

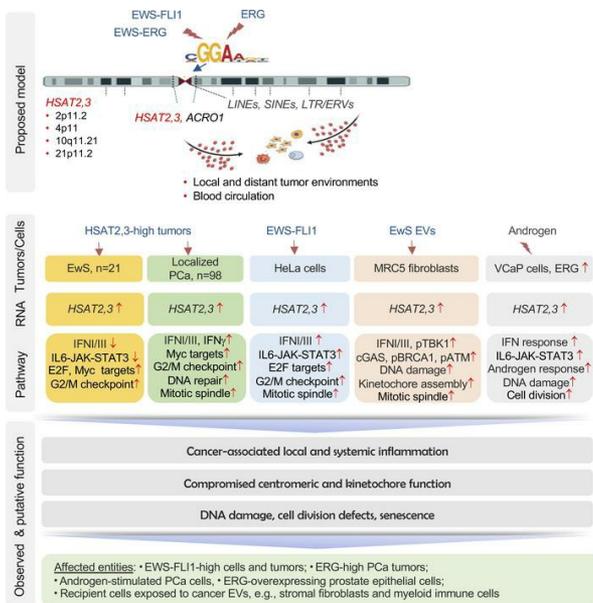
# Oncogenic ETS fusions promote DNA damage and proinflammatory responses via pericentromeric RNAs in extracellular vesicles

Peter Ruzanov, ... , Lincoln D. Stein, Poul H. Sorensen

*J Clin Invest.* 2024. <https://doi.org/10.1172/JCI169470>.

Research In-Press Preview Inflammation

## Graphical abstract



Find the latest version:

<https://jci.me/169470/pdf>



## **Oncogenic ETS fusions promote DNA damage and proinflammatory responses via pericentromeric RNAs in extracellular vesicles**

Peter Ruzanov<sup>1†</sup>, Valentina Evdokimova<sup>1†</sup>, Manideep C. Pachva<sup>2,3</sup>, Alon Minkovich<sup>1</sup>, Zhenbo Zhang<sup>1</sup>, Sofya Langman<sup>2,3</sup>, Hendrik Gassmann<sup>4</sup>, Uwe Thiel<sup>4</sup>, Marija Orlic-Milacic<sup>1</sup>, Syed H. Zaidi<sup>1</sup>, Vanya Peltekova<sup>1</sup>, Lawrence E. Heisler<sup>1</sup>, Manju Sharma<sup>5</sup>, Michael E. Cox<sup>5</sup>, Trevor D. McKee<sup>6,7</sup>, Mark Zaidi<sup>7,8</sup>, Eve Lapouble<sup>9</sup>, John D. McPherson<sup>1,10</sup>, Olivier Delattre<sup>9,11</sup>, Laszlo Radvanyi<sup>1,12‡</sup>, Stefan E. G. Burdach<sup>2,4,13,14‡</sup>, Lincoln D. Stein<sup>1,15‡\*</sup> and Poul H. Sorensen<sup>2,3‡\*</sup>

<sup>1</sup>Ontario Institute for Cancer Research, Toronto, ON, Canada

<sup>2</sup>Department of Molecular Oncology, British Columbia Cancer Research Centre, and

<sup>3</sup>Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada

<sup>4</sup>Department of Pediatrics, Children's Cancer Research Center, Kinderklinik München Schwabing, School of Medicine, Technical University of Munich, Germany

<sup>5</sup>Vancouver Prostate Centre, Vancouver, Canada

<sup>6</sup>STTARR Innovation Centre, Radiation Medicine Program, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada;

<sup>7</sup>Pathomics Inc., Toronto ON

<sup>8</sup>Department of Medical Biophysics, University of Toronto, Canada;

<sup>9</sup>Unité Génétique Somatique (UGS), Institut Curie, Centre Hospitalier Paris, France

<sup>10</sup>Department of Biochemistry and Molecular Medicine, University of California Davis Comprehensive Cancer Center, Sacramento, California 95817

<sup>11</sup>Diversity and Plasticity of Childhood tumors, INSERM U830, Institut Curie Research Center, PSL Research University, Paris, France

<sup>12</sup>Department of Immunology, University of Toronto,

<sup>13</sup>CCC München Comprehensive Cancer Center, DKTK German Cancer Consortium Munich

<sup>14</sup>Institute of Pathology, Translation Pediatric Cancer Research Action, School of Medicine, Technical University of Munich, Germany

<sup>15</sup>Department of Molecular Genetics, University of Toronto.

†These authors contributed equally to this work.

‡LR, SEGB, LDS and PHS share senior authorship.

\*Co-corresponding authors.

**Keywords:** Ewing sarcoma, prostate cancer, EWS::FLI1, ERG, extracellular vesicles, pericentromeric heterochromatin, repeat RNA, HSAT, HERV-K, cGAS, cancer-associated inflammation, DNA damage

## Abstract

Aberrant expression of ETS transcription factors characterizes numerous human malignancies. Many of these proteins, including EWS::FLI1 and EWS::ERG fusions in Ewing sarcoma (EwS) and TMPRSS2::ERG in prostate cancer (PCa), drive oncogenic programs via binding to GGAA repeats. We report here that both EWS::FLI1 and ERG bind and transcriptionally activate GGAA-rich pericentromeric heterochromatin. The respective pathogen-like *HSAT2* and *HSAT3* RNAs, together with *LINE*, *SINE*, *ERV* and other repeat transcripts, are expressed in EwS and PCa tumors, secreted in extracellular vesicles (EVs) and are highly elevated in plasma of EwS patients with metastatic disease. High *HSAT2,3* levels in EWS::FLI1 or ERG expressing cells and tumors were associated with induction of G2/M checkpoint, mitotic spindle and DNA damage programs. These programs were also activated in EwS EV-treated fibroblasts, coincident with accumulation of *HSAT2,3* RNAs, proinflammatory responses, mitotic defects, and senescence. Mechanistically, *HSAT2,3*-enriched cancer EVs induced cGAS-TBK1 innate immune signaling and formation of cytosolic granules positive for double-strand RNAs, RNA-DNA and cGAS. Hence, aberrantly expressed ETS proteins derepress pericentromeric heterochromatin, yielding pathogenic RNAs which transmit genotoxic stress and inflammation to local and distant sites. Monitoring *HSAT2,3* plasma levels and preventing their dissemination may thus improve therapeutic strategies and blood-based diagnostics.

## Introduction

EwS is a poorly differentiated and highly aggressive childhood and adolescent malignancy of bone and soft tissues, thought to originate from primitive bone marrow-derived mesenchymal stem cells (1, 2). Unlike the majority of adult cancers, EwS has extremely low mutation rates and is characterized by single chromosomal translocations, most frequently t(11;22)(q24;q12) and t(21;22)(q22;q12), which are found in ~85% and 10% of EwS cases, respectively (2-4). The resultant fusion proteins, EWS::FLI1 and EWS::ERG, consist of the N-terminal transactivation domain of the RNA-binding protein EWS linked to the C-terminal DNA-binding domains of FLI1 or ERG (1-3). Both FLI1 and ERG belong to the E26 transformation-specific (ETS) transcription factor family and exhibit high degree of similarity, including virtually identical DNA-binding domains (5, 6). Aside from EwS, ETS factors are also fused or aberrantly expressed in many adult solid and haematological malignancies, including melanoma, leukemia and carcinomas of prostate, breast, ovaries, lungs, pancreas and colon (5). In prostate cancer (PCa), up to 50% of cases are characterized by *TMPRSS2::ERG* gene fusions, in which ERG expression is driven by the androgen-responsive *TMPRSS2* promoter (5, 7, 8).

Both EWS::FLI1 and EWS::ERG in EwS, and aberrantly expressed ERG in PCa can bind to GGAA repeats in the promoters and enhancers of coding genes and within microsatellites, causing widespread chromatin remodeling and transcriptional rewiring (5, 6, 9-13). Despite their histogenetic differences, EwS and PCa display many similarities, including evidence of genotoxic stress, such as impaired DNA damage repair pathways, accumulation of R-loops and enhanced sensitivity to poly (ADP-ribose) polymerase (PARP) inhibitors, similar to BRCA1/2-deficient breast cancer (5, 14-16). Lastly, both EwS and PCa belong to the class of immunologically cold solid tumors and are characterized by persistent inflammation and the immunosuppressive tumor microenvironment (TME), which facilitates tumor progression and resistance to chemo and immunotherapy (17). Identifying and targeting mechanisms driving genotoxic stress and inflammation in these cancers may improve clinical outcomes.

Our prior work showed that EVs released by EwS cells inhibit T cell activation and induce proinflammatory phenotypes in CD14<sup>+</sup> and CD33<sup>+</sup> blood monocytes (18, 19). Here, we demonstrate that this effect is likely mediated by repeat RNAs, including human satellite 2 and 3 (*HSAT2,3*), long and short interspersed elements (*LINEs* and *SINEs*) and endogenous retroviruses (*LTR/ERVs*). We show that *HSAT2,3* expression is directly driven by EWS::FLI1 and ERG in EwS and PCa cells, respectively, and may represent one of the earliest events triggering DNA damage and inflammation in tumor cells and the tumor microenvironment (TME).

## Results

*EwS EVs are enriched with retroelement and pericentromeric RNAs.* Our previous finding that EwS EVs induce antiviral innate immune pathways in healthy donor-derived monocytes (18) prompted us to investigate the potential role of the RNAs they carry. Total EV preparations were purified from the conditioned medium of a panel of EwS cell lines by filtration and differential centrifugation (Supplemental Figure 1A). Their purity was confirmed by nanoparticle tracking, RNA profiling and immunoblotting, and was consistent with the characteristics of small EVs (20-22), including their size (40-200 nm), enrichment with small (<200 nt) RNAs and lack of major cellular ribosomal 18S and 28S RNAs (Supplemental Figure 1, B and C). Compared to parental cells, purified EVs were also enriched with classical exosome markers Syntenin 1 and CD81, weakly positive for microvesicle markers Annexin A1 and ARRDC1, and negative for common cellular contaminants Calnexin, Hsp90, Histone H2A,  $\beta$ -Tubulin and the autophagosome marker LC3B (Supplemental Figure 1D).

We next analyzed the EV RNA contents by whole transcriptome RNA sequencing (RNAseq), initially testing three EwS cell lines positive for type 1 (TC32, TC71) or type 3 (A4573) EWS::FLI1 fusion proteins (Figure 1A and Supplemental Table 1). RNAseq analysis of EV RNAs compared to the respective parental cells showed a ~2 fold decrease of protein coding transcripts and a robust enrichment with *7SL/SRP*, *LINE*, *SINE* and *LTR/ERV* and satellite repeat RNAs, with repeat RNA/mRNA ratios 1.7 and 0.5 in EVs and cells, respectively (Figure 1, B and C). These patterns were characteristic of EVs from all three EwS cell lines (Figure 1B), and were previously reported in other tumor EVs, including melanoma, brain and breast cancers (23-25).

Notably, *HSAT2,3* RNAs were almost exclusively detected in EVs but not in the corresponding parental cell lines, which instead exhibited high levels of *ALR* RNA from large centromeric alpha satellite repeats (Figure 1D). These RNAs originated from the GGAAT-rich pericentromeric regions defined by tandemly repeated ~23- to 26-bp consensus motifs (*HSAT2*) or by highly divergent GGAAT/ATTCC simple repeats (*HSAT3*) (26-28). Despite their sequence variability, we were able to uniquely map ~60% of *HSAT2,3* reads using Burrows-Wheeler Aligner (BWA) default parameters, with *HSAT2* predominantly mapping to 10q11.21 and 16q11.2, and *HSAT3* to multiple pericentromeric regions (Figure 1E). Using additional cell lines, we found that EVs from A673 and TC32 EwS cells were highly enriched with *HSAT2,3* RNAs, while EVs from normal diploid MRC5 fibroblasts were negative (Figure 1E). This was also verified by reverse transcription droplet-based digital polymerase chain reaction (RT-ddPCR), which showed high abundance of *HSAT2*, *HSAT3* and *7SL/SRP* RNAs in EVs from TC32 and A673 cells, but not from non-transformed HS-5 fibroblast-like stromal cells or HMEC-1 and HUVEC endothelial cells (Figure 1F). Together, these data uncover extensive expression of locus-specific

*HSAT2,3* RNAs in EwS cells, and their enrichment in EVs, conjointly with a variety of retroelement and other repeat RNAs.

*Repeat RNAs are the most abundant species in EwS patients' plasma.* To establish the potential clinical relevance of the above findings, we purified and characterized EVs from plasma of newly diagnosed therapy-naïve EwS patients with localized or metastatic disease (designated the EW cohort) and age-matched healthy donors (Supplemental Tables 2, A and B). Purified plasma EVs were within a standard size range (40-200 nm), positive for CD63 and CD81 exosome markers and enriched with small (<200 nt) RNAs (Supplemental Figure 2, A-D). Whole transcriptome RNAseq analysis of uniquely mapped non-ribosomal reads revealed significant differences between EwS patients (n=12) and healthy age-matched donors (n=7): of total RNA content, ~61-66% RNAs in plasma EVs from EwS patients were derived from repeats, compared to ~13% in healthy donors (p = 0.012, Wilcoxon signed-rank sum test; Figure 2A). Similar to EVs purified from EwS cell lines, 7SL/SRP, simple and low complexity repeat RNAs were the most abundant repeat RNAs in plasma EVs of all EwS patients (Figure 2B). Retroelement (*LINE*, *SINE*, *LTR/ERV*) and pericentromeric (*ACRO1* and *HSAT2,3*) RNAs were highly upregulated, especially in patients with metastasis (Figure 2, B and C). These results were validated by RT-ddPCR using the additional plasma EV specimens from 20 EwS patients (designated the TUM cohort) and 49 age-matched healthy donors (Supplemental Tables 2, A and B), which showed a significant increase of *ACRO1*, *HSAT2* and *HERV-K* envelope (*HERV-Kenv*) RNAs in plasma of EwS patients vs healthy donors (p< 0.05, Wilcoxon signed-rank test; Figure 2D). Overall, EVs from EwS patients exhibited higher repeat RNA/mRNA ratio (~2.7-2.9 vs 0.3 in healthy donors), revealing a striking difference between EwS patients and healthy donors.

*HSAT2,3 are expressed in EwS tumors.* Among different satellite RNAs, *HSAT2* and, especially, *HSAT3* were the most abundant species in plasma EVs from EwS patients with metastatic disease (Figure 3A). Similar to EwS EVs from cell lines, *HSAT2,3* reads from EwS patients' plasma EVs were mapped to multiple genomic loci including 1q12, 2p11.2, 4p11, 10q11.21 and 21p11.2, and covered both sense and antisense strands of *HSAT2,3* arrays, adjacent to the centromeric alpha satellite repeat ALR (Figure 3B and Supplemental Figure 3). Analysis of published RNAseq tumor datasets from 21 EwS patients (29) confirmed the expression of *HSAT2,3* RNAs in primary and metastatic tumors, including cases with *EWS::FLI1* (n=16) and *EWS::ERG* (n=5) fusions, and their origins from the 4p11, 10q11.21 and 16q11.2 *HSAT2,3* loci (Figure 3, C and D). Differential gene expression analysis revealed a sharp distinction between *HSAT2,3*-high vs -low tumors, with 2,131 mRNAs upregulated and 570 downregulated (log<sub>2</sub> fold change ≥ 2, false discovery rate (FDR) < 0.25; Figure 3E). Gene set enrichment analysis (GSEA) of *HSAT2,3*-high vs -low tumors showed a significant enrichment with cell cycle and

proliferation pathways (hallmark *MYC*, *E2F*, G2/M checkpoint; FDR < 0.2; Figure 3F) and downregulation of inflammatory and interferon (IFN) pathways (FDR < 0.05; Figure 3G).

Using commercial probes, we next analyzed *HSAT2* and *HERV-K* RNA expression in matching primary and metastatic tumors from three EwS patients by ViewRNA in situ hybridization (ISH). In all three cases, *HSAT2* was significantly upregulated in metastatic sites compared to primary tumors ( $p < 0.05$ , unpaired two-tailed equal variance t-test), while *HERV-K* showed an opposite trend (Supplemental Figure 4, A and B). Overall, *HERV-K* levels were substantially higher than those of *HSAT2* RNAs, which could be due to the preferential secretion of *HSAT2* in EVs or due to detection limitations, given that only a fraction of these highly polymorphic RNAs could be detected with any particular probe. Hence, *HSAT2,3* RNAs are expressed in primary and metastatic EwS tumors with EWS::FLI1 and EWS::ERG fusions, and their expression is linked to cell cycle defects and reduced immune responsiveness.

*EWS::FLI1 preferentially binds to HSAT3 pericentromeric chromatin and activates HSAT2,3 RNA expression in non-EwS cells.* Knowing that EWS::FLI1 and EWS::ERG bind to core GGA(A/T) motifs with at least four consecutive GGAA repeats (11, 13), we next assessed if EWS::FLI1 can bind to GGAA-rich pericentromeric heterochromatin using published chromatin immunoprecipitation sequencing (ChIPseq) datasets from A673 EwS cells (10). Because pericentromeric regions are usually excluded from the analysis due to the high noise-to-signal ratio (13), we assembled the unmapped EWS::FLI1 ChIPseq reads into 200-700-nt contigs, to enhance selectivity and reduce false positives. This identified ~20 high-quality contigs mapping to *HSAT2,3* repeats on 1q12, 4p11, 10q11.21 and 21p11.2, albeit only the *HSAT3* binding sites were significantly enriched compared to the input control ( $p = 0.0037$ , Wilcoxon rank sum test; Figure 4, A and B). Analysis of the respective contig sequences with BLAST showed strong homology (>95%) to previously established *HSAT2,3* subfamily-specific 24-mers (26), and identified GAATGGAAT as the highest-scoring recurring sequence motif (Figure 4C).

To further explore the possibility that EWS::FLI1 directly binds to pericentromeric heterochromatin and induces expression of *HSAT2,3* RNAs, we used HeLa cells as they are known to be *HSAT2*-negative (30, 31) and permissive for EWS::FLI1 expression (32). Transient expression of EWS::FLI1 in these cells followed by RNAseq showed that *HSAT2* and *HSAT3* RNAs were strongly upregulated in EWS::FLI1 expressing cells compared to vector controls, while other satellite RNAs were either downregulated (*SATR1*, *SATR2*, *BSR* and *CER*) or not substantially changed (*SST1*, *ACRO1*, *ALR*, *MSR1*; Figure 4, D and E). The effect was notable only at 72h post-transfection, despite EWS::FLI1 protein levels reaching maximum at much earlier time points (24 and 48h), suggesting a relatively slow transcriptional activation of *HSAT2,3* heterochromatin in these settings. Similar to *HSAT2,3*-high EwS tumors, GSEA analysis of coding transcripts in EWS::FLI1 expressing vs vector HeLa cells revealed upregulation of cell cycle and DNA damage pathways, including G2/M checkpoint, PI3K-AKT-mTOR,

E2F, c-MYC, mitotic spindle, reactive oxygen species and IL-6-JAK-STAT3 inflammatory signaling categories (Figure 4F and Supplemental Figure 5). These results identify *HSAT3* as a preferential EWS::FLI1 binding site within pericentromeric heterochromatin and demonstrate upregulation of *HSAT2,3* RNA upon transient expression of EWS::FLI1 in HeLa cells, concomitant with activation of cell cycle, DNA damage and proinflammatory pathways.

*ERG binds to pericentromeric chromatin and activates HSAT2,3 expression and DNA damage pathways in PCa cells.* Given the structural and functional similarities between FLI1 and ERG (5, 6) and the reported ability of ERG and other ETS factors to activate transcription in PCa cells via binding to GGAA repeats in a complex with the wild-type EWS protein (12), we next investigated if transcriptional activation of *HSAT2,3* pericentromeric heterochromatin may also occur in ERG-driven cancers, particularly PCa, where ERG is overexpressed in more than 50% of cases (5, 8). We thus used the *TMPRSS2::ERG* fusion gene-positive VCaP cell line, where endogenous ERG expression can be induced by natural (dihydrotestosterone, DHT) or synthetic (methyltrienolone, R1881) androgens via androgen receptor (AR) binding to the *TMPRSS2* promoter (7). Similar to EWS::FLI1, analysis of published ERG ChIPseq datasets (9) yielded *HSAT3* (but not *HSAT2*) contigs which mapped to the 4p11, 10q11.21 and 21p11.2 *HSAT2,3* loci. The occupancy of these binding sites was significantly increased in DHT-treated (18h DHT) ERG-high vs untreated (0h DHT) ERG-low cells ( $p \leq 0.001$ ; two-way ANOVA; Figure 5A). The top-scoring ERG-specific 43-mer contained multiple continuously repeated GAAT sequences (Figure 5B). Among lower-scoring motifs, we also identified GAATGGAAT, suggesting that ERG and EWS::FLI1 binding sites may at least partially overlap.

To examine if *HSAT2,3* expression in PCa cells is ERG-dependent, VCaP cells were transfected with ERG-targeting or scrambled control (ctrl) siRNAs followed by treatment with R1881 for 48h, to induce the endogenous ERG expression (Figure 5C, top panel). As expected, ERG protein levels were elevated in R1881-stimulated (ERG-high) compared to unstimulated (ERG-low) cells, and were diminished upon treatment with ERG-targeting siRNA (ERG KD; Figure 5C, bottom panel). Coincident with the ERG increase, RNAseq analysis showed significant upregulation of *HSAT2,3* (but not *ALR*) in ERG-high vs ERG-low VCaP cells (~60-fold;  $p = 0.028$ , paired two-tailed t-test; Figure 5D), with the majority of reads mapping to the 2p11.2- and 4p11-*HSAT3* and 10q11.21-*HSAT2,3* loci (Figure 5E). Expression of *HSAT2,3* RNAs was likely driven by ERG and not by androgen, given that their levels were reduced in ERG-KD cells despite treatment with R1881 (Figure 5, D and E). For unclear reasons, other satellite RNAs were diminished in ERG-high and even more so in ERG-KD cells (Figure 5D), which may reflect their interdependent expression within centromeric chromatin (33). Both *HSAT2* and *HSAT3* RNAs were also detected in EVs and their levels increased by >4-fold in EVs from ERG-high vs ERG-low cells ( $p < 0.001$ , t-test), while *HERV-K* and *7SL/SRP* remained unchanged (Figure 5F). As expected,

differential gene expression analysis of ERG-high vs ERG-low VCaP cells identified upregulation of canonical androgen (*KLK2*, *KLK3/PSA*, *FKBP5*, *STEAP4*, *NKX3-1*, *TMPRSS2*) and ERG (*PLAT*, *SERPINE1/PAI-1*, *ADAMTS1*, *EZH2*) signature genes (Supplemental Figure 6A). Moreover, genes involved in proinflammatory responses and immunosuppression (*IL-10*, *IDO1*, *SOCS1*, *PDCD1/PD-1*, *IL-6*), DNA damage repair (*BRCA1*, *H2AFX*, *XRCC2*) and centrosome and kinetochore assembly networks (*KIF20A*, *KIF20B*, *KIF2C*, *PLK2*, *PLK3*, *CDC25B*, *CDC20*, *AURKA*, *AURKB*, *CENPL*, *CENPU*) were also upregulated (Supplemental Figure 6, B and C). GSEA of ERG-high vs -low cells confirmed significant upregulation of numerous known PCa-associated and inflammatory pathways, including *TNFA*, *KRAS*, *WNTB-CATENIN*, and *IL-6-JAK-STAT3* networks (FDR < 0.1; Supplemental Figure 6D).

To test if *HSAT2,3* expression can be induced by ERG in non-transformed cells, we examined normal prostate (PNT1B and RWPE1) and benign prostatic hyperplasia (BPH1) cell lines transduced to express ERG or vector alone, as previously described (34). We found that compared to vector alone, *HSAT2* levels were significantly upregulated in all three ERG-expressing cell lines (Figure 5G), and were also detected in EVs, alongside with *HERV-K* (Figure 5H). However, *HSAT3* RNAs were detected only in ERG-BPH-1 cells (Figure 5G) and not in EVs, suggesting that ERG alone is not sufficient to activate *HSAT3* expression in non-cancerous prostate cells.

To examine *HSAT2,3* RNA expression in tumor settings and to identify co-activated gene expression programs, we interrogated RNAseq datasets from 98 treatment-naïve patients with localized PCa (35). This showed that *HSAT3* and, to a lesser extent, *HSAT2* were expressed in PCa tumors but their levels were substantially lower than those of *ALR* (Supplemental Figure 7A), resembling their expression patterns in EwS tumors (Figure 3C). Elevated levels of *HSAT2,3* RNAs positively correlated with tumor grade (Gleason score 6 vs 7-8) and a slightly increased risk of biochemical recurrence (Supplemental Figure 7, A and B), but none of them reached statistical significance. We also did not find a significant correlation with any of the most frequent somatic mutations detected in these tumors, including *ZFH3*, *APC*, *CDKN1B*, *RNF43*, *TP53* and *ATM* genes (Supplemental Figure 7C). Similar to EwS tumors (Figure 3F), GSEA analysis of *HSAT2,3*-high vs -low PCa tumors revealed evidence of cell cycle dysregulation and genotoxic stress, as manifested by significant upregulation of *E2F*, *MYC*, G2/M checkpoint, mitotic spindle and DNA damage repair networks (FDR < 0.05; Figure 5I), while downregulated pathways did not reach statistical significance (Figure 5J). In contrast to EwS however, the IFN response was significantly upregulated in *HSAT2,3*-high PCa tumors (FDR < 0.1; Figure 5I), and thus immunogenicity of *HSAT2,3*-expressing tumors may be cancer type/stage and context-dependent.

*HSAT RNAs exhibit pathogen-like features and are disseminated in the TME.* Previous studies have indicated that *HSAT2,3* repeats in the human genome possess unusual hydrodynamic and

hydrogen bonding properties (36), and that the *HSAT2* RNAs exhibit pathogen-like CpG and UpA motif usage and can propagate via reverse transcription and re-insertion into the genome (37, 38). We found that indeed, *HSAT2* and, especially, *HSAT3* RNAs could be distinguished from cellular mRNAs based on asymmetrical distribution of GC and AU nucleotides (GC/AU skew; Figure 6A) and can thus be recognized by innate immune sensors as “non-self”. Moreover, they are likely present in EVs (but not in cells) in a double-stranded (ds) form, given equal ratios of *HSAT2,3* RNAseq reads in sense and antisense orientations (Supplemental Figure 8), which may further increase their immunogenicity. Indeed, using dsRNA-specific J2 antibodies, we were able to immunoprecipitate *HSAT2,3* and *HERV-K* RNAs from the conditioned medium from three different EwS cell lines (Figure 6B). Smaller but a significant fraction of these RNAs was also immunoprecipitated with S9.6 antibodies, suggesting the presence of RNA-DNA hybrids.

To investigate if these RNAs can be transferred to stromal cells in the TME, we used a previously developed xenograft model of EwS, where TC32 cells were implanted under the renal capsules of NOD/SCID mice (39). Xenografted TC32 tumors exhibited typical EwS morphology of small blue tumor cells with enlarged nuclei (Figure 6C). ViewRNA-ISH with commercial probes for human *HSAT2* and *HERV-K* RNAs confirmed their expression in tumor cells, some of which contained large *HSAT2*-positive nuclear aggregates (Figure 6D). Using the *HSAT2* probes in combination with probes for mouse immune cells (*Cd45*) or fibroblasts (*Fsp-1*), we identified *HSAT2*-positive mouse cells in proximity to areas of tumor beds, while mouse epithelial cells and more distant *Fsp-1*<sup>+</sup> fibroblasts in adjacent normal kidney tissues were negative (Figure 6, E and F). Quantitative image analysis showed that more than 20% of tumor-infiltrating *Cd45*<sup>+</sup> immune cells and *Fsp-1*<sup>+</sup> fibroblasts were positive for human *HSAT2* RNAs (Figure 6, G and H), supporting their transfer from tumor to stromal cells.

We next examined the potential impact of EwS EVs on stromal fibroblasts and myeloid cells, using MRC5 fibroblasts and MUTZ-3 myeloid cells as in vitro models. After the 6-h exposure to EwS EVs, both MRC5 and MUTZ-3 cells secreted type-I and type III IFNs (IFN $\alpha$ , IFN $\beta$  and IL-28A/IFN $\lambda$ 2), with TC32 EVs having the strongest effect (Supplemental Figure 9A-C). In addition, MUTZ-3 cells secreted proinflammatory and immunosuppressive cytokines, including IL-8, IL-10, IL-20 and IL-35 (Supplemental Figure 9C). Prolonged exposure ( $\geq 72$ h) to TC32 EVs increased the number of cells with mitotic defects (Supplemental Figure 9, D and E) and positivity for senescence-associated  $\beta$ -galactosidase (Supplemental Figure 9, F and G), reminiscent of abnormalities caused by ectopic *HSAT2* expression (31). Therefore, *HSAT2,3* RNAs exhibit pathogen-like dinucleotide usage, can be transmitted to stromal immune cells and fibroblasts in the TME, and may contribute to EwS EV-induced proinflammatory responses, mitotic abnormalities, and senescence in adjacent stromal fibroblasts.

*HSAT-enriched EwS EVs activate cGAS and DNA damage pathways in recipient cells.* Despite the highly immunogenic properties, *HSAT2,3* expression did not activate inflammatory responses in EwS cells and tumors (Figure 3G), which could be due to their preferential secretion in EVs and/or their nuclear retention in tumor cells. We also considered a possibility that *HSAT2,3* RNAs in tumor cells may be sequestered in nuclear paraspeckles that are known to capture aberrant RNAs and regulate innate immunity (40, 41). Using paraspeckle *NEAT1* RNA as a positive control, ViewRNA fluorescence in situ hybridization (FISH) showed that *HSAT2* RNAs are indeed localized to the nucleus and nuclear envelope in A673, TC32 and A4573 EwS cells, but they are not co-localized with *NEAT1* (Supplemental Figure 10), and thus are not trapped in paraspeckles.

In contrast to tumor cells, *HSAT2* RNAs were detected on the plasma membrane and in the cytosol of MRC5 cells treated with TC32 EVs, but not with MRC5 EVs (Figure 7A). To determine if *HSAT2* and other pathogen-like RNAs in EwS EV-treated cells may engage cytosolic innate immune sensors, we focused on the cyclic GMP-AMP synthase (cGAS)-TANK binding kinase 1 (TBK1) pathway known to be activated by pathogen- and self-derived dsDNAs (42, 43), including those of retroelement origins (44). Immunofluorescence microscopy with cGAS and S9.6 (or J2) antibodies revealed large cytosolic granules positive for cGAS, RNA-DNA and dsRNAs in MRC5 cells treated with TC32 or A673 EwS EVs (Figure 7, B and C), which could be indicative of cGAS activation (45).

To further verify EwS EV-mediated RNA transfer to MRC5 cells and to identify activated pathways, we performed RNAseq analysis of TC32 EV-treated and mock control MRC5 cells (Figure 7D, top panel). *HSAT2,3* RNAseq reads were detected exclusively in TC32 EV-treated cells and mapped to the 2p11.2, 4p11, 10q11.21, 16q11.2 and 21p11.2 loci (Figure 7D, bottom panel), supporting their EwS cell origin. Also, consistent with the observed secretion of proinflammatory cytokines and IFNs by TC32 EV-treated MRC5 cells (Supplemental Figure 9B), analysis of differentially expressed coding genes in TC32 EV-treated vs mock MRC5 cells revealed upregulation of IFN-I/III and proinflammatory response, including *cGAS/MB21D1*, the IFN regulatory factor 7 (*IRF7*) and proinflammatory cytokines *IL8*, *IL32* and *TNFRSF14* (Figure 7E and Supplemental Table 3A). Remarkably however, the largest group of upregulated genes (54 of 297) were those encoding centromere and kinetochore complex components (“*CENPA/NDC80*” module), key regulators of spindle assembly and chromatin condensation (“*PLK1/NEK2*” module) and DNA damage repair proteins (“*BRCA1/TOP2A*” module), as shown by gene ontology (GO) and Reactome network analysis (Supplemental Figure 11 and Supplemental Table 3B), supporting links between *HSAT2,3* and cell cycle, kinetochore assembly, DNA damage and innate immune pathways.

Activation of the above pathways in MRC5 cells treated with A673 or TC32 EVs was further confirmed by immunoblotting. In contrast to MRC5 cells treated with MRC5 EVs or with the dsRNA

mimetic poly(I:C), EwS EVs activated DNA damage (pATM and pBRCA1) and innate immune (cGAS and pTBK1) responses, which persisted over the 72-h time course (Figure 7F). Likewise, compared to EVs from vector control cells, the *HSAT2*-enriched EVs from ERG-expressing prostate cells, especially BPH1, upregulated cGAS, pTBK1, pATM, pBRCA1 and  $\gamma$ H2AX levels in MRC5 cells (Supplemental Figure 12). In summary, reactivation of pericentromeric heterochromatin in cancers with aberrantly expressed ETS proteins, including EWS::FLI1 and ERG, results in expression and dissemination of *HSAT2,3* RNAs, which together with *HERV-K* and other retroelement RNAs present in EVs may activate DNA damage and cGAS-pTBK innate immune signaling in the recipient immune cells and fibroblasts, driving local and systemic inflammation (Figure 7G).

## Discussion

One of the biggest obstacles for successful immunotherapy in the majority of solid human malignancies is low immunogenicity of the tumor and the proinflammatory immunosuppressive TME (17). Our study demonstrates that reactivation of constitutively silent heterochromatin may reinforce oncogenic gene expression programs and may also contribute to a paradoxical co-occurrence of inflammation and immunosuppression. In particular, we show that elevated expression of *HSAT2,3* RNAs in EwS and PCa as well as in EWS::FLI1 expressing HeLa and in ERG-high prostate cells was associated with activation of proinflammatory signaling, DNA damage, G2/M checkpoint and mitotic spindle/kinetochore maintenance gene expression programs, many of which are known to be dysregulated in both cancers and, in some cases, are directly driven by EWS::FLI1 or ERG (14-16, 46). Moreover, these programs were also activated in EwS EV-treated fibroblasts, contemplating a possibility that *HSAT2,3* RNAs can mediate genotoxic effects of these oncogenic proteins in bystander host cells. The remarkable similarities of gene expression programs in *HSAT2,3*-high cells and tumors could be attributed to the important roles of peri/centromeric RNAs in genome architecture, kinetochore assembly and mitosis (33, 47, 48). This could also be due to their ability to directly bind and sequester BRCA1 and other proteins that are involved in heterochromatin maintenance, replication, transcription, and splicing (30, 31, 49).

Mechanistically, we show that activation of *HSAT2,3* pericentromeric heterochromatin may occur through the direct binding of EWS::FLI1 (or ERG) to *HSAT3* repeats, primarily those on chromosomes 4p11 and 10q11.21, and less so, on 2p11.2 and 21p11.2. While the exact mechanism remains to be established, preferential binding to *HSAT3* repeats could be due to their enrichment with consecutively repeated GGAAT motifs, which are high affinity binding sites for both EWS::FLI1 and ERG in a complex with EWS (5, 6, 12, 13). Binding to *HSAT3* may then activate the adjacent *HSAT2* repeats, given their proximity in certain genomic locations, particularly on chromosomes 2p11.2 and 10q11.21. It cannot be excluded however that *HSAT2* repeats can be directly activated by overexpressed ERG in certain

cellular contexts, as hinted by its more promiscuous consensus binding motifs and by upregulation of *HSAT2* (but not *HSAT3*) in ERG-high BPH1, PNT1 and RWPE1 non-transformed prostate cells.

Our work also shows that EwS and PCa tumor cells have developed the ability to release *HSAT2,3* RNAs in EVs. A potential mechanism may involve their sorting in EVs in a form of dsRNAs and RNA-DNA hybrids, and retention of ssRNAs in nuclei of tumor cells. This could also explain a lack of immune response in *HSAT2,3* expressing EwS cells compared to EwS EV-treated cells, where cytosolic accumulation of *HSAT2* RNAs was associated with vigorous IFN-I/III and proinflammatory responses, activation of cGAS-pTBK1 signaling and a buildup of cytosolic RNA-DNA and dsRNA-positive cGAS aggregates. The cGAS-TBK1 innate immune pathway is emerging as a key driver of cancer-associated chronic inflammation and immunosuppression (42, 50). Upon engagement with cytosolic DNA, cGAS forms liquid phase droplets, where it becomes activated (43, 45). cGAS can also bind to RNA-DNA hybrids including retroelements (44), and can restrict RNA virus infections (51). It is plausible that similar mechanisms may operate in the host cells upon exposure to cancer EVs, as a result of cytosolic accumulation of *HSAT2,3* and retroelement RNAs, some of which may be able to reverse transcribe (38, 52, 53). Alternatively, cGAS activation in cancer EV-treated cells can be driven by DNA damage and leakage of genomic DNA to the cytosol, which may also lead to senescence, as observed in this study and by others (42, 43).

In other cancers and senescence-associated disorders, aberrant activation of pericentromeric chromatin has been linked to *BRCA1* mutations (49, 54) and malfunction of various epigenetic modifiers, including histone deacetylase SIRT6 (55) and DNA methyltransferases DNMT1 and DNMT3B (56). *HSAT2* expression was observed in colorectal, pancreatic and ovarian carcinomas, and was linked to immunosuppression (25, 57). It is thus tempting to speculate that pericentromeric chromatin activation is a convergent point for multiple oncogenic drivers in diverse human malignancies. Our current and previous (18, 19) studies indicate that this process may also extend to normal host cells in the TME and systemic circulation, causing DNA damage, proinflammatory responses, senescence and immunosuppressive phenotypes (Figure 7G). In this regard, our investigation argues against the proposed strategy to induce expression of *ERVs*, *LINEs* and other repeat RNAs (so-called “viral mimicry” state) as means of increasing tumor immunogenicity (58), as this may ultimately lead to chronic inflammation, immunosuppression and senescence. Quite the contrary, in order to reactivate the immune system, further investigation should be focused on developing “containment” strategies, to prevent their expression in tumor cells, dissemination in EVs and “infection” of target cells.

## Methods

*Sex as a biological variable.* The EwS patients' cohorts included 18 males and 20 females. Sex was not considered as a biological variable in this study; similar results were obtained for both sexes. Relevant clinical information (age, gender, tumor site and stage) is shown in Supplemental Table 2, A and B.

*Plasma EV isolation.* Plasma specimens were kept at -80°C. After thawing, plasma (~1-2 ml/patient) was cleared by two-step centrifugation at 2,000g for 10 min and at 10,000g for 20 min, to remove cell debris and aggregates. Supernatants were diluted 1:4 with ice-cold phosphate buffered saline (PBS), passed through 0.45 µm filters and centrifuged for 90 min, 4°C at 100,000g (S110-AT rotor, ThermoFisher). To ensure removal of plasma contaminants, EV-containing pellets were resuspended in 3 ml of PBS and pulled down at 100,000g for 60 min. Pellets were dissolved in PBS and examined using the nanoparticle tracking system (NS300 NTA).

*Cell lines and EV isolation.* Cell lines and growth conditions are listed in Supplemental Table 1. For EV isolation, cells were cultured on 15-cm plates (10 x 10<sup>6</sup> cells/plate) in growth medium containing 2% EV-depleted FBS (System Biosciences) for 18-24h. Conditioned medium (CM) from the same number of cells was collected and cleared by sequential centrifugation at 2,000g for 10 min and 10,000g for 20 min, passed through 0.45 µm filters and concentrated on Amicon Ultra-15mL-30K tubes (Millipore). Concentrated CM was then diluted 1:1 with PBS and centrifuged at 100,000g for 90 min. Crude EVs were resuspended in PBS and pelleted again at 100,000g for 60 min. EV pellets were either directly used for RNA isolation with the mirVana RNA kit (AM1561, ThermoFisher), or resuspended in PBS for downstream analyses.

*RNA immunoprecipitation (IP) from the conditioned medium.* CM was precleared by filtering and sequential low-speed centrifugation, concentrated on Amicon Ultra tubes, and supplemented with the EV lysis buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5% Triton X-100, 0.1% SDS) in a total volume of 2 ml. The CM was then pre-cleared with protein G magnetic beads (S1430S, New England Biolabs) at 4°C for 1h, and incubated for 3h with 5 µl of S9.6 (ENH001, Kerfast) or J2 (ES2001) antibodies, or normal mouse IgG (sc-2025, Santa Cruz) immobilized on 30 µl of protein G magnetic beads. After removal of unbound fractions, magnetic beads were washed with the 140 mM NaCl EV lysis buffer, and bound RNAs were eluted with 2 µg/ml proteinase K (19131, Qiagen) and isolated with the mirVana kit (AM1561, ThermoFisher) using 1 ng of MS2 RNA (10165948001, Sigma) as a spike-in control.

*Reverse transcription (RT)-PCR.* Purified RNAs were treated with Turbo DNase (AM1907, ThermoFisher) and quantified using the Qubit RNA high sensitivity kit (Q32852, ThermoFisher). RNA integrity was analyzed using the Agilent RNA Pico 6000 kit on a 2100 Bioanalyzer instrument (Agilent Technologies). cDNA was synthesized using SuperScript III Reverse Transcriptase (18080044,

ThermoFisher), 100-500 ng of DNase-treated RNA and random hexamer primers. Quantitative PCR (qPCR) was performed on the ViiA 7 Real-Time PCR instrument (ThermoFisher) using the PowerUp SYBR Green Master Mix (A25777, ThermoFisher) and the respective primer sets (Supplemental Table 4). RT-negative controls were included in all qPCR reactions. Expression of target transcripts was quantified by  $\Delta\Delta C_t$  method using normalized  $C_t$  values. qPCR products were analyzed by 1.5% agarose gel electrophoresis and confirmed by Sanger sequencing. Droplet-based digital PCR (ddPCR) was performed on the QX100 Droplet Digital PCR system (Bio-Rad) using the QX200™ ddPCR™ EvaGreen Supermix (1864033, Bio-Rad). PCR reactions were incubated at 95°C, 5 min, followed by 40 cycles of 95°C for 30 s and 58°C for 1 min, with final incubations at 4°C for 5 min and at 90°C for 5 min. The average number of accepted droplets for the valid measurement results was ~17,000.

*HeLa and VCaP cell culture assays.* Transient transfection of HeLa cells was performed on 6-well plates ( $0.5 \times 10^6$  cells/well) using the Lipofectamine 2000 reagent and the p3xFLAG-CMV-10-EWS::FLI1 plasmid (59), a gift from Beat Schäfer (University Children's Hospital, Switzerland). Cells were collected at 24, 48, and 72h post-transfection and analyzed by immunoblotting and RNAseq. VCaP cells were grown in phenol red free DMEM supplemented with 10% charcoal stripped fetal bovine serum for 24h prior stimulation with 1 nM R1881/methyltrienolone (Sigma) for 48h. Where indicated, cells were transfected with On-target plus SMART pool siRNAs (Dharmacon), targeting human ERG (L-003886-00-0005) or non-targeting control (Src: d-001810-01-05) using DharmaFECT transfection reagent (Dharmacon). Cells and the respective CM were collected 48h after treatment with R1881 and siRNAs, and used for immunoblotting and RNA isolation.

*Immunoblotting.* Whole cell extracts and EVs were prepared in lysis buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.8, 1 mM DTT, 0.25% NP-40) supplemented with complete protease inhibitor cocktail (Roche Diagnostics), resolved on SDS-10% polyacrylamide gel (10-25  $\mu$ g protein/lane) and transferred onto nitrocellulose membrane (0.45  $\mu$ m; 162-0115, Bio-Rad). The membrane was incubated for 30 min in 5% non-fat dry milk/TBST (150 mM NaCl, 20 mM Tris-HCl, pH 7.8, 0.1% Tween-20) blocking buffer followed by overnight incubation at 4°C with primary antibodies listed in Supplemental Table 5. The membranes were washed with TBST, incubated with horseradish peroxidase-conjugated secondary antibodies at 1:5,000 dilution for 40 min at room temperature, washed with TBST, developed with ECL reagents and documented by digital imaging using ChemiDoc MP Imaging system (Bio-Rad).

*Bio-Plex proinflammatory cytokine profiling.* MRC5 fibroblasts or MUTZ-3 myeloid cells were treated for 6h with 25-50  $\mu$ l of EVs ( $\sim 2.5 \times 10^9$  particles/50,000 cells/well in 24-well plates) or mock control (PBS). The medium was replaced with the fresh one containing 2% FBS, and cells were grown for another 18h. The CM was collected and analyzed using the magnetic bead-based multiplex

immunoassay Bio-Plex 200 system and the 37-plex human inflammation panel (171AL001M, Bio-Rad). Standard and sample cytokine concentrations were calculated using the Bio-Rad Bio-Plex 200 software. Background-corrected signals for each cytokine were normalized to the mean signal intensity obtained for each sample and were plotted using the mean and standard deviation from three biological replicates.

*Senescence-associated  $\beta$ -galactosidase (SA- $\beta$ -gal) staining.* MRC5 cells were plated in 24-well plates at low confluency (25,000 cells/well), treated with 50  $\mu$ l of EVs ( $5 \times 10^9$  particles) or PBS in 500  $\mu$ l of 2% FBS medium for 24h, and then grown for 3-5 days in a fresh 10% FBS medium. Senescent cells were detected with the senescence  $\beta$ -gal staining kit (9860, Cell Signaling).

*Immunofluorescence (IF) microscopy and ViewRNA-FISH.* MRC5 and EwS cells were plated on coverslips in 24-well plates and treated with EVs ( $\sim 5 \times 10^9$  particles/50,000 cells/well). For IF, cells were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100 and blocked with 5% bovine serum albumin (BSA) for 1h. Staining was done using  $\alpha$ -Tubulin (3873, Cell Signaling) and Aurora A (14475, Cell Signaling) antibodies (Supplemental Table 5) overnight at 4°C, followed by 1-h incubation with Alexa Fluor-488 and -594 conjugated anti-mouse and anti-rabbit secondary antibodies. ViewRNA-FISH was done using the ViewRNA Cell Plus Assay Kit (88-19000-99, ThermoFisher) and probe sets for *HSAT2* (VA6-19493-06) and *NEAT1* (VA4-20324). Coverslips were mounted with ProlongGold antifade reagent with DAPI (P36935, Invitrogen). Images were acquired using a Zeiss confocal microscope with x63 oil immersion lens.

*ViewRNA in situ hybridization (ISH).* Archived formalin-fixed paraffin-embedded (FFPE) sections of EwS tumors were provided by Katja Specht and Katja Steiger (Institute of Pathology, Technical University of Munich School of Medicine, Munich). FFPE sections of mouse xenografts of TC32 EwS cells were from a previous study (39), provided by Amal El-Naggar (University of British Columbia, Vancouver). ViewRNA-ISH was performed using the tissue 2-Plex assay kit (QVT0012, ThermoFisher) and probe against human *HERV-Kpol* (VF1-18759-06) and *HSAT2* (VA6-19493-06), and mouse *S100a4/Fsp-1* (VB1-18990-06) and *Ptprc/Cd45* (VB1-12657-06) RNAs. Briefly, 5  $\mu$ m FFPE sections were baked and deparaffinized, followed by heat pretreatment and protease digestion. Target probes were then hybridized for 2h at 40°C followed by ViewRNA signal amplification. Signal detection was performed sequentially, using Fast Blue and Fast Red substrates, and slides were then counterstained with hematoxylin. Signal was considered specific if normal kidney tissue and no-probe controls showed less than 1 dot per 20 cells. Regions of interest were captured at x100 magnification using a Leica DM2000 microscope and LAS software with a pixel resolution of 0.04  $\mu$ m/pixel. Stain separation of the fast blue, fast red and nonspecific background signals were achieved using a python implementation of the independent component analysis approach (60). The resulting deconvoluted images were analysed using a custom rule set built in the object-oriented machine learning software Developer XD (Definiens

AG, Munich, Germany). The rule set first separated nuclear regions from nonspecific signal using a weighted combination of all isolated stain intensities, with watershed and object reshaping algorithms being employed to identify and separate closely-spaced nuclei. Blue and red spots were then detected with the fast blue and fast red channels, respectively. The resulting cell objects were exported as a table listing the number and type of spots within each cell. The number of cells containing at least one red spot, one blue spot, or both were tabulated as a fraction of total cells detected.

*Whole transcriptome RNAseq.* Total cellular RNAs from EwS cell lines (100-500 ng) and EV RNAs (20-200 ng) were depleted of ribosomal RNA using the RiboZero rRNA removal kit (20020596, Illumina) and used to construct strand-specific RNAseq libraries with the TruSeq Stranded Total RNA library prep. The quality of libraries was confirmed using the Agilent High Sensitivity DNA assay, and those of sufficient quality and quantity were pooled in sets of four and sequenced using the HiSeq PE cluster kit v4 on the Illumina HiSeq 2500 sequencer, generating on average ~30-100 million paired-end 100-bp reads/sample. RNAseq libraries from HeLa and VCaP cells were prepared using the KAPA RNA HyperPrep Kit with RiboErase (HMR) (8098140702, Roche) and sequenced to about 100 million paired-end 150-bp reads/sample on an Illumina NovaSeq sequencer.

*RNAseq data analysis of repeat transcripts.* Paired-end strand-specific 100 or 125-bp reads were aligned to human genome build 37 (hg19/GRCh37) using the Burrows-Wheeler Aligner (BWA; v.0.7.13) with subsequent filtering of ribosomal and tRNA reads, followed by removal of secondary alignments and reads with mapping score less than 30. Transcript expression was quantified as the number of reads per million (RPM). Based on hg19-compliant RepeatMasker annotations (<https://www.repeatmasker.org>), reads were assigned to different repeat classes using bedtools suite 2.29.2. *HSAT3* was defined by simple GAATG/CATTC repeats larger than 1 kb and located in pericentromeric regions. To quantify expression of different repeat classes, overlapping annotations were resolved, giving priority to features in a hierarchical order: miRNA > piRNA > lncRNA/lincRNA > RepeatMasker annotations. The annotated reads were partitioned into two large groups: “Non-repeat” transcripts, represented by protein coding (including 5'- and 3'-UTRs and introns) and non-coding (pseudogenes, snRNAs, miRNAs and lncRNAs), and “Repeat” transcripts. The latter group was further subdivided into *SINEs*, *LINES*, DNA and *LTR/ERVs*, low complexity, simple (microsatellite) and satellite (pericentromeric and centromeric repeats). The results were plotted using the R statistical environment (<https://www.r-project.org>). Plots were generated using the ggplot2 package.

*Coding gene alignments, differential expression and Gene Set Enrichment Analysis (GSEA).* For protein coding genes, reads were aligned to the human genome (release 87 for GRCh37) with STAR v2.6.0c (61). Counts for genes were derived with the SUBREAD suite (2.0.6) featureCounts tool (62) using GENCODE annotations (release 19). Differential expression analysis of protein coding genes was

performed with edgeR 3.42.4 tools (63). Transcripts showing 2-fold difference and p-value < 0.05 were considered as differentially expressed. For GSEA, RNAseq read counts were normalized with DESeq2 (64) followed by calculation of enrichment score using the GSEA v.4.2.3 and Molecular Signature Database (MSigDB) version 7.5.1 (<http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp>). Normalized enrichment score (NES) was used as a measure of coordinated changes in hallmark gene pathways with “Gene Set” as the permutation type. The permutation number was set at 1000, FDR values of less than 0.25 (or normalized p-value < 0.01 for pathways which did not pass FDR threshold) were considered significant. Leading edge genes were identified for each significantly enriched pathway, and their contribution to the signal was assessed and visualized by Z-score scaling. Heatmaps were generated using the pheatmap R statistical programming package.

*Gene Ontology (GO) and Reactome Functional Interaction (FI) network analysis.* GO analysis and functional enrichment in the molecular function, biological process and cellular component categories were performed using GOrilla application (65). Highly enriched “Chromatin-remodeling”, “Cellular response to DNA damage” and “Cell division” GO categories were further analyzed using the Reactome (<https://reactome.org>) database V71 and Reactome FIViz 2019 Cytoscape plugin (66), to identify network modules of highly connected genes and their significantly enriched (FDR-adjusted p-value < 0.05) biological processes.

*ChIPseq data analysis.* Publicly available ERG ChIPseq datasets (GSE28950) (9) were extracted as FASTQ files using Gene Expression Omnibus tools. The EWS::FLI1 ChIPseq data (46) were retrieved from the BAM files using Picard v.2.9.0 and converted to FASTQ files. The reads were mapped to hg19 (GRCh37.p13) reference assembly using Bowtie2 aligner (v.2.3.4.1). Peaks were called using MACS2 (v2.2.6) default settings, with  $Q \leq 0.05$  as the threshold. Unmapped reads were assembled into contigs using Trinity v.2.6.6. Quality of the assembled contigs was assessed with utility scripts from trinity suite and limited to the size of 700 bases or fewer. To establish their identities and locations, the contigs were aligned against the 24-mer *HSAT2,3* motif reference database (26) and RepeatMasker-annotated satellite repeats using BLAST tools.

*Data availability.* RNAseq datasets generated from plasma EVs of 12 EwS patients and 7 healthy donors, cell lines and their EVs were deposited at the Sequence Read Archive (SRA) under the accession number PRJNA548159. Values for all data points in graphs are reported in the Supporting Data Values file.

*Statistical analysis.* All statistical analyses were conducted using R statistical environment v4.3.2 (<https://www.R-project.org/>). Additional packages used for data analysis and visualization included Bioconductor (<https://www.bioconductor.org>) and CRAN (<https://cran.r-project.org/web/packages/>). Normal distribution of data in RNAseq and ddPCR datasets was confirmed with Shapiro-Wilk test. To

correct for multiple comparisons, the unpaired t-test on log-transformed values were adjusted using Benjamini-Hochberg (BH) test. For data that failed a normality test, the Wilcoxon signed-rank test was applied. For comparisons among more than two groups of samples, one-way ANOVA or Welch F test (in cases when homoscedasticity condition was violated) were applied. To check for associations between biochemical recurrence (BCR) rates and *HSAT2,3* expression, survival analysis was performed using the R Survival package (v3.5–7) with 0.1 as cutoff. Cox proportional hazards model was then fit using patients groups divided by median *HSAT2,3* abundance using BCR as endpoint. Data visualization was performed using the survminer R package (0.4.9).

*Study approval.* Blood plasma specimens were obtained from EwS patients enrolled in GenEwing (the EW cohort; Institute Curie, Paris, France) and NEO-IDENT (the TUM cohort; the Technical University of Munich, Germany) studies, and from age-matching healthy donors undergoing routine checkups at the Technical University of Munich. Each cohort received the approvals from the respective institutional review boards, Comité de Protection des Personnes, Ile-de-France (09-12037) and the Ethikkommission der Medizinischen Fakultät der Technischen Universität München (2562/09 and 649/20S). Protocols for clinical sample handling and processing were reviewed and approved by the University of Toronto (35129 and 36089).

## **Author Contributions**

P.R. provided computational and statistical analysis, data interpretation and figure preparation; V.E. designed the study, performed experiments and wrote the manuscript; M.C.P, Z.Z. and H.G. purified EVs and performed RT-ddPCR and immunoblotting; A.M. performed immunofluorescence microscopy; S.L. contributed to transfection experiments; H.G., U.T., E.L., O.D. and S.E.G.B. provided plasma samples and interpreted the data; M.O.M. performed Reactome data analysis; S.H.Z. and V.P. provided intellectual and technical support; M.S. and M.E.C. provided VCaP and ERG expressing prostate cells, conducted experiments and interpreted the data; T.D.M. and M.Z. conducted quantitative image spot analysis; L.E.H., J.D.M. and L.D.S. supervised RNAseq and data analysis; L.R., L.D.S., M.E.C., S.E.G.B. and P.H.S. secured funding, reviewed and edited the manuscript. All authors reviewed and approved the manuscript. The first co-authorship is shared between P.R. and V.E. based on their equal contribution and mutual agreement.

## **Acknowledgements**

We thank all patients and their families, and the personnel of the Genetic Somatic Unit and the Institute of Pathology for their help in collecting and bio-banking blood and tumor specimens. We thank B.D. Crompton and K. Stegmaier for sharing their RNAseq EwS tumor datasets, B. Schäfer for the EWS::FLI1

expressing plasmid, A.M. El-Naggar for mouse xenograft and K. Specht and K. Steiger for EwS tumor sections. This work was supported by funding from the Government of Ontario (to L.R. and L.D.S.), CIHR Foundation grant FDN-143280, the British Columbia Cancer Foundation through generous donations from Team Finn and other generous riders in the Ride to Conquer Cancer (to P.H.S.), Terry Fox Research Institution New Frontiers Program Project SRN 1062 (to M.E.C), Cura Placida Children's Cancer Research Foundation grants (to V.E., S.E.G.B. and P.H.S.), the St. Baldrick's Foundation (Martha's BEST Grant for All, 663113) and the 'Du musst kämpfen' Wohltätigkeitsinitiative (Access to Systems Biology-Based Individualized Cell Therapies for Adolescents with Refractory Pediatric Cancer, AdoRe, 200310stb, to S.E.G.B.). U.T. is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; # 501830041), the "Zukunft Gesundheit e.V." and the Dr. Sepp and Hanne Sturm Memorial Foundation. U.T. and H.G. are funded by the Dr. Robert Pflieger Foundation.

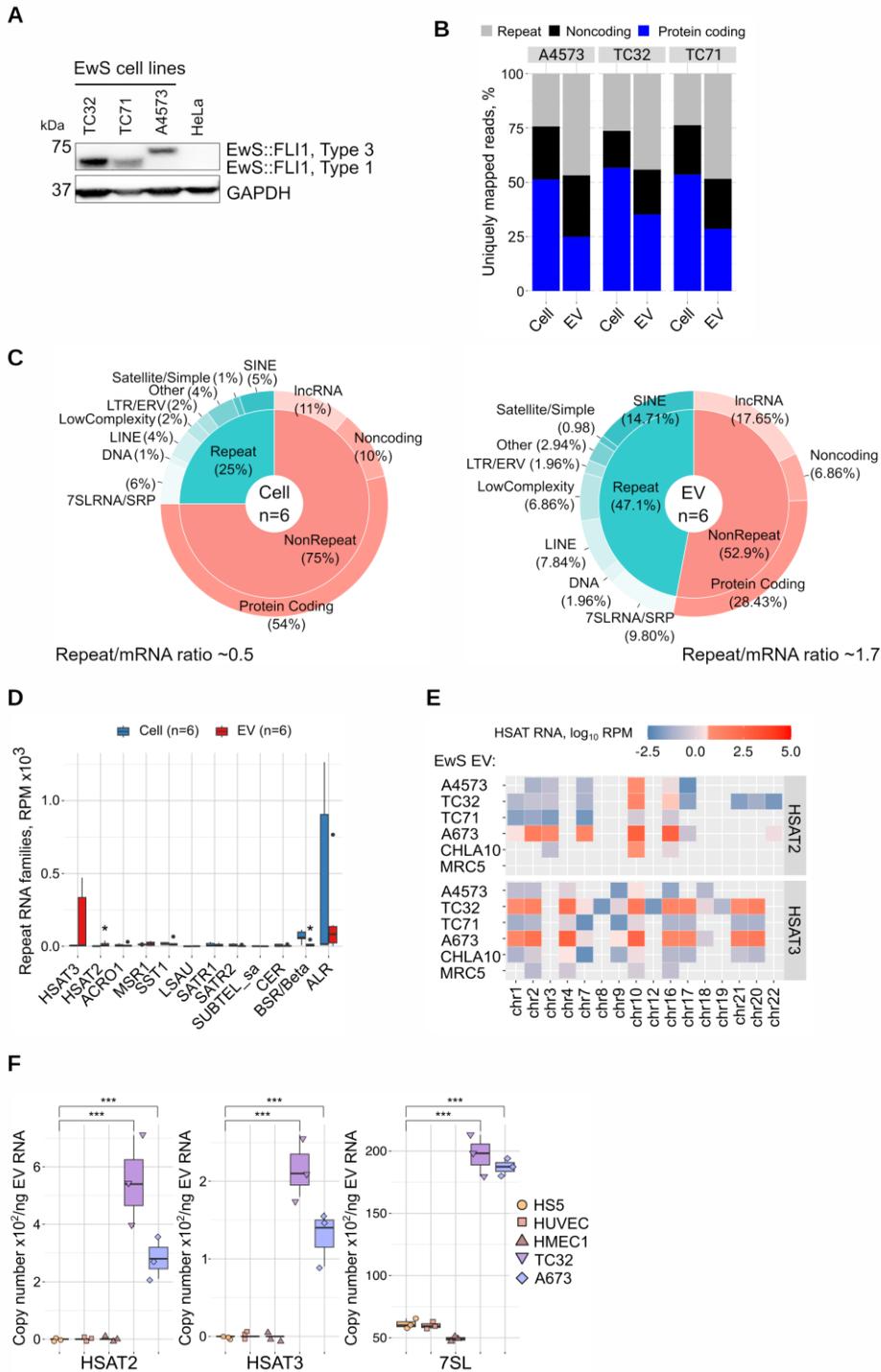
## REFERENCES

1. Grunewald TGP, Cidre-Aranaz F, Surdez D, Tomazou EM, de Alava E, Kovar H, Sorensen PH, Delattre O, and Dirksen U. Ewing sarcoma. *Nature reviews Disease primers*. 2018;4(1):5.
2. Riggi N, Suvà ML, and Stamenkovic I. Ewing's Sarcoma. *The New England journal of medicine*. 2021;384(2):154-64.
3. Sorensen PH, Lessnick SL, Lopez-Terrada D, Liu XF, Triche TJ, and Denny CT. A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. *Nat Genet*. 1994;6(2):146-51.
4. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, Rouleau G, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*. 1992;359(6391):162-5.
5. Sizemore GM, Pitarresi JR, Balakrishnan S, and Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer*. 2017;17(6):337-51.
6. Hollenhorst PC, McIntosh LP, and Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem*. 2011;80(437-71).
7. Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, Cao Q, Prensner JR, Rubin MA, Shah RB, et al. Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia*. 2008;10(2):177-88.
8. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, NY)*. 2005;310(5748):644-8.
9. Chng KR, Chang CW, Tan SK, Yang C, Hong SZ, Sng NY, and Cheung E. A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. *The EMBO journal*. 2012;31(12):2810-23.
10. Tomazou EM, Sheffield NC, Schmidl C, Schuster M, Schonegger A, Datlinger P, Kubicek S, Bock C, and Kovar H. Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell reports*. 2015;10(7):1082-95.
11. Orth MF, Surdez D, Faehling T, Ehlers AC, Marchetto A, Grossetête S, Volckmann R, Zwijnenburg DA, Gerke JS, Zaidi S, et al. Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation. *Cell reports*. 2022;41(10):111761.
12. Kedage V, Selvaraj N, Nicholas TR, Budka JA, Plotnik JP, Jerde TJ, and Hollenhorst PC. An Interaction with Ewing's Sarcoma Breakpoint Protein EWS Defines a Specific Oncogenic Mechanism of ETS Factors Rearranged in Prostate Cancer. *Cell reports*. 2016;17(5):1289-301.
13. Guillon N, Tirode F, Boeva V, Zynovyev A, Barillot E, and Delattre O. The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. *PLoS one*. 2009;4(3):e4932.
14. Gargi G YJ. EWS-FLI1 Regulates Genotoxic Stress Response in Ewing Sarcoma. *J Cancer Biol Res*. 2015;3(2):1063.
15. Burdak-Rothkamm S, Mansour WY, and Rothkamm K. DNA Damage Repair Deficiency in Prostate Cancer. *Trends Cancer*. 2020;6(11):974-84.
16. Gorthi A, Romero JC, Loranc E, Cao L, Lawrence LA, Goodale E, Iniguez AB, Bernard X, Masamsetti VP, Roston S, et al. EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature*. 2018;555(7696):387-91.
17. Evdokimova V, Gassmann H, Radvanyi L, and Burdach SEG. Current State of Immunotherapy and Mechanisms of Immune Evasion in Ewing Sarcoma and Osteosarcoma. *Cancers*. 2022;15(1).
18. Gassmann H, Schneider K, Evdokimova V, Ruzanov P, Schober SJ, Xue B, von Heyking K, Thiede M, Richter GHS, Pfaffl MW, et al. Ewing Sarcoma-Derived Extracellular Vesicles Impair Dendritic Cell Maturation and Function. *Cells*. 2021;10(8).

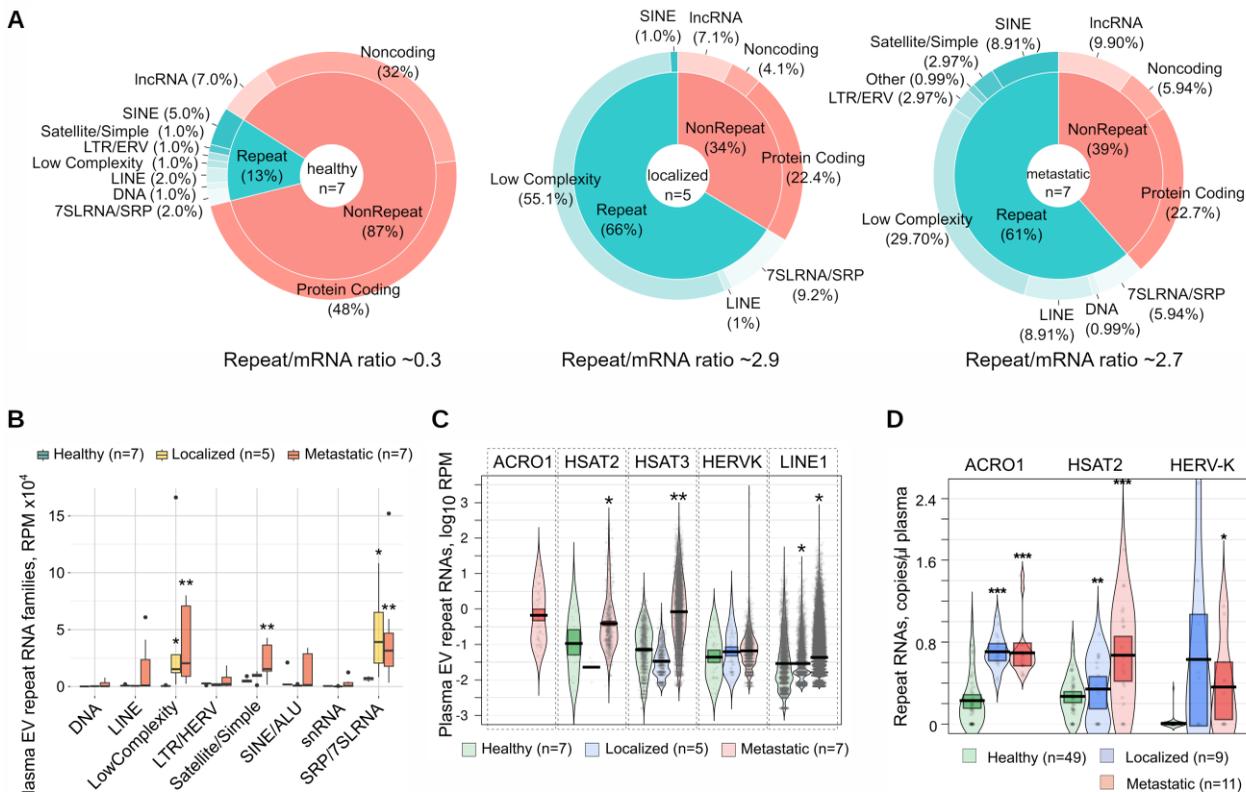
19. Evdokimova V, Ruzanov P, Gassmann H, Zaidi SH, Peltekova V, Heisler LE, McPherson JD, Orlic-Milacic M, Specht K, Steiger K, et al. Exosomes transmit retroelement RNAs to drive inflammation and immunosuppression in Ewing Sarcoma. *bioRxiv*. 2019:806851.
20. Thery C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, Antoniou A, Arab T, Archer F, Atkin-Smith GK, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. 2018;7(1):1535750.
21. Mateescu B, Kowal EJ, van Balkom BW, Bartel S, Bhattacharyya SN, Buzas EI, Buck AH, de Candia P, Chow FW, Das S, et al. Obstacles and opportunities in the functional analysis of extracellular vesicle RNA - an ISEV position paper. *J Extracell Vesicles*. 2017;6(1):1286095.
22. Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, Liebler DC, Ping J, Liu Q, Evans R, et al. Reassessment of Exosome Composition. *Cell*. 2019;177(2):428-45.e18.
23. Nabet BY, Qiu Y, Shabason JE, Wu TJ, Yoon T, Kim BC, Benci JL, DeMichele AM, Tchou J, Marcotrigiano J, et al. Exosome RNA Unshielding Couples Stromal Activation to Pattern Recognition Receptor Signaling in Cancer. *Cell*. 2017;170(2):352-66.e13.
24. Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, and Skog J. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nature communications*. 2011;2(180).
25. Porter RL, Sun S, Flores MN, Berzolla E, You E, Phillips IE, Kc N, Desai N, Tai EC, Szabolcs A, et al. Satellite repeat RNA expression in epithelial ovarian cancer associates with a tumor-immunosuppressive phenotype. *The Journal of clinical investigation*. 2022;132(16).
26. Altemose N, Miga KH, Maggioni M, and Willard HF. Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS Comput Biol*. 2014;10(5):e1003628.
27. Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, and Abrusan G. Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics*. 2008;9(533).
28. Altemose N. A classical revival: Human satellite DNAs enter the genomics era. *Seminars in cell & developmental biology*. 2022;128(2-14).
29. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, Kiezun A, Carter SL, Shukla SA, Mehta SS, et al. The genomic landscape of pediatric Ewing sarcoma. *Cancer discovery*. 2014;4(11):1326-41.
30. Hall LL, Byron M, Carone DM, Whitfield TW, Pouliot GP, Fischer A, Jones P, and Lawrence JB. Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. *Cell reports*. 2017;18(12):2943-56.
31. Landers CC, Rabeler CA, Ferrari EK, D'Alessandro LR, Kang DD, Malisa J, Bashir SM, and Carone DM. Ectopic expression of pericentric HSATII RNA results in nuclear RNA accumulation, MeCP2 recruitment, and cell division defects. *Chromosoma*. 2021;130(1):75-90.
32. García-Domínguez DJ, Hontecillas-Prieto L, León EA, Sánchez-Molina S, Rodríguez-Núñez P, Morón FJ, Hajji N, Mackintosh C, and de Álava E. An inducible ectopic expression system of EWSR1-FLI1 as a tool for understanding Ewing sarcoma oncogenesis. *PloS one*. 2020;15(6):e0234243.
33. Hall LE, Mitchell SE, and O'Neill RJ. Pericentric and centromeric transcription: a perfect balance required. *Chromosome Res*. 2012;20(5):535-46.
34. Becker-Santos DD, Guo Y, Ghaffari M, Vickers ED, Lehman M, Altamirano-Dimas M, Oloumi A, Furukawa J, Sharma M, Wang Y, et al. Integrin-linked kinase as a target for ERG-mediated invasive properties in prostate cancer models. *Carcinogenesis*. 2012;33(12):2558-67.
35. Chen S, Huang V, Xu X, Livingstone J, Soares F, Jeon J, Zeng Y, Hua JT, Petricca J, Guo H, et al. Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell*. 2019;176(4):831-43.e22.

36. Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, and Moyzis RK. Highly conserved repetitive DNA sequences are present at human centromeres. *Proceedings of the National Academy of Sciences of the United States of America*. 1992;89(5):1695-9.
37. Tanne A, Muniz LR, Puzio-Kuter A, Leonova KI, Gudkov AV, Ting DT, Monasson R, Cocco S, Levine AJ, Bhardwaj N, et al. Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(49):15154-9.
38. Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, MacKenzie OC, Brannigan BW, Wittner BS, Jung H, et al. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(49):15148-53.
39. El-Naggar AM, Veinotte CJ, Cheng H, Grunewald TG, Negri GL, Somasekharan SP, Corkery DP, Tirode F, Mathers J, Khan D, et al. Translational Activation of HIF1alpha by YB-1 Promotes Sarcoma Metastasis. *Cancer cell*. 2015;27(5):682-97.
40. Taiana E, Ronchetti D, Todoerti K, Nobili L, Tassone P, Amodio N, and Neri A. LncRNA NEAT1 in Paraspeckles: A Structural Scaffold for Cellular DNA Damage Response Systems? *Non-coding RNA*. 2020;6(3).
41. McCluggage F, and Fox AH. Paraspeckle nuclear condensates: Global sensors of cell stress? *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2021;43(5):e2000245.
42. Kwon J, and Bakhoun SF. The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer discovery*. 2020;10(1):26-39.
43. Hopfner KP, and Hornung V. Molecular mechanisms and cellular functions of cGAS-STING signalling. *Nature reviews Molecular cell biology*. 2020;21(9):501-21.
44. Mathavarajah S, and Dellaire G. LINE-1: an emerging initiator of cGAS-STING signalling and inflammation that is dysregulated in disease. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. 2023.
45. Du M, and Chen ZJ. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science (New York, NY)*. 2018;361(6403):704-9.
46. Bilke S, Schwentner R, Yang F, Kauer M, Jug G, Walker RL, Davis S, Zhu YJ, Pineda M, Meltzer PS, et al. Oncogenic ETS fusions deregulate E2F3 target genes in Ewing sarcoma and prostate cancer. *Genome Res*. 2013;23(11):1797-809.
47. Talbert PB, and Henikoff S. Transcribing Centromeres: Noncoding RNAs and Kinetochores Assembly. *Trends Genet*. 2018;34(8):587-99.
48. Ninomiya K, Yamazaki T, and Hirose T. Satellite RNAs: emerging players in subnuclear architecture and gene regulation. *The EMBO journal*. 2023;42(18):e114331.
49. Zhu Q, Hoong N, Aslanian A, Hara T, Benner C, Heinz S, Miga KH, Ke E, Verma S, Soroczynski J, et al. Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol Cell*. 2018;70(5):842-53.e7.
50. Samson N, and Ablasser A. The cGAS-STING pathway and cancer. *Nature cancer*. 2022;3(12):1452-63.
51. Webb LG, and Fernandez-Sesma A. RNA viruses and the cGAS-STING pathway: reframing our understanding of innate immune sensing. *Current opinion in virology*. 2022;53(10):1206.
52. Kassiotis G. The Immunological Conundrum of Endogenous Retroelements. *Annual review of immunology*. 2023;41(99-125).
53. Russ E, and Iordanskiy S. Endogenous Retroviruses as Modulators of Innate Immunity. *Pathogens (Basel, Switzerland)*. 2023;12(2).
54. Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, and Verma IM. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature*. 2011;477(7363):179-84.

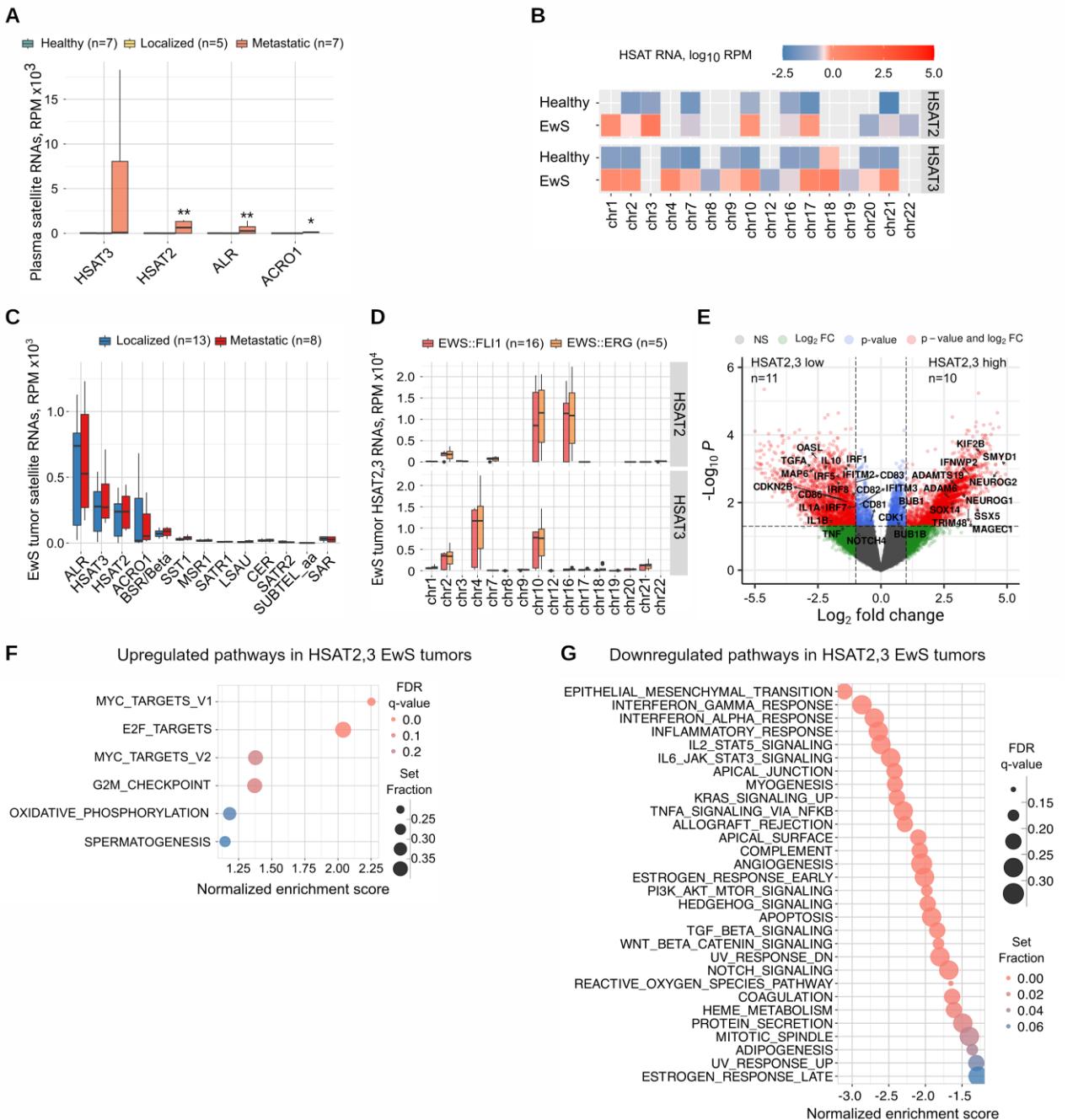
55. Tasselli L, Xi Y, Zheng W, Tennen RI, Odrowaz Z, Simeoni F, Li W, and Chua KF. SIRT6 deacetylates H3K18ac at pericentric chromatin to prevent mitotic errors and cellular senescence. *Nat Struct Mol Biol.* 2016;23(5):434-40.
56. Cruickshanks HA, McBryan T, Nelson DM, Vanderkraats ND, Shah PP, van Tuyn J, Singh Rai T, Brock C, Donahue G, Dunican DS, et al. Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol.* 2013;15(12):1495-506.
57. Solovyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, Rosenberg JE, Bhardwaj N, Ting DT, and Greenbaum BD. Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. *Cell reports.* 2018;23(2):512-21.
58. Ishak CA, Classon M, and De Carvalho DD. Deregulation of Retroelements as an Emerging Therapeutic Opportunity in Cancer. *Trends Cancer.* 2018;4(8):583-97.
59. Boro A, Prêtre K, Rechfeld F, Thalhammer V, Oesch S, Wachtel M, Schäfer BW, and Niggli FK. Small-molecule screen identifies modulators of EWS/FLI1 target gene expression and cell survival in Ewing's sarcoma. *Int J Cancer.* 2012;131(9):2153-64.
60. Alsubaie N, Trahearn N, Raza SE, Snead D, and Rajpoot NM. Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation. *PLoS one.* 2017;12(1):e0169875.
61. Dobin A, and Gingeras TR. Optimizing RNA-Seq Mapping with STAR. *Methods in molecular biology (Clifton, NJ).* 2016;1415(245-62).
62. Liao Y, Smyth GK, and Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England).* 2014;30(7):923-30.
63. Robinson MD, McCarthy DJ, and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England).* 2010;26(1):139-40.
64. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology.* 2014;15(12):550.
65. Eden E, Navon R, Steinfeld I, Lipson D, and Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10(48).
66. Wu G, and Haw R. Functional Interaction Network Construction and Analysis for Disease Discovery. *Methods in molecular biology (Clifton, NJ).* 2017;1558(235-53).



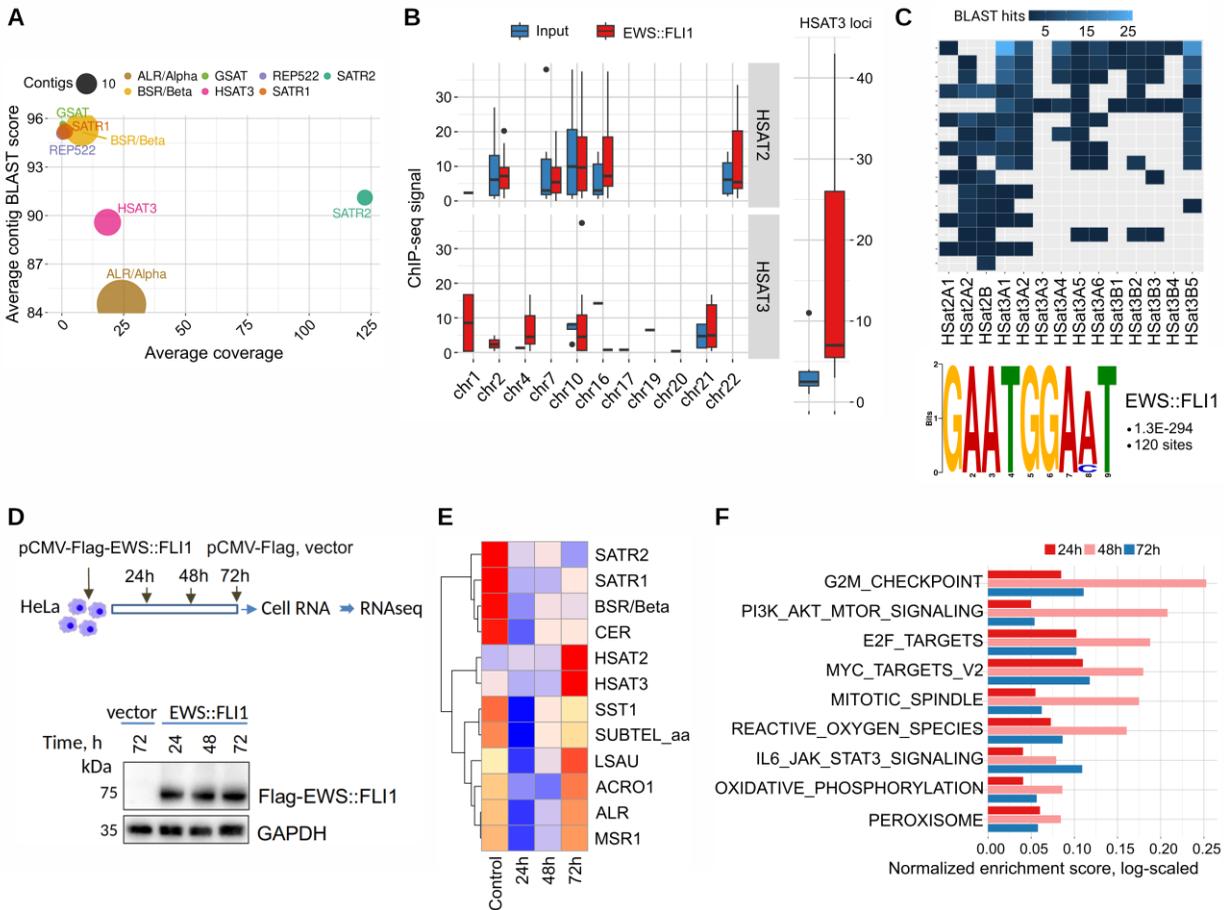
**Figure 1. EwS EVs are enriched with retroelement and pericentromeric RNAs.** (A) Immunoblotting analysis of the endogenous EWS::FLI1 protein expression in EwS and HeLa cell extracts using FLI1 antibodies and GAPDH as a loading control. (B) Percentage of RNAseq reads mapped to “Repeats” (RepeatMasker-annotated elements), “Noncoding” (pseudogenes, long and small ncRNAs), and “Protein-coding” (mRNAs, 5’ and 3’ UTRs, and introns) in each EwS cell line and the respective EVs from two biological replicates. (C) Pie charts representing averaged values for each repeat and non-repeat category detected in EwS cells and EVs from B. Comparisons made between repeat RNA content in EVs vs cells;  $p=0.032$ , Wilcoxon signed-rank test. (D) Representation of each satellite family member, shown as reads per million mapped reads (RPM). Comparisons EVs vs cells;  $*p<0.01$ , Wilcoxon test. (E) Heatmaps show chromosomal localization of HSAT2,3 RNAseq reads detected in EVs from EwS and MRC5 cell lines. Expression values were log<sub>10</sub>-scaled. Each square represents an average expression from two replicates. Gray color indicates no reads. (F) RT-ddPCR analysis of indicated RNAs in EVs from non-transformed (HS-5, HUVEC, HMEC-1) and EwS (TC32 and A673) cell lines. Data are mean  $\pm$  SEM; comparisons to HS-5 EVs,  $**p<0.01$ ,  $***p<0.001$ , two-way ANOVA.



**Figure 2. Satellite and retroelement RNAs are the most abundant species in blood plasma of EwS patients. (A)** RNAseq analysis of plasma EV RNAs from EwS patients (EW cohort; n=12) and age-matched healthy donors (n=7). Pie charts represent averaged values for each repeat and nonrepeat category. Comparisons made between EwS patients vs healthy donors;  $p = 0.012$ , Wilcoxon signed-rank test. **(B)** RepeatMasker-annotated families from **A** shown as mean RPM  $\pm$  SEM; EwS patients vs healthy donors,  $*p < 0.05$ ,  $**p < 0.01$ , Wilcoxon signed-rank test. **(C)** Violin plots of representative RNAs from **B**; EwS patients vs healthy donors, adjusted  $*p < 1.36e-11$ ,  $**p < 2.00e-41$ . Note that *ACRO1* was not detected in healthy donors and patients with localized disease. **(D)** RT-ddPCR analysis of indicated RNAs in plasma of EwS patients (TUM cohort, n=20) vs healthy donors (n=49); adjusted  $*p < 0.05$ ,  $**p = 0.0002$ ,  $***p < 2.00e-8$ . Variance between groups was analyzed with the Welch F test,  $p = 4.27e-12$  (*ACRO1*),  $p = 0.0038$  (*HSAT2*) and  $p = 0.0158$  (*HERV-Kenv*). P values in **C, D** were adjusted for multiple comparisons using Benjamini-Hochberg method.

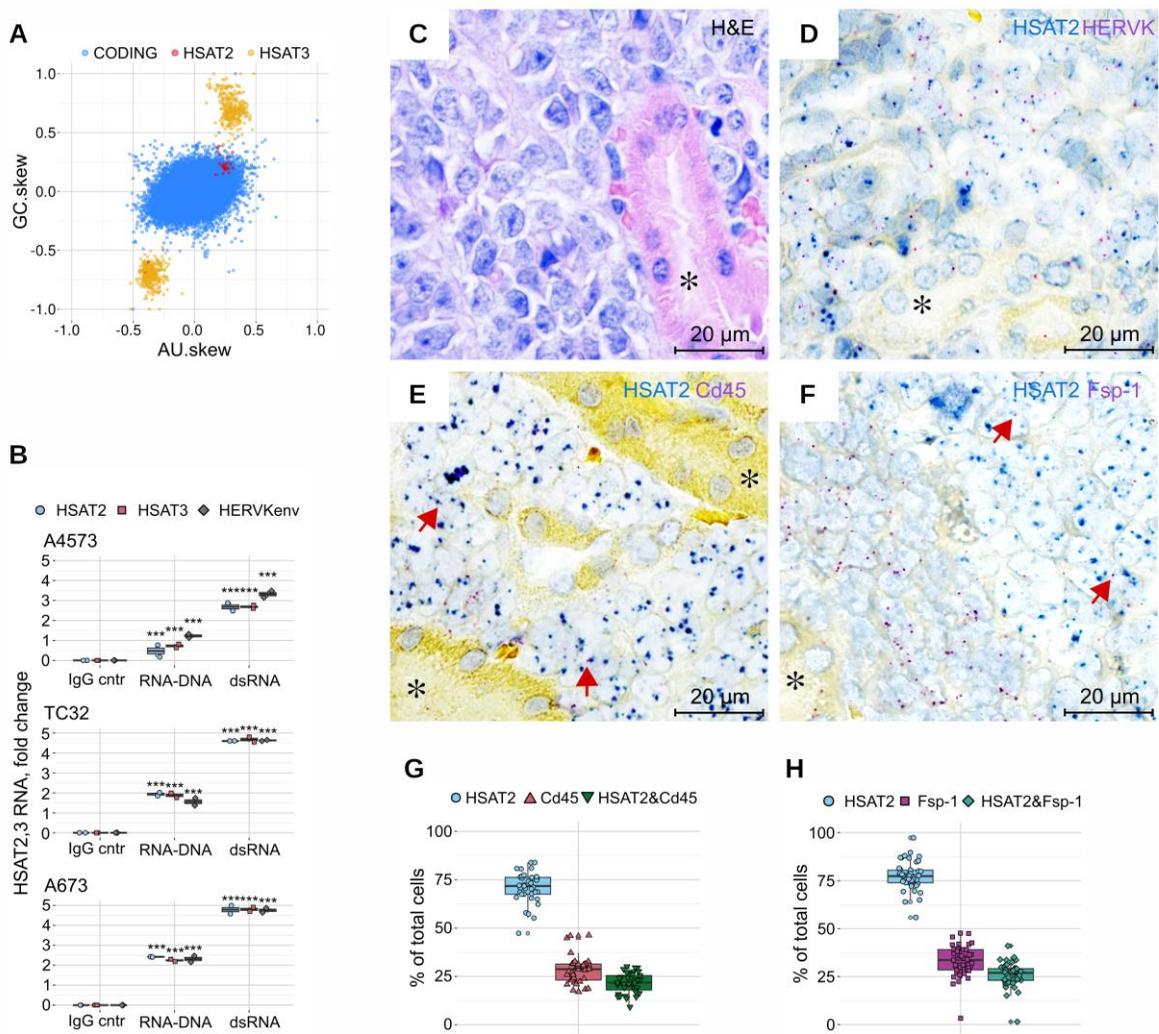


**Figure 3. Elevated *HSAT2,3* expression in EwS tumors is associated with upregulation cell cycle checkpoint and downregulation of immune response genes.** (A) Representation of satellite RNAs in plasma of EwS patients vs healthy donors. Data are mean RPM  $\pm$  SEM; \* $p < 0.05$ , \*\* $p < 0.01$ , Wilcoxon signed-rank test. (B) Chromosomal distribution of uniquely mapped *HSAT2,3* reads in plasma of patients with localized or metastatic disease. Cumulative read counts are shown. (C-G) Analysis of EwS tumor datasets ( $n=21$ ). Representation of satellite RNAs (C) and chromosomal distribution of uniquely mapped *HSAT2,3* RNAseq reads in EwS tumors positive for EWS::FLI1 or EWS::ERG (D). Data in C and D are mean RPM  $\pm$  SEM. (E) Volcano plot,  $\log_2$  fold change vs  $-\log_{10}$  p-value of differentially expressed genes in *HSAT2,3*-high vs -low EwS tumors. Red dots represent genes which passed  $p \leq 0.05$  thresholds and changed  $> 2$ -fold. (F, G) GSEA of up- or down-regulated pathways in *HSAT2,3*-high EwS tumors, ranked based on  $\log_2$  fold change from E. Circle size indicates gene set size, and circle color indicates the FDR adjusted q value.

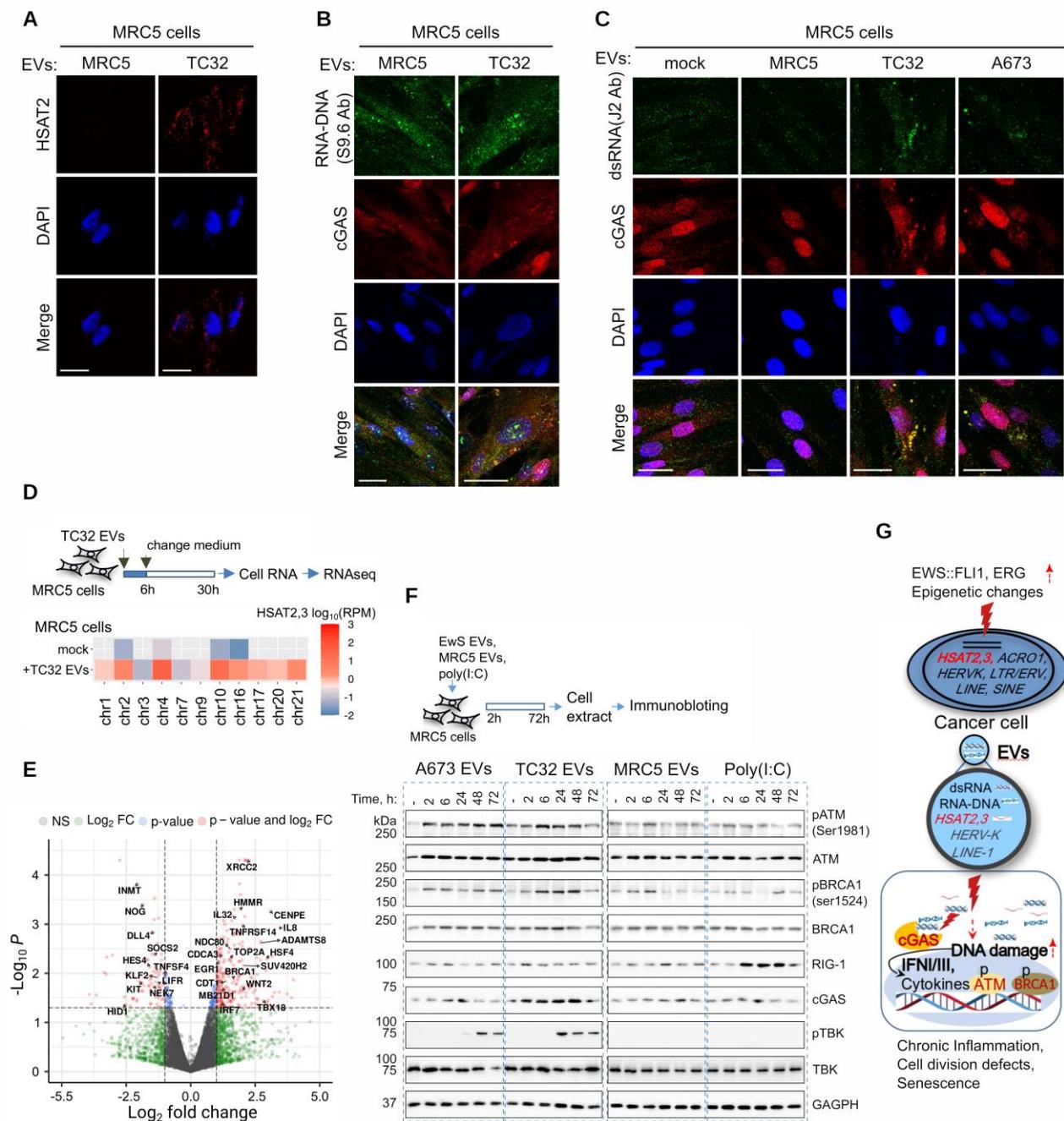


**Figure 4. EWS::FLI1 preferentially binds to *HSAT3* pericentromeric chromatin and activates *HSAT2,3* expression in non-Ews cells.** (A) Detection of EWS::FLI1 binding sites in centromeric regions by alignment with the 200-700-nt contigs assembled from the unmapped EWS::FLI1 ChIPseq reads; number of binding peaks overlapping a particular repeat element is reflected by bubble size. Peaks of highest quality (average MACS2 peak score > 5) are shown. (B) Chromosomal distribution of the *HSAT2,3* contigs (left panel) and total number of the identified *HSAT3* loci (right panel) from the EWS::FLI1 ChIPseq dataset vs input fractions ( $p = 0.0037$ , Wilcoxon rank-sum test). Contigs with alignment identity  $\geq 95\%$  were analyzed. (C) Identification of EWS::FLI1 binding motifs within pericentromeric regions by alignment to the 24-mer *HSAT2,3* motif libraries (top) and the highest-scoring recurring sequence motif identified by MEME tools (bottom). (D) Outline of the HeLa transient transfection experiment (top) and immunoblotting using Flag antibodies and GAPDH as a loading control (bottom). (E) Heatmap of satellite RNA expression; color indicates normalized RPM values scaled and centered by row. Each time point is average value from two biological replicates. (F) Bar plot of pathway enrichment in EWS::FLI1 expressing vs vector control HeLa cells. Log-scaled normalized enrichment score for the pathways which passed 0.25 FDR threshold is shown.





**Figure 6. *HSA2,3* RNAs exhibit pathogen-like features and are disseminated in the tumor microenvironment. (A)** CpG or UpA motif usage in *HSA2,3* RNAs compared to coding transcriptome. The GC and AU skew is calculated as  $(A-U)/(A+U)$  and  $(G-C)/(G+C)$  ratio, and the results are plotted in R using ggplot2 package. **(B)** RT-qPCR analysis of RNAs immunoprecipitated from the CM using IgG control, S9.6 (RNA-DNA) and J2 (dsRNA) antibodies. Each bar represents two technical replicates of one of two independent experiments. Values normalized to the exogenously added spike-in *MS2* RNA  $\pm$  SD. Fold change relative to IgG control; unpaired t-tests with BH adjustment, \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . **(C)** Hematoxylin and eosin (H&E) staining of mouse renal subcapsular TC32 EwS tumor xenografts **(D-F)** ViewRNA-ISH staining with the indicated probes, counterstained with hematoxylin. Red arrows, co-localization of human *HSA2* (blue dots) with mouse *Fsp-1* or *Cd45* (purple dots) in a single cell. Asterisks, adjacent normal mouse kidney parenchyma. **(G, H)** Data quantification from **E, F** using Developer XD machine learning software. Data are mean  $\pm$  SD. Scale bars, 20  $\mu$ m.



**Figure 7. *HSAT* RNA-enriched EwS EVs activate innate immune and DNA damage pathways in the recipient cells. (A)** ViewRNA-FISH with *HSAT2* probes and **(B, C)** immunofluorescence imaging of MRC5 cells treated with TC32 or MRC5 EVs for 24h, counterstained with DAPI. Scale bars, 10  $\mu$ m. **(D)** Experimental outline and heatmap of chromosomal distribution of *HSAT2,3* RNAseq reads from MRC5 cells treated with PBS (mock) or TC32 EVs. Gray color indicates no reads. **(E)** Volcano plot,  $\log_2$  fold change vs  $-\log_{10}$  p-value of differentially expressed genes in TC32 EV-treated MRC5 cells vs mock. Red dots show genes which passed  $p \leq 0.05$  thresholds and changed  $>2$ -fold. **(F)** Immunoblotting pathway analysis of MRC5 cells treated with mock (-), EwS or MRC5 EVs, or with 5 ng/ml poly(I:C). **(G)** A proposed model of pericentromeric chromatin activation in EwS and PCa, and the effects on stromal cells. Dissemination of *HSAT2,3* and other pathogen-like repeat RNAs in cancer EVs and their accumulation in stromal fibroblasts and immune cells induces DNA damage and cGAS-pTBK1 signaling. Prolonged activation of these pathways due to continuous exposure to cancer EVs, ongoing reverse transcription or unresolved DNA damage may lead to local and systemic chronic inflammation, mitotic defects and senescence.