplementary Figure 16. ORA dot plot for ORF1p-predicted (adjusted P < 0.05) gained 1q cytobands in LSCC.

Dot plot showing top 10 GOBP pathways from ORA for all genes located on ORF1p-predicted (adjusted P < 0.05) gained 1q cytobands in LSCC.

REFERENCES

1. Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., de Lima, L.G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science *376*, eabk3112. https://doi.org/10.1126/science.abk3112.

2. Boissinot, S., Chevret, P., and Furano, A.V. (2000). L1 (LINE-1) Retrotransposon Evolution and Amplification in Recent Human History. Molecular Biology and Evolution *17*, 915–928. https://doi.org/10.1093/oxfordjournals.molbev.a026372.

3. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H. (1996). High Frequency Retrotransposition in Cultured Mammalian Cells. Cell *87*, 917–927. https://doi.org/10.1016/S0092-8674(00)81998-4.

4. Kubo, S., Seleme, M. del C., Soifer, H.S., Perez, J.L.G., Moran, J.V., Kazazian, H.H., and Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. Proceedings of the National Academy of Sciences *103*, 8036–8041. https://doi.org/10.1073/pnas.0601954103.

5. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian Jr., H.H. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280–5285. https://doi.org/10.1073/pnas.0831042100.

6. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159–1170. https://doi.org/10.1016/j.cell.2010.05.021.

7. Thawani, A., Florez Ariza, A.J., Nogales, E., and Collins, K. (2024). Template and target-site recognition by human LINE-1 in retrotransposition. Nature *626*, 186–193. https://doi.org/10.1038/s41586-023-06933-5.

8. Feng, Q., Moran, J.V., Kazazian, H.H., and Boeke, J.D. (1996). Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. Cell *87*, 905–916. https://doi.org/10.1016/S0092-8674(00)81997-2.

9. Cashen, B.A., Naufer, M.N., Morse, M., McCauley, M.J., Rouzina, I., Jones, C.E., Furano, A.V., and Williams, M.C. (2024). L1-ORF1p nucleoprotein can rapidly assume distinct conformations and simultaneously bind more than one nucleic acid. Nucleic Acids Research *52*, 14013–14029. https://doi.org/10.1093/nar/gkae1141.

10. Baldwin, E.T., van Eeuwen, T., Hoyos, D., Zalevsky, A., Tchesnokov, E.P., Sánchez, R., Miller, B.D., Di Stefano, L.H., Ruiz, F.X., Hancock, M., et al. (2023). Structures, functions and adaptations of the human LINE-1 ORF2 protein. Nature 2023 626:7997 *626*, 194–206. https://doi.org/10.1038/s41586-023-06947-z.

11. Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41–48. https://doi.org/10.1038/ng1223.

12. Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., and Kazazian, H.H. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes & Development *23*, 1303–1312. https://doi.org/10.1101/GAD.1803909.

13. Belancio, V.P., Roy-Engel, A.M., Pochampally, R.R., and Deininger, P. (2010). Somatic expression of LINE-1 elements in human tissues. Nucleic Acids Research *38*, 3909–3922. https://doi.org/10.1093/nar/gkq132.

14. Mendez-Dorantes, C., and Burns, K.H. (2023). LINE-1 retrotransposition and its deregulation in cancers: implications for therapeutic opportunities. Genes & Development *37*, 948–967. https://doi.org/10.1101/GAD.351051.123.

15. Phokaew, C., Kowudtitham, S., Subbalekha, K., Shuangshoti, S., and Mutirangura, A. (2008). LINE-1 methylation patterns of different loci in normal and cancerous cells. Nucleic Acids Res *36*, 5704–5712. https://doi.org/10.1093/nar/gkn571.

16. Lavia, P., Sciamanna, I., and Spadafora, C. (2022). An Epigenetic LINE-1-Based Mechanism in Cancer. Int J Mol Sci *23*. https://doi.org/10.3390/ijms232314610.

17. Ariumi, Y. (2016). Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition. Front. Chem. *4*. https://doi.org/10.3389/fchem.2016.00028.

18. Protasova, M.S., Andreeva, T.V., and Rogaev, E.I. (2021). Factors Regulating the Activity of LINE1 Retrotransposons. Genes 2021, Vol. 12, Page 1562 *12*, 1562. https://doi.org/10.3390/GENES12101562.

19. Luqman-Fatah, A., and Miyoshi, T. (2022). Human LINE-1 retrotransposons: impacts on the genome and regulation by host factors. Genes Genet Syst. https://doi.org/10.1266/ggs.22-00038.

20. Tiwari, B., Jones, A.E., Caillet, C.J., Das, S., Royer, S.K., and Abrams, J.M. (2020). p53 directly represses human LINE1 transposons. Genes Dev *34*, 1439–1451. https://doi.org/10.1101/gad.343186.120.

21. Crow, M.K. (2010). Long interspersed nuclear elements (LINE-1): Potential triggers of systemic autoimmune disease. Autoimmunity *43*, 7–16. https://doi.org/10.3109/08916930903374865.

22. Zhao, Y., Oreskovic, E., Zhang, Q., Lu, Q., Gilman, A., Lin, Y.S., He, J., Zheng, Z., Lu, J.Y., Lee, J., et al. (2021). Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. Nature Immunology 2021 22:10 *22*, 1219–1230. https://doi.org/10.1038/s41590-021-01027-8.

23. Zhao, K., Du, J., Peng, Y., Li, P., Wang, S., Wang, Y., Hou, J., Kang, J., Zheng, W., Hua, S., et al. (2018). LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. Journal of Autoimmunity *90*, 105–115. https://doi.org/10.1016/j.jaut.2018.02.007.

24. Li, X., and Liu, N. (2025). Advances in understanding LINE-1 regulation and function in the human genome. Trends in Genetics. https://doi.org/10.1016/j.tig.2025.04.011.

25. Ishak, C.A., Marhon, S.A., Tchrakian, N., Hodgson, A., Loo Yau, H., Gonzaga, I.M., Peralta, M., Lungu, I.M., Gomez, S., Liang, S.-B., et al. (2025). Chronic Viral Mimicry Induction following p53 Loss Promotes Immune Evasion. Cancer Discovery, OF1–OF25. https://doi.org/10.1158/2159-8290.CD-24-0094.

26. Fiori, M.E., Di Franco, S., Villanova, L., Bianca, P., Stassi, G., and De Maria, R. (2019). Cancer-associated fibroblasts as abettors of tumor progression at the crossroads of EMT and therapy resistance. Mol Cancer *18*, 1–16. https://doi.org/10.1186/s12943-019-0994-2.

27. Solovyeva, A.I., Afanasev, R.V., Popova, M.A., and Enukashvily, N.I. (2024). Dysregulation of Transposon Transcription Profiles in Cancer Cells Resembles That of Embryonic Stem Cells. Current Issues in Molecular Biology *46*, 8576–8599. https://doi.org/10.3390/cimb46080505.

28. Mita, P., Wudzinska, A., Sun, X., Andrade, J., Nayak, S., Kahler, D.J., Badri, S., LaCava, J., Ueberheide, B., Yun, C.Y., et al. (2018). LINE-1 protein localization and functional dynamics during the cell cycle. Elife *7*. https://doi.org/10.7554/eLife.30058.

29. Sun, S., You, E., Hong, J., Hoyos, D., Priore, I.S.D., Tsanov, K.M., Mattagajasingh, O., Gioacchino, A.D., Marhon, S.A., Chacon-Barahona, J., et al. (2024). Cancer cells restrict immunogenicity of retrotransposon expression via distinct mechanisms. Immunity *0*. https://doi.org/10.1016/j.immuni.2024.10.015.

30. Simon, M., Van Meter, M., Ablaeva, J., Ke, Z., Gonzalez, R.S., Taguchi, T., De Cecco, M., Leonova, K.I., Kogan, V., Helfand, S.L., et al. (2019). LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. Cell Metabolism *29*, 871-885.e5. https://doi.org/10.1016/j.cmet.2019.02.014.

31. Mita, P., Sun, X., Fenyö, D., Kahler, D.J., Li, D., Agmon, N., Wudzinska, A., Keegan, S., Bader, J.S., Yun, C., et al. (2020). BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. Nat Struct Mol Biol *27*, 179–191. https://doi.org/10.1038/s41594-020-0374-z.

32. Kong, Y., Rose, C.M., Cass, A.A., Williams, A.G., Darwish, M., Lianoglou, S., Haverty, P.M., Tong, A.J., Blanchette, C., Albert, M.L., et al. (2019). Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nature Communications 2019 10:1 *10*, 1–14. https://doi.org/10.1038/s41467-019-13035-2.

33. Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D., et al. (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet *52*, 306–319. https://doi.org/10.1038/s41588-019-0562-0.

34. Solovyov, A., Behr, J.M., Hoyos, D., Banks, E., Drong, A.W., Thornlow, B., Zhong, J.Z., Garcia-Rivera, E., McKerrow, W., Chu, C., et al. (2025). Pan-cancer multi-omic model of LINE-1 activity reveals locus heterogeneity of retrotransposition efficiency. Nat Commun *16*, 2049. https://doi.org/10.1038/s41467-025-57271-1.

35. Jung, H., Choi, J.K., and Lee, E.A. (2018). Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. Genome Research *28*, 1136–1146. https://doi.org/10.1101/GR.231837.117/-/DC1.

36. Karkas, R., Abdullah, K.S.A., Kaizer, L., Ürmös, Á., Raya, M., Tiszlavicz, L., Pankotai, T., Nagy, I., Mátés, L., and Sükösd, F. (2024). LINE-1 ORF1p is a Promising Biomarker in Cervical Intraepithelial Neoplasia Degree Assessment. International Journal of Gynecological Pathology. https://doi.org/10.1097/PGP.0000000000001035.

37. Pisanic, T.R., Asaka, S., Lin, S.-F., Yen, T.-T., Sun, H., Bahadirli-Talbott, A., Wang, T.-H., Burns, K.H., Wang, T.-L., and Shih, I.-M. (2019). Long Interspersed Nuclear Element 1 Retrotransposons Become Deregulated during the Development of Ovarian Cancer Precursor Lesions. The American Journal of Pathology *189*, 513–520. https://doi.org/10.1016/j.ajpath.2018.11.005.

38. Ardeljan, D., Wang, X., Oghbaie, M., Taylor, M.S., Husband, D., Deshpande, V., Steranka, J.P., Gorbounov, M., Yang, W.R., Sie, B., et al. (2019). LINE-1 ORF2p expression is nearly imperceptible in human cancers. Mobile DNA *11*, 1. https://doi.org/10.1186/s13100-019-0191-2.

39. Ardeljan, D., Taylor, M.S., Ting, D.T., and Burns, K.H. (2017). The Human Long Interspersed Element-1 Retrotransposon: An Emerging Biomarker of Neoplasia. Clin Chem *63*, 816–822. https://doi.org/10.1373/clinchem.2016.257444.

40. Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M.S., Hruban, R.H., Iacobuzio-Donahue, C.A., Maitra, A., Torbenson, M.S., et al. (2014). Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. The American Journal of Pathology *184*, 1280–1286. https://doi.org/10.1016/J.AJPATH.2014.01.007.

41. Sato, S., Gillette, M., de Santiago, P.R., Kuhn, E., Burgess, M., Doucette, K., Feng, Y., Mendez-Dorantes, C., Ippoliti, P.J., Hobday, S., et al. (2023). LINE-1 ORF1p as a candidate biomarker in high grade serous ovarian carcinoma. Scientific Reports 2023 13:1 *13*, 1–15. https://doi.org/10.1038/s41598-023-28840-5.

42. You, E., Patel, B.K., Rojas, A.S., Sun, S., Danaher, P., Ho, N.I., Phillips, I.E., Raabe, M.J., Song, Y., Xu, K.H., et al. (2025). LINE-1 ORF1p Mimics Viral Innate Immune Evasion Mechanisms in Pancreatic Ductal Adenocarcinoma. Cancer Discovery, OF1–OF20. https://doi.org/10.1158/2159-8290.CD-24-1317.

43. Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A.P., Aguet, F., Akiyama, Y., Anand, S., Birger, C., Cao, S., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. Cancer Cell *41*, 1397–1406. https://doi.org/10.1016/J.CCELL.2023.06.009.

44. McKerrow, W., Wang, X., Mendez-Dorantes, C., Mita, P., Cao, S., Grivainis, M., Ding, L., LaCava, J., Burns, K.H., Boeke, J.D., et al. (2022). LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. Proc Natl Acad Sci U S A *119*. https://doi.org/10.1073/pnas.2115999119.

45. Xiang, R., Ma, L., Yang, M., Zheng, Z., Chen, X., Jia, F., Xie, F., Zhou, Y., Li, F., Wu, K., et al. (2021). Increased expression of peptides from non-coding genes in cancer proteomics datasets suggests potential tumor neoantigens. Commun Biol *4*, 1–12. https://doi.org/10.1038/s42003-021-02007-2.

46. Achanta, P., Steranka, J.P., Tang, Z., Rodic, N., Sharma, R., Yang, W.R., Ma, S., Grivainis, M., Huang, C.R.L., Schneider, A.M., et al. (2016). Somatic retrotransposition is infrequent in glioblastomas. Mob DNA *7*, 22. https://doi.org/10.1186/s13100-016-0077-5.

47. Carreira, P.E., Ewing, A.D., Li, G., Schauer, S.N., Upton, K.R., Fagg, A.C., Morell, S., Kindlova, M., Gerdes, P., Richardson, S.R., et al. (2016). Evidence for L1-associated DNA rearrangements and negligible L1 retrotransposition in glioblastoma multiforme. Mobile DNA *7*, 21. https://doi.org/10.1186/s13100-016-0076-6.

48. Sun, Z., Zhang, R., Zhang, X., Sun, Y., Liu, P., Francoeur, N., Han, L., Lam, W.Y., Yi, Z., Sebra, R., et al. (2022). LINE-1 promotes tumorigenicity and exacerbates tumor progression via stimulating metabolism reprogramming in non-small cell lung cancer. Molecular Cancer *21*, 147. https://doi.org/10.1186/s12943-022-01618-5.

49. Benard, A., van de Velde, C.J.H., Lessard, L., Putter, H., Takeshima, L., Kuppen, P.J.K., and Hoon, D.S.B. (2013). Epigenetic status of LINE-1 predicts clinical outcome in early-stage rectal cancer. Br J Cancer *109*, 3073–3083. https://doi.org/10.1038/bjc.2013.654.

50. Sha, D., Jin, Z., Budczies, J., Kluck, K., Stenzinger, A., and Sinicrope, F.A. (2020). Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. Cancer Discovery *10*, 1808–1825. https://doi.org/10.1158/2159-8290.CD-20-0522.

51. Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al. (2020). Proteogenomic

Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell *183*, 1436-1456.e31. https://doi.org/10.1016/j.cell.2020.10.036.

52. Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell *177*, 1035-1049.e19. https://doi.org/10.1016/j.cell.2019.03.030.

53. McDermott, J.E., Arshad, O.A., Petyuk, V.A., Fu, Y., Gritsenko, M.A., Clauss, T.R., Moore, R.J., Schepmoes, A.A., Zhao, R., Monroe, M.E., et al. (2020). Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. Cell Reports Medicine *1*, 100004. https://doi.org/10.1016/j.xcrm.2020.100004.

54. Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic Characterization of Endometrial Carcinoma. Cell *180*, 729-748 e26. https://doi.org/10.1016/j.cell.2020.01.026.

55. Huang, C., Chen, L., Savage, S.R., Eguez, R.V., Dou, Y., Li, Y., da Veiga Leprevost, F., Jaehnig, E.J., Lei, J.T., Wen, B., et al. (2021). Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell *39*, 361-379.e16. https://doi.org/10.1016/j.ccell.2020.12.007.

56. Li, Y., Lih, T.-S.M., Dhanasekaran, S.M., Mannan, R., Chen, L., Cieslik, M., Wu, Y., Lu, R.J.-H., Clark, D.J., Kołodziejczak, I., et al. (2023). Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness. Cancer Cell *41*, 139-163.e17. https://doi.org/10.1016/j.ccell.2022.12.001.

57. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.-W., Reva, B., et al. (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200-225.e35. https://doi.org/10.1016/j.cell.2020.06.013.

58. Satpathy, S., Krug, K., Jean Beltran, P.M., Savage, S.R., Petralia, F., Kumar-Sinha, C., Dou, Y., Reva, B., Kane, M.H., Avanessian, S.C., et al. (2021). A proteogenomic portrait of lung squamous cell carcinoma. Cell *184*, 4348-4371.e40. https://doi.org/10.1016/J.CELL.2021.07.016.

59. Cao, L., Huang, C., Zhou, D.C., Hu, Y., Lih, T.M., Savage, S.R., Krug, K., Clark, D.J., Schnaubelt, M., Chen, L., et al. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell *184*, 5031-5052.e26. https://doi.org/10.1016/j.cell.2021.08.023.

60. Campbell, J.D., Yau, C., Bowlby, R., Liu, Y., Brennan, K., Fan, H., Taylor, A.M., Wang, C., Walter, V., Akbani, R., et al. (2018). Genomic, Pathway Network, and

Immunologic Features Distinguishing Squamous Carcinomas. Cell Reports *23*, 194-212.e6. https://doi.org/10.1016/J.CELREP.2018.03.063.

61. Dou, Y., Katsnelson, L., Gritsenko, M.A., Hu, Y., Reva, B., Hong, R., Wang, Y.T., Kolodziejczak, I., Lu, R.J.H., Tsai, C.F., et al. (2023). Proteogenomic insights suggest druggable pathways in endometrial carcinoma. Cancer Cell *41*, 1586-1605.e15. https://doi.org/10.1016/J.CCELL.2023.07.007.

62. Mosaddeghi, P., Farahmandnejad, M., and Zarshenas, M.M. (2023). The role of transposable elements in aging and cancer. Biogerontology *24*, 479–491. https://doi.org/10.1007/S10522-023-10028-Z/FIGURES/3.

63. Gázquez-Gutiérrez, A., Witteveldt, J., Heras, S.R., and Macias, S. (2021). Sensing of transposable elements by the antiviral innate immune system. RNA *27*, 735–752. https://doi.org/10.1261/rna.078721.121.

64. Zhen, Z., Chen, Y., Wang, H., Tang, H., Zhang, H., Liu, H., Jiang, Y., and Mao, Z. (2023). Nuclear cGAS restricts L1 retrotransposition by promoting TRIM41-mediated ORF2p ubiquitination and degradation. Nat Commun *14*, 8217. https://doi.org/10.1038/s41467-023-43001-y.

65. Mathavarajah, S., and Dellaire, G. (2024). LINE-1: an emerging initiator of cGAS-STING signalling and inflammation that is dysregulated in disease. Biochem. Cell Biol. *102*, 38–46. https://doi.org/10.1139/bcb-2023-0134.

66. Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R.L., Li, D., Adney, E.M., Jiang, H., Burns, K.H., Chait, B.T., et al. (2013). Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. Cell *155*, 1034–1048. https://doi.org/10.1016/J.CELL.2013.10.021.

67. Zernia, S., Ettefa, F., Sil, S., Koeman, C., Deplazes-Lauber, J., Freitag, M., Holt, L.J., and Stigler, J. (2025). LINE-1 ribonucleoprotein condensates bind DNA to enable nuclear entry during mitosis. Science Advances *11*, eadt9318. https://doi.org/10.1126/sciadv.adt9318.

68. Geng, X., Geng, Z., Li, H., Zhang, Y., Li, J., and Chang, H. (2020). Over-expression of TFB2M facilitates cell growth and metastasis via activating ROS-Akt-NF-κB signalling in hepatocellular carcinoma. Liver International *40*, 1756–1769. https://doi.org/10.1111/liv.14440.

69. Yu, K., Tian, H., and Deng, H. (2020). PPM1G restricts innate immune signaling mediated by STING and MAVS and is hijacked by KSHV for immune evasion. Science Advances *6*, eabd0276. https://doi.org/10.1126/sciadv.abd0276.

70. Ebaid, A.M., Zakaria, M.A., Mekkawy, E.M., Nageeb, R.S., Elfwakhry, R.M., Seleem, D.A., Shabana, M.A., and Esawy, M.M. (2024). VPS13C and STING expression in neuropsychiatric systemic lupus erythematosus: unveiling an unbreached territory. Lupus Sci Med *11*, e001271. https://doi.org/10.1136/lupus-2024-001271.

71. Zhang, L., Wei, X., Wang, Z., Liu, P., Hou, Y., Xu, Y., Su, H., Koci, M.D., Yin, H., and Zhang, C. (2023). NF-κB activation enhances STING signaling by altering microtubule-mediated STING trafficking. Cell Reports *42*, 112185. https://doi.org/10.1016/j.celrep.2023.112185.

72. Abe, T., and Barber, G.N. (2014). Cytosolic-DNA-Mediated, STING-Dependent Proinflammatory Gene Induction Necessitates Canonical NF-κB Activation through TBK1. Journal of Virology *88*, 5328–5341. https://doi.org/10.1128/jvi.00037-14.

73. Karin, M. (1999). How NF-κB is activated: the role of the IκB kinase (IKK) complex. Oncogene *18*, 6867–6874. https://doi.org/10.1038/sj.onc.1203219.

74. Liu, T., Zhang, L., Joo, D., and Sun, S.-C. (2017). NF-κB signaling in inflammation. Sig Transduct Target Ther *2*, 17023. https://doi.org/10.1038/sigtrans.2017.23.

75. Chen, R., Ishak, C.A., and De Carvalho, D.D. (2021). Endogenous Retroelements and the Viral Mimicry Response in Cancer Therapy and Cellular Homeostasis. Cancer Discovery *11*, 2707–2725. https://doi.org/10.1158/2159-8290.CD-21-0506.

76. Cahn, A.R., Bhardwaj, N., and Vabret, N. (2022). Dark genome, bright ideas: Recent approaches to harness transposable elements in immunotherapies. Cancer Cell *40*, 792–797. https://doi.org/10.1016/j.ccell.2022.07.003.

77. Li, Y., Porta-Pardo, E., Tokheim, C., Bailey, M.H., Yaron, T.M., Stathias, V., Geffen, Y., Imbach, K.J., Cao, S., Anand, S., et al. (2023). Pan-cancer proteogenomics connects oncogenic drivers to functional states. Cell *186*, 3921-3944.e25. https://doi.org/10.1016/J.CELL.2023.07.014.

78. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: Digitally portraying the tissue cellular heterogeneity landscape. Genome Biology *18*, 1–14. https://doi.org/10.1186/S13059-017-1349-1/FIGURES/4.

79. Ye, Y., Gaugler, B., Mohty, M., and Malard, F. (2020). Plasmacytoid dendritic cell biology and its role in immune-mediated diseases. Clinical & Translational Immunology *9*, e1139. https://doi.org/10.1002/cti2.1139.

80. Künzli, M., and Masopust, D. (2023). CD4+ T cell memory. Nat Immunol *24*, 903–914. https://doi.org/10.1038/s41590-023-01510-4.

81. Raphael, I., Joern, R.R., and Forsthuber, T.G. (2020). Memory CD4+ T Cells in Immunity and Autoimmune Diseases. Cells *9*, 531. https://doi.org/10.3390/cells9030531.

82. Wu, Y., and Zhou, B.P. (2010). TNF-α/NF-κB/Snail pathway in cancer cell migration and invasion. Br J Cancer *102*, 639–644. https://doi.org/10.1038/sj.bjc.6605530.

83. Liu, J., Zhang, X., Cheng, Y., and Cao, X. (2021). Dendritic cell migration in inflammation and immunity. Cell Mol Immunol *18*, 2461–2471. https://doi.org/10.1038/s41423-021-00726-4.

84. Knouse, K.A., Davoli, T., Elledge, S.J., and Amon, A. (2017). Aneuploidy in Cancer: Seq-ing Answers to Old Questions. Annual Review of Cancer Biology *1*, 335–354. https://doi.org/10.1146/annurev-cancerbio-042616-072231.

85. Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*. https://doi.org/10.1126/SCIENCE.AAF8399/SUPPL_FILE/AAF8399-DAVOLI-SM.PDF.

86. Han, G., Yang, G., Hao, D., Lu, Y., Thein, K., Simpson, B.S., Chen, J., Sun, R., Alhalabi, O., Wang, R., et al. (2021). 9p21 loss confers a cold tumor immune microenvironment and primary resistance to immune checkpoint therapy. Nat Commun *12*, 5606. https://doi.org/10.1038/s41467-021-25894-9.

87. Avigan, Z.M., Mitsiades, C.S., and Laganà, A. (2025). The role of 1q abnormalities in multiple myeloma: Genomic insights, clinical implications, and therapeutic challenges. Seminars in Hematology *62*, 20–30. https://doi.org/10.1053/j.seminhematol.2024.10.001.

88. Ting Hsiung, D., Marsit, C.J., Houseman, E.A., Eddy, K., Furniss, C.S., McClean, M.D., and Kelsey, K.T. (2007). Global DNA Methylation Level in Whole Blood as a Biomarker in Head and Neck Squamous Cell Carcinoma. Cancer Epidemiology, Biomarkers & Prevention *16*, 108–114. https://doi.org/10.1158/1055-9965.EPI-06-0636.

89. Qiao, H., Yin, H., Feng, Y., and Tang, H. (2022). Pan-cancer analysis reveals the relationship between RCSD1 immune infiltration and clinical prognosis in human tumors. Front. Immunol. *13*. https://doi.org/10.3389/fimmu.2022.1008778.

90. Dexiu, C., Xianying, L., Yingchun, H., and Jiafu, L. (2022). Advances in CD247. Scandinavian Journal of Immunology *96*, e13170. https://doi.org/10.1111/sji.13170.

91. Al Barashdi, M.A., Ali, A., McMullin, M.F., and Mills, K. (2021). Protein tyrosine phosphatase receptor type C (PTPRC or CD45). J Clin Pathol *74*, 548–552. https://doi.org/10.1136/jclinpath-2020-206927.

92. MangiavacchiArianna, MorelliGabriele, ReppeSjur, Saera-VilaAlfonso, LiuPeng, EggerschwilerBenjamin, ZhangHuoming, BensaddekDalila, A, C., GomezCarolina, M., et al. (2024). LINE-1 RNA triggers matrix formation in bone cells via a PKR-mediated inflammatory response. The EMBO Journal. https://doi.org/10.1038/s44318-024-00143-z.

93. Peng, Y., Guo, J., Sun, T., Fu, Y., Zheng, H., Dong, C., and Xiong, S. (2020). USP39 Serves as a Deubiquitinase to Stabilize STAT1 and Sustains Type I IFN–Induced Antiviral Immunity. J Immunol *205*, 3167–3178. https://doi.org/10.4049/jimmunol.1901384.

94. Song, Q., Yang, Y., Jiang, D., Qin, Z., Xu, C., Wang, H., Huang, J., Chen, L., Luo, R., Zhang, X., et al. (2022). Proteomic analysis reveals key differences between squamous cell carcinomas and adenocarcinomas across multiple tissues. Nat Commun *13*, 4167. https://doi.org/10.1038/s41467-022-31719-0.

95. Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic Acids Research *41*, 7401–7419. https://doi.org/10.1093/nar/gkt512.

96. Saeed, A.F.U.H., Ruan, X., Guan, H., Su, J., and Ouyang, S. (2020). Regulation of cGAS-Mediated Immune Responses and Immunotherapy. Advanced Science *7*, 1902599. https://doi.org/10.1002/advs.201902599.

97. Yang, R., Yu, S., Xu, T., Zhang, J., and Wu, S. (2022). Emerging role of RNA sensors in tumor microenvironment and immunotherapy. Journal of Hematology & Oncology *15*, 43. https://doi.org/10.1186/s13045-022-01261-z.

98. Solovyov, A., Vabret, N., Arora, K.S., Snyder, A., Funt, S.A., Bajorin, D.F., Rosenberg, J.E., Bhardwaj, N., Ting, D.T., and Greenbaum, B.D. (2018). Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. Cell Reports *23*, 512–521. https://doi.org/10.1016/J.CELREP.2018.03.042.

99. Taylor, M.S., Wu, C., Fridy, P.C., Zhang, S.J., Senussi, Y., Wolters, J.C., Cajuso, T., Cheng, W.-C., Heaps, J.D., Miller, B.D., et al. (2023). Ultrasensitive detection of circulating LINE-1 ORF1p as a specific multi-cancer biomarker. Cancer Discovery. https://doi.org/10.1158/2159-8290.CD-23-0313.

100. Vylegzhanina, A.V., Bespalov, I.A., Novototskaya-Vlasova, K.A., Hall, B.M., Gleiberman, A.S., Yu, H., Leontieva, O.V., Leonova, K.I., Kurnasov, O.V., Osterman, A.L., et al. (2023). Cancer Relevance of Circulating Antibodies Against LINE-1 Antigens in Humans. Cancer Research Communications *3*, 2256–2267. https://doi.org/10.1158/2767-9764.CRC-23-0289.

101. de Santiago, P.R., Sato, S., Zhang, S.J., Dougher, M.C., Devins, K.M., Bilecz, A.J., Rayamajhi, S., Mingo, G., Rendulich, H.S., Feng, Y., et al. (2025). LINE-1 ORF1p expression occurs in clear cell ovarian carcinoma precursors and is a candidate blood biomarker. npj Precis. Onc. *9*, 1–12. https://doi.org/10.1038/s41698-025-00849-1.

102. Wang, L.-B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and

metabolomic characterization of human glioblastoma. Cancer Cell *39*, 509-528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.

103. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

104. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218. https://doi.org/10.1038/nature12213.

105. McKerrow, W., and Fenyö, D. (2020). L1EM: a tool for accurate locus specific LINE-1 RNA quantification. Bioinformatics *36*, 1167–1173. https://doi.org/10.1093/bioinformatics/btz724.

106. Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics *20*, 1466–1467. https://doi.org/10.1093/bioinformatics/bth092.

107. Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., Genomes Project, C., and Devine, S.E. (2017). The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res *27*, 1916–1929. https://doi.org/10.1101/gr.218032.116.

108. Zhu, Y., Orre, L.M., Zhou Tran, Y., Mermelekas, G., Johansson, H.J., Malyutina, A., Anders, S., and Lehtiö, J. (2020). DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis*. Molecular & Cellular Proteomics *19*, 1047–1057. https://doi.org/10.1074/mcp.TIR119.001646.

109. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation *2*, 100141. https://doi.org/10.1016/J.XINN.2021.100141.

110. Damotte, D., Warren, S., Arrondeau, J., Boudou-Rouquette, P., Mansuet-Lupo, A., Biton, J., Ouakrim, H., Alifano, M., Gervais, C., Bellesoeur, A., et al. (2019). The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. Journal of Translational Medicine *17*, 1–10. https://doi.org/10.1186/S12967-019-2100-3/FIGURES/3.

111. Bhuva, D.D., Cursons, J., and Davis, M.J. (2020). Stable gene expression for normalisation and single-sample scoring. Nucleic Acids Research *48*, e113. https://doi.org/10.1093/nar/gkaa802.

112. Onomoto, K., Onoguchi, K., and Yoneyama, M. (2021). Regulation of RIG-I-like receptor-mediated signaling: interaction between host and viral factors. Cell Mol Immunol *18*, 539–555. https://doi.org/10.1038/s41423-020-00602-7.

113. Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model (Springer).

114. Kassambara, A., Kosinski, M., and Biecek, P. (2024). survminer: Drawing Survival Curves using “ggplot2.”

115. Dolgalev, I. (2025). msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. Version R package version 25.1.0.9000.

116. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. Bioconductor. http://bioconductor.org/packages/org.Hs.eg.db/.

117. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

118. Kassambara, A. (2025). ggpubr: “ggplot2” Based Publication Ready Plots.

119. Gu, Z. (2022). Complex heatmap visualization. iMeta *1*, e43. https://doi.org/10.1002/imt2.43.