# Phospho-IWS1-dependent *U2AF2* splicing is cell-cycle-regulated, promotes proliferation and predicts poor prognosis of EGFR-mutant lung adenocarcinoma.

Georgios I. Laliotis[1,2,3,4,16], Evangelia Chavdoula[1,2], Maria D. Paraskevopoulou[5,6], Abdul D. Kaba[1,2], Alessandro La Ferlita[1,2,7], Satishkumar Singh[8], Vollter Anastas[1,2,9], Salvatore Alaimo[7], Arturo Orlacchio[1,2], Keith A. Nair II[1,2], Vasiliki Taraslia[4,10], Ioannis Vlachos[11,12], Marina Capece[1,2], Artemis Hatzigeorgiou[11], Dario Palmieri[1,2], Christos Tsatsanis[13,14], Lalit Sehgal[6], David P. Carbone[15], Vincenzo Coppola[1,2] and Philip N. Tsichlis[1,2,16]

[1]The Ohio State University, Department of Cancer Biology and Genetics, Columbus, OH, 43210, USA, [2]The Ohio State University Comprehensive Cancer Center-Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, Columbus, OH, 43210, USA, [3]University of Crete, School of Medicine, Heraklion Crete, 71500, Greece, [4]Sidney Kimmel Comprehensive Cancer Center and Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD 21205 USA, [5]Molecular Oncology Research Institute, Tufts Medical Center, Boston, MA, 02111, USA, [6]Takeda Pharmaceuticals U.S.A, Boston, MA, 02139, USA, [7]Department of Clinical and Experimental Medicine, Bioinformatics Unit, University of Catania, Catania, 95131, Italy, [8]College of Medicine, Department of Hematology, The Ohio State University, Columbus, OH 43210, [9]Tufts Graduate School of Biomedical Sciences, Program in Genetics, Boston, MA, 02111, USA, [10]Center of Basic Research, Biomedical Research Foundation of the Academy of Athens, 11527, Greece, [11]DIANA-Lab, Hellenic Pasteur Institute, Athens, 11521, Greece, [12]Beth Israel-Deaconess Medical Center, Department Of Pathology, Harvard Medical School, Boston, MA 02215, USA, [13]Department of Clinical Chemistry-Biochemistry, School of Medicine, University of Crete, Heraklion 71110, Crete, Greece, [14]Institute for Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas, Heraklion 70013, Greece, [15]Department of Internal Medicine, Division of Medical Oncology, The Ohio State University Medical Center, Columbus, OH, 43210, USA



Running Title: IWS1 and *U2AF2* RNA splicing in EGFR mutant lung adenocarcinoma.

[16]These authors jointly supervised this work : Philip N. Tsichlis (Philip.tsichlis@osumc.edu) (lead contact) and Georgios I. Laliotis (glaliot7@gmail.com)

**Abstract**

Our previous studies have shown that IWS1 is a phosphorylation target of AKT in human lung cancer. RNA-seq studies using lung adenocarcinoma cells lacking IWS1, identified an exon 2 deficient splice variant of *U2AF2*, affecting U2AF65 RS domain and its binding with Prp19. This depends on phosphorylated IWS1 and the assembly of LEDGF/SRSF1 splicing complexes, via H3K36 trimethylation through the cell cycle. Inhibition of the pathway results in Sororin downregulation, a target and regulator of ERK, G2/M phase arrest, impaired cell proliferation and tumor growth, more pronounced in *EGFR* mutant cells. Analysis of lung adenocarcinoma samples revealed that the IWS1 phosphorylation pathway correlates with advanced stage, relapse, metastasis and poor survival in patients harboring *EGFR* mutations. This work highlights the instrumental role of the AKT/p-IWS1 axis to the epigenetic regulation of RNA processing and oncogenic signals, proposing it as a drug target in *EGFR* mutant lung adenocarcinoma.

## Introduction

Cells use a plethora of molecular mechanisms to precisely regulate mRNA expression (Glisovic T. et al., 2008[1]). These co- and post-transcriptional processes are orchestrated by a complex RNA-binding protein network. (Jewer M. et al., 2012[2]). One of the main mechanisms of this regulation is alternative RNA splicing. Most human genes harbor introns which are removed during pre-mRNA splicing (Pan Q. et al., 2008[3]), a common mechanism employed by eukaryotic cells to generate multiple transcripts and expand their functional network of the generated protein products. Thus, regulation of alternative RNA splicing affects cellular fate and function and controls the pathobiology of many diseases, including cancer (Paronetto M.P. et al., 2016[4], Zhang X. et al., 2016[5]). Its importance has been demonstrated in different types of cancer, including non-small cell lung cancer (NSCLC) (Coomer A.O. et al., 2019[6]).



Non-small cell lung cancer (NSCLC) is the 2nd most common cancer, with more than 250,000 new cases per year in the US and extremely poor prognosis with <5% 5-year survival (Siegel et al., 2020[7]). NSCLC is further divided into lung adenocarcinoma, squamous cell carcinoma and large cell carcinoma based on histological features. (Yuan M et al, 2019[8]). Notably, up to 69% of NSCLCs harbor mutations in the Epidermal Growth Factor Receptor (*EGFR*) and Kirsten Rat Sarcoma (*KRAS*) genes, affecting major downstream oncogenic signal such as the PI3K/AKT and ERK/MAPK pathway (Sequist et al., 2020[9]).

Following our original observations linking the alternative RNA splicing of *FGFR2* with NSCLC (Sanidas et al., 2014[10]), several additional splicing events have been linked to the biology of NSCLC including Bcl-$X_L$, CD44, Androgen Receptor (AR), HLA-G and PKM, by promoting cell survival, metastasis, chemoresistance, immune surveillance and metabolic advantage, respectively (Li Z. et al., 2016[11], Todaro M. et al., 2014[12], Oltean S. et al., 2014[13],

Calabretta S. et al., 2016[14]). Furthermore, another mechanism utilized by lung cancer cells involves the re-programming of alternative RNA splicing through the introduction of de novo mutations which create new splice sites and alternatively spliced isoforms (Vo N.S. et al., 2018[15], Giaj Levra M. et al., 2014[16]). This knowledge has led to an ongoing phase I clinical trial to determine the effects of a modulator of the core splicing function, H3B-8800, in cancer patients (Seiler M., 2018[17]). Thus, deeper comprehension of the molecular regulation of RNA splicing in lung cancer, and other cancer types, is a necessity for proper design of anti-tumor therapies.

Mechanistically, alternative RNA splicing is regulated co-transcriptionally and through chromatin modifications in the body of transcribed genes. When the rate of transcription is low, it increases the probability for exons that are not efficiently spliced, to be spliced out of the mature transcript (Listerman et al., 2012[18]). Chromatin modifications are recognized by readers of epigenetic marks, which orchestrate the assembly of molecular complexes that bind to, and functionally regulate RNA-associated enhancers or repressors of splicing (*cis*-acting RNA sequences). This alters the rate of assembly and the composition of spliceosomal complexes (*trans*-acting RNA Binding Proteins-RBPs) (Luco et al., 2010[19], Pradeepa et al., 2012[20], Oltean S. et al., 2014[10]). Cancer cells utilize this machinery in order to manipulate alternative RNA splicing and promote tumorigenesis. Well studied families of RBPs include the Serine/Rich (SR proteins) and heterogeneous nuclear ribonucleoproteins (hNRNPs). (Chen M. et al., 2009[21]). SR proteins bind RNA exonic splicing enhancers (ESE) and intronic splicing enhancers (ISE) and usually promote exon inclusion. (Obeng et al., 2019[22]). On the contrary, hnRNPs bind to exonic splicing silencers (ESS) and intronic splicing silencers (ISS) and in most cases lead to exon skipping (Martinez-Contreras et al., 2007[23]). Several reports have shown the role of these RBPs in the regulation of alternative RNA splicing and tumor growth in cancer cells, including Serine Rich Specific Factor 1

(SRSF1) and Polypyrimidine Tract Binding Protein (PTBP1) (de Miguel et al., 2014[24], Chen M et al., 2012[25], Jin W et al., 2003[26]).

One cellular function, which interfaces with the RNA splicing machinery, is the cell cycle regulation (Dominguez et al., 2016[27]). Progression through the cell cycle depends on periodic changes of RNA metabolism, which can be achieved by multiple mechanisms, one of which is the periodic modulation of RNA splicing (Dominguez et al., 2016[27]). On the other hand, alterations in RNA splicing, are regulated by periodic shifts in the expression and the activity of known cell cycle regulators, such as the Aurora kinases (Moore et al., 2010[28]), the large RS domain-containing protein SON (Ahn et al., 2011[29]), the RNA recognition motif (RRM)-containing protein TgRRM1 (Suvorova et al, 2013[30]) and CLK1 (Dominguez et al., 2016[27]), demonstrating the bidirectional relationship of these cellular functions.

We have previously shown that the transcription elongation factor IWS1 and the AKT3 kinase play a central role in the regulation of the alternative RNA splicing of *FGFR2*, by promoting the exclusion of exon IIIb. This exclusion event depends on the phosphorylation of IWS1 by AKT3 at Ser720/Thr721. IWS1 binds SPT6, which is attached to the phosphorylated Ser2 of the C-terminal domain (CTD) of RNA polymerase II. SETD2, a Histone H3 trimethyl-transferase, binds to the IWS1/Spt6 complex only when IWS1 is phosphorylated at Ser720/Thr721. SETD2 bound to the phosphorylated and CTD-associated IWS1, trimethylates histone H3 at K36 during transcription. The Histone H3K36me3 marks in the body of actively-transcribed target *FGFR2* gene are recognized by MRG15, which interacts with PTB, leading to exclusion of exon IIIb in the mature *FGFR2* mRNA transcript, affecting cell invasion, migration, cell proliferation and tumor growth in mouse xenografts models. More importantly, we had shown robust expression of IWS1 phosphorylation in 21 out of 24 NSCLC samples, which correlates with AKT phosphorylation, introducing the role of IWS1 in lung cancer (Sanidas et al., 2014[10]).

Here, to gain insight on the role of IWS1 phosphorylation in the pathogenesis of lung cancer and the epigenetic regulation of alternative RNA splicing genome-wide, we performed RNA-seq studies in lung adenocarcinoma cells NCI-H522, in which IWS1 was knocked down or replaced by its phosphorylation site mutant. The results revealed an RNA splice variant of the splicing factor U2 Associated Factor A2 (*U2AF2*), which lacks exon 2 upon loss of phosphorylated IWS1. U2AF65, its protein product, as a splicing factor binds to the Polypyrimidine (Py)-tract/3' splice site and initiates spliceosome assembly, by promoting the interaction between U2 snRNP and the branchpoint (Shen H. et al., 2004[31]). Several reports have illustrated the role of U2AF65 in lung cancer. Specifically, cancer-associated mutations of *U2AF2* in lung cancer patients alter the affinity of its RNA-Binding Domain with Py-tract (Glasser E. et al., 2017[32]). Furthermore, it has been shown that stabilization of U2AF65 by OTUB2 promotes Warburg effect and tumorigenesis in NSCLC, pointing out the importance of this splicing factor in NSCLC. (Li J. et al., 2018[33]). In this report, our goal was to further investigate the molecular mechanisms that control exon inclusion phenotype induced by IWS1 phosphorylation, to characterize this splice variant and to highlight its significance in lung carcinogenesis.



## Results

### IWS1 expression and phosphorylation regulate alternative mRNA splicing.

We have previously reported that IWS1 phosphorylation at Ser 720/Thr721, primarily by AKT3, resulted in the exclusion of exon IIIb of *FGFR2* in the human NSCLC cell lines NCI-H522 and NCI-H1299 (Sanidas et al., 2014[10]). To explore the molecular mechanisms driving the IWS1 phosphorylation-dependent RNA splicing and gene expression, we performed RNA-Seq to examine the transcriptome of shControl, shIWS1, shIWS1/wild type IWS1 rescue (shIWS1/WT-R) and shIWS1/phosphorylation site IWS1 mutant rescue

(shIWS1/MT-R) NCI-H522 cells. To further address the role of IWS1 phosphorylation in the regulation of RNA processing, alongside with these samples, we also engineered shIWS1/IWS1 phosphomimetic mutant S720D/T721E rescue (shIWS1/IWS1 DE MT-R) NCI-H522 and NCI-H1299 cells (Sanidas et al., 2014[10]), which were used to validate the putative results of the RNA-seq experiment. Notably, all these rescue clones are engineered to be shIWS1-resistant (Sanidas et al., 2014[10]). First, we confirmed the significant down-regulation of IWS1 expression after the transduction of the cells with lentiviral shIWS1 construct and the rescue of IWS1 knock-down with the Flag-tagged wild type, phosphorylation IWS1 site mutant and phosphomimetic IWS1 mutant (**Fig. 1a**). Differential expression analysis of the RNA-seq data, identified 1,621 and 562 differentially expressed genes between shControl and shIWS1 and between shIWS1/WT-R and shIWS1/MT-R cells, respectively ($p \leq 0.01$, FDR ≤0.2). 340 genes were detected as differentially expressed in both comparisons, following the same direction in their expression (**Fig. S1a, S1b**). Moreover, 19 out of the FDR-ranked top 100 differentially expressed genes, in shControl versus shIWS1 cells, were also differentially expressed in shIWS1/WT-R versus shIWS1/MT-R, with all of them being regulated similarly in both comparisons (up- or down-regulated) (**Fig. S1c**). Gene Set Enrichment analysis (Subramanian et al., 2005[34]), revealed significant enrichment of genes involved in RNA metabolism and regulation of RNA processing (**Fig. S1d**).

Using an unbiased approach, we performed differential exon usage analysis of the RNA-seq data, DEXseq (Anders et al., 2012[35]), to determine the effect of IWS1 expression and phosphorylation on RNA splicing. This analysis identified 1,434 (corresponding to 851 genes) and 436 (corresponding to 273 genes) differentially employed exons between shControl versus shIWS1 and shIWS1/WT-R versus shIWS1/MT-R cells, respectively ($p \leq 0.05$). The 1,796 differentially expressed genes and the 692 genes with differential exon usage in shIWS1 versus shControl cells, exhibited an overlap of 165 genes ($p \leq 0.05$). Similarly, the 858 differentially expressed genes and the 230 genes with differential exon

usage, between shIWS1/MT-R and shIWS/WT-R cells revealed an overlap of 44 genes (**Fig. 1b**).

Our earlier studies have shown that the IWS1 expression and phosphorylation promote exon exclusion in the *FGFR2* gene (Sanidas et al., 2014[10]). Therefore, we analysed the exon usage data, and we observed that the most common event associated with the expression and phosphorylation of IWS1, was exon inclusion (**Fig. 1c).** Functional analysis of the IWS1-regulated alternative spliced genes, focused on GO-biological processes, identified significant enrichment of genes involved in RNA processing and RNA splicing, among the top biological processes regulated by IWS1-dependent alternative RNA splicing (**Fig. 1d, 1e**). It is remarkable that the results were similar with the results generated from the functional analysis of the IWS1-dependent differential gene expression. Therefore, these findings imply that the effect of IWS1 on RNA processing may be direct or indirect. The indirect effect may be due to the IWS1 expression and phosphorylation-dependent differential regulation of genes involved in RNA processing.



Validation of the RNA-seq, using RT-PCR, showed several examples of alternative RNA splicing of genes which are regulated by IWS1 and IWS1-phosphorylation, via exon inclusion (**Fig. 1, S1g-S1j**). One of these events is the inclusion of exon 2 in the mature mRNA transcript of *U2AF2*, the gene encoding the splicing factor U2AF65. Whereas the predominant *U2AF2* mRNA transcript in shControl and shIWS1/WT-R cells contains exon 2, the predominant transcript in shIWS1 and shIWS1/MT-R cells, is a transcript lacking exon 2. To validate these results, we examined the ratio of exons 2 and 3 (E2/E3 *U2AF2*) by RT-PCR in NCI-H522 and NCI-H1299 shControl, shIWS1, shIWS1/WT-R, shIWS1/MT-R and shIWS1/DE MT-R cells. The results revealed a decrease in the E2/E3 ratio in shIWS1 and shIWS1/MT-R cells, relative to shControl cells, which was rescued in shIWS1/DE MT-R (**Fig. 1f, 1g**). The decrease in the E2/E3 ratio in shIWS1 and shIWS1/MT-R, relative to shControl cells, and the subsequent rescue of the ratio in shIWS1/DE MT-R cells, was confirmed by

quantitative RT-PCR (**Fig. S1e upper panel**). In parallel experiments, qRT-PCR in the same cells revealed increased IIIb/IIIc *FGFR2* ratio in shIWS1 and shIWS1/MT-R cells, relative to shControl cells, which was rescued in shIWS1/DE MT-R, confirming our previous findings on the effect of IWS1 phosphorylation on *FGFR2* alternative RNA splicing (Sanidas et al., 2014[10]) (**Fig. S1e lower panel).** Notably, the knockdown of IWS1 or the rescue with the non-phosphorylatable site mutant (shIWS1/MT-R) did not significantly change the total expression of *U2AF2* or the inclusion of *U2AF2* exon 3 in NCI-H522 and NCI-H1299 cells. (**Fig. S1f**).

To establish whether IWS1 is directly involved in the molecular complex that regulates *U2AF2* RNA splicing, in NCI-H522 and NCI-H1299 cells transduced with shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R, we performed chromatin immunoprecipitation (ChIP) assays. The results revealed that both IWS1 WT and S720A/T721A bind equally well to exon 2 and exon 3 of *U2AF2* (**Fig. 1h**). Therefore, IWS1 phosphorylation controls *U2AF2* exon 2 alternative RNA splicing by regulating events occurring after the binding of IWS1 to chromatin.



**IWS1 phosphorylation-dependent mRNA splicing of *U2AF2* is regulated by serum and IGF-1 via AKT3.**

IWS1 is phosphorylated by AKT3, and to a lesser extent by AKT1, at Ser720/Thr721 (Sanidas et al., 2014[10]). Since the phosphorylation and activity of the AKT isoforms can be induced by tyrosine-kinase signals, we questioned whether IGF-1 stimulation of serum starved NCI-H522 and NCI-H1299 cells, promotes *U2AF2* exon 2 inclusion along with the expected AKT activation and IWS1 phosphorylation. The results showed that the E2/E3 *U2AF2* and IIIb/IIIc *FGFR2* ratio, which was used as a control, indeed parallels the activation of AKT and the subsequent phosphorylation of IWS1 (**Fig. 2a, S2a**). To address whether the *U2AF2* exon 2 inclusion depends on AKT activation, we treated NCI-H522 and

NCI-H1299 cells growing in serum-containing complete media with 5 μM of the pan-AKT inhibitor MK2206, a dose that fully inhibits all AKT isoforms (Sanidas et al., 2014[10]) .The results confirmed that MK2206 inhibits both AKT (T308 and S473) and IWS1 phosphorylation (S720) and *U2AF2* exon 2 inclusion (**Fig. 2b, S2b upper panel).** Consistently, the same treatment promoted the inclusion of exon IIIb in *FGFR2* mRNA (**Fig. S2b lower panel**). To determine whether it is the AKT3 isoform, which is responsible for the observed effects of AKT inhibition on the *U2AF2* alternative RNA splicing, we transduced NCI-H522 and NCI-H1299 cells with shAKT1, shAKT2 and shAKT3 along with shControl, and we examined its effects on the alternative RNA splicing of IWS1 targets, *U2AF2* exon 2 and *FGFR2* exon IIIb, using RT-PCR and qRT-PCR. Interestingly, the results showed that inhibition of AKT1 and AKT3 phenocopied the treatment with MK2206 (**Fig. 2c, S2c)**. Notably, the effect of AKT1 inhibition was less pronounced, as expected, given that IWS1 is primarily phosphorylated by the AKT3 kinase. These data provide evidence that the physiological regulation of alternative RNA splicing of *U2AF2* by external stimuli depends on the phosphorylation of IWS1 by AKT3, and to a lesser extent, by AKT1.



**_U2AF2_ Exon 2 inclusion, induced by IWS1 phosphorylation at Ser720/Thr721, depends on histone H3K36 trimethylation by SETD2.**

We have previously reported that IWS1 phosphorylation by AKT promotes the exclusion of exon IIIb from the mature *FGFR2* mRNA transcript, via a process that depends on Histone H3K36 trimethylation by SETD2, and that the latter is recruited to the CTD of RNA Pol II by phosphorylated IWS1 (Sanidas et al., 2014[10]). To determine whether the *U2AF2* exon 2 inclusion phenotype is also dependent on Histone H3K36 trimethylation, we performed ChIP assays in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells, addressing the abundance of H3K36me3 marks on exons 2 and 3 of *U2AF2*, as well as on the *U2AF2* transcriptional start site (TSS) and the *GAPDH* gene, which

were used as controls. The results revealed that the knockdown of IWS1 or the replacement with the phosphorylation-site mutant significantly decreased the abundance of H3K36me3 on *U2AF2* exon 2 and 3 (**Fig. 3a**). Parallel ChIP assays showed the decreased H3K36me3 abundance on *FGFR2* exons IIIb and IIIc in the same cells, which was used as control (**Fig. S3a**). More importantly, inhibition of AKT, using MK2206, phenocopied the H3K36me3 pattern of phosphorylation-deficient IWS1 mutant (**Fig. 3b**). Since IWS1-phosphorylation recruits SETD2 in C-terminal domain of RNA-pol II, we examined the binding of SETD2 on *U2AF2, FGFR2* and *GAPDH* gene in the same NCI-H522 and NCI-H1299 cells, transduced using a SETD2 lentiviral construct, with a Hemagglutinin (HA)-tag. The results showed a pattern of SETD2 binding in both IWS1 splicing targets *U2AF2* and *FGFR2*, which parallels the IWS1 phosphorylation-dependent pattern of H3K36me3 abundance (**Fig. 3c, Fig. S3b**), suggesting that H3K36 trimethylation is mediated by SETD2.



The preceding data suggested that the enzymatically active SETD2 contributes to the IWS1 phosphorylation-dependent regulation of the *U2AF2* alternative RNA splicing. The role of catalytically-active SETD2 was confirmed by SETD2 knockdown and rescue experiments in NCI-H522 and NCI-H1299 cells. Using RT-PCR and qRT-PCR, we observed that the knockdown of SETD2 phenocopies the knockdown of IWS1 on the *U2AF2* and *FGFR2* RNA splicing, an effect rescued by the wild type SETD2, but not by the SETD2 methyltransferase mutant R1625C, which globally silences H3K36me3 (Hacker et. al., 2016[36]) (**Fig. 3d, Fig. S3c**). We conclude that the catalytically-active SETD2 is indeed required for the inclusion of exon 2 in the mature *U2AF2* mRNA.

Recently, and after the completion of the ChIP experiments, we carried out ChIP-Seq experiments, addressing the binding of IWS1 and SETD2 and the distribution of H3K36me3 marks genome-wide in NCI-H522 shIWS1/WT-R and shIWS1/MT-R cells. The unbiased data on the abundance of these markers in the *U2AF2* gene were in general agreement with the

ChIP data described above. Specifically, IWS1 was found to bind *U2AF2* E2 independent of its phosphorylation, but SETD2 binding and H3K36me3 abundance on *U2AF2* E2 increased only when IWS1 was phosphorylated (**Fig. 3e**). Furthermore, the same analysis confirms our previous results, by showing the increased H3K36me3 abundance on *FGFR2* exon IIIb only when IWS1 was phosphorylated (**Fig. S3d**) (Sanidas et al., 2014[10]).

If the recruitment of SETD2 in C-terminal domain of RNA-pol II depends on the phosphorylation of IWS1, as our data indicate, one would expect that ectopic overexpression of SETD2 would fail to rescue the *U2AF2* alternative RNA splicing phenotype in shIWS1 and shIWS1/MT-R cells. This was confirmed by experiments addressing the rescue of *U2AF2* alternative RNA splicing in NCI-H522 and NCI-H1299 shCon, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells transduced with a SETD2 lentiviral construct, with a Hemagglutinin (HA)-tag (**Fig. S4a**). The failure of SETD2 to rescue the phenotype supports the model of SETD2 recruitment by phosphorylated IWS1.



However, histone methylation is a dynamic step-by-step process. (Greer et al., 2012[37]) and while SETD2 is the only known H3K36 trimethyl-transferase in mammalian cells (Hyun K. et al., 2017[38]), there are several Lysine methyltransferases, which catalyze mono- or di-methylation of Histone H3 at K36 and may influence the SETD2 output. Transfection of the NCI-H522 and NCI-H1299 cells with siRNAs targeting a set of methyltransferases that are known to catalyze Histone H3K36 mono and di-methylation (NSD1, NSD2 and NSD3), or only di-methylation (SMYD2 and ASHL1) (Lucio-Eterovic et al., 2010[339], Q. Qiao et al., 2010[40], Berdasco et al., 2009[41] , Rahman et al., 2011[42], Brown M.A et al., 2006[43], Gregory G.D. et al., 2007[44]), revealed no contribution to the regulation of the alternative RNA splicing of *U2AF2* (**Fig. S4b**).

Conversely, Histone methylation is a reversible process (Bannister et al., 2002[45]). The Histone H3K36me3 marks are erased by KDM4A and KDM4C, two members of the

KDM4 JmjC domain-histone demethylase family (Berry, W.L et al., 2013[46], Cloos, P.A et al., 2007[47], Li, W et al., 2011[48]). Ectopic expression of KDM4A, KDM4B and KDM4C using lentiviral constructs in NCI-H522 and NCI-H1299 cells showed that none of these three altered the IWS1 phosphorylation-dependent alternative RNA splicing pattern of *U2AF2* (**Fig. S4c**). Overall, these data indicate that the reported demethylation of H3K36me3 by members of KDM4/JMJD2 Histone demethylase family does not play a significant role in our model.

**The regulation of the alternative RNA splicing of the *U2AF2* exon 2 by IWS1 phosphorylation, depends on the p52 isoform of the H3K36me3 reader LEDGF.**

Based on our preceding studies, the regulation of the *FGFR2* alternative RNA splicing by IWS1 phosphorylation depends on the reading of the Histone H3K36me3 marks by MRG15 (Sanidas et al., 2014[10]). To determine whether MRG15 is also the reader of the IWS1-dependent alternative RNA splicing of *U2AF2*, we knocked down MRG15 in both NCI-H522 and NCI-H1299 cells. Using RT-PCR and qRT-PCR to monitor the alternative splicing of *U2AF2* in these cells, revealed that it is independent of MRG15 (**Fig. 4a Left panels**). In agreement with this result, the knockdown of the splicing repressor and binding partner of MRG15, PTB, also had no effect on the RNA splicing of *U2AF2* (**Fig. 4a Right panels**), results not surprising since IWS1 phosphorylation mediates exon inclusion signals in our model. In parallel quantitative RT-PCR experiments, the knockdown of MRG15 and PTB in NCI-H522 and NCI-H1299, increased the *FGFR2* IIIb/IIIc, as expected (**Fig. S5a**) (Luco et al., 2010[19], Sanidas et al., 2014[10]). To further validate these observations, we knocked down PTBP2 in NCI-H522 and NCI-H1299, a paralog of PTB, that has been reported to rescue PTB functions (Cheung et al., 2009[49]). First, the western blotting in NCI-H522 and NCI-H1299 shPTB cells showed no upregulation of PTBP2 upon loss of PTB (**Fig. 4a**). Furthermore, quantitative RT-PCR results revealed that the regulation of *U2AF2*

RNA splicing is independent of PTBP2, with an expected increase of *FGFR2* IIIb/IIIc ratio (**Fig. S5b**).

To identify the factor that controls the IWS1 phosphorylation-dependent exon inclusion of *U2AF2* exon 2, we transfected NCI-H522 and NCI-H1299 cells with siRNAs of the known H3K36me3 readers PHF1 (Cai L. et al., 2013[50]), BRPF1 (Vezzoli et al., 2010[51]), MSH6 (Li F. et al., 2014[52]), GLYR-1 (Vermuelen et al., 2010[53]) and LEDGF (Pradeepa et al., 2012[20]), along with a control siRNA. After validating the significant downregulation of the expression by Western blotting, monitoring the effects of these transfections on E2/E3 *U2AF2* ratio using RT-PCR and qRT-PCR, revealed that only the knockdown of LEDGF phenocopied the knockdown of IWS1 on the alternative RNA splicing of *U2AF2*, suggesting that LEDGF is the sole H3K36me3 reader responsible for the observed phenotype (**Fig. 4b, S5c**). To confirm this conclusion and to determine which isoform of LEDGF may be responsible for the phenotype, we used a lentiviral shRNA construct to knockdown LEDGF, targeting the 3' UTR, and we rescued the knockdown by transducing the cells with lentiviral constructs of the alternatively spliced LEDGF isoforms, p75 and p52, with a V5-tag (**Fig. 4c, upper panel**). Monitoring the effects of these transductions by RT-PCR and qRT-PCR, revealed that only the p52 isoform rescues the shLEDGF-induced U2AF2 alternative RNA splicing phenotype. More importantly, the A51P mutant of p52-LEDGF, which cannot bind Histone H3K36me3 (Shun et al., 2008[54]), failed to rescue the splicing defect, suggesting that the *U2AF2* exon inclusion depends on the reading of the H3K36me3 marks by p52/LEDGF (**Fig. 4c, S5d**). Notably, the knockdown of LEDGF did not affect the *FGFR2* alternative RNA splicing pattern (**Fig. S5e**), suggesting that although the IWS1 phosphorylation-dependent RNA splicing targets are regulated via H3K36me3, the complexes recruited upon IWS1 phosphorylation, differ.

The preceding data provide strong genetic evidence that p52/LEDGF regulates the *U2AF2* alternative RNA splicing by reading the IWS1 phosphorylation-dependent Histone H3K36me3 marks. To validate these observations, chromatin Immunoprecipitation assays for the V5-tagged p52/LEDGF isoform, confirmed that the p52/LEDGF binds *U2AF2* exons 2 and 3 only when IWS1 is phosphorylated, observations consistent with SETD2 binding and H3K36me3 abundance (**Fig. 4d, S5f**).

If the p52 isoform of LEDGF regulates the alternative RNA splicing of *U2AF2* by reading the Histone H3K36me3 mediated by IWS1-phosphorylation, as suggested by the preceding data, one would expect that overexpression of p52/LEDGF would not rescue the *U2AF2* alternative RNA splicing phenotype in shIWS1 and shIWS1/MT-R cells. This was confirmed by experiments addressing the *U2AF2* RNA splicing in NCI-H522 and NCI-H1299 shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells along with a V5-tagged p52/LEDGF construct (**Fig. S5g**). Altogether, these data support the model based on which p52/LEDGF regulates the alternative RNA splicing of *U2AF2* by reading the Histone H3K36 trimethylation on the body of *U2AF2* gene, induced by IWS1 phosphorylation at S720/T721, mediated by SETD2.



**The p52 isoform of LEDGF regulates the alternative RNA splicing of *U2AF2*, via its interaction with the RNA splicing factor SRSF1.**

It has been reported that the p52/LEDGF isoform re-localizes in the nucleus, through signals mediated from its unique C-terminal domain, and interacts with the splicing factor SRSF1, affecting its distribution to alternatively spliced genes. (Pradeepa et al., 2012[20]). To investigate a potential role of SRSF1 in *U2AF2* alternative RNA splicing, we knocked down SRSF1 and we showed that its loss phenocopies the knockdown of IWS1. The dependence of the *U2AF2* exon 2 splicing on SRSF1, which was suggested by this result, was confirmed

by rescue experiments with wild type SRSF1 (**Fig. 4e, S6a**). Consistent with the LEDGF results, quantitative RT-PCR experiments showed that the knockdown of SRSF1 did not affect the RNA splicing pattern of *FGFR2* (**Figure S6a, right lower panel**). We therefore conclude that SRSF1 regulates the alternative RNA splicing of the *U2AF2* exon 2. More importantly, SRSF1 did not rescue the *U2AF2* RNA splicing phenotype induced by shIWS1 or shIWS1/MT-R (**Fig. S6b**), suggesting that it does not function independently. Instead, we hypothesized that it provides the link between IWS1 phosphorylation-dependent chromatin modifications and p52/LEDGF binding to the RNA splicing machinery. To test this hypothesis, we performed ChIP experiments addressing the binding of SRSF1 to *U2AF2* exons 2 and 3 in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells, transduced with a V5-tagged SRSF1 lentiviral construct. The *U2AF2* TSS and GAPDH exon 3 regions were again used as controls. The results confirmed that the binding of SRSF1 to exons 2 and 3 parallels the SETD2 and p52/LEDGF binding along with H3K36me3 abundance and occurs only when IWS1 is phosphorylated (**Fig. 4f, S6c**), providing support to the proposed hypothesis.



Based on the preceding data, we hypothesize that the binding of SRSF1 to the chromatin-associated p52/LEDGF should bring it into proximity with the nascent pre-mRNA, facilitating their interaction. Analysis of the *U2AF2* mRNA sequence, using the web-based pipeline RBP-map (Paz et al., 2014[55]), identified four potential SRSF1 binding sites (2 in *U2AF2* exon 2 and 2 in exon 3) (**Fig. S6e**), providing additional support to this hypothesis. Furthermore, similar analysis of the validated IWS1 splicing targets (*SLC12A2, IFT88, STXBP1, C1qTNF6*), revealed putative SRSF1 binding sites in both the spliced and adjacent exon (**Fig. S6f, S6g, S6h, S6i**), suggesting a potential universal role of SRSF1 on the IWS1 phosphorylation-dependent exon inclusion phenotype. To experimentally address the

proposed model, we carried out RNA-IP (RIP) experiments in the same shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells, focusing on the binding of SRSF1 on the *U2AF2* exon 2, intron 2 and exon 3. The results confirmed that SRSF1 binds primarily to exon 2, but only in the shControl and shIWS/WT-R cells, which parallels its chromatin binding to the H3K36me3-bound p52/LEDGF (**Fig. 4g, S6d**).

Based on the data presented in Figure 3 and Figure 4, we propose that IWS1 phosphorylation by AKT is required for the recruitment of SETD2 to CTD of RNA Pol II. The latter promotes the trimethylation of Histone H3 at K36 on the body of *U2AF2* and perhaps other target genes. This facilitates the recruitment of the histone H3K36me3-interacting protein p52/LEDGF and its binding partner SRSF1, resulting in exon inclusion of *U2AF2* exon 2 (**Fig. 4h**).



**The U2AF65 splice variant lacking exon 2, does not interact with Prp19.**

The predominant splice variant of the *U2AF2* mRNA that we identified in shIWS1 and shIWS1/MT-R cells, lacks exon 2, which encodes part of the N-terminal RS domain of U2AF65 (**Fig 5a**). This domain is responsible for the interaction of U2AF65 with several factors that contribute to mRNA splicing, 3' cleavage and polyadenylation (Millevoi et al., 2006[58], Shen H, et al., 2004[59]). One of these factors is Prp19, a key component of the seven-member ubiquitin ligase complex Prp19C (S.P Chan et al., 2003[60], S.P Chan et al., 2005[61], R. Hogg et al., 2010[62], Chanarat S. et al., 2013[63]).

Using co-immunoprecipitation experiments in HEK-293T cells engineered to express the exon-2 containing U2AF65α or the exon 2-excluded U2AF65β variants, with a V5-tag, we confirmed that only the U2AF65α splice variant interact with the endogenous Prp19 (**Fig. 5b**). More importantly, consistent with the previous results, immunoprecipitation of

endogenous U2AF65 in NCI-H522 and NCI-H1299 shControl, shIWS1, shIWS1/WT-R and shIWS1-MT-R cells, revealed impaired U2AF65-Prp19 interaction in shIWS1 and shIWS1/MT-R cells, which express primarily U2AF65β, while maintaining interaction with its U2-heterodimer partner, U2AF35, and SF3B1, a U2 snRNP protein (**Fig. 5c, S7a**). These data confirmed that the interaction of U2AF65 with Prp19 depends on the sequence encoded by *U2AF2* exon 2, whose inclusion is regulated by IWS1 phosphorylation.

**The splicing of the *U2AF2* mRNA downstream of IWS1 phosphorylation, regulates the mRNA splicing and expression of *CDCA5*.**

Based on previous reports, U2AF65 binds RNA Pol II, leading to an U2AF65-dependent recruitment of Prp19 to the newly-synthetized pre-mRNA and promoting proper co-transcriptional splicing activation (C.J David et. al., 2011[64]). Furthermore, it had been shown that U2AF65 and Prp19 are necessary for the pre-mRNA splicing and accumulation of Sororin, a component of the cohesin complex, which is encoded by *CDCA5* (Watrin et. al., 2014[65]). Given that aberrant splicing of *U2AF2* in shIWS1 and shIWS1/MT-R cells resulted in the loss of the interaction between U2AF65 and Prp19, we hypothesized that the pre-mRNA splicing of *CDCA5* in these cells will be impaired. We therefore employed qRT-PCR to examine the ratio of spliced and unspliced *CDCA5* mRNA in the same cells. The results confirmed that the mRNA splicing of *CDCA5* was impaired in both the shIWS1 and shIWS1/MT-R cells, with no significant change in the spliced/unspliced ratios of the control *GUSB* mRNA (Watrin et. al., 2014[65]) (**Fig. 5d, S7b, S7c upper panels).** More importantly, the splicing defect was rescued by U2AF65α but not by U2AF65β. (**Fig. 5e, S7d upper panels**).

To determine how IWS1 phosphorylation regulates *CDCA5* RNA splicing, we performed RNA Immunoprecipitation (RIP) experiments in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells. The results confirmed

that both spliced variants of U2AF65 bind equally well *CDCA5* pre-mRNA, as well as the control *GUSB* pre-mRNA, as expected (**Fig. 5d, 5e, S7c, S7d middle panels**) (Shen H. et al., 2004[31]). However, Prp19 binding to the same pre-mRNA regions of *CDCA5* was significantly impaired in shIWS1 and shIWS1/MT-R cells, which predominantly express the U2AF65β isoform (**Fig. 5d, S7c lower panels**). More importantly, the impaired Prp19 binding to *CDCA5* pre-mRNA was rescued by U2AF65α, but not U2AF65β (**Fig. 5e, S7d lower panels**).

Given that only spliced mRNAs are transported out of the nucleus, we also examined the abundance of cytosolic *CDCA5* mRNA in shControl, shIWS1, shIWS1/WT-R, shIWS1/MT-R and shIWS1 rescues with U2AF65α or U2AF65β. To this extent, we fractionated the nuclear and cytosolic compartments of these NCI-H522 and NCI-H1299 cells. Then, using western blotting, we confirmed the validity of the fractionation by the expected expression of Lamin A/C and GAPDH only in the nuclear and cytosolic protein compartment, respectively (**Fig. S7e**). The results also confirmed that the mature *CDCA5* mRNA in the cytoplasm was downregulated as expected, and that the RNA transport deficiency is rescued by U2AF65α, but not U2AF65β (**Fig. 5f**).

To determine whether the *CDCA5* RNA splicing defect in shIWS1 and shIWS1/MT-R NCI-H522 and NCI-H1299 cells prevents the expression of its protein product, Sororin, we examined the its expression along with the expression of IWS1, p-IWS1, U2AF65 and Prp19, by western blotting. The results confirmed that the expression of Sororin parallels IWS1 phosphorylation as it was impaired in shIWS1 and shIWS1/MT-R NCI-H522 and NCI-H1299 cells. More importantly, the low expression of Sororin in shIWS1 was rescued by U2AF65α, but not by U2AF65β (**Fig. 5g, S7f upper panels**). Taken together, these data confirm that IWS1 phosphorylation via the alternative RNA splicing of *U2AF2*, is ultimately responsible for the regulation of the mRNA splicing of *CDCA5*.

**Sororin and p-ERK form a positive feedback loop, which is activated by IWS1 phosphorylation and promotes the expression of CDK1 and Cyclin B1.**

It has been previously shown that the downregulation of Sororin leads to reduced ERK phosphorylation, in human Colorectal Cancer (CRC) and human Hepatocellular Carcinoma (HCC) (Wang J., et al., 2018[66], A. Zhen et. al., 2019[67]). Since the IWS1 phosphorylation-dependent *U2AF2* alternative RNA splicing regulates the abundance of Sororin, we hypothesized reduced ERK phosphorylation in NCI-H522 and NCI-H1299 shIWS1 and shIWS1/MT-R cells. This was confirmed by experiments showing that ERK phosphorylation at Y202/T204 was reduced in shIWS1 and shIWS1/MT-rescue cells, while total ERK was not. More importantly, the reduction of p-ERK levels in shIWS1 cells was rescued by U2AF65α, but not by U2AF65β (**Fig. 5g, S7f lower panels**), confirming the p-IWS1/Sororin/p-ERK axis in lung adenocarcinoma.

Sororin is a known target of ERK, which phosphorylates the protein at S79 and S209 (Nguyen et. al., 2010[68]). Interestingly, whereas the wild type Sororin and the Sororin phosphomimetic mutant S79E/T209E (Sororin DM-E) rescued the phosphorylation of ERK in shIWS1 cells, the S79A/S209A (Sororin DM-A) mutant did not (**Fig 5g and S7f Lower panels**). This suggests the existence of a Sororin-ERK phosphorylation positive feedback loop in lung adenocarcinoma, which is controlled by the AKT-dependent IWS1 phosphorylation, and the *U2AF2* alternative RNA splicing.

Based on previous observations, down-regulation of Sororin leads to impaired expression of CDK1 and Cyclin B1 (Wang J., et al., 2018[66], A. Zhen et. al., 2019[67]). To this extent, we hypothesized that IWS1 phosphorylation regulates the expression of CDK1 and Cyclin B1, through the regulation of *U2AF2* RNA splicing and the abundance of Sororin. Consistent with the preceding data, western blotting revealed reduced levels of CDK1 (Y15) phosphorylation, CDK1 and Cyclin B1 levels in shIWS1 and shIWS1/MT-R cells. More

importantly, the downregulation of these molecules in shIWS1 cells was rescued by U2AF65α, but not by U2AF65β, wild type Sororin and Sororin DM-E but not by Sororin DM-A (**Fig. 5g, S7f lower panels**). These data indicate that the Sororin-ERK phosphorylation feedback loop, downstream of IWS1 phosphorylation, controls the expression of CDK1 and Cyclin B1, through a process depending on *U2AF2* alternative RNA splicing.

**IWS1 phosphorylation by AKT3 promotes cellular transformation through *U2AF2* alternative RNA splicing and Sororin expression.**

Based on the proceeding data, IWS1 phosphorylation maintains the establishment of a Sororin/ERK phosphorylation feedback loop in lung adenocarcinoma, through the epigenetic regulation of *U2AF2* alternative RNA splicing. Given that ERK phosphorylation is a well-known signal for cellular transformation (Guo et al., 2020[69], Wee et al., 2017[70]), we then questioned whether the AKT/IWS1 pathway promotes transformation of immortalized Human Bronchial epithelial cells (HBEC-hTERT). To this extent, we overexpressed myristoylated AKT3 (myr-AKT3), the phosphomimetic AKT3 T305/S472DD (AKT3-DD), phosphomimetic IWS1 S720D/T721E (IWS1 DE), the two spliced U2AF65 isoforms (U2AF65α, U2AF65β) and several forms of Sororin (Sororin wt, Sororin DM-E, Sororin DM-A) (**Fig. S8a**). Consistent with the ERK phosphorylation pattern, cell transformation assays showed effective transformation of HBEC-hTERT induced by the active AKT3 (myr-AKT3, AKT3-DD) and IWS1 (IWS1 DE). More importantly, the downstream pathway of U2AF65α spliced isoform and active Sororin (Sororin wt, Sororin DM-E) induced cellular transformation, but the exon 2 deficient U2AF65β and inactive Sororin mutant (Sororin DM-A), did not (**Fig. S8b, S8c**). Collectively, these results confirm that IWS1 phosphorylation promotes cellular transformation through the epigenetic regulation of *U2AF2* alternative RNA splicing and Sororin/ERK loop.

**IWS1 phosphorylation controls the ERK phosphorylation levels, in cells harboring *EGFR* and *KRAS* mutations, through *U2AF2* RNA splicing.**

*EGFR* and *KRAS* are frequently mutated in human lung adenocarcinoma and promote downstream pro-oncogenic signals through the phosphorylation of ERK (Guo et al., 2020[69], Wee et al., 2017[70]). We therefore addressed the role of IWS1 phosphorylation and *U2AF2* alternative RNA splicing on Sororin and ERK phosphorylation in lung adenocarcinoma cells harboring *KRAS* (A549 and NCI-H460) and *EGFR* (NCI-H1975, PC-9, NCI-H1650) mutations. The results showed that *U2AF2* alternative RNA splicing pattern is independent of the mutational status (**Fig. 5h, upper and lower left panels**). More importantly, the knockdown of IWS1 downregulated the expression of Sororin and ERK phosphorylation (**Fig. 5h, upper panel**). These data were striking, because they showed that the effect of Sororin on ERK phosphorylation is dominant over the effect of *KRAS* and *EGFR* mutations. Surprisingly, the reduction of ERK phosphorylation induced by the loss of IWS1 phosphorylation was more robust in *EGFR* as opposed to the *KRAS* mutated cells (**Fig. 5h, lower right panels)**. Altogether, this data suggests that IWS1 phosphorylation controls the *U2AF2* alternative RNA splicing independent of the cellular mutational status and shows robust reduction of the Sororin/ERK phosphorylation axis in lung adenocarcinoma cell lines, harboring *EGFR* mutations.



**IWS1 phosphorylation promotes cell proliferation by controlling the Sororin/ERK phosphorylation feedback loop, through the *U2AF2* RNA splicing pattern.**

Given that the Sororin/ERK phosphorylation positive feedback loop, activated by IWS1 phosphorylation, promotes the expression of CDK1 and Cyclin B1, we hypothesized that the stimulation of cell proliferation by IWS1, we had observed previously (Sanidas et al., 2014[10]), may depend on the activation of the *U2AF2* RNA splicing-regulated CDCA5/ERK phosphorylation loop. To address this hypothesis, we examined the rate of proliferation of

shControl, shIWS1, shIWS1/Sororin WT, shIWS1/Sororin DM-A and shIWS1/Sororin DM-E NCI-H522 and NCI-H1299 cells growing under standard culture conditions. The results showed that cell proliferation was inhibited by shIWS1. More importantly, the inhibition was rescued by wild type Sororin and Sororin DM-E but not by the Sororin DM-A. Significantly, the proliferative effect of the Sororin DM-E was more robust than the proliferative effect of wild type (wt) Sororin **(Fig. 6a).** To further support these data, similar proliferation assays were performed in NCI-H522 and NCI-H1299 shControl, shIWS1, shIWS1/U2AF65α and shIWS1/U2AF65β. The results revealed that the proliferation phenotype in shIWS1 cells was rescued by U2AF65α, but not by U2AF65β, confirming that the regulation of cell proliferation depends on the *U2AF2* RNA splicing pattern (**Fig. 6b**).

As we pointed out above, the loss of phosphorylated IWS1 results in more robust down-regulation of ERK phosphorylation in *EGFR* as opposed to *KRAS* mutated lung adenocarcinoma cells (**Fig. 5g**), suggesting that the *EGFR* mutated cells would be more sensitive to the loss of IWS1. This was confirmed by experiments showing reduced cell proliferation of A549 (*KRAS* mutant) and NCI-H1975 (*EGFR* mutant) shIWS1 cells (**Fig 6c**). More importantly, in parallel with the data of the reduction of ERK phosphorylation, *EGFR* mutated cell line NCI-H1975, demonstrates the highest reduction in cell proliferation induced by the loss of IWS1 (**Fig. S9a**). Furthermore, the observed inhibition of cell proliferation in A549 and NCI-H1975 shIWS1 cells was rescued by U2AF65α, but not by U2AF65β, suggesting that the *U2AF2* RNA splicing affects cell proliferation in *KRAS* and *EGFR* mutated cells (**Fig. 6d**). Additional support was provided by Western Blot analysis showing that in all cell lines, the cell proliferation marker PCNA was downregulated in shIWS1 cells, an effect rescued by U2AF65α, but not the U2AF65β (**Fig. S9b**). Notably, shIWS1-induced downregulation was more robust in *EGFR* (NCI-H1975, NCI-H1650) as opposed to *KRAS* mutated (A549, NCI-H460) cells (**Fig. S9b**)**.** Altogether, these data suggest that IWS1

phosphorylation controls cell proliferation through the regulation Sororin/ERK phosphorylation loop mediated by *U2AF2* RNA splicing.

To determine whether IWS1 phosphorylation stimulates cell proliferation by promoting progression through the G2/M phase of the cell cycle, we stained log phase cultures of shControl, shIWS1, shIWS1/U2AF65α and shIWS1/U2AF65β NCI-H522, NCI-H1299, A549 and NCI-H1975 cells with Propidium Iodide (PI) and we examined the percentage of cells in different phases of the cell cycle by flow-cytometry. The results confirmed that the knockdown of IWS1 induced prolonged G2/M phase of the cell cycle, an effect rescued by U2AF65α, but not by U2AF65β, confirming the role of IWS1 in cell cycle progression through *U2AF2* RNA splicing (**Fig. 6e**). Consistent with the proliferation data, the G2/M arrest was less pronounced in the *KRAS* mutant cell line, A549 **(Fig. 6e).** Altogether, data presented in figures 5 and 6 suggest that the *EGFR* mutated lung adenocarcinoma cell lines are more sensitive than *KRAS* mutant cells upon the loss of IWS1 phosphorylation.



**IWS1 phosphorylation and *U2AF2* splice variant fluctuate during progression through the cell cycle.**

The preceding findings suggest that an RNA splicing event, regulated by the AKT-mediated phosphorylation of IWS1, plays a critical role in cell cycle progression. Given that the expression and activity of molecules critically involved in the regulation of the cell cycle, tend to fluctuate as cells transit from one cell cycle phase to the next (Kent et al., 2019[71]), we examined the expression and phosphorylation of IWS1, the *U2AF2* alternative RNA splicing pattern and the expression of SETD2, Sororin and the trimethylation of histone H3K36 in G1, S and G2/M in NCI-H1299 cells. To this end, using carboxyfluorescein succinimidyl ester (CFSE)-like DNA dyes, cells were stained in G1, S and G2/M phase and they were fractionated by FACS sorting (Begum et al, 2013[72]). Western blotting of these

fractions confirmed that IWS1, phospho-IWS1, SETD2, Histone H3K36me3 and Sororin are indeed upregulated in S and G2/M. As specific cell cycle phase markers, Cyclin E1 (G1 phase), CDC25A (S phase) and phosphorylated Histone H3 (S10) were used, validating the integrity of the fractionation **(Fig. 6f lower panels)**. More importantly, IWS1 expression and phosphorylation occured more during S phase, whereas Sororin, SETD2 and Histone H3K36me3 levels were more robust during G2/M phase, as previously reported (Watrin et al., 2014[65], Dronamraju et al., 2018[73]) **(Fig. 6f upper panels).** Furthermore, using RNA extracts from the same cells, RT-PCR experiments revealed an *U2AF2* RNA exon inclusion pattern which parallels the expression of IWS1 during the cell cycle **(Fig. 6f).** Moreover, qRT-PCR confirmed that the expression of *IWS1* and *CDCA5* in the RNA level parallels their protein expression, with no change of total *U2AF2* levels (**Fig. 6g**).

Chromatin Immuno-Cleavage (ChIC) experiments revealed increased binding of IWS1 on *U2AF2* exon 2 and 3 during S and G2/M phase, with significantly more binding during S phase (**Fig. 6h upper panel**). Consistently, the binding of SETD2, the abundance of H3K36me3 and the E2/E3 *U2AF2* ratio parallel the levels and the binding of IWS1 on *U2AF2* gene, although total SETD2 and H3K36me3 are more abundant during G2/M phase, confirming the strong dependence of the pathway on IWS1 (**Fig. 6h, 6i**).

Overall, these data provide strong evidence for the model in which IWS1, whose expression increases during S and G2/M, recruits SETD2 to CTD of RNA-Pol II, following phosphorylation by AKT. RNA-Pol II-associated SETD2 trimethylates Histone H3 at K36 during transcriptional elongation, resulting in the assembly of p52/LEDGF-SRSF1 complexes which promote the inclusion of exon 2 in the *U2AF2* mRNA. Subsequently, the U2AF65α isoform along with Prp19, facilitate the proper splicing of *CDCA5* pre-mRNA, leading to Sororin accumulation during S and G2/M phase. Finally, Sororin forms a positive feedback loop with ERK phosphorylation, and controls the progression through the cell cycle in lung adenocarcinoma. The events described in this model define a molecular mechanism of a

splice variant which is epigenetically regulated during the cell cycle, feeding back into its regulation (**Fig. 6j, S9c**).

**IWS1 phosphorylation controls tumor growth *in vivo*, by regulating the *U2AF2*/Sororin/ERK axis.**

Based on our preceding studies, the loss of IWS1 and IWS1 phosphorylation impair tumor growth in a mouse xenograft model (Sanidas et al., 2014[10]). To further validate and expand this observation, we knocked down IWS1 in NCI-H1299, A549 and NCI-H1975 cells and we injected them subcutaneously, along with shControl cells, in immunocompromised NSG mice. The animals were monitored for tumor development and were sacrificed 4 (NCI-H1299 and NCI-H1975) and 6 (A549) weeks later, respectively. The results revealed that the knockdown of IWS1 reduced tumor weight and volume (**Fig. 7a**). To further demonstrate the role of *U2AF2* alternative RNA splicing in tumor growth, we repeated the mouse xenografts models by injecting NCI-H1299 cells transduced with shIWS1 and rescued with the two U2AF65 isoforms, U2AF65α and U2AF65β, along with shControl cells (**Fig. S10a**). The results showed that the reduced tumor volume phenotype in shIWS1-derived tumors was rescued in U2AF65α, but not U2AF65β-derived tumors, confirming the role of the *U2AF2* RNA splicing pattern on tumor growth (**Fig. 7b**).

Analysis of the percentage reduction of tumor characteristics in the various mouse xenografts cell lines models, showed that the reduction of tumor growth induced by the loss of IWS1 was less pronounced in A549 cells, harboring *KRAS* mutations (**Fig. 7c**). The weak growth reduction of tumours derived from shIWS1 A549 cells, paralleled the weak inhibition of ERK phosphorylation (**Fig 5h**), cell proliferation (**Fig. S9a**) and cell cycle progression (**Fig. 6e**) induced by the knockdown of IWS1 in these cells.

The preceding data confirmed that the knockdown of IWS1 reduces tumor growth. We then proceeded to address the mechanism of the growth inhibition, which had not been

adequately addressed in our earlier studies. First, we confirmed the efficiency of the IWS1 knockdown, by probing western blots of tumor cell lysates with anti-IWS1 and anti-phospho-IWS1 (S720) antibodies (**Fig. 7d**). Following this, we employed RT-PCR and qRT-PCR on RNA derived from these tumors, to address the usage of exon 2 in the *U2AF2* mRNA. The results confirmed that the knockdown of IWS1 has no effect on the total *U2AF2* mRNA levels (**Fig. S10b**), but promotes the exclusion of exon 2 from *U2AF2* mRNA (**Fig. 7d upper panels, S10b**). Probing both tumor lysates and tissue sections with antibodies to regulators and targets of the Sororin/ERK feedback loop, confirmed that the activity of the loop is reduced in tumors derived from shIWS1 cells (**Fig. 7D lower panels, S10C**). Measuring the abundance of the PCNA and Ki-67 proliferation markers by western blotting (PCNA), or Immunohistochemistry (Ki-67), also confirmed the reduced expression of these markers in the shIWS1 tumors (**Fig. 7d lower panels, 7e**).



Quantitative analyses of the western blot (p-ERK and PCNA) and IHC data (Ki-67) (**Fig. 7f**), showed that the reduction of all these markers was less pronounced in the *KRAS* mutant A549 tumors, suggesting that *KRAS* mutant cells are more resistant to the loss of IWS1. Altogether, these results are in agreement with the less pronounced effect of shIWS1 in the tumor growth in A549-derived tumors (**Fig. 7c**) and with the less pronounced effect of shIWS1 in p-ERK levels (**Fig. 5h**), cell proliferation in culture and degree of G2/M arrest in these cells (**Fig. 6**).

**The AKT3/p-IWS1/U2AF2 axis is active in human lung adenocarcinoma and affects the clinical outcome of *EGFR* mutant patients.**

To determine whether the pathway activated by IWS1 phosphorylation leading to Sororin expression and ERK phosphorylation is active in human lung adenocarcinoma (LUAD), we examined the expression and phosphorylation of several components of this axis, along with the alternative RNA splicing pattern of *U2AF2* in a set of 40 human LUAD

samples. For 30 of these tumors, Normal Adjacent Tissue (NAT) was also available and it was tested on parallel with the matching tumor sample (**Fig. 8a**). More importantly, the exon 2-containing *U2AF2* transcript was more abundant in tumor versus normal samples, with no change in total levels of *U2AF2* (**Fig. S11a**). Overall, these results confirmed that the pathway was active in the tumors, but not in the NAT.

Human LUAD frequently harbors *KRAS* or *EGFR* mutations and data presented in this report suggest that lung adenocarcinoma cells harboring *EGFR* mutations are more sensitive to the loss of IWS1, as opposed to *KRAS* mutant cells (**Fig. 5, Fig. 6, Fig. 7**). We therefore used monoclonal antibodies, which selectively recognize *KRAS* and *EGFR* mutant forms (Kim et al., 2015[75]), and we identified tumors harboring these mutations (**Fig. 8a**). Comparison of the correlations coefficients of IWS1 and phosphorylated IWS1 with the E2/E3 *U2AF2* and all the components of the downstream Sororin/ERK phosphorylation pathway, revealed strong correlations in the entire cohort with more robust correlations in the *EGFR* than the *KRAS* mutant tumors (**Fig. 8b**). Strikingly, the levels of IWS1 phosphorylation and abundance of exon 2-containing *U2AF2* isoform exhibit strong correlation with more aggressive clinical stage and defines poor survival in the entire patients' and the *EGFR* mutant cohort, but not in the *KRAS* mutated group (**Fig. 8c, 8d**).

The preceding data were confirmed by IHC, using sections of a commercially available tissue microarray (TMA) of 50 LUAD with paired NAT. The TMA samples were probed with antibodies against p-IWS1, Sororin, p-ERK, p-CDK1 and EGFR E746-A750 deletion (**Fig. S11b, S11c**). The results confirmed that the pathway is more active in the tumors, compared to the NAT samples **(Figure S11d)**. More importantly, IWS1 phosphorylation expression exhibits strong correlations with the components of the Sororin/ERK axis, with more robust correlations in the *EGFR* mutant tumors (**Fig. S11e**). Consistently, the results also showed that the activity of the pathway correlates with clinical stage and histological grade **(Fig. S11f, S11g)**. In addition to confirming the western blot

data in our set of LUAD patients, the IHC data also demonstrate that the activity of the pathway can be monitored clinically in human lung adenocarcinoma patients by IHC.

The data generated from the analysis of our LUAD samples and the TMA were confirmed by data in publicly available databases. Analysis of the data derived from Tumor Cancer Genome Atlas (TCGA), revealed correlations between *IWS1* or *SRSF1*, *U2AF2* E2 and other components of the IWS1 phosphorylation pathway (**Fig. 8e**). The results also showed that the *U2AF2* E2/E3 ratio and the expression of *CDCA5* mRNA were significantly upregulated in TCGA LUAD patients with high *IWS1*, compared to the low *IWS1* group (**Fig. S11h**). More importantly, *IWS1* expression correlates with more aggressive tumor stage in tumors harboring *EGFR* but not *KRAS* mutations (**Fig. S11i**).

Based on recent reports, 55% of lung adenocarcinoma patients relapse and up to 40% of them develop metastatic disease (Popper et al., 2016[76], West et al., 2020[77]). Given that the IWS1 phosphorylation controls oncogenic signals and plays a critical role in patients with lung adenocarcinoma, as suggested in this report, we questioned whether the IWS1-mediated axis is associated to clinical relapse or metastasis in LUAD patients. Analysis of a relapse-related molecular signature database GSE13213 (Tomida et al., 2009[78]), revealed increased expression of the IWS1/Sororin axis in relapsing patients harboring *EGFR*, but not with *KRAS* mutations (**Fig. 8f, S11j**). We therefore conclude that the activation of the IWS1 phosphorylation pathway may also contribute to the relapse of lung adenocarcinoma, especially with *EGFR* mutations. Furthermore, analysis of RNA-seq data derived from dissected brain metastatic lesions from lung adenocarcinoma patients (GSE141685), revealed increased expression of the components of IWS1/Sororin axis in the metastatic as opposed to primary lung adenocarcinoma lesion (**Fig. 8g**). Importantly, the *U2AF2* E2/E3 ratio was also elevated in the metastatic tumours, while the IIIb/IIIc ratio of the *FGFR2* mRNA was reduced (**Fig. S11k and S11l**), in agreement with our earlier observations, showing that IWS1 phosphorylation promotes *FGFR2* exon IIIb skipping

(Sanidas et al, 2014[10]). To further elucidate the role of IWS1 phosphorylation in metastasis in lung adenocarcinoma, analysis of single-cell RNA-Seq (scRNA-seq) data from a recent Patient-Derived LUAD Metastatic model (Laughney et al., 2020[79]), revealed increased expression and robust correlation of IWS1 and CDCA5 in the incipient metastasis providing evidence of the role of IWS1 in the metastatic potential of lung adenocarcinoma (**Fig. 8g**). Moreover, analysis of the RNA-Seq data in the TCGA LUAD dataset, which contains information on cancer-associated mutations, confirmed the link between *IWS1* expression and metastatic disease, but also showed that *IWS1* is again upregulated in tumours derived from patients with metastatic disease, and harboring *EGFR*, but not *KRAS* mutations (**Fig. S11m**).

As expected from the preceding data, *IWS1* expression, *U2AF2* exon 2 inclusion and *FGFR2* exon IIIb exclusion, are indicators of poor prognosis in patients with lung adenocarcinomas in the TCGA LUAD dataset, harboring *EGFR*, but not *KRAS* mutations (**Fig. 8h**). In addition, the survival of patients with lung adenocarcinomas in the same dataset, expressing high levels of *IWS1*, was reduced if the tumours also harbored mutated *EGFR* (**Fig. S11n**). Confirmatory to these observations, was the finding that the synergy between *IWS1* expression and *EGFR* mutations, was also observed in the GSE13213/GSE26939 dataset (**Fig. S11o**).

Finally, analysis of Whole Exome Sequence (WES) lung adenocarcinoma datasets (Ding et al., 2008[81], Imielinksi et al., 2012[82], Rivzi et al., 2015[83], Jordan et al., 2017[84], Chen et al., 2020[85]) via cBioportal (Cerami et al., 2012[86], Gao et al., 2013[87]) for mutations and copy-number alterations, revealed the occurrence of mutations, fusions and amplification of the AKT/IWS1 pathway described in this report, in patients with Lung Adenocarcinoma (**Figure S12a**). Notably, the same analysis revealed the existence of missense and splice site mutations of *U2AF2* on RS domain sequences, along with previously described

cancer-associated mutations of the U2AF65 RRM domain, G176E/V, known to affect RNA splicing (Glasser E. et al., 2017[32], Maji et al., 2020[88]) (**Fig. S12b**). Furthermore, consistently with all previous data, *AKT3* and *CDCA5* expression is associated with poor prognosis in patients with lung adenocarcinomas in the TCGA LUAD dataset (**Fig. S12c**).

Altogether, these data come in agreement with the *in vitro* (**Fig. 5 and Fig. 6**) and *in vivo* (**Fig. 7**) data provided in this report, that indicate that the p-AKT/p-IWS1 axis through the epigenetic regulation of *U2AF2* alternative RNA splicing and oncogenic signals, defines poor survival in lung adenocarcinoma patients harboring *EGFR* mutations.



## Discussion

### Regulation of the Sororin/ERK phosphorylation loop from IWS1 phosphorylation

Our studies implicate IWS1 phosphorylation as a regulator of oncogenic signals in lung adenocarcinoma through alternative RNA splicing. We report a *U2AF2* splice variant in lung adenocarcinoma which lacks exon 2 upon loss of phosphorylated IWS1. This signaling pathway is initiated by the AKT3-dependent phosphorylation of IWS1, which induces the epigenetic regulation of *U2AF2* alternative RNA splicing, through H3K36me3-mediated signals (**Fig. 3**). The shift in the alternative RNA pattern controls the interaction of U2AF65 with Prp19. These two factors regulate the pre-mRNA splicing of *CDCA5* and the expression of its protein product, Sororin, co-transcriptionally (**Fig. 5**). Our data provide insight for the existence of a Sororin/ERK phosphorylation feedback loop in lung adenocarcinoma, which is regulated by the alternative RNA splicing of *U2AF2*. Notably, the Sororin-dependent regulation of ERK phosphorylation by IWS1 is dominant over the ERK-activating *EGFR* mutations and has a major impact in the biology of these cells harboring such mutations **(Fig. 5 and Fig. 6)**. Sororin is a member of the cohesin complex, a seven member protein complex which plays a central role in sister chromatid cohesion and higher chromosomal

architecture during meiosis (Ishiguro et al., 2019[89]). This molecular Sororin/ERK phosphorylation switch controls cell proliferation and the progression through the G2/M phase of the cell cycle (**Fig. 6**). Defects in Sororin expression lead to the activation of spindle assembly checkpoint, delaying the progression through the G2/M phase (Dreier et al., 2011[90]). Given that loss of IWS1 reduced the expression of Sororin, this could explain the G2/M phase delay observed in shIWS1 cells (**Fig. 6**). How the ERK-phosphorylated Sororin promotes the phosphorylation and activation of ERK is currently unknown. Our working hypothesis is that the phosphorylation and activation of ERK is due to signals induced by the interaction of Sororin with its partners in the cohesin complex. If this is the case, the cell may use this mechanism to sense the successful progression from prometaphase to metaphase, in order to activate a molecular switch. This enhances the phosphorylation of Sororin, facilitating progression through the G2/M phase of the cell cycle. In addition, data presented in this report show that the loss of IWS1 leads to impaired expression of CDK1 and Cyclin B1, a phenotype regulated through the Sororin/ERK phosphorylation loop (**Fig. 5**). On the other hand, our data indicate strong correlation of *CDCA5*, *CDK1* and *CCNB1* (Cyclin B1) mRNA levels in LUAD patients, derived from TCGA database (**Fig. 8**). These observations indicate that the induction of Cyclin B1 and CDK1 by Sororin is most likely at the level of transcription. A potential mechanism for this transcriptional regulation, controlled by IWS1, was suggested by earlier studies showing that the Cohesin complex interacts with the Mediator complex. The Mediator-Cohesin complexes are loaded by the NIPBL-Cohesin loading factor to enhancers and core promoters of target genes. Enhancer and core promoter-associated complexes promote loop formation between these segments, and regulate transcription (Kagey et al., 2010[91]). The contribution of this and other mechanisms on the regulation of *CCNB1* and *CDK1* expression is under investigation.

**Regulation of RNA processing in multiple layers**

Another important conclusion, based on the data presented in this report, is that RNA splicing is a process regulated at multiple levels. IWS1 phosphorylation directly regulates the alternative RNA splicing of *U2AF2*, via a H3K36me3-mediated mechanism (**Fig. 3**) and thus, introducing a new layer of RNA splicing regulation. This shift in alternative RNA splicing pattern affects the binding and formation of the U2AF65/Prp19 interacting complex, required for the efficient splicing of *CDCA5*, which ultimately controls pro-oncogenic ERK phosphorylation signals (**Fig. 5, Fig. 6**). The reason for the multilayered control of RNA splicing by a single RNA splicing regulator could be that this allows a limited number of available pathways to converge, in different combinations, on a large number of RNA splicing events and differentially regulate them. To this extent, our data provide insight on the importance of the AKT/IWS1 phosphorylation axis on the fine-tuning of the global epigenetic regulation of RNA splicing cellular program, governing cell cycle regulation and tumor progression. Another important point raised from our data, is that the removal of RS domain of U2AF65, demonstrates a gene-specific indirect effect on intron removal events, through the regulation and recruitment of Prp19 on the target introns. Although U2AF65 is a general splicing factor affecting the branchpoint assembly (Shen H. et al., 2004[31]), earlier observations on the regulation of *CDCA5* pre-mRNA, revealed that components of Prp19 Complex (Prp19 and Cdc5L) along with U2AF65, affected the intron removal of *CDCA5* pre-mRNA, but not the removal of intronic regions in other genes, such as *GUSB* (Watrin et. al., 2014[65]), demonstrating a selective effect of the U2AF65-Prp19 complex on the intronic regions of target genes. Consistently, our RIP data demonstrate the binding of Prp19 in the intronic areas of *CDCA5*, but not *GUSB*, when U2AF65α (exon 2 containing isoform) is expressed (**Fig. 5**), further supporting the selectivity of the latter complex on the target

genes. The genome-wide distribution of U2AF65 and Prp19 upon loss of phosphorylated IWS1 and RS domain exclusion, will be addressed in future studies.

**IWS1 phosphorylation and *U2AF2* splicing are regulated through the cell cycle, feeding back to its regulation, enhancing tumor proliferation.**

Given that the cell cycle is an integrated system and that cell cycle regulatory mechanisms tend to also be cell cycle-regulated, we examined and confirmed that IWS1 expression and phosphorylation fluctuate through the cell cycle, regulating *U2AF2* RNA splicing. The mechanism of the expression of IWS1 is under investigation. The increased activity of the pathway may also occur due to cell cycle-dependent changes in the activity of AKT. Earlier reports have shown that CDK2-Cyclin A2 complex is activated in S phase and phosphorylates AKT at Ser477/Thr479 (Liu et al., 2014[75]), which further enhances the activity of IWS1 pathway. Furthermore, our results show that the epigenetic complexes assembled upon the phosphorylation of IWS1 on *U2AF2* gene, are regulated in a cell-cycle specific manner. These results come in agreement with previous reports stating that spliceosomal assembly is also regulated through the cell-cycle, implying an extra layer of alternative RNA splicing regulation (Karamysheva et al., 2015[92], Hofmann et al., 2010[93]). More importantly, our results reveal that IWS1 phosphorylation epigenetically controls the dynamics of a *U2AF2* splice variant through the cell cycle and regulates the levels of Sororin, which subsequently feeds back to the regulation of the cell cycle and tumor proliferation, through enhancement of ERK oncogenic signals (**Fig. 6**).

**Interpretation of the same histone modification signal through different epigenetic complexes, affecting lung adenocarcinoma development.**

Data presented in this report showed that signals originate from AKT phosphorylation may regulate alternative RNA splicing of target genes by different mechanisms, through IWS1 phosphorylation. Our previous findings had shown that IWS1 phosphorylation regulates the alternative RNA splicing of *FGFR2* by promoting the exclusion of exon IIIb from the mature transcript (Sanidas et al., 2014[10]). Here we show that, similar to the *FGFR2*, the exon inclusion of *U2AF2* exon 2, is also under the control of the SETD2-dependent Histone H3K36 trimethylation in the body of the actively transcribed genes. However, the reader of the Histone H3K36me3 mark is the p52 isoform of LEDGF which interacts with the RNA-binding protein SRSF1, as opposed to the MRG15/PTB complex. Furthermore, we also provide evidence of a more global involvement of SRSF1 in the IWS1 phosphorylation-dependent exon inclusion phenotype (**Fig. S6**). Therefore, although the transduction signal and the regulatory chromatin modification mark is the same for both the exon inclusion and exclusion signals, the effector complexes assembled around H3K36me3, that are responsible for the effects on RNA splicing, differ. Interestingly, both the IWS1 phosphorylation splicing targets affect lung adenocarcinoma development. On the one hand, the *FGFR2* exon exclusion pattern promotes invasion and EMT (Sanidas et al., 2014[10]) and to the other hand, *U2AF2* exon inclusion pattern promotes oncogenic and survival signals, converging towards more aggressive phenotype and clinical profile in lung adenocarcinoma patients, especially with *EGFR* mutations. This isoform specific aggressive clinical phenotype, comes in agreement with recent reports in the literature describing the existence of a cancer-specific splicing addiction network, in which cancer-related isoforms, in our case

*U2AF2* and *FGFR2*, manipulate and maintain oncogenic and invasion signals (Bonnal et al., 2020[94], Wang and Aifantis, 2020[95]).

**The clinical impact of the p-IWS1/*U2AF2* axis in *EGFR* mutant lung adenocarcinoma.**

Throughout this report, we provide evidence that IWS1 phosphorylation, through the epigenetic regulation of *U2AF2* alternative RNA splicing, represents an important oncogenic signal in *EGFR* mutant lung adenocarcinoma. Based on consistent results in cell lines (**Fig. 6**), animals (**Fig. 7**) and human tumor samples (**Fig. 8**), we showed the major impact of IWS1 phosphorylation on proliferation, tumor growth and, eventually, clinical outcome in *EGFR* mutant lung adenocarcinoma patients. Since, the PI3K/AKT axis is triggered by the *EGFR* mutated landscape, and as our data suggest, the activity of the p-IWS1/*U2AF2*/Sororin axis would be increased (**Fig. 8**), further inducing the activation and maintenance of the tumorigenic ERK phosphorylation signals. To this extent, signals derived from the AKT/IWS1 phosphorylation pathway facilitates and maintains the oncogenic addiction of these tumors which depends on ERK phosphorylation. Furthermore, our data indicate that IWS1 phosphorylation and the RNA splicing pattern of its targets, *U2AF2* and *FGFR2*, selectively correlates not only with tumor grade, stage and patient survival, but also with metastasis and with relapse (**Fig. 8**). Based on these findings, we propose two translational applications for the IWS1 phosphorylation pathway described in this report. First, *EGFR* mutant lung adenocarcinoma patients may benefit from a dual inhibition of EGFR Tyrosine Kinase Inhibitors (TKI) and specific AKT1/3 inhibitors (not currently investigated) or dual AKT/MEK inhibition (MK2206-AZD6244 : NCT01306045), limiting the ERK oncogenic addiction. Based on our recent results, IWS1 phosphorylation also regulates the expression of type I IFNs (Laliotis et al., 2020[96], Laliotis et al., 2021[97], Laliotis et al., 2021[98]). Furthermore, based on a recent report, the type I IFN signaling is a major determinant of EGFR TKI-sensitivity (Gong et al., 2020[97]). Given that the AKT/p-IWS1 axis

regulates type I IFN response, the proposed combined EGFR/AKT inhibition could determine the response of *EGFR* mutant LUAD patients through that pathway.

Secondly, the IWS1 phosphorylation-dependent cancer splicing addiction can be manipulated by synergistic use of EGFR TKI, inhibitors of the AKT/IWS1 axis and the use of highly isoform-specific antisense oligonucleotides (Splice Switching Oligonucleotides-SSO) and pharmacologic modulators of splicing machinery (Obeng et al., 2019[98]), which are currently under clinical trials (NCT03901469, NCT02711956, NCT02268552, NCT02908685).

Collectively, our results suggest that IWS1 phosphorylation by AKT acts as an epigenetic switch that, through H3K36me3-mediated signals, regulates RNA splicing in lung adenocarcinoma maintaining the oncogenic and splicing addiction, and may serve as a precision-medicine marker and important drug target in *EGFR* mutant lung adenocarcinoma.



## Methods

### Cells, Culture conditions, Growth factors and inhibitors.

NCI-H522, NCI-H1299, A549, NCI-H460, NCI-H1975, PC-9, NCI-H1650 and HBEC-hTERT cells were grown in Roswell Park Memorial Institute 1640 medium (Sigma-Millipore, Cat No. D8758) and HEK-293T cells were grown in Dulbecco's modified Eagle's medium (Sigma-Millipore, Cat No. D5796) supplemented with penicillin/streptomycin (Corning, Cat No. 30-002-CI), nonessential amino acids (Corning, Cat No. 25-025-CI), glutamine (Corning, Cat No. 25-005-CI), plasmocin 2.5ng/uL (Invivogen, Cat No. ant-mpp) and 10% fetal bovine serum. Cells were used for up to 5 passages. Cell lines were also periodically checked for mycoplasma, using the PCR mycoplasma detection kit (ABM, Cat No. G238). All experiments were carried out in mycoplasma-free cultures. IGF-1 (Cell Signaling

Technologies, Cat. No. 8917) (20ng/ml), was used to stimulate NCI-H522 or NCI-H1299 cells that had been serum-starved for 24 hours. The stimulation occurred for 4 hours. To inhibit AKT in cells growing in complete media, we treated them with the AKT inhibitor MK2206 (MERCK) (5 μM) for 4 hours. At this concentration, MK2206 inhibits all three AKT isoforms.

**Transfections and infections**

Retroviral constructs were packaged by transient transfection of these constructs in HEK-293T cells, in combination with ecotropic (Eco-pac) or amphotropic (Ampho-pac) packaging constructs. Lentivirus constructs were also packaged in HEK-293T cells by transient transfection of the constructs in combination with the packaging constructs psPax2 (Addgene #12260) and pMΔ2.G (Addgene #12259). Transfections were carried out using 2x HEPES Buffered Saline (Sigma, Cat. No 51558) and $CaCl_2$ precipitation.

After 48 hours of transient transfections of HEK-293T cells, the supernatant were collected and filtered. Infections were carried out in the presence of 8 μg/ml polybrene (Sigma, Cat. No. 107689). Depending on the selection marker in the vector, 48 hours after the infection, cells were selected for resistance to puromycin (Gibco, Cat. No. A11138) (10 μg/ml), G-418 (Cellgro, Cat. No. 30-234) (500μg/ml), or blasticidin (Gibco, Cat. No A1113903) (5 μg/ml). Cells infected with multiple constructs, were selected for infection with the first construct, prior to the next infection.

Transfection of lung adenocarcinoma cell lines with siRNAs (20 nM final concentration) were carried out, using the Lipofectamine 3000 Transfection Reagent (Invitrogen, Cat. No. 13778) and Opti-MEM Reduced Serum Medium (Gibco, Cat. no. 11058021), according to the manufacturer's protocol.

**Western Blotting**

Cells were lysed using a RIPA lysis buffer (50 mM Tris (pH 7.5), 0.1% SDS, 150 mM NaCl, 5 mM EDTA, 0.5% Sodium deoxycholate, 1% NP-40 and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails (Thermofisher, Cat. No 78444). Lysates were sonicated twice for 30 seconds and clarified by centrifugation at 18,000×*g* for 15 min at 4ºC. The clarified lysates were electrophoresed (20µg protein per lane) in SDS-PAGE. Electrophoresed lysates were transferred to polyvinylidene difluoride (PVDF) membranes (EMD Millipore Cat No. IPVH00010) in 25 mM Tris and 192 mM glycine. Following blocking with 5% nonfat dry milk in TBS and 0.1% Tween-20, the membranes were probed with antibodies (at the recommended dilution), followed by horseradish peroxidase-labeled secondary antibodies (1:2500), and they were developed with Pierce ECL Western Blotting Substrate (Thermo Scientific, cat. no 32106). The antibodies used are listed in Table S1.



**Image acquisition and figure preparation**

For the western blotting images, the acquisition was performed in Li-Cor Fc Odyssey Imaging System (LI-COR Biosciences, Lincoln, NE) using the 700 nm (protein ladder detection), 800 nm (reduced background and increased sensitivity) and chemi luminescent (protein bands) detection using a linear acquisition method. For the DNA agarose gels, the acquisition was performed in Li-Cor Fc Odyssey Imaging System using the 600 nm channel using a linear acquisition method. Identical approach was followed for all the images presented in this report to ensure unbiased analysis. In both conditions, the images were exported in high-quality image files (600 dpi png files) and further imported in Adobe Illustrator 2020 (Adobe, San Jose, CA) for figures preparation. The summary figures were designed in Bio Render (https://biorender.com).

**Library Preparation and Sequencing**

Total RNA was isolated from shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI H522 cells, using the PureLink RNA Kit (Invitrogen Cat No 12183018A), according to manufacturer's protocol. RNA samples were analyzed on Advanced Analytical Fragment Analyzer, using an RNA kit for integrity check and quantification. 100 – 500 ng of total RNA from each sample used as input for library preparation using the Illumina TruSeq stranded mRNA Library Preparation Kit (Cat. No. RS-122-2101) and they were individually indexed. Resulting libraries were quantified on Fragment Analyzer using a next generation sequencing (NGS) kit and the libraries of all the samples were pooled in equal molar concentration. The pooled library was sequenced on an Illumina HiSeq 2500 platform with Rapid V2 chemistry and paired-end 100 bases format. Sequencing results were demultiplexed with bcl2fastq and compressed, and demultiplexed fastq file pairs from each sample were used for analysis. The whole procedure was performed in Tufts University Core Genomic Facility (TUCF-http://tucf-genomics.tufts.edu).



**RNA-seq analysis**

Paired End sequencing (2x100 bp) of Poly-A+ selected RNAs was realized for shControl, shIWS1, shIWS1/WT-IWS1 rescue and shIWS1/phosphorylation site mutant IWS1 rescue NCI-H522 cells. All RNA-Seq experiments were performed in duplicates in a HiSeq 2500 system in the TUCF Genomics facility of Tufts University (http://tucf-genomics.tufts.edu). The average depth of sequenced samples was 37,5 M (±5M fragments). Data pre-processing and alignment was conducted as previously described (Vlahos et al, 2016[99]). RNA-Seq libraries were checked using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and sequence contaminants were detected and removed using an in-house

developed algorithm and additional software such as the Kraken suite and Cutadapt (Martin et al., 2011[100]). Paired end reads were aligned against the human reference genome (GRCh38/hg38) with GSNAP spliced aligners (Wu et al., 2011[101]). For gene and transcript annotation we utilized Ensembl v85 reference database (Aken et al., 2017[102]).

**Differential Gene expression and alternative splicing**

Gene and exon-level expressions were calculated by counting reads overlapping meta-gene and exon features using featureCounts (Liao et al., 2014[103]). DESeq (Anders et al., 2010[104]) and DEXSeq (Anders et al., 2012[105]) were employed to test for differential gene expression and differential exon usage, respectively.

*Differential gene expression analysis.* We performed differential gene expression analysis with R package DESeq, which utilizes a generalized linear model (GLM) and is applied directly to raw read counts. We identified 1,357 genes differentially modulated (p-value≤0.01, FDR≤0.2) in shIWS1 against shControl, and 417 genes differentially regulated (p-value≤0.01, FDR≤0.2) in shIWS1 plus S720A/T721A IWS1 MT-rescue juxtaposed with shIWS1 plus WT-rescue.

*Differential Exon usage.* In order to detect differentially spliced (DS) transcripts we applied DEXSeq. DEXSeq employs a GLM to model the differential exon-usage between sample groups. Our analysis resulted in 1,434 differentially employed exons assigned to 851 genes in the shIWS1 versus shControl (FDR≤0.05). Pairwise comparison of shIWS1 plus S720A/T721A *IWS1* MT-rescue with shIWS1 plus WT-rescue revealed 436 differentially utilized exonic regions, affecting 273 genes (FDR≤0.05).

Detailed list of differentially expressed genes and differential exon-usage events is provided in supplementary files. In both DESeq and DEXSeq analyses, false discovery rate (FDR) was controlled with the Benjamini-Hochberg method (Benjamini et al., 1995[106]).

Notably, 44 genes in shIWS1 plus S720A/T721A IWS1 MT-rescue were differentially expressed and alternatively spliced. In the comparison of shIWS1 versus shControl groups, we observed 165 genes common in DESeq and DEXSeq analyses.

**Gene-Set enrichment analysis**

GSEA was performed using the GSEA v2.0.13 software. All gene set files for this analysis were obtained from GSEA website (www.broadinstitute.org/gsea/). The Enrichment map was used for visualization of the GSEA results. Enrichment score (ES) and False discovery rate (FDR) value were applied to sort pathways enriched after gene set permutations were performed 1000 times for the analysis.



**Functional analysis of alternative RNA splicing events**

DEXSeq Log2 Fold Change values of differentially alternative splicing genes for NCI-H522 shIWS1 vs shControl and H522 shIWS1/MT-R vs shIWS1/WT-R were imported in the RStudio framework (V 3.5.2) for the Gene Ontology (GO) analysis. Specifically, GO analysis was performed by using the Bioconductor GOfuncR (Grote S. et al., 2020[107]) and annotating the alternative splicing genes to their biological processes. For each biological process, the number of associated genes and combined score, which is the absolute value of the sum of the Log2 Fold Change values of each gene associated with a specific biological process, were also calculated.

**RT-PCR and qRT-PCR**

Total cell RNA was extracted using the PureLink RNA Kit (Invitrogen, Cat. No 12183018A). cDNA was synthesized from 1.0 µg of total RNA, using oligo-dT priming and the QuantiTect Rev. Transcription Kit (QIAGEN, Cat No. 205310). Gene and exon expressions were quantified by quantitative real time RT-PCR, performed in triplicates, using the iTaq™

Universal SYBR® Green Super mix (Biorad, Cat No. 1725121) and a StepOne Plus qRT-PCR machine (Thermofisher). Data was normalized to hGAPDH or human 18S rRNA, which was used as an internal control. The primer sets used for all the real time PCR assays throughout this report are listed on the Table S2.

**Chromatin Immunoprecipitation and ChIP-Seq**

For the Chromatin Immunoprecipitation, we used $2x10^6$ cells. Proteins were initially cross-linked to DNA with 1% formaldehyde (Sigma, Cat. No F8775) for 15 minutes in 37°C. The cross-linking reaction was stopped by using 0.125M Glycine for 5 minutes in room temperature. Cells were lysed with Nuclear lysis buffer (50 mM Tris (pH 8.0), 10 mM EDTA, 0.5% SDS), in the presence of fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails (Thermofisher, Cat. No 78444) and incubated on ice for 10 minutes. Then, the cellular lysates were sonicated to shear DNA to 300- to 500-bp fragments. Following the sonication the lysates were clarified with centrifugation at 18,000g for 15 minutes in 4°C. The samples were then diluted to 5mL volume with IP Dilution buffer (16.7 mM Tris (pH 8.0), 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS). A fraction of each sample was then precleared with protein A and salmon sperm DNA-bound agarose beads (Cell Signaling Technologies, Cat. No 9863), for 1 hour in 4°C. Following overnight incubation with the immunoprecipitating antibody (Supplementary Table S1) or the Rabbit Isotype Control (Thermofisher, Cat. No 10500C) and 4h of incubation with Pierce™ Protein A/G Magnetic Beads (Thermofisher, Cat. No 88803) at 4°C, the immunoprecipitates were subjected to multiple washes. The washes were performed using Low Salt Wash Buffer (20 mM Tris(pH 8.0), 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS), High Salt Wash Buffer (20 mM Tris (pH 8.0), 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), LiCl Wash Buffer (10 mM Tris(pH 8.0), 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1%(w/v) deoxycholic acid) and TE buffer (10mM Tris (pH 8.0), 1mM EDTA). DNA recovered after reversion of the

protein-DNA cross-links with NaCl and incubation with proteinase K. Subsequently, it was extracted with DNA Purification Buffers and Spin Columns (Cell Signaling Technologies, Cat. No 14209). Real-time PCR using different sets of primers (Supplementary Table S2) to amplify the *U2AF2*, *FGFR2* and *GAPDH* genomic loci was carried out in the immunoprecipitated DNA, as well as in the 2% input and IgG-immunoprecipitated DNA, by using the iTaq™ Universal SYBR® Green Super mix (Biorad, Cat No. 1725121) and a StepOne Plus qRT-PCR machine (Thermofisher). The data were analysed using the analysis substrate file provided online by Sigma-Aldrich, calculating the fold enrichment. (https://www.sigmaaldrich.com/technical-documents/articles/biology/chip-qpcr-data-analysis.html). The detailed protocol and buffer preparation can be found in the online protocols depository (Laliotis et al., 2020[108]).



**ChIP-Seq library preparation and sequencing**

ChIP-seq libraries were generated using NEB Next® Ultra™ II DNA Library Prep Kit for Illumina® (New England Laboratories, Cat. No E7645) using standard manufacturer's protocol. The sequencing was performed on a HiSeq 2500 following Illumina protocols paired-end 100 bases format. Sequencing results were demultiplexed with bcl2fastq and compressed, and demultiplexed fastq file pairs from each sample were used for analysis. All the procedures were performed in the DNA Sequencing Center of Brigham Yount University. (Provo, UT) (https://biology.byu.edu/dnasc)

**ChIP-Seq analysis**

All ChIP-Seq experiments were performed in duplicates, and average depth of sequenced samples was 49M (±5M fragments). ChIP-Seq libraries were quality checked using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and sequence contaminants were detected and removed using Cutadapt (Martin et al., 2011[100]). Paired end reads were

aligned against the human reference genome (GRCh38/hg38) using Bowtie (version 2.2.6) (with default parameters). Peak discovery was performed with HOMER (version 4.6) (Heinz et al., 2010[109]). Sonicated input was used as a control for peak discovery. Snapshots of genomic areas were created using the Integrative Genomic Viewer (IGV) of the Broad Institute. (https://software.broadinstitute.org/software/igv/home) (Robinson et al., 2011[110])



**siRNAs, shRNAs, expression constructs and direct-site mutagenesis**

The origin of the siRNAs, shRNAs, and expression constructs is described in Table S3. The U2AF65 isoform α and β were cloned in pENTR/D-TOPO cloning vector (Invitrogen, Cat. No. 45-0218), with PCR-based techniques, using cDNA substrate derived from NCI-H522 shControl and NCI-H522 shIWS1 cells, respectively. To ensure successful amplification and separation, the PCR products were run in 1% agarose gel and were gel-purified using the NucleoSpin Gel and PCR Clean-Up kit (M&N, Cat. No. 740609.50). Following cloning in the entry vector, the two clones were recombined with pLx304 V5-DEST (Addgene #25890) using standard Clonase II LR mix (Thermofisher, Cat No 11791100). The Gateway LR reaction was incubated at room temperature overnight. The pDORN221 p52/LEDGF was obtained by DNAsu Plasmid Repository (DNAsu Plasmid Repository Clone : HsCD00000034). The p75/LEDGF isoform was cloned in pENTR/D-TOPO cloning vector (Invitrogen, Cat. No. 45-0218), with PCR-based techniques, using cDNA substrate derived from NCI-H522 shControl cells, with similar approach as U2AF65 isoforms. The pDONR201 AKT3 was obtained by DNAsu Plasmid Repository (DNAsu Plasmid Repository Clone : HsCD00005031). It was then cloned in pENTR/D-TOPO cloning vector, with PCR-based

techniques as described above. All of these constructs were then recombined with pLx304 V5-DEST with similar approach as U2AF65 isoforms. Site-directed mutagenesis for all the constructs used in this report were carried out using PCR-based techniques. The primers used for the site-directed mutagenesis and subcloning are described in Table S2. Briefly, using PCR-based techniques with overlapping set of primers for the targeted mutational site, the PCR product was incubated for 4 hours with 2uL DpnI (NEB, Cat. No R0176), at 37°C, for cleavage of the methylated PCR product. For the removal of the excess buffers, the DpnI-cleaved PCR product was purified using the NucleoSpin Gel and PCR Clean-Up kit. Subsequently, 5uL of the purified product was transformed using DH5-alpha Electrocompetent E.coli (NEB, Cat. No. C2986) on a Eppendorf Eporator® (Eppendorf, Cat. No. 4309000027). All constructs were sequenced in Genomic Shared Resource (GSR) of The Ohio State University(https://cancer.osu.edu/for-cancer-researchers/resources-for-cancer-researchers/shared-resources/genomics), prior to use.



**Immunoprecipitation**

Cells were initially lysed using a cytosolic Lysis Buffer 1 (LB1), Triton X-100 lysis buffer (50 mM Tris-HCL (pH 7.5), 20 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.25% Triton X-100, 10% Glycerol, 1mM DTT) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails (Thermofisher, Cat. No 78444) and were rotated for 10 minutes at 4°C. Then the lysates were clarified by centrifugation at 14,000 × *g* for 6 min. The supernatant, containing the cytosolic protein fraction, was aspirated and stored for future use. The precipitated nuclear fraction was further processed by using Lysis Buffer 2 (LB2) (10 mM Tris-HCL (pH 7.5), 20 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1mM DTT) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails. Subsequently, the lysates were clarified by centrifugation at

12,000 × *g* for 6 min. Following the discard of the supernatant fraction, the precipitated nuclei was further lysed with Lysis Buffer 3 (LB3) (10 mM Tris-HCL (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.5mM EGTA, 0.1% (w/v) sodium deoxycholate, 0.5% (v/v) N-lauroylsarcosine) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails. Then, the lysates were sonicated and clarified by centrifugation at 21,000 × *g* for 15 min. Pierce™ Protein A/G Magnetic Beads (Thermofisher, Cat. No 88803) were washed 3 times by LB3 for 5' minutes each. Following overnight incubation with the immunoprecipitating antibody (Supplementary Table S1) or the Mouse Isotype Control (Thermofisher, Cat. No 10400C) at 4°C, the conjugates were subjected to multiple washes with LB3 and 300uL of the clarified lysates were added in the Magnetic beads-Antibody conjugates, followed by overnight incubation at 4°C. The agarose bead-bound immunoprecipitates were washed five times with LB3, 5 min each time, at 4°C and they were electrophoresed (20μg protein per lane) in SDS-PAGE. Electrophoresed lysates or immunoprecipitates were transferred to polyvinylidene difluoride membranes in 25 mM Tris, 192 mM glycine. Following blocking with 5% nonfat dry milk in TBS and 0.1% Tween-20, the membranes were probed with antibodies (at the recommended dilution), followed by horseradish peroxidase-labeled secondary antibodies (1:2500), and they were developed with Pierce ECL Western Blotting Substrate (Thermo Scientific, cat. no 32106). In order to reduce the signal of the IgG heavy and light chains, the V5-tagged and endogenous U2AF65 was immunoprecipitated using monoclonal antibody produced in mouse and the immunoblotting for the detection of U2AF65 was performed with monoclonal antibody against V5-tag or U2AF65, raised in rabbit. The antibodies used are listed in Table S1. The detailed protocol and buffer preparation can be found in the online protocols depository (Laliotis et al., 2020[111], Mohammed et al., 2016[112]).

**RNA Immunoprecipitation**

For the RNA Immunoprecipitation, proteins were initially cross-linked to DNA with 1% formaldehyde (Sigma, Cat. No F8775) for 15 minutes in 37ºC. The cross-linking reaction was stopped by using 0.125M Glycine for 5 minutes in room temperature. Cells were scraped in the presence of 1mL 1x Phosphate Buffered Saline (PBS)-Nuclear Isolation Buffer (1.28M sucrose, 40mM Tris-HCl, 20mM $MgCl_2$ , 4% Triton-X 100)-$H_20$ (1:1:3 ratio). Following two additional washes, cell were lysed with RIP buffer (150mM KCl, 25mM Tris-HCl, 5mM EDTA, 0.5mM DTT, 0.5% NP-40), in the presence of fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails (Thermofisher, Cat. No 78444) and RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher, Cat. No. 10777019) and incubated on ice for 10 minutes. The lysates were clarified by centrifugation for 30' in 12,000xg at 4ºC. A fraction of each sample was then precleared with protein A and salmon sperm DNA-bound agarose beads (Cell Signaling Technologies, Cat. No 9863), for 1 hour in 4ºC. Following overnight incubation with the immunoprecipitating antibody (Supplementary Table S1) or the IgG isotype control (Rabbit Isotype Control (Thermofisher, Cat. No 10500C or Mouse Isotype Control (Thermofisher, Cat. No 10400C)) and 4h of incubation with Pierce™ Protein A/G Magnetic Beads (Thermofisher, Cat. No 88803) at 4°C, the immunoprecipitates were subjected to multiple washes. Beads were then washed four times using the RIP buffer and the RNA-protein complexes were eluted in 100uL RIP buffer and the RNA was recovered by reverse cross-linking at 70°C and proteinase K incubation at 55°C. The RNA was then extracted using phenol-chloroform-isoamyl alcohol mixture and glycogen (20mg/mL) (Thermofisher Cat No. R0561) precipitation overnight in -80°C. The RNA material was reverse-transcribed with random hexamers. Real-time PCR using different sets of primers

(Supplementary Table S2) to amplify the U2AF2, CDCA5 and GUSB pre-mRNA intron and exonic regions was carried out, in triplicates, in the immunoprecipitated material, as well as in the 2% input and IgG-immunoprecipitated RNA, by using the iTaq™ Universal SYBR® Green Super mix (Biorad, Cat No. 1725121) and a StepOne Plus qRT-PCR machine (Thermofisher). The data were analysed using the analysis substrate file provided online by Sigma-Aldrich.

(https://www.sigmaaldrich.com/technical-documents/articles/biology/chip-qpcr-data-analysis.html). SNRNP-70 binding in the human U1 snRNP gene, using the primers F : 5'-GGG AGA TAC CAT GAT CAC GAA GGT-3', R : 5'-CCA CAA ATT ATG CAG TCG AGT TTC CC-3', was used as the control for RNA IPs. The detailed protocol and buffer preparation can be found in the online protocols depository (Laliotis et al., 2020[113]).



**Subcellular Fractionation**

$5x10^6$ cells were washed 2 times with ice-cold 1x PBS and trypsinized. After centrifugation at 1,200 x g for 5 minutes, the pellet was resuspended in 1mL 1x PBS and aliquoted in two equal fractions. In the first fraction, the cells were lysed using a cytosolic Lysis Buffer 1 (LB1), Triton X-100 lysis buffer (50 mM Tris-HCL (pH 7.5), 20 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.25% Triton X-100, 10% Glycerol, 1mM DTT) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails (Thermofisher, Cat. No 78444) and were rotated for 10 minutes at 4°C. Then the lysates were clarified by centrifugation at 14,000 × *g* for 6 min. The supernatant, containing the cytosolic protein fraction, was aspirated and labeled appropriately for downstream applications. The precipitated nuclear fraction was further processed by using Lysis Buffer 2 (LB2) (10 mM Tris-HCL (pH 7.5), 20 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1mM DTT) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails. Subsequently, the lysates were clarified by centrifugation at 12,000 × *g* for 6 min.

Following the discard of the supernatant fraction, the precipitated nuclei was further lysed with Lysis Buffer 3 (LB3) (10 mM Tris-HCL (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.5mM EGTA, 0.1% (w/v) sodium deoxycholate, 0.5% (v/v) N-lauroylsarcosine) and fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails. Then, the lysates were sonicated and clarified by centrifugation at 21,000 × *g* for 15 min. The expression of Lamin A/C and GAPDH were determined by immunoblotting for validation of the fractionation of the nuclear and cytosolic protein compartment, respectively. In the second fraction, the cells were washed twice with TD buffer (135mM NaCl, 5mM KCl, 0.7mM $Na_2HPO_4$, 25mM Tris-HCl) and then lysed using TD/1% NP-40/RVC (Ribonucleoside-Vanadyl Complex, NEB, Cat. No. S1402) in the presence of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher, Cat. No. 10777019). Following 10 minutes ice-incubation and centrifugation at 21,000 x g for 1 minute, the supernatant, which contains the cytosolic part, was aspirated and kept on ice. The nuclear fraction was further washed with TD/0.5% NP-40/RVC twice. Then, the RNA from both fractions was isolated using Trizol and phenol-chloroform-isoamyl alcohol mixture and was ethanol precipitated overnight in -80°C. cDNA was synthesized from 1.0 µg of total RNA, using oligo-dT priming and the QuantiTect Rev. Transcription Kit. Cytosolic *CDCA5* mRNA expression was quantified by quantitative real time RT-PCR, performed in triplicates, using the iTaq™ Universal SYBR® Green Super mix and a StepOne Plus qRT-PCR machine. Data was normalized to *hGAPDH*, which was used as an internal control. The primer sets used for all the real time PCR assays throughout this report are listed on the Table S2.



**Cell Proliferation assay**

Cells were harvested by trypsinization, counted either using a manual hemocytometer or TC-20™ automated cell counter (Biorad, Hercules, CA), and plated evenly in 12 well tissue culture plates in 3 biological replicates for each condition (Empty Vector, shIWS1, shIWS1

CDCA5 WT rescue, shIWS1 CDCA5 S79A/S209A mutant rescue, shIWS1 CDCA5 S79E/S209E mutant rescue). Given different growth parameters for each cells, cells were plated according to their expected growth: H522 plated at 8000 cells/well, H1299 plated at 5000 cells/well, H1975 plated at 5000 cells/well, and A549 plated at 8000 cells/well. Photomicrographs were taken every 6 hours using an Incucyte live cell imager (Essen Biosciences, Ann Arbor, MI) depending on growth parameters of each cell line (NCI-H522 7 days, NCI-H1299 7 days, A549 7 days and NCI-H1975 12 days total acquisition). Images were taken and analyzed using the Incucyte confluence masking software (Essen Biosciences, Ann Arbor, MI) where the area of cellular confluence was identified, labeled, and measured in relation to total area to determine confluence. Software labeling of confluence was optimized for each cell line to ensure proper labeling of the entire cellular area with minimal to no background labeling. The same analysis parameters were used across different conditions within the same cell line to ensure unbiased analysis.



**Cell Transformation assay**

Cell transformation assays in immortalized HBEC hTERT were performed using the Cell Transformation Assay Kit-Colorimetric (Abcam Cat No. ab235698). Based on manufacturer's protocol, two layers of agarose were made (base and top layer). Prior to the initiation of the experiment, we performed a cell-dose curve by using 7 serial dilutions of cells (2-fold) and incubating them for 4 hours in 37°C with WST working solution. After that , the absorbance at 450nm was determined and the cell-dose curve was calculated ($y = \alpha x + \beta$), using linear regression on GraphPad Prism 8.4. In order to perform the assay, after solidification of the base agarose layer, $2.5x10^4$ HBEC hTERT cells per condition were mixed with top agarose layer in 10x DMEM solution and plated in 96-well plate, in triplicates along with blank wells. The cells were then plated for 7 days in 37°C and monitored for colony formation. After 7 days, the cells were imaged in the Incucyte live cell imager using the 20x lens. Then, the

cells were incubated for 4 hours on WST working solution at 37°C. The absorbance at 450nm was determined with a plate reader. Regarding the analysis, the average of the blank wells was subtracted from all the readings of the experimental conditions. Then, the final number of the transformed cells was calculated by inserting the corrected values in the cell-dose curve created prior to the experiment.

**FACS analysis**

The cells were plated in equal numbers and they were harvested from semi-confluent cultures 48 hours later. Semi-confluent cell cultures were harvested by trypsinization. The cellular pellet was resuspended in 700uL 1x PBS and fixed in 2.8mL ethanol, overnight at -20ºC. Following two washes with 1x PBS, the fixed cells were stained with Propidium Iodide mix (Propidium Iodide (1:2500) (Invitrogen, Cat. No. P3566), 0.1 mg/mL RNAse A (Invitrogen, Cat. No. 12091-039), 0.05% Triton-X) and incubated in the dark at 37ºC for 30 minutes. Subsequently, the cells were analysed on the BD FACS Calibur (BD Biosciences, San Jose, CA). All the experiments were performed in three biological replicates. All the raw values of the analysis can be found on Table S4. The analysis was performed in the Flow Cytometry Shared Resource (FCSR) of the Ohio State University (https://cancer.osu.edu/for-cancer-researchers/resources-for-cancer-researchers/shared-resources/flow-cytometry)



**Cell Sorting**

$2x10^6$ NCI-H1299 cells were cultured in standard conditions. Cells were harvested by trypsinization, counted either using a manual hemocytometer or TC-20™ automated cell counter (Biorad, Hercules, CA). Then, the cells were resuspended in DMEM volume for a final amount of $5x10^5$ cells/mL. The cells were stained by adding 2uL/mL Vybrant™ DyeCycle™ Ruby Stain (Thermo Fisher, Cat. No. V10309) and incubated in the dark at 37ºC

for 30 minutes. Subsequently, the cells were analysed on the BD FACS Aria III (BD Biosciences, San Jose, CA). Cellular fractions enriched for cells in G1, S and G2/M were harvested for protein, RNA and Chromatin. Regarding the protein extraction, cells were lysed using a RIPA lysis buffer and processed as described in the 'Western Blotting' section. Total cell RNA was extracted using the PureLink RNA Kit and further processed for the expression of IWS1, CDCA5, U2AF2 nad E2/E3 U2AF2 ratio as described in the 'RT-PCR and qRT-PCR' section. For Chromatin studies, we performed Chromatin Immuno-Cleavage (ChIC) studies, which will be described in the following section. The antibodies and primer sets used are described in table S1 and S2, respectively.



**Chromatin Immuno-Cleavage (ChIC)**

NCI-H1299 cells were sorted, as described above. Out of the cell fractions enriched for G0/G1, S and G2/M phase, $5x10^4$ cells were washed several times with wash buffer (20mM HEPES (pH 7.5), 150mM NaCl, 0.5mM Spermidine) in the presence of fresh 1x Halt™ Protease and Phosphatase Inhibitor Cocktails. Magnetic Biomag Plus Concanavalin A Beads (Bangs Laboratories, Cat. No. BP531) were activated with multiple washes using a binding buffer (20mM HEPES-KOH (pH 7.9), 10mM KCl, 1mM $CaCl_2$, 1mM $MnCl_2$). Prior to use, the immunoprecipitation antibodies (Supplementary Table S1) or the Rabbit Isotype Control (Thermofisher, Cat. No 10500C), were diluted in 1:50 dilution in 50 uL antibody buffer (2mM EDTA (pH 8.0), 0.1% (wt/vol) digitonin diluted in wash buffer). Then, the activated beads were resuspended with the antibody buffer, containing the immunoprecipitated antibody, and mixed with the cell fraction. Following overnight incubation at 4°C, the immunoprecipitates were subjected to multiple washes with the wash buffer. Similarly to the primary immunoprecipitating antibody, the Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) secondary antibody (Antibodies-Online, Cat. No. ABIN101961) was diluted in 1:50 dilution in 50uL antibody buffer, was mixed with the immunoprecipitates and incubated

for 4 hours at 4°C. Subsequently, the immunoprecipitates were subjected to multiple washes with the wash buffer and mixed with the CUTANA™ pAG-MNase (EpiCypher, Cat No. SKU: 15-1116) at 700 ng/mL. The targeted digestion was activated with 100mM $CaCl_2$ and occurred by incubation on ice for 30 minutes. The reaction was terminated with addition of 2x stop buffer (340mM NaCl, 20mM EDTA (pH 8.0), 4mM EGTA, 0.1% (wt/vol) digitonin, 0.2 mg RNAse A, 0.02 mg Glycogen) and the chromatin fragments were released by incubation at 37°C for 10 minutes. Subsequently, the chromatin fragments were extracted with DNA Purification Buffers and Spin Columns (Cell Signaling Technologies, Cat. No 14209). Real-time PCR using different sets of primers (Supplementary Table S2) to amplify the *U2AF2* genomic loci was carried out in the immunoprecipitated DNA, as well as in the IgG-immunoprecipitated DNA, by using the iTaq™ Universal SYBR® Green Super mix (Biorad, Cat No. 1725121) and a StepOne Plus qRT-PCR machine (Thermofisher). The data were analysed using the analysis substrate file provided online by Sigma-Aldrich, calculating the fold enrichment. (https://www.sigmaaldrich.com/technical-documents/articles/biology/chip-qpcr-data-analysis.html). This is based on the previously published protocol fo ChIC assays. (Skene et al., 2018[114])



**Ethics statement**

All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of the Ohio State University. IACUC protocol number 2018A00000134; P.I.: Lalit Sehgal, IACUC protocol number 2018A00000126; P.I.: Philip Tsichlis.

**Tumor xenograft**

The indicated cell numbers, $2x10^6$ cells (NCI-H1299 cell line), $5x10^6$ (A549 cell line) and $1x10^7$ (NCI-H1975 cell line) per injection, were mixed with 30% Matrigel (Corning, Cat. No.

356231) in PBS in a total volume of 200 µl and implanted subcutaneously into either side (left side for the shControl and right side for the shIWS1 group) of immunocompromised 6 week old NSG (NOD.*Cg-Prkdc*$^{scid}$ *Il2rg*$^{tm1Wjl}$/SzJ) mice. The mice were monitored every 3 days and the size of the tumors was measured using a digital caliper. The tumor volume was calculated by use of the modified ellipsoid formula: $\overline{V} = \frac{1}{2} v \times s^{2}$ (1/2 Length × Width$^2$). The mice were sacrificed 4 (NCI-H1299 and NCI-H1975 cell line) and 6 (A549 cell line) weeks after the injection. The tumors were removed and their weights were measured. One part of the dissected tumors was immediately snap frozen in liquid nitrogen prior to RNA and protein isolation, and a second part was fixed overnight in 10% (v/v) formalin (Sigma, Cat. No. HT501640), transferred to 70% EtOH and then embedded in paraffin at the Comparative Pathology & Mouse Phenotyping Shared Resource of the Ohio State University Comprehensive Cancer Center, prior to immunohistochemistry (IHC) staining.



**IHC staining**

Sections of 5 µm from the paraffin embedded mouse tumors or slides of the lung adenocarcinoma tissue array (US Biomax, LC1504) were heated to 55°C for 20 min prior to deparaffinization in xylene (Fisher scientific, Cat. No. X3F-1GAL). The slides were then rehydrated through graded ethanol concentrations up to distilled water. The endogenous peroxidase activity was blocked at RT by a 10 min incubation in the final developmental 3% $H_2O_2$ (Fisher Scientific, Cat. No. H325500) in PBS (pH 7.4), followed by antigen retrieval at 80 °C for 30min using the Citrate Buffer, pH 6.0, Antigen Retriever (Sigma, Cat. No. C9999). The Vectastain Elite ABC Universal kit peroxidase (Horse Anti-Mouse/Rabbit IgG) (Vector Laboratories, Cat. No. PK-6200) was used for blocking and incubation with the primary and secondary antibody according to the manufacturer's instructions. Briefly, the blocking serum (normal goat serum) was applied and the slide was incubated for 20 min at RT followed by

incubation with the primary antibody diluted in PBS with 2.5% normal serum O/N at 4 °C. The slide was then rinsed with PBS for 5 min and incubated with the biotinylated Universal Antibody for 30 min at RT. After a 5 min wash with PBS the slide was incubated for 30 min with the Vectastain Elite ABC reagent followed by a 2-10 min incubation with a DAB peroxidase substrate solution (Vector Laboratories, Cat. No. SK-400) according to the manufacturer's instructions. The slide was then washed in tap water and covered with the DPX mounting medium (Sigma, Cat. No. 06522). The primary antibodies are listed in table S1.

**RNA and protein isolation from the mouse xenograft tumors**

50-100 mg of the frozen mouse xenograft samples were homogenized using 1ml Trizol reagent (Thermofisher Scientific, Cat. No. 15596026). RNA and protein were extracted according to the manufacturer's instructions. Briefly, phase separation was performed by adding 200 µl of chloroform (Sigma, Cat. No. C2432) per 1ml of Trizol followed by centrifugation at 12,000 x *g* for 15 min at 4°C. The RNA of the clear upper aqueous phase was transferred into a new tube. The remaining organic phase was stored O/N at 4 °C for subsequent isolation of the proteins. **RNA extraction**: RNA was precipitated by mixing the samples of the aqueous phase with 0.5 ml isopropanol (Fisher scientific, Cat. No. A416P-4), incubation at RT for 15 min and spin at 12,000 *x g* for 10 min at 4 °C. The RNA pellet was washed with 75% ethanol followed by spin at 7,500 *x g* for 5 min at 4 °C and dissolved in 30 µl DEPC-treated water (IBI Scientific, Cat. No. IB42210). **Protein extraction:** 0.3 ml of 100% ethanol was added to the interphase-organic phase and the samples were centrifuged at 2,000 × *g* for 5 min at 4 °C. The proteins were then precipitated by adding 1.5 ml of isopropanol to the phenol-ethanol supernatant, incubation at RT for 10 min and spin at 12,000 × *g* for 10 min at 4 °C. The protein pellet was washed 3 times in a solution of 0.3 M guanidine hydrochloride (Sigma, Cat. No. SRE0066) in 95% ethanol. During each wash

cycle, the pellet was stored in the wash solution for 20 min at RT followed by centrifugation at 7,500 x *g* for 5 min at 4°C. The pellet was then washed by adding 100% ethanol followed by a 20 min incubation at RT and spin at 7,500 × *g* for 5 min at 4 °C. The protein pellet was dissolved in 200 µl 1% SDS with protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific, Cat. No. 78444). Any insoluble material was removed by centrifugation at 10,000 x *g* for 10 min at 4°C. Both protein and RNA materials from the tumors were further processed with immunoblotting and qRT-PCR, respectively, for the expression of various targets described in this report. The list of antibodies and primers used can be found in table S1 and table S2.



**Imaging**

All images were captured on the Nikon eclipse 50i microscope with attached Axiocam 506 color camera using the ZEN 2.6 blue edition software (Zeiss). Image processing and analysis was further performed using the ImageJ software, as described in the following section.

**Analysis of Ki-67 IHC signal**

Imaging files were imported to ImageJ (Schneider et al., 2012[115]). Using the native freehand gesture of the software, the signal from the granular areas of the tumor slides was detected and this signal number was divided to the total area of the selected slide section. For each slide, at least 5 different areas of the tumor were scanned. The average value of the signal divided to its different area, was the final value of the analysis. Identical approach and settings were followed for all the images to ensure unbiased analysis.

**Analysis of Tissue Microarrays IHC signal**

Similar to the mouse IHC analysis, images derived from 50 paired with NAT, lung adenocarcinoma samples were imported to image J and processed with the same analysis procedure. Since the tissue arrays contain two sections from each patient, the average of the two values was used for further expression correlation and clinical analysis.

**Human Tumor Samples.**

Thirty Lung Adenocarcinoma samples with matching normal adjacent tissue were obtained from the Tissue Bank of The Ohio State University, under the universal consenting and biobanking protocol, Total Cancer Care (TCC). TCC is the single protocol used by the Oncology Research Information Exchange Network (ORIEN), which was formed through a partnership between the OSUCCC – James and Moffitt Cancer Center in Tampa, FL. For more information, please advise the Biospecimen Core Services facility of The Ohio State university Comprehensive Cancer Center (https://cancer.osu.edu/for-cancer-researchers/resources-for-cancer-researchers/shared-resources/biospecimen-services) and ORIEN project (https://cancer.osu.edu/for-cancer-researchers/resources-for-cancer-researchers/orien). Ten additional lung adenocarcinomas without matching normal tissue had been obtained earlier from the tissue bank of Tufts Medical Center, after written consent and review of the bioethics board committee. The latter had been used also in an earlier study on the role of IWS1 in NSCLC (Sanidas et al., 2014[10]). The tumor samples in this study were provided as unidentified samples. Frozen tissues were dissected on dry ice into smaller pieces, in order to perform protein and RNA extraction. Tissues were then transferred into chilled 5mL polystyrene round bottom tubes. Protein was extracted by adding ice-cold 500uL NIH lysis buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% Triton-X and 5 mM EDTA) and

homogenizing the tissues using an electric homogenizer, leaving the tube on ice. Homogenized samples were moved into chilled microcentrifuge tubes, incubated for 40 min on ice and then centrifuged at 16,000x g for 20 minutes at 4°C. The supernatant was collected in a fresh tube and placed on ice for protein quantification, performed using BioRad Bradford Reagent (Biorad, Cat. No. 5000001). RNA was extracted by grinding tissue samples as before, in 1ml of Trizol Reagent (Thermo Fisher, Cat. No. 15596026) and then performing the extraction as per manufacturer's instructions. Ten unpaired lung adenocarcinoma samples were described previously (Sanidas et al., 2014[10]).



**LUAD Samples Correlation analysis**

The western blot images, the qRT-PCR *U2AF2* E2/E3 ratio derived from the 40 LUAD patients cohort and the tissue microarray IHC signals were used for correlation analysis. The WB images for the various components of the IWS1/Sororin pathway were imported to ImageJ and the intensity of the bands were measured. The normalized to tubulin values, *U2AF2* E2/E3 ratio and the IHC signals, were imported to GraphPad Prism 8.4 and the correlation coefficients were calculated using simple linear regression. Following this, these correlation values were visualized with a heat map. The exact values and statistics from the correlation analysis, can be found in Supplementary Table S5.

**LUAD Samples clinical characteristics and survival analysis analysis**

For the WB LUAD samples and TMA samples, the clinical stage information was available. For the analysis, the patients were grouped into Stage I and Stage II/III groups. The expression of IWS1 phosphorylation, *U2AF2* E2/E3 ratio and the IHC signal of the IWS1/Sororin axis were imported and visualized in GraphPad Prism 8.4. For the TMA samples the histological grade was available, the patients were grouped into Grade 2 and Grade 3 groups and the same approach was followed. The statistics were performed as

described in the corresponding sections. For the 30 paired WB LUAD samples, the survival data were available. Based on the present or not of *KRAS* or *EGFR* mutations, the patients were subgrouped into *KRAS* or *EGFR* mutated groups, along with the entire cohort. Regarding the cut-off of IWS1 expression or *U2AF2* E2/E3 ratio on the above patients, the high patient group was defined as expression or ratio higher than the mean of the samples, plus one standard deviation (x+1s) and the low group as expression or ratio lower than that mark. The visualization of the data and the statistics were performed in GraphPad Prism 8.4 with a Kaplan-Meier method (Goel et al., 2010[116]) and logrank p statistics.



**TCGA analysis**

TCGA data were downloaded from https://portal.gdc.cancer.gov/. Overall 658 TCGA-LUAD patients (all stages), 162 TCGA-LUAD (all stages) with *KRAS* mutations and 67 TCGA-LUAD with *EGFR* mutations were obtained with their survival and clinical data (clinical stage, Metastasis, Survival). A total of 516 out of 658 had RNA-seq data. For the expression of SRSF1, 128 out of 658 had RNA-Seq data. Figures for the correlation analysis were generated using the visualization tools of the Xena browser (http://xena.ucsc.edu/.) and GraphPad Prism 8.4. The clinical stage and Metastasis analysis for the TCGA-LUAD patients was performed as described above. The Survival analysis was performed in GraphPad Prism 8.4 with a Kaplan-Meier method and log rank p statistics, with similar approach to the cut-off decision as described above.

**Microarrays datasets**

Lung adenocarcinoma gene expression was retrieved from two different publicly available Agilent microarray datasets present on GEO (GSE13213 (Tomida et al., 2009[78]), GSE26939 (Wilkerson et al., 2012[117]) by using Log Ratio values calculated by the authors and imported in the RStudio framework (V 3.5.2) for the gene selection (IWS1, CDCA5, CDC2, CCNB1).

Log Ratios of the red and green processed signals (i.e.background-corrected and dye-normalized) were calculated, as well as a log ratio error and a p-value for each feature. For gene expression 2-color arrays, these calculations assess the confidence you can have that the gene is or is not differentially expressed. For these datasets, the mutational status of *KRAS* and *EGFR* was available along with clinical stage, relapse and survival information. The expression of components of the IWS1/Sororin pathway in the GSE13213 was visualized as a heat map using the z-scores of the expression in the individual patients. The survival information for both the GSE13213 and GSE26939 datasets were pooled and performed as described above. All the figures and statistics were conducted in GraphPad Prism 8.4.



**Brain Metastasis RNA-seq analysis**

Expression values form 6 dissected brain metastasis derived from lung adenocarcinoma patients were obtained from RNA-seq data provided in the publicly available dataset GSE141685. The FPKM values were downloaded from the provided analysis. The exon specific expression of *U2AF2* E2,E3 and *FGFR2* IIIb, IIIc is described in the following section. As primary lung adenocarcinoma tumors, we used RNA-seq data from 24, randomly selected, TCGA-LUAD patients independent of their mutational, clinical or survival status. The TCGA ID of these patients along with their clinical information can be found in Supplementary Table S6. The results were visualized as heat maps in GraphPad Prism 8.4

**Exon-specific normalized expression**

The exon expression profile of TCGA Lung Adenocarcinoma samples and GSE141685 was measured experimentally using the Illumina HiSeq 2000 RNA Sequencing platform. Exons were mapped onto the human genome coordinates using UCSC Xena unc_RNAseq_exon probeMap and exon-level transcription estimation was presented in RPKM values (Reads

Per Kilobase of exon model per Million mapped reads). Log2 (RPKM+1) exon expression matrix was then imported in the R Studio framework (V 3.5.2) for exon (exon 2 and 3 of *U2AF2*, exon IIIb and IIIc of *FGFR2*) and sample selection.

**scRNA-Seq Analysis and t-SNE**

The normalized counts from scRNA-Seq data derived from a mouse lung adenocarcinoma metastatic model (GSE123903) (Laughney et al., 2020[79]), were retrieved and analyzed for *IWS1* and *CDCA5* expression. The data were visualized using the Barnes–Hut approximate version of *t*-SNE (Maaten et al., 2008[118]) (https://github.com/lvdmaaten/bhtsne)



**Code availability**

All the code used for the analysis in this report is derived from previously published reports. It is also explained and cited in the appropriate materials and methods or supplementary experimental procedures sections.

**Data availability**

All the raw data underlying figures 1-8 (uncropped gel images, qPCR, FACS, plates reader and proliferation data), processed counts of RNA-Seq, Junction-Seq splicing platform for visualization of the exon usage results and microscope images derived from this report have been deposited in the Mendeley Dataset in two independent publicly available datasets (Laliotis et al., 2020[119,120]). Specific *P* values are also included in these datasets. The main uncropped full scans are provided as source data with the paper. All the RNA-seq and ChIP-Seq data in this report have been deposited in the Gene Expression Omnibus (GEO), under the superseries accession number GSE166955 (RNA-seq : GSE166953, ChIP-Seq : GSE165964).

## Statistics and reproducibility

The experiments in Fig. 1f-g, S1e-j, 2a-c, S2a-c, 3a-d, S3a-c, S4a-c, 4a-g, S5a-g, S6a-d, 5b-h, S7a-f, S8b-c, 6a-i, S9a-b were performed at least in 3 independent biological experiments. The data in figure 7 (mouse xenografts) were performed once, using 5 mice/group. The analysis of LUAD samples in 8a was performed two times. The IHC staining of xenografts derived tumors in S10c and of the patient derived tissue-microarrays, was performed once with the antibodies and techniques outlined in the methods section. All the statistical analysis was performed in GraphPad Prism, as described in the corresponding section. All the statistical analysis reports can be found in the Mendeley dataset where the source data of this report were deposited. (Laliotis et al., 2020[119])



## Acknowledgments

The authors wish to thank all the members of the Tsichlis Lab for helpful discussions. We also thank Drs Carlo M. Croce, Michael Freitas, Wayne Miles, Raphael Pollock, Amanda Toland, Samir Achaya, Joal Beane and Esmerina Tilli for reviewing the manuscript before the submission. This work was supported by the NIH grant R01 CA186729 to P.N.T., the NIH grant R01 CA198117 to P.N.T and V.C, by start-up funds from the OSUCCC to P.N.T,, from the National Center for Advancing Translational Sciences grant KL2TR002734 to L.S. G.I.L is supported by the Pelotonia Post-Doctoral fellowship from OSUCCC.

## Author Contributions

**G.I.L.** Conceptualization, overall experimental design. Performed all the experiments, analyzed the data, prepared the figures, performed and analyzed the ChIP-seq and wrote the manuscript. **E.C.** Designed and performed the mouse xenografts, protein and RNA extraction of mouse tumors, performed the mouse xenografts and human TMA staining

studies and picture acquisition and edited the manuscript **M.D.P.** Bioinformatics analyses of RNA-seq data. **A.C.** Performed RT-PCR experiments. **A.L.F.** Bioinformatics analyses of Microarray and TCGA data. **S.S.** Assisted in FACS experiments **V.A.** Analyzed the proliferation experiments. **K.A.N** Performed RT-PCR experiment and assisted in site-directed mutagenesis **S.A.** Bioinformatics analyses of scRNA-Seq, Overall Bioinformatics supervision. **A.O.** Extraction of the tumor samples and edited the manuscript **V.T.** Performed the RNA-seq. **I.V.** Bioinformatics analyses of RNA-seq data. **M.C.** Extraction of the tumor samples. **A.H.** Bioinformatics analyses of RNA-seq data. **D.P.** Experimental design and edited the manuscript. **C.T.** Advised on the design of experiments. **L.S.** Advised on the design of experiments. **D.P.C.** Advised on the biology of lung cancer and on the design of experiments, provided cell lines and reagents. **V.C.** Contributed to overall experimental design. **P.N.T.** Overall experimental design, manuscript writing and editing.



## Competing Interests

The authors declare no competing interests.

## Corresponding authors

Correspondence to : Philip N. Tsichlis and Georgios I. Laliotis

## References

1. Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G., 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*, *582*(14), pp.1977-1986.
2. Jewer, M., Findlay, S.D. and Postovit, L.M., 2012. Post-transcriptional regulation in cancer progression. *Journal of cell communication and signaling*, *6*(4), pp.233-248.

3. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, *40*(12), p.1413.
4. Paronetto, M.P., Passacantilli, I. and Sette, C., 2016. Alternative splicing and cell survival: from tissue homeostasis to disease. *Cell Death & Differentiation*, *23*(12), pp.1919-1929.
5. Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F. and Black, D.L., 2016. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell*, *166*(5), pp.1147-1162.
6. Coomer, A.O., Black, F., Greystoke, A., Munkley, J. and Elliott, D.J., 2019. Alternative splicing in lung cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*.
7. Siegel, R.L., Miller, K.D. and Jemal, A., 2016. Cancer statistics, 2016. *CA: a cancer journal for clinicians*, *66*(1), pp.7-30.
8. Yuan, M., Huang, L.L., Chen, J.H., Wu, J. and Xu, Q., 2019. The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. *Signal Transduction and Targeted Therapy*, *4*(1), pp.1-14.
9. Sequist, L.V., Neal, J.W., Jett, J.R. and Ross, M.E., 2016. Personalized, genotype directed therapy for advanced non-small cell lung cancer. *UpToDate, UpToDate*.
10. Sanidas, I., Polytarchou, C., Hatziapostolou, M., Ezell, S.A., Kottakis, F., Hu, L., Guo, A., Xie, J., Comb, M.J., Iliopoulos, D. and Tsichlis, P.N., 2014. Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Molecular cell*, *53*(4), pp.577-590.
11. Li, Z., Li, Q., Han, L., Tian, N., Liang, Q., Li, Y., Zhao, X., Du, C. and Tian, Y., 2016. Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines. *Oncology reports*, *35*(2), pp.1013-1019.

12. Todaro, M., Gaggianesi, M., Catalano, V., Benfante, A., Iovino, F., Biffoni, M., Apuzzo, T., Sperduti, I., Volpe, S., Cocorullo, G. and Gulotta, G., 2014. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell stem cell*, *14*(3), pp.342-356.
13. Oltean, S. and Bates, D.O., 2014. Hallmarks of alternative splicing in cancer. *Oncogene*, *33*(46), pp.5311-5318.
14. Calabretta, S., Bielli, P., Passacantilli, I., Pilozzi, E., Fendrich, V., Capurso, G., Delle Fave, G. and Sette, C., 2016. Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. *Oncogene*, *35*(16), pp.2031-2039.
15. Vo, N.S., Reynolds, S.M., Zhao, Y., Climente-González, H., Chai11, S., Wang, F., Varghese, R., Huang, M., Liang, W.W., Wyczalkowski, M.A. and Sengupta, S., Systematic Analysis of Splice-Site-Creating Mutations in Cancer.
16. Giaj Levra, M., Andre Olaussen, K., Novello, S. and Soria, J.C., 2014. PARP inhibitors: an interesting pathway also for non-small cell lung cancer?. *Current pharmaceutical design*, *20*(24), pp.3875-3882.
17. Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., Agrawal, A.A., Caleb, B., Csibi, A., Sean, E. and Fekkes, P., 2018. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature medicine*, *24*(4), p.497.
18. Listerman, I., Sapra, A.K. and Neugebauer, K.M., 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nature structural & molecular biology*, *13*(9), p.815.
19. Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M. and Misteli, T., 2010. Regulation of alternative splicing by histone modifications. *Science*, *327*(5968), pp.996-1000.

20. Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R. and Bickmore, W.A., 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS genetics*, *8*(5).

21. Chen, M. and Manley, J.L., 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology*, *10*(11), pp.741-754.

22. Obeng, E.A., Stewart, C. and Abdel-Wahab, O., 2019. Altered RNA Processing in Cancer Pathogenesis and Therapy. *Cancer discovery*, *9*(11), pp.1493-1510.

23. Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T. and Chabot, B., 2008. 8 hnRNP proteins and splicing control. *Advances in Experimental Medicine & Biology*, *623*, p.123.

24. de Miguel, F.J., Sharma, R.D., Pajares, M.J., Montuenga, L.M., Rubio, A. and Pio, R., 2014. Identification of alternative splicing events regulated by the oncogenic factor SRSF1 in lung cancer. *Cancer research*, *74*(4), pp.1105-1115.

25. Chen, M., David, C.J. and Manley, J.L., 2012. Concentration-dependent control of pyruvate kinase M mutually exclusive splicing by hnRNP proteins. *Nature structural & molecular biology*, *19*(3), p.346.

26. Jin, W., Bruno, I.G., Xie, T.X., Sanger, L.J. and Cote, G.J., 2003. Polypyrimidine tract-binding protein down-regulates fibroblast growth factor receptor 1 α-exon inclusion. *Cancer research*, *63*(19), pp.6154-6157.

27. Dominguez, D., Tsai, Y.H., Weatheritt, R., Wang, Y., Blencowe, B.J. and Wang, Z., 2016. An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife*, *5*, p.e10288.

28. Moore, M.J., Wang, Q., Kennedy, C.J. and Silver, P.A., 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell*, *142*(4), pp.625-636.

29. Ahn, E.Y., DeKelver, R.C., Lo, M.C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.D. and Zhang, D.E., 2011. SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Molecular cell*, *42*(2), pp.185-198.

30. Suvorova, E.S., Croken, M., Kratzer, S., Ting, L.M., de Felipe, M.C., Balu, B., Markillie, M.L., Weiss, L.M., Kim, K. and White, M.W., 2013. Discovery of a splicing regulator required for cell cycle progression. *PLoS Genet*, *9*(2), p.e1003305.

31. Shen, H. and Green, M.R., 2004. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Molecular cell*, *16*(3), pp.363-373.

32. Glasser, E., Agrawal, A.A., Jenkins, J.L. and Kielkopf, C.L., 2017. Cancer-associated mutations mapped on high-resolution structures of the U2AF2 RNA recognition motifs. *Biochemistry*, *56*(36), pp.4757-4761.

33. Li, J., Cheng, D., Zhu, M., Yu, H., Pan, Z., Liu, L., Geng, Q., Pan, H., Yan, M. and Yao, M., 2019. OTUB2 stabilizes U2AF2 to promote the Warburg effect and tumorigenesis via the AKT/mTOR signaling pathway in non-small cell lung cancer. *Theranostics*, *9*(1), p.179.

34. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), pp.15545-15550.

35. Anders, S., Reyes, A. and Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. *Nature Precedings*, pp.1-1.

36. Hacker, K.E., Fahey, C.C., Shinsky, S.A., Chiang, Y.C.J., DiFiore, J.V., Jha, D.K., Vo, A.H., Shavit, J.A., Davis, I.J., Strahl, B.D. and Rathmell, W.K., 2016. Structure/Function analysis of recurrent mutations in SETD2 protein reveals a critical

and conserved role for a SET domain residue in maintaining protein stability and histone H3 Lys-36 trimethylation. *Journal of Biological Chemistry*, *291*(40), pp.21283-21295.

37. Greer, E.L. and Shi, Y., 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, *13*(5), p.343.
38. Hyun, K., Jeon, J., Park, K. and Kim, J., 2017. Writing, erasing and reading histone lysine methylation. *Experimental & molecular medicine*, *49*(4), p.e324.
39. Lucio-Eterovic, A.K., Singh, M.M., Gardner, J.E., Veerappan, C.S., Rice, J.C. and Carpenter, P.B., 2010. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences*, *107*(39), pp.16952-16957.
40. Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D. and Xu, R.M., 2011. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *Journal of Biological Chemistry*, *286*(10), pp.8361-8368.
41. Berdasco, M., Ropero, S., Setien, F., Fraga, M.F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N.K., Rahman, N. and Esteller, M., 2009. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proceedings of the National Academy of Sciences*, *106*(51), pp.21830-21835.
42. Rahman, S., Sowa, M.E., Ottinger, M., Smith, J.A., Shi, Y., Harper, J.W. and Howley, P.M., 2011. The Brd4 extra terminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Molecular and cellular biology*, *31*(13), pp.2641-2652.
43. Brown, M.A., Sims, R.J., Gottlieb, P.D. and Tucker, P.W., 2006. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine

36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Molecular cancer*, *5*(1), p.26.

44. Gregory, G.D., Vakoc, C.R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E. and Blobel, G.A., 2007. Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Molecular and cellular biology*, *27*(24), pp.8466-8479.
45. Bannister, A.J., Schneider, R. and Kouzarides, T., 2002. Histone methylation: dynamic or static?. *Cell*, *109*(7), pp.801-806.
46. Berry, W.L. and Janknecht, R., 2013. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer research*, *73*(10), pp.2936-2942.
47. Cloos, P.A., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K.H. and Helin, K., 2006. The putative oncogene GASC1 demethylates tri-and dimethylated lysine 9 on histone H3. *Nature*, *442*(7100), p.307.
48. Li, W., Zhao, L., Zang, W., Liu, Z., Chen, L., Liu, T., Xu, D. and Jia, J., 2011. Histone demethylase JMJD2B is required for tumor cell proliferation and survival and is overexpressed in gastric cancer. *Biochemical and biophysical research communications*, *416*(3-4), pp.372-378.
49. Cheung, H.C., Hai, T., Zhu, W., Baggerly, K.A., Tsavachidis, S., Krahe, R. and Cote, G.J., 2009. Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain*, *132*(8), pp.2277-2288.
50. Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D. and Strahl, B.D., 2013. An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Molecular cell*, *49*(3), pp.571-582.
51. Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Göttgens, B. and Bycroft, M., 2010. Molecular basis of histone

H3K36me3 recognition by the PWWP domain of Brpf1. *Nature structural & molecular biology*, *17*(5), pp.617-619.

52. Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W. and Li, G.M., 2013. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell*, *153*(3), pp.590-600.
53. Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G. and Mann, M., 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, *142*(6), pp.967-980.
54. Shun, M.C., Botbol, Y., Li, X., Di Nunzio, F., Daigle, J.E., Yan, N., Lieberman, J., Lavigne, M. and Engelman, A., 2008. Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *Journal of virology*, *82*(23), pp.11555-11567.Singh, P.K., Plumb, M.R., Ferris, A.L., Iben, J.R., Wu, X., Fadel, H.J., Luke, B.T., Esnault, C., Poeschla, E.M., Hughes, S.H. and Kvaratskhelia, M., 2015. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes & development*, *29*(21), pp.2287-2297.
55. Ferris, A.L., Wu, X., Hughes, C.M., Stewart, C., Smith, S.J., Milne, T.A., Wang, G.G., Shun, M.C., Allis, C.D., Engelman, A. and Hughes, S.H., 2010. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proceedings of the National Academy of Sciences*, *107*(7), pp.3135-3140.
56. Paz, I., Kosti, I., Ares Jr, M., Cline, M. and Mandel-Gutfreund, Y., 2014. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic acids research*, *42*(W1), pp.W361-W367.

57. Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N.H., Raimer, A., Sun, S., Jensen, M.A., Hua, Y. and Allain, F.H.T., 2015. SRSF1-regulated alternative splicing in breast cancer. *Molecular cell*, *60*(1), pp.105-117.

58. Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M. and Vagner, S., 2006. An interaction between U2AF 65 and CF Im links the splicing and 3′ end processing machineries. *The EMBO journal*, *25*(20), pp.4854-4864.

59. Shen, H. and Green, M.R., 2004. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Molecular cell*, *16*(3), pp.363-373.

60. Chan, S.P., Kao, D.I., Tsai, W.Y. and Cheng, S.C., 2003. The Prp19p-associated complex in spliceosome activation. *Science*, *302*(5643), pp.279-282.

61. Chan, S.P. and Cheng, S.C., 2005. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *Journal of Biological Chemistry*, *280*(35), pp.31190-31199.

62. Hogg, R., McGrail, J.C. and O'Keefe, R.T., 2010. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing.

63. Chanarat, S. and Sträßer, K., 2013. Splicing and beyond: the many faces of the Prp19 complex. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1833*(10), pp.2126-2134.

64. David, C.J., Boyne, A.R., Millhouse, S.R. and Manley, J.L., 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. *Genes & development*, *25*(9), pp.972-983.

65. Watrin, E., Demidova, M., Watrin, T., Hu, Z. and Prigent, C., 2014. Sororin pre‐mRNA splicing is required for proper sister chromatid cohesion in human cells. *EMBO reports*, *15*(9), pp.948-955.

66. Wang, J., Xia, C., Pu, M., Dai, B., Yang, X., Shang, R., Yang, Z., Zhang, R., Tao, K. and Dou, K., 2018. Silencing of CDCA5 inhibits cancer progression and serves as a prognostic biomarker for hepatocellular carcinoma. *Oncology reports*, *40*(4), pp.1875-1884.

67. Shen, A., Liu, L., Chen, H., Qi, F., Huang, Y., Lin, J., Sferra, T.J., Sankararaman, S., Wei, L., Chu, J. and Chen, Y., 2019. Cell division cycle associated 5 promotes colorectal cancer progression by activating the ERK signaling pathway. *Oncogenesis*, *8*(3), p.19.

68. Nguyen, M.H., Koinuma, J., Ueda, K., Ito, T., Tsuchiya, E., Nakamura, Y. and Daigo, Y., 2010. Phosphorylation and activation of cell division cycle associated 5 by mitogen-activated protein kinase play a crucial role in human lung carcinogenesis. *Cancer research*, *70*(13), pp.5337-5347.

69. Guo, Y.J., Pan, W.W., Liu, S.B., Shen, Z.F., Xu, Y. and Hu, L.L., 2020. ERK/MAPK signalling pathway and tumorigenesis. *Experimental and Therapeutic Medicine*, *19*(3), pp.1997-2007.

70. Wee, P. and Wang, Z., 2017. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers*, *9*(5), p.52.

71. Kent, L.N. and Leone, G., 2019. The broken cycle: E2F dysfunction in cancer. *Nature Reviews Cancer*, *19*(6), pp.326-338.

72. Begum, J., Day, W., Henderson, C., Purewal, S., Cerveira, J., Summers, H., Rees, P., Davies, D. and Filby, A., 2013. A method for evaluating the use of fluorescent dyes to track proliferation in cell lines by dye dilution. *Cytometry Part A*, *83*(12), pp.1085-1095.

73. Dronamraju, R., Jha, D.K., Eser, U., Adams, A.T., Dominguez, D., Choudhury, R., Chiang, Y.C., Rathmell, W.K., Emanuele, M.J., Churchman, L.S. and Strahl, B.D.,

2018. Set2 methyltransferase facilitates cell cycle progression by maintaining transcriptional fidelity. *Nucleic acids research*, *46*(3), pp.1331-1344.

74. Liu, P., Begley, M., Michowski, W., Inuzuka, H., Ginzberg, M., Gao, D., Tsou, P., Gan, W., Papa, A., Kim, B.M. and Wan, L., 2014. Cell-cycle-regulated activation of Akt kinase by phosphorylation at its carboxyl terminus. *Nature*, *508*(7497), pp.541-545.
75. Kim, C.H., Kim, S.H., Park, S.Y., Yoo, J., Kim, S.K. and Kim, H.K., 2015. Identification of EGFR mutations by immunohistochemistry with EGFR mutation–specific antibodies in biopsy and resection specimens from pulmonary adenocarcinoma. *Cancer research and treatment: official journal of Korean Cancer Association*, *47*(4), p.653.
76. Popper, H.H., 2016. Progression and metastasis of lung cancer. *Cancer and Metastasis Reviews*, *35*(1), pp.75-91.
77. West, H.J., Shaw, A.T. and Schild, S.E., 2018. Brain metastases in non-small cell lung cancer.
78. Tomida, S., Takeuchi, T., Shimada, Y., Arima, C., Matsuo, K., Mitsudomi, T., Yatabe, Y. and Takahashi, T., 2009. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *Journal of clinical oncology*, *27*(17), pp.2793-2799.
79. Laughney, A.M., Hu, J., Campbell, N.R., Bakhoum, S.F., Setty, M., Lavallée, V.P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S. and Allaj, V., 2020. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nature Medicine*, *26*(2), pp.259-269.
80. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. and Antipin, Y., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.

81. Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B. and Fulton, L., 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, *455*(7216), pp.1069-1075.

82. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A. and Sougnez, C., 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*, *150*(6), pp.1107-1120.

83. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S. and Miller, M.L., 2015. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science*, *348*(6230), pp.124-128.

84. Jordan, E.J., Kim, H.R., Arcila, M.E., Barron, D., Chakravarty, D., Gao, J., Chang, M.T., Ni, A., Kundra, R., Jonsson, P. and Jayakumaran, G., 2017. Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer discovery*, *7*(6), pp.596-609.

85. Chen, J., Yang, H., Teo, A.S.M., Amer, L.B., Sherbaf, F.G., Tan, C.Q., Alvarez, J.J.S., Lu, B., Lim, J.Q., Takano, A. and Nahar, R., 2020. Genomic landscape of lung adenocarcinoma in East Asians. *Nature Genetics*, *52*(2), pp.177-186.

86. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. and Antipin, Y., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.

87. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. and Cerami, E., 2013. Integrative analysis of

complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, *6*(269), pp.pl1-pl1.

88. Maji, D., Glasser, E., Henderson, S., Galardi, J., Pulvino, M.J., Jenkins, J.L. and Kielkopf, C.L., 2020. Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. *Journal of Biological Chemistry*, pp.jbc-RA120.
89. Ishiguro, K.I., 2019. The cohesin complex in mammalian meiosis. *Genes to Cells*, *24*(1), pp.6-30.
90. Dreier, M.R., Bekier, M.E. and Taylor, W.R., 2011. Regulation of sororin by Cdk1-mediated phosphorylation. *Journal of cell science*, *124*(17), pp.2976-2987.
91. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S. and Taatjes, D.J., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), pp.430-435.
92. Karamysheva, Z., Díaz-Martínez, L.A., Warrington, R. and Yu, H., 2015. Graded requirement for the spliceosome in cell cycle progression. *Cell Cycle*, *14*(12), pp.1873-1883.
93. Hofmann, J.C., Husedzinovic, A. and Gruss, O.J., 2010. The function of spliceosome components in open mitosis. *Nucleus*, *1*(6), pp.447-459.
94. Bonnal, S.C., López-Oreja, I. and Valcárcel, J., 2020. Roles and mechanisms of alternative splicing in cancer—implications for care. *Nature Reviews Clinical Oncology*, pp.1-18.
95. Wang, E. and Aifantis, I., 2020. RNA Splicing and Cancer. *Trends in Cancer*.
96. Georgios I. Laliotis, Adam D. Kenney, Evangelia Chavdoula, Arturo Orlacchio, Abdul K. Kaba, Alessandro La Ferlita, Vollter Anastas, Christos Tsatsanis, Joal D. Beane, Lalit Sehgal, Vincenzo Coppola, Jacob S. Yount, Philip N. Tsichlis, bioRxiv 2020.12.26.424461, IWS1 phosphorylation by AKT kinase controls the

nucleocytoplasmic export of type I IFNs and the sensitivity of lung adenocarcinoma cells to oncolytic viral infection, through *U2AF2* RNA splicing. *https://doi.org/10.1101/2020.12.26.424461*

97. Georgios I. Laliotis, Adam D. Kenney, Evangelia Chavdoula, Arturo Orlacchio, Vollter Anastas, Alessandro La Ferlita, Abdul Kaba, Lalit Sehgal, Vincenzo Coppola, Jacob S. Yount and Philip N. Tsichlis, Sensitivity of cancer cells to oncolytic viruses is defined by IWS1 phosphorylation dependent epigenetic regulation of U2AF2 splicing and nucleocytoplasmic export of type I IFN transcripts, Cancer Immunol Res February 1 2021 (9) (2 Supplement) PO088; **DOI:**10.1158/2326-6074.TUMIMM20-PO088
98. Georgios I. Laliotis, Evangelia Chavdoula, Abdul Kaba, Alessandro La Ferlita, Vollter Anastas, Arturo Orlacchio, Lalit Sehgal, David P. Carbone, Vincenzo Coppola and Philip N. Tsichlis, The inhibition of IWS1 phosphorylation promotes genomic instability, the cGAS/STING pathway activation and PD-L1 levels, through the U2AF2 alternative RNA splicing and Sororin expression, Cancer Immunol Res February 1 2021 (9) (2 Supplement) PO010; **DOI:**10.1158/2326-6074.TUMIMM20-PO010
99. Gong, K., Guo, G., Panchani, N., Bender, M.E., Gerber, D.E., Minna, J.D., Fattah, F., Gao, B., Peyton, M., Kernstine, K. and Mukherjee, B., 2020. EGFR inhibition triggers an adaptive response by co-opting antiviral signaling pathways in lung cancer. *Nature Cancer*, *1*(4), pp.394-409.
100. Obeng, E.A., Stewart, C. and Abdel-Wahab, O., 2019. Altered RNA Processing in Cancer Pathogenesis and Therapy. *Cancer discovery*, *9*(11), pp.1493-1510.
101. Vlachos, I. S. et al. DIANA-mirExTra v2.0: uncovering microRNAs and transcription factors with crucial roles in NGS expression data. *Nucleic Acids Res.* **44**, W128–W134 (2016).

102. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

103. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).

104. Aken B.L., Achuthan P., Akanni W., Amode M.R., Bernsdorff F., Bhai J., Billis K., Carvalho-Silva D., Cummins C., Clapham P.et al. Ensembl 2017. Nucleic Acids Res. 2017; 45:D635–D642.1.

105. Liao Y, et al. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics , 2014, vol. 30 (pg. 923-930)

106. Anders S., Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

107. Anders, S., Reyes, A. and Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. *Nature Precedings*, pp.1-1.

108. Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B. 1995;57:289–300.

109. Grote, S., 2018. GOfuncR: Gene ontology enrichment using FUNC. *R package version*, *1*(0).

110. Georgios I Laliotis, Philip N. Tsichlis 2020. Effective identification of DNA-bound protein complexes using Chromatin Immunoprecipitation. Protocols.io, dx.doi.org/10.17504/protocols.io.bjj6kkre

111. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589 (2010).

112. Robinson, J., Thorvaldsdóttir, H., Winckler, W. *et al.* Integrative genomics viewer.*Nat Biotechnol* 29, 24–26 (2011)

113. Georgios I. Laliotis 2020. Effective Identification of Protein-Protein Interaction using RIME-IP. Protocols.io, https://dx.doi.org/10.17504/protocols.io.bqx8mxrw

114. Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S. and D'santos, C.S., 2016. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nature protocols*, *11*(2), p.316.

115. Georgios I Laliotis, Philip N. Tsichlis 2020. Effective identification of RNA-binding proteins using RNA Immunoprecipitation. Protocols.io, dx.doi.org/10.17504/protocols.io.bjpbkmin

116. Skene, P.J., Henikoff, J.G. and Henikoff, S., 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature protocols*, *13*(5), p.1006.

117. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, *9*(7), pp.671-675.

118. Goel, M.K., Khanna, P. and Kishore, J., 2010. Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research*, *1*(4), p.274.

119. Wilkerson, M.D., Yin, X., Walter, V., Zhao, N., Cabanski, C.R., Hayward, M.C., Miller, C.R., Socinski, M.A., Parsons, A.M., Thorne, L.B. and Haithcock, B.E., 2012. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PloS one*, *7*(5).

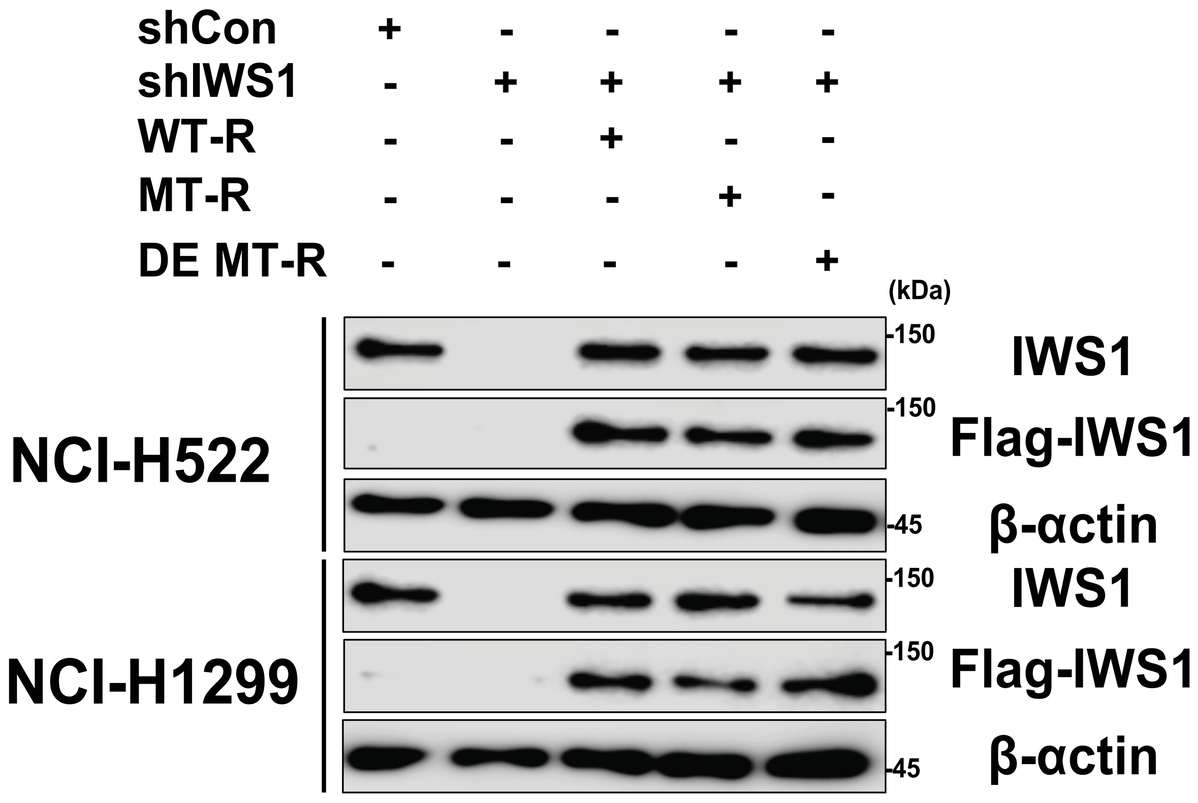
120. Maaten, L.V.D. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, *9*(Nov), pp.2579-2605.

121. Laliotis, Georgios I.; Chavdoula, Evangelia; Paraskevopoulou, Maria D.; Kaba, Abdul; La Ferlita, Alessandro; Singh, Satish; Anastas, Vollter; Nair II, Keith A.; Orlacchio, Arturo; Taraslia, Vasiliki; Vlachos, Ioannis; Capece, Marina; Hatzigeorgiou, Artemis; Palmieri, Dario; Tsatsanis, Christos; Alaimo, Salvatore; Sehgal, Lalit; Carbone, David P.; Coppola, Vincenzo; Tsichlis, Philip N. (2020), "IWS1 phosphorylation promotes cell proliferation and predicts poor prognosis in EGFR

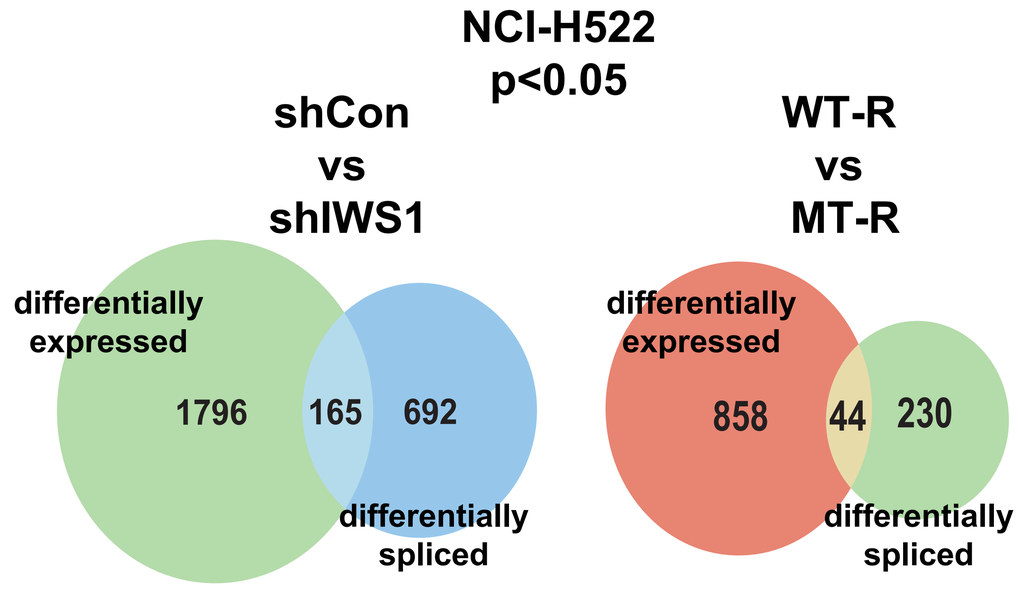
mutant lung adenocarcinoma patients, through the cell cycle-regulated U2AF2 RNA splicing.", Mendeley Data, V4, **DOI:** 10.17632/y82yx8gfkj.4

122. Laliotis, Georgios; Chavdoula, Evangelia; Tsichlis, Philip N. (2020), "IWS1 phosphorylation promotes cell proliferation and predicts poor prognosis in EGFR mutant lung adenocarcinoma patients, through the cell cycle-regulated U2AF2 RNA splicing_images_depository", Mendeley Data, V1, doi:10.17632/ts44bfg45k.1

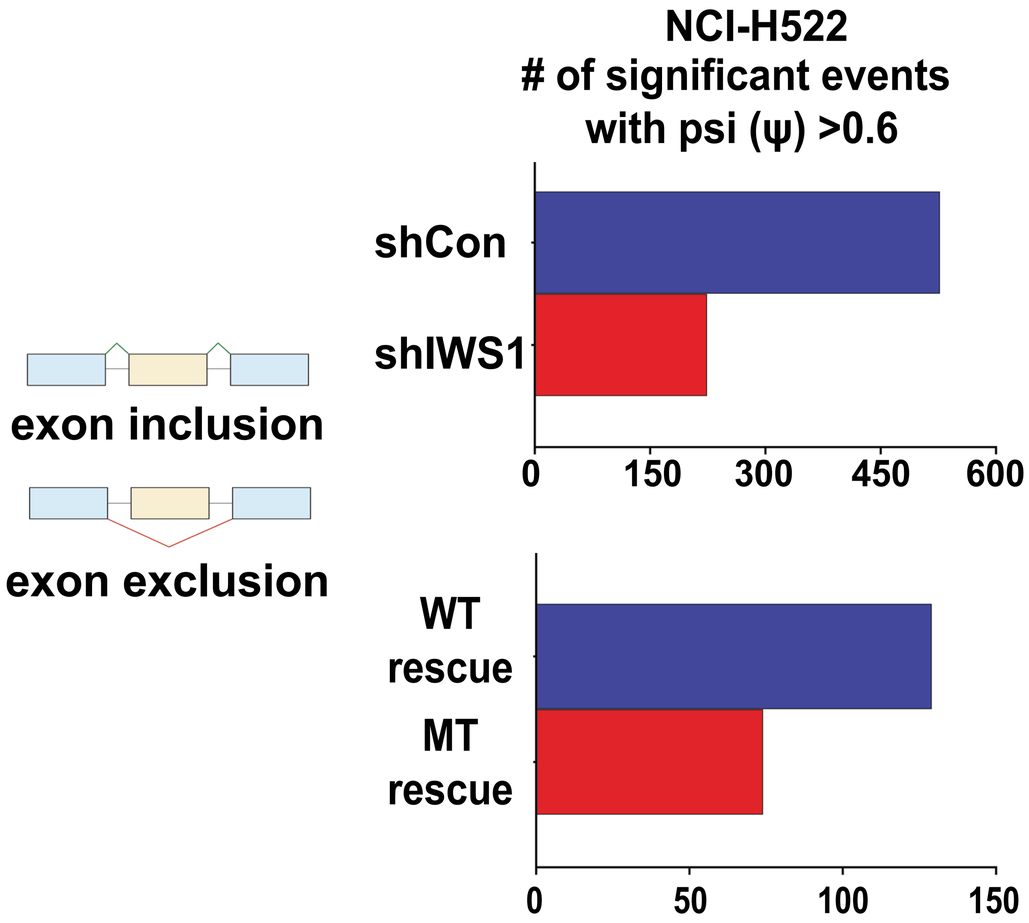
A



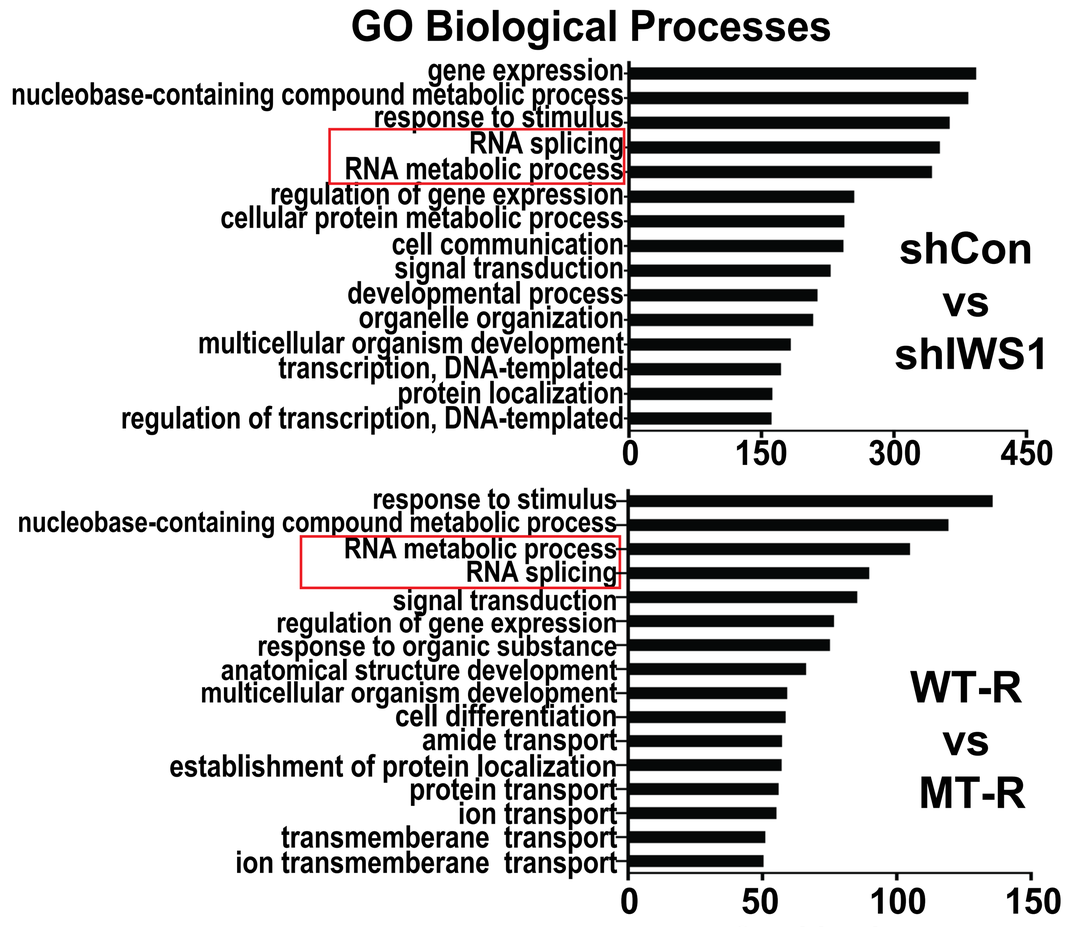

B



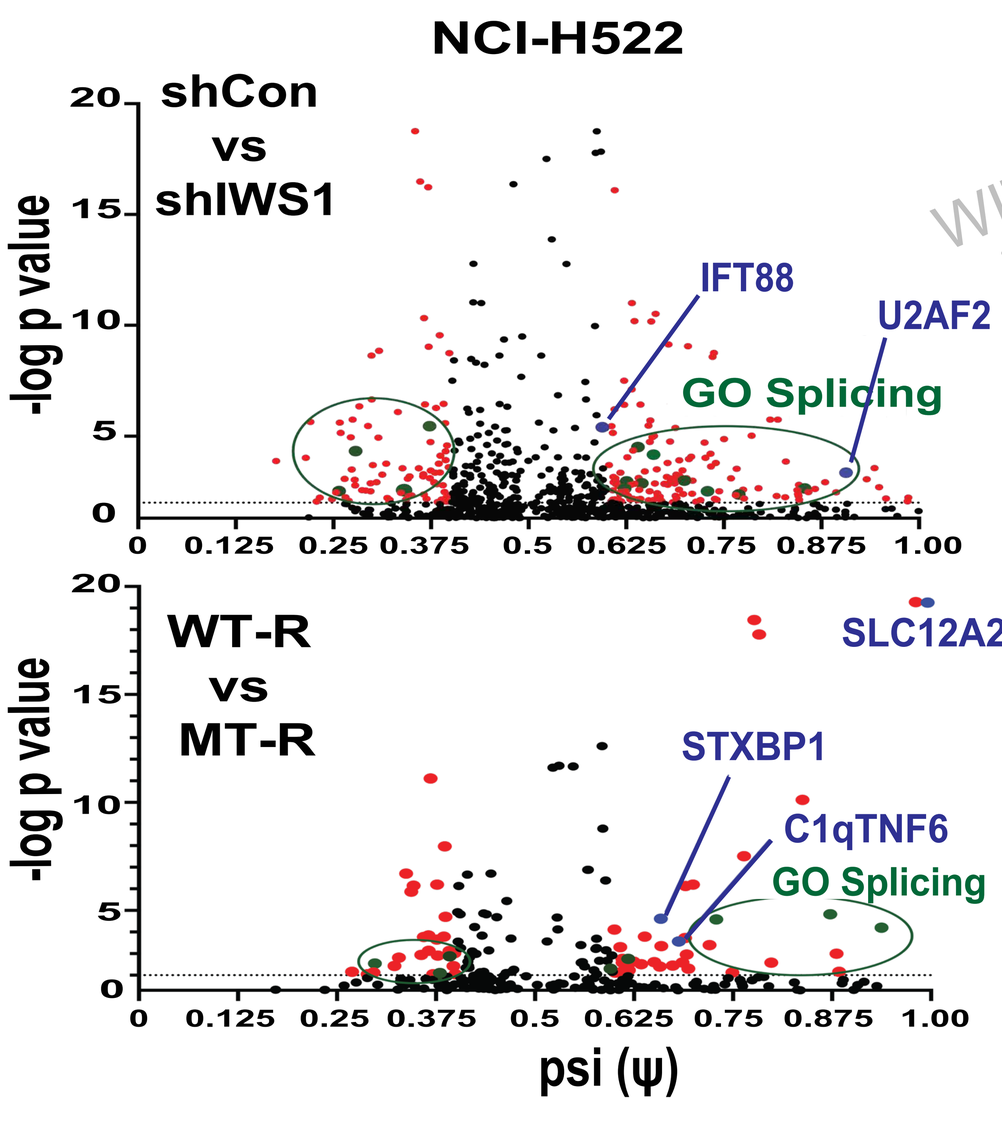

C



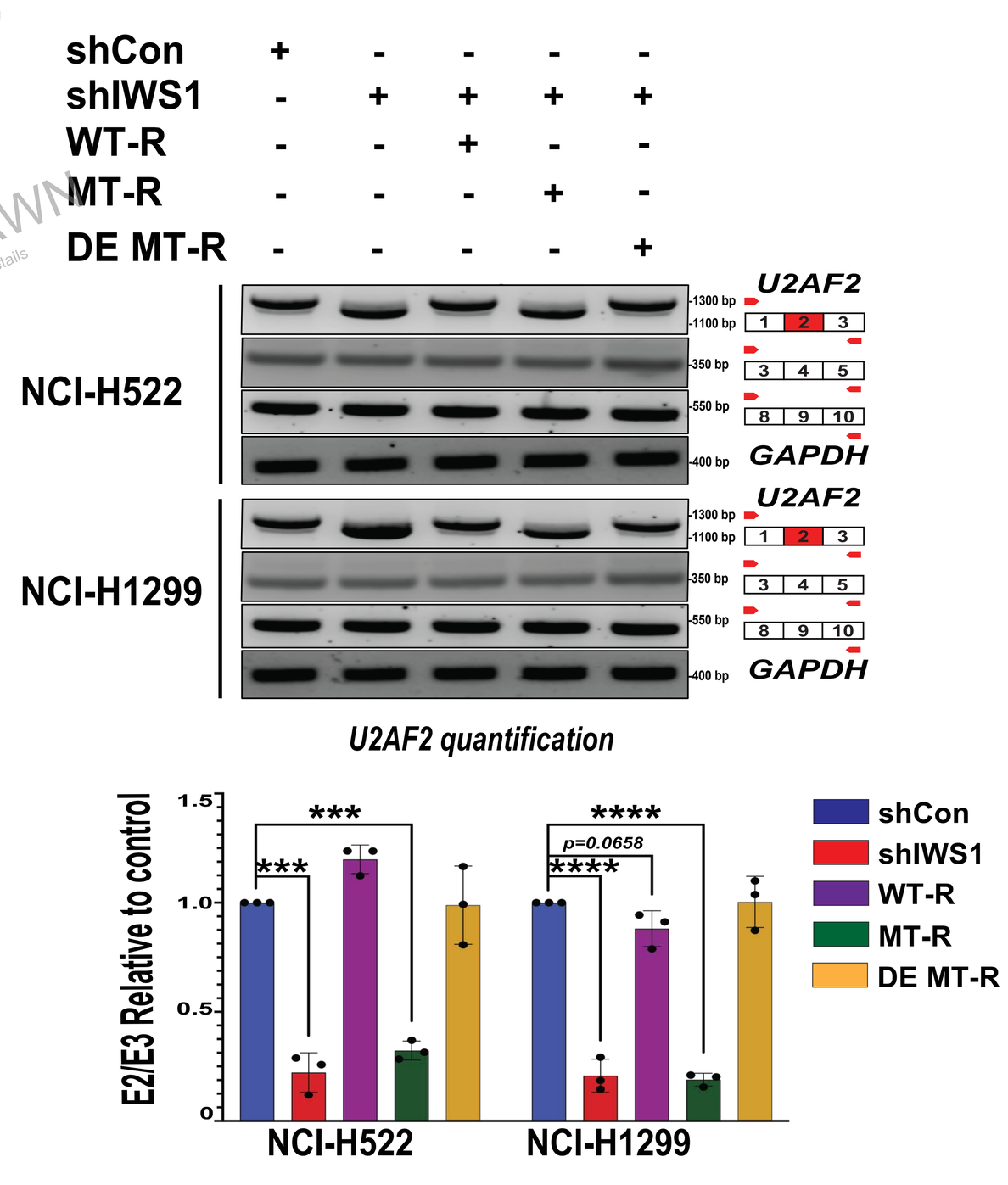

D




E




F




G


H

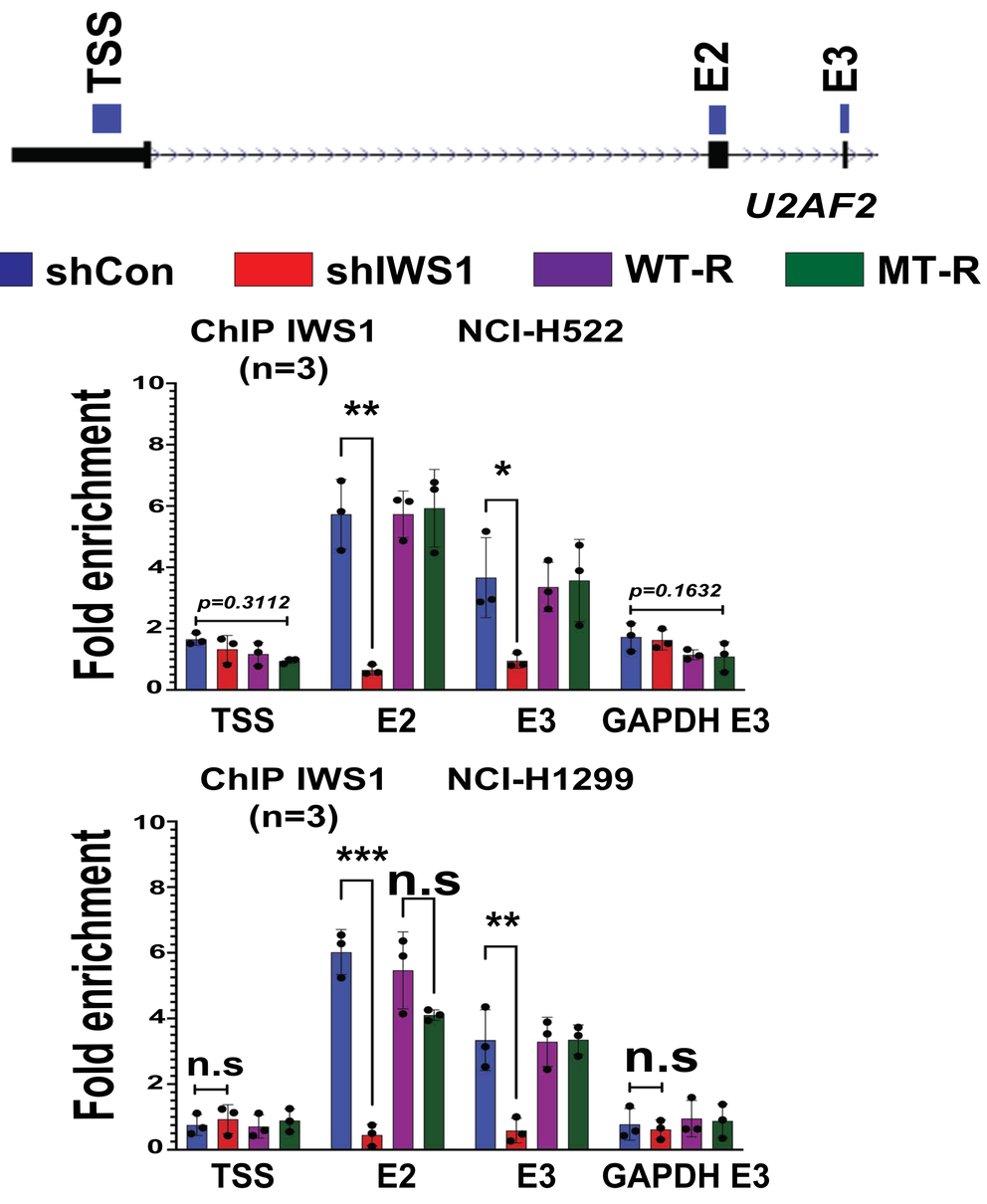


**Figure 1. IWS1 expression and/or phosphorylation regulate alternative mRNA splicing**

**A.** Western blots of lysates of NCI-H522 and NCI-H1299 cells, transduced with the indicated constructs, were probed with anti-IWS1, anti-Flag-tag and anti-β-actin (control) antibodies.

**B.** Overlaps between differentially-expressed and differentially spliced genes in shControl and shIWS1 or shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells ($q<0.05$) are illustrated, using Venn diagrams.

**C.** Bar graphs show the number of alternative splicing events with exon inclusion with a percentage spliced in (psi/ψ) >0.6 in shControl versus shIWS1 (upper panel) and shIWS1/WT-R versus shIWS1/MT-R (lower panel) NCI-H522 cells. The comparisons were limited to alternative splicing events with a p value < 0.05. This shows that IWS1 phosphorylation promotes exon inclusion (see diagram on the left) in multiple alternatively spliced genes.

**D.** GO analysis, based on differences in the abundance of alternative splicing events in shControl versus shIWS1 (upper panel) and shIWS1/WT-R versus shIWS1/MT-R (lower panel) NCI-H522 cells. Comparisons were limited to alternative splicing events whose abundance changes significantly with the expression or phosphorylation of IWS1($p<0.05$). Red boxes highlight gene sets involved in the regulation of RNA processing.

**E.** Volcano plots of the -log p value vs the exon inclusion levels of all the exon inclusion and exon exclusion alternative splicing events, detected by DEXseq in the comparisons between shControl and shIWS1 (upper panel), or shIWS1/WT-R and shIWS1/MT-R (lower panel) NCI-H522 cells. The statistically significant events ($p<0.05$) with a percentage spliced in (psi/ψ) level of >0.6 or <0.4 are shown in red. Statistically significant events in genes in the GO functions RNA splicing or RNA metabolic processes are shown in green. Alternatively-spliced IWS1 targets validated in this report are shown in blue.

(*U2AF2, IFT88, SLC12A2, STXBP1, C1qTNF6)*. The validation of these through RT-PCR is shown in Figure S1.

**F.** *IWS1 phosphorylation regulates U2AF2 alternative mRNA splicing.* (Upper panel) RT-PCR of *U2AF2*, using oligonucleotide primers that map in exons 1 and 3, exons 3 and 5 or exons 8 and 10 (*U2AF2* expression control). RNA transcripts containing or lacking exon 2 are distinguished based on the size of the amplified cDNA derived from these transcripts. RT-PCR was carried out using RNA derived from shControl, shIWS1 shIWS1/WT-R, shIWS1/MT-R and shIWS1/DE MT-R NCI-H522 and NCI-H1299 cells. *hGAPDH* was used as the loading control. (Lower panel) The relative abundance of the amplified cDNA bands in the upper panel was determined, and the ratio between exons E2 and E3 was calculated relative to the ratio in shControl cells, which was given the arbitrary value of 1.

**G.** Sequencing chromatograms showing the junction between exons 1 and 2 and exons 1 and 3 in the two alternatively spliced *U2AF2* RNA transcripts.

**H.** *IWS1 binds exons 2 and 3, but not the transcription start site (TSS) of the U2AF2 gene.* (Upper) UCSC browser snapshot showing exons 1, 2 and 3 of the human *U2AF2* gene. The map position of the PCR primer sets used in the ChIP experiments in this figure is indicated by blue marks. (Lower) ChIP assays addressing the binding of IWS1 on the *U2AF2* and *GAPDH* genes in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells. Bars show the mean fold enrichment (anti-IWS1 IP, vs IgG control IP) in IWS1 binding, in shIWS1 relative to shControl cells or in shIWS1/MT-R relative to shIWS1/WT-R cells ±SD. Data were normalized relative to the input (2%)

All assays were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test)

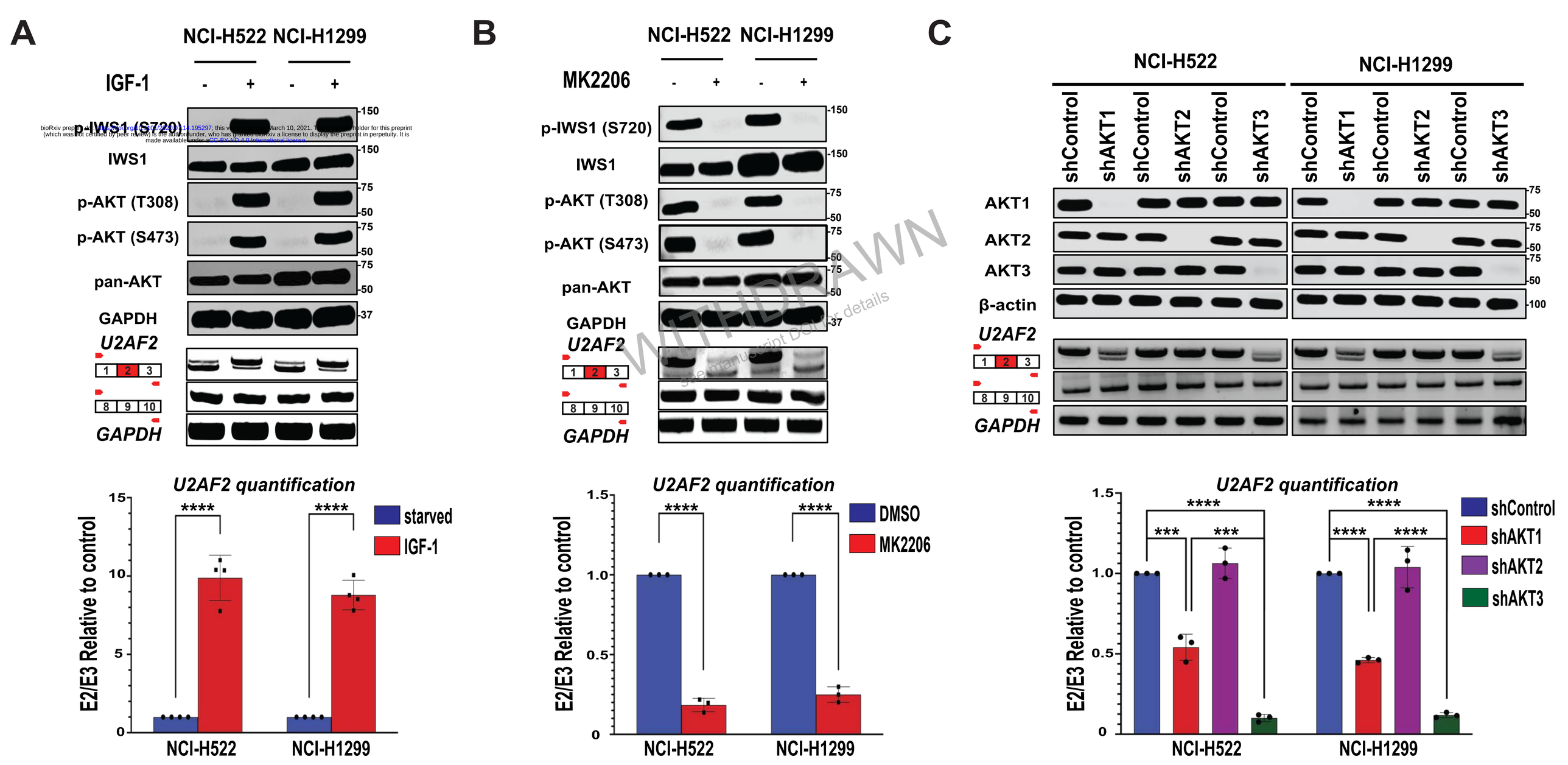


**Figure 2. IWS1 phosphorylation-dependent mRNA splicing of *U2AF2* is regulated by serum and IGF-1-induced signals transduced by AKT3.**

**A.** *IGF-1 promotes the IWS1-dependent U2AF2 exon 2 (E2) inclusion.* (Upper panel) Following serum starvation for 24 hours, NCI-H522 and NCI-H1299 cells were stimulated with IGF-1 or harvested. Four hours later, the stimulated cells were lysed and all the lysates were probed with the indicated antibodies. (Middle panel) RT-PCR with mRNA derived from the cells shown in the upper panel and oligonucleotide primers mapping to exons 1 and 3 of the *U2AF2* gene, show that IGF1-induced signals promote *U2AF2* exon 2 inclusion. (Lower panel) The *U2AF2* gene *E2/E3* ratio was calculated following quantification of the RT-PCR products in the middle panel. The bars show this ratio (mean± SD) in IGF1-stimulated NCI-H522 and NCI-H1299 cells relative to the ratio in untreated serum-starved cells.

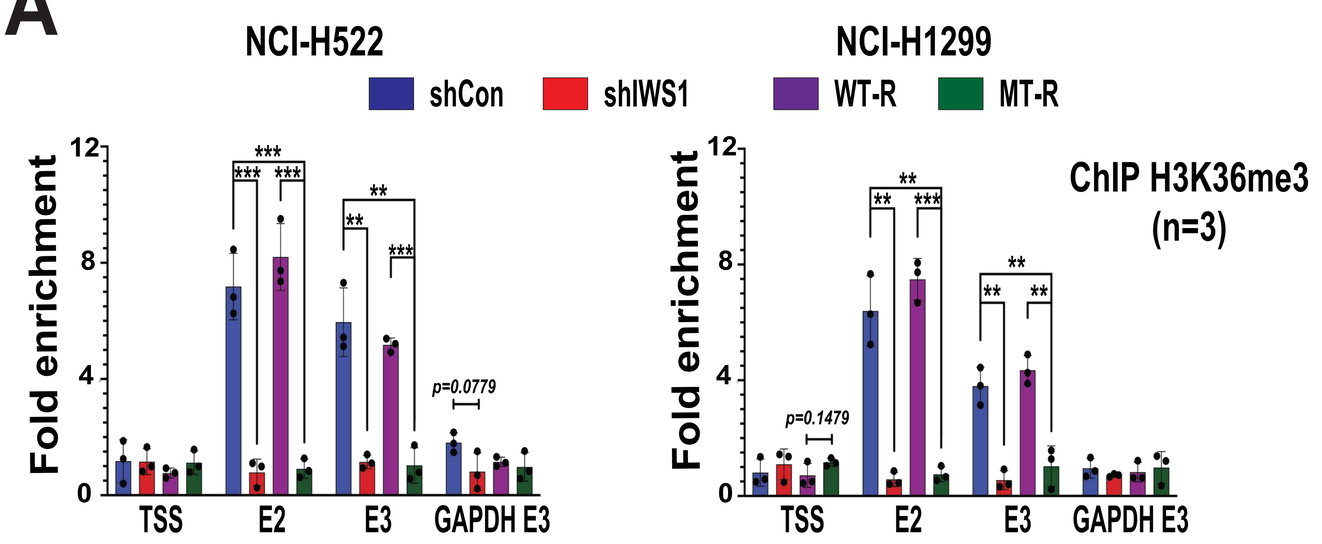
**B.** *U2AF2 exon 2 (E2) inclusion depends on AKT kinase.* (Upper panel) NCI-H522 and NCI-H1299 cell lysates, harvested following a 4 hr treatment with MK2206 (5μM) or DMSO, were probed with the indicated antibodies. (Middle panel) RT-PCR reactions, using mRNA derived from the cells in the upper panel and oligonucleotide primers mapping in *U2AF2* exons 1 and 3 show that inhibiting AKT inhibits the inclusion of exon 2 in mature *U2AF2* mRNA transcripts in both cell lines. (Lower panel) The *U2AF2* mRNA *E2/E3* ratio was calculated following quantification of the RT-PCR products in the middle panel. The bars show this ratio (mean± SD) in MK2206-treated (5μM) NCI-H522 and NCI-H1299 cells relative to the ratio in untreated cells.

**C.** *U2AF2 exon 2 (E2) inclusion depends on AKT3***.** (Upper panel) NCI-H522 and NCI-H1299 cell lysates, transduced with shControl, shAKT1, shAKT2 and shAKT3 were probed with the indicated antibodies. (Middle panel) RT-PCR reactions, using mRNA derived from the cells in the upper panel and oligonucleotide primers mapping in *U2AF2* exons 1 and 3 show that knocking down AKT3, and to a lesser extent AKT1, inhibits the inclusion of exon 2 in mature

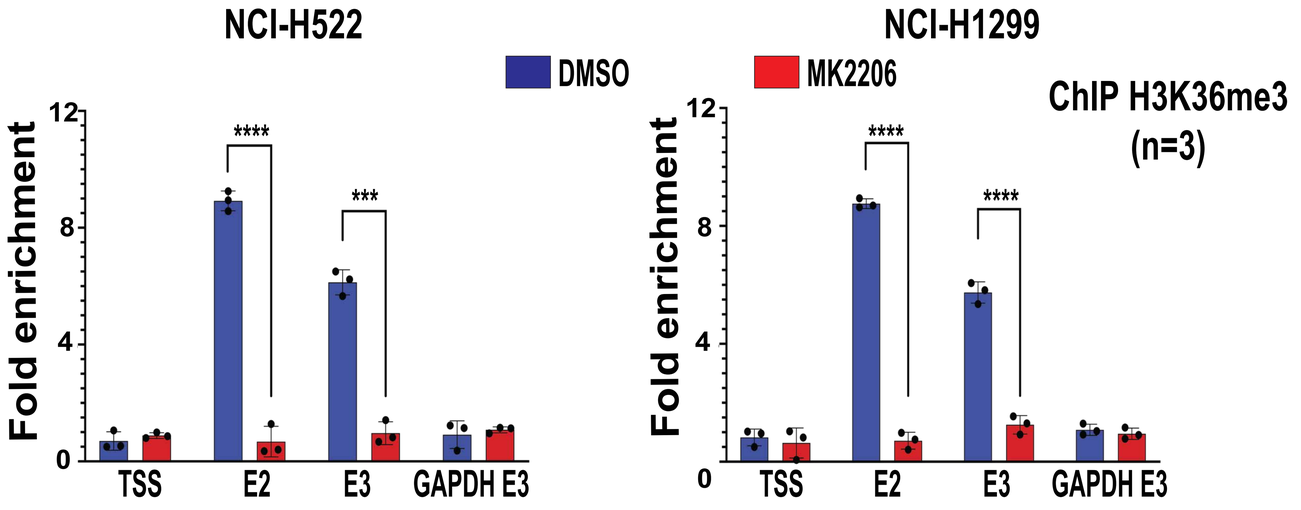
*U2AF2* mRNA transcripts in both cell lines. (Lower panel) The *U2AF2* mRNA *E2/E3* ratio was calculated following quantification of the RT-PCR products in the middle panel. The bars show this ratio (mean± SD) in shAKT1, shAKT2 and shAKT3 NCI-H522 and NCI-H1299 cells relative to the ratio in shControl cells.

All experiments in this figure were done in triplicate, on three biological replicates. n.s: non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

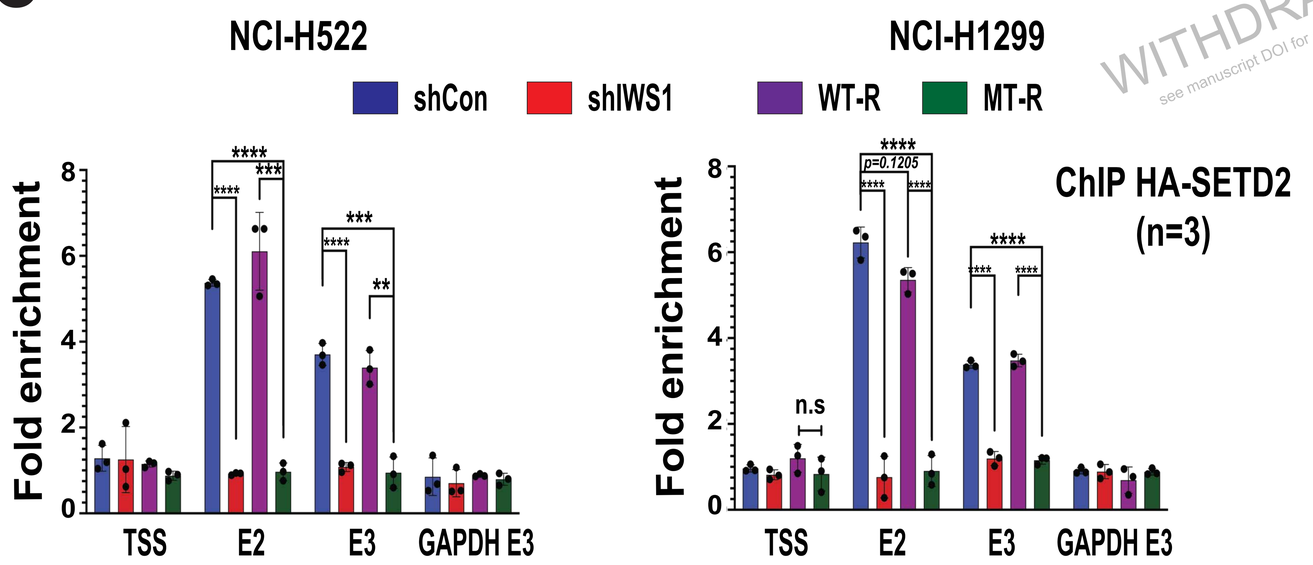
A



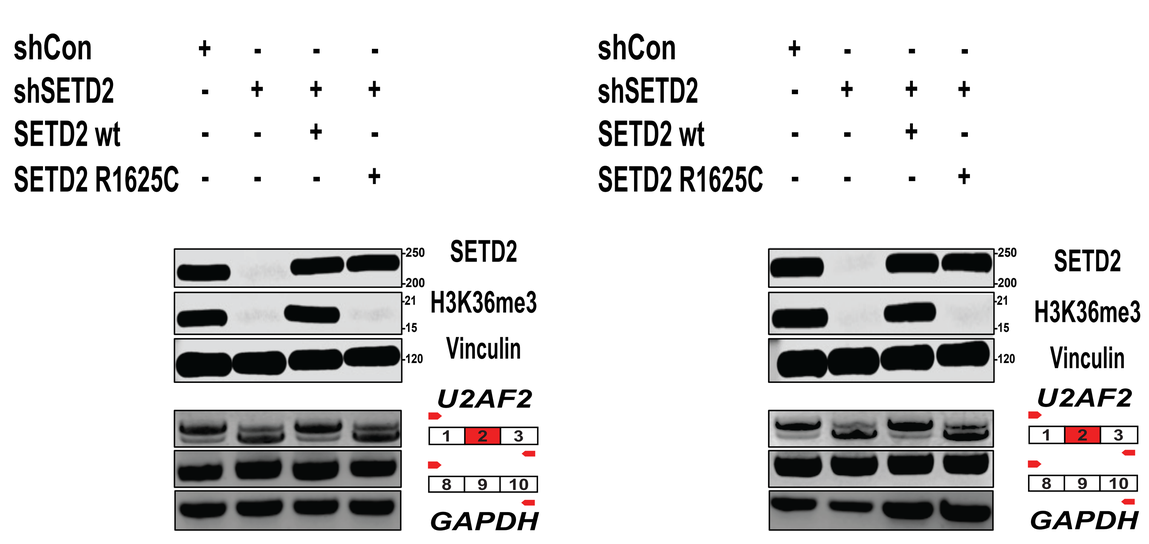

B

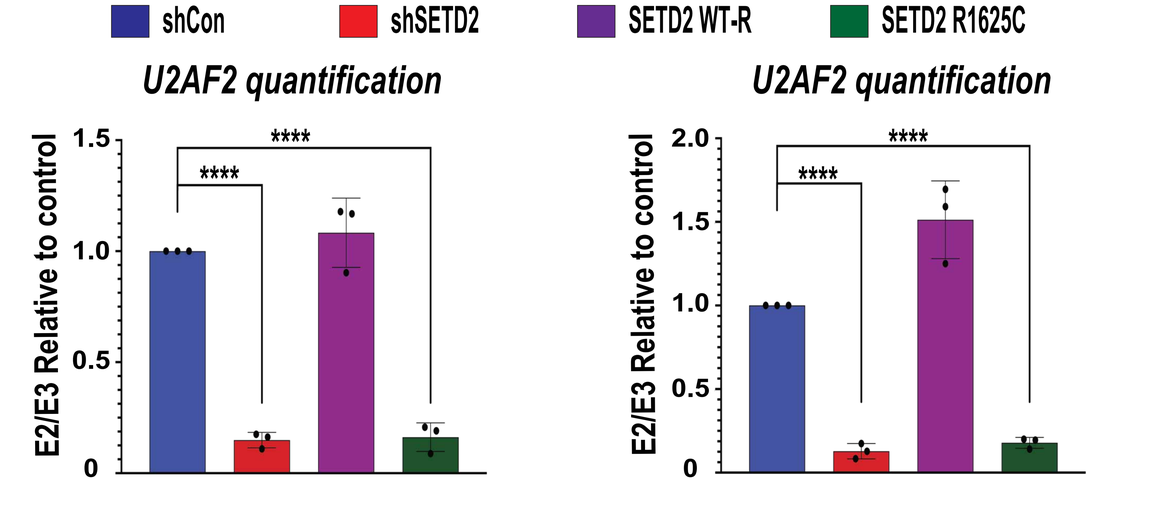


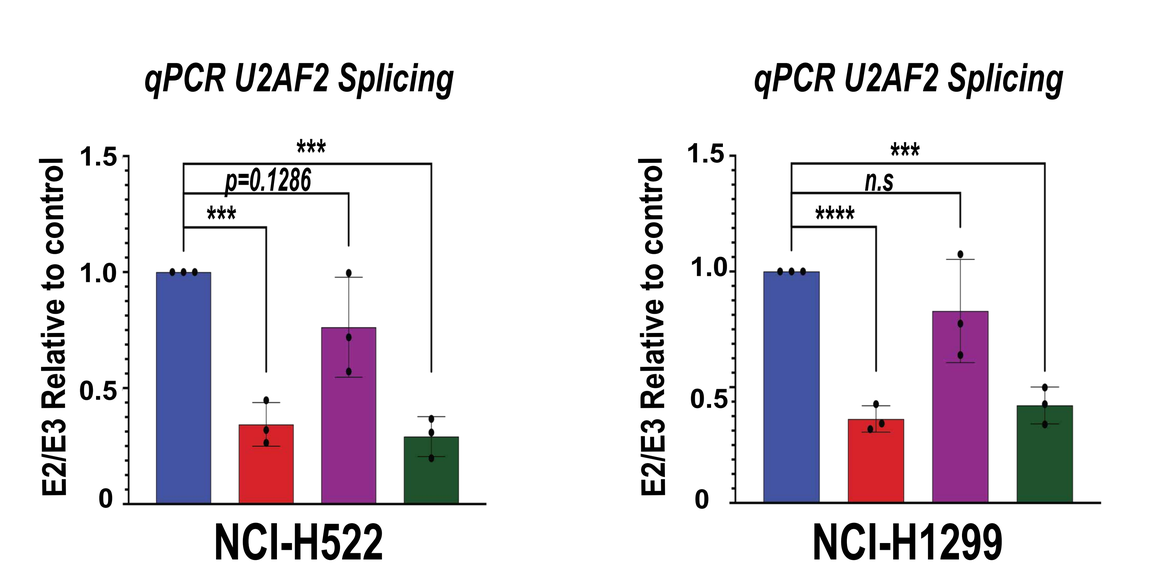

C




D




E

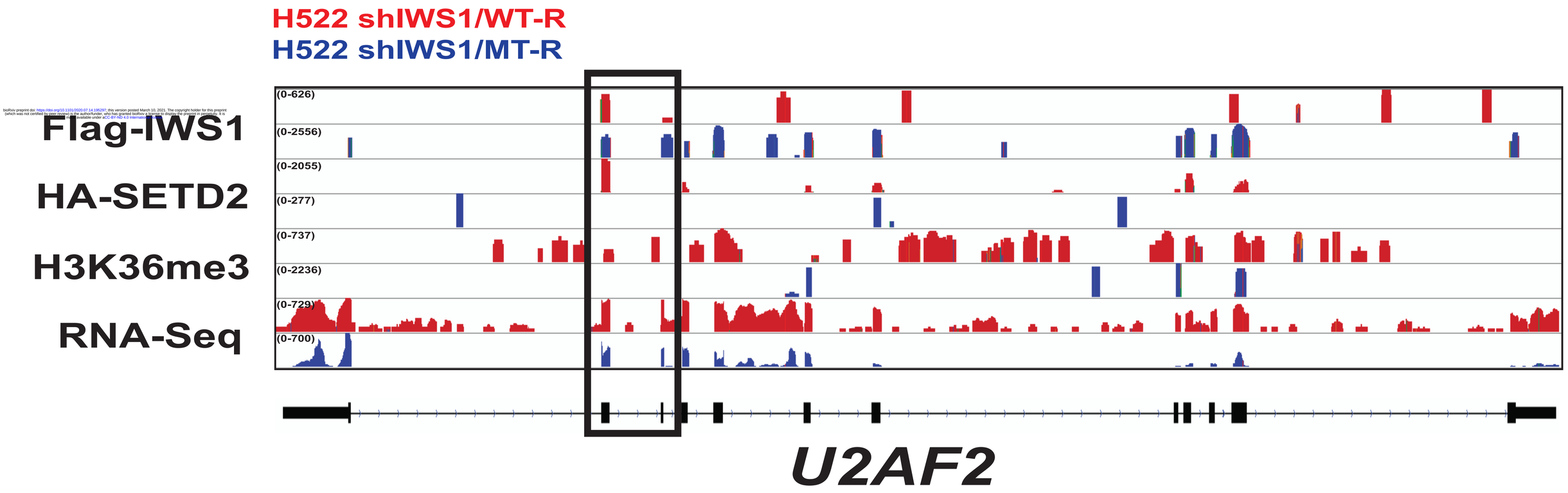


**Figure 3. *U2AF2* Exon 2 inclusion, induced by IWS1 phosphorylation at Ser720/Thr721, depends on H3K36 trimethylation by SETD2.**

**A.** ChIP assays showing the abundance of H3K36me3 on the *U2AF2* and *GAPDH* genes in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 (left) and NCI-H1299 cells (right). Bars show the mean fold enrichment in H3K36me3 (anti-H3K36me3 IP, vs IgG control IP) in the indicated regions of the *U2AF2* gene, in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%).

**B.** *AKT inhibition interferes with the H3K36me3 trimethylation in the U2AF2 gene*. ChIP assays showing the abundance of H3K36me3 on the *U2AF2* and *GAPDH* genes in NCI-H522 (left) and NCI-H1299 cells (right), before and after the treatment with MK2206 (5μM) or DMSO. The bars show the mean fold enrichment of H3K36me3 (anti-H3K36me3 IP, vs IgG control IP) in the indicated regions of the *U2AF2* gene, in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%).

**C.** ChIP assays showing the binding of HA-SETD2 to the indicated regions of the *U2AF2* and *GAPDH* genes, in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 (left) and NCI-H1299 cells (right), transduced with a lentiviral HA-SETD2 construct. The bars show the mean fold enrichment in SETD2 binding (anti-HA IP, vs IgG control IP) in the indicated regions of the *U2AF2* gene, in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%). The expression of HA-SETD2 is presented in Figure S4A.

**D.** *The IWS1-regulated alternative RNA splicing of U2AF2, depends on enzymatically active SETD2*. (Upper panel) Lysates of NCI-H522 (left) and NCI-H1299 (right) cells transduced with shControl or shSETD2 lentiviral constructs and shSETD2 cells rescued with wild type

SETD2, or the catalytically inactive SETD2 R1625C mutant. were probed with the indicated antibodies. RT-PCR, using RNA derived from these cells and oligonucleotide primers mapping in *U2AF2* exons 1 and 3, revealed that the inclusion of exon 2 in the mature U2AF2 mRNA transcripts depends on the expression of enzymatically active SETD2. (Middle panel) The RT-PCR products in the experiment in the upper panel were quantified. Bars show the E2/E3 ratio relative to the shControl, which was given the value of 1. (Lower panel) The ratio of the *U2AF2* exons 2 and 3 was determined by quantitative RT-PCR, using the mRNA isolated from the cells in the upper panel. Again, bars show the E2/E3 ratio relative to the shControl.

Error bars indicate SD. All experiments in this figure were done on three biological replicates, in triplicate. n.s: non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

**E.** Snapshots of the integrative genomic viewer showing the distribution of Flag-IWS1, HA-SETD2, and histone H3K36me3 marks, within the *U2AF2* gene, obtained from a ChIP-Seq analysis, along with the distribution of RNA reads obtained from the RNA-seq analysis of shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells. Scale represents reads per million (RPM). Snapshots of peaks detected in both biological replicates are shown. The black box outlines *U2AF2* exons 2 and 3 and the adjacent regions.

**Figure 4. The regulation of the alternative splicing of the *U2AF2* exon 2 by IWS1 phosphorylation, depends on the p52 isoform of the H3K36me3 reader LEDGF and its splicing partner SRSF1.**

A. *The H3K36me3 reader MRG15 and its binding partner PTB, play no role in the regulation of U2AF2 alternative RNA splicing.* (Upper panel) Western blots of lysates of NCI-H522 and NCI-H1299 cells transduced with shControl and shMRG15 (left) or shControl and shPTB (right) were probed with the indicated antibodies. RT-PCR, using RNA isolated from these cells and *U2AF2* exon 1 and exon 3 or exon 8 and 10 oligonucleotide primers shows that the knockdown of neither MRG15 nor PTB affects the alternative splicing of *U2AF2*. (Middle and lower panel) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the E2/E3 ratio in shMRG15 and shPTB NCI-H522 and NCI-H1299 cells, relative to the shControl ± SD.

B. *The alternative splicing of the U2AF2 exon 2 is regulated by the H3K36me3 reader LEDGF.* (Upper Panel) The listed H3K36me3 readers were efficiently knocked down by siRNA transfection of NCI-H522 cells, as determined by probing western blots of transfected cell lysates with the indicated antibodies. RT-PCR, using RNA isolated from these cells and *U2AF2* exon 1 and exon 3 or exon 8 and 10 oligonucleotide primers shows that only the knockdown of LEDGF affects the alternative splicing of *U2AF2*. (Middle and lower panel) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the E2/E3 ratio in NCI-H522 cells, after the knockdown of the indicated readers, relative to the siControl ± SD.

C. *The alternative RNA splicing of the U2AF2 exon 2 is regulated by the p52, but not the p75 isoform of LEDGF.*(Upper panel) Schematic representation of the p52 and p75 isoforms of LEDGF showing the position of the PWWP domain (green), the AT hook-like domain (red), the unique to p52 8 aa long C-terminal domain (black) and the unique to p75

integrase binding domain (IBD) (blue) (Ferris et al., 2010[55]). (Middle panel) Western blots of lysates of shControl, shLEDGF, shLEDGF/LEDGF-p75-R,shLEDGF/LEDGF-p52-WT-R,shLEDGF/LEDGF-p52A51P-R NCI-H522 cells, were probed with the indicated antibodies. RT-PCR, using RNA isolated from these cells and *U2AF2* exons 1 and 3 or exons 8 and 10 oligonucleotide primers shows that only the wild type p52 isoform of LEDGF rescues the alternative splicing of *U2AF2* in shLEDGF-transduced cells. The p75 isoform and the p52 mutant A51P, which does not bind H3K36me3, failed to rescue. (Lower panels) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the E2/E3 ratio in shLEDGF-transduced NCI-H522 cells, before and after the rescue with the indicated constructs, relative to the shControl ± SD.

D. *The recruitment of p52/LEDGF to the U2AF2 gene depends on IWS1 phosphorylation.* ChIP assays showing the binding of p52/LEDGF to the *U2AF2* and *GAPDH* genes in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells, transduced with a lentiviral V5-p52/LEDGF construct. The bars show the mean fold enrichment in p52/LEDGF binding (anti-V5 IP, vs IgG control IP) to the indicated regions of the *U2AF2*gene, in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%). The expression of V5-p52/LEDGF is shown in figure S5G.

E. *The alternative RNA splicing of the U2AF2 exon 2 depends on SRSF1.* (Upper panel) Western blots of lysates of shControl, shSRSF1 and shSRSF1/SRSF1 WT-R NCI-H522 cells, were probed with the indicated antibodies. RT-PCR, using RNA isolated from these cells and *U2AF2* exons 1 and 3 or exons 8 and 10 oligonucleotide primers, shows that the knockdown of SRSF1 promotes a shift in the RNA splicing pattern of *U2AF2*, which favors the exclusion of exon 2 from the mature transcripts, and that wild type SRSF1 rescues the shift. (Lower panels) The RT-PCR results were confirmed by quantitative RT-PCR. Bar

graphs show the E2/E3 ratio in shSRSF1-transduced NCI-H522 cells, before and after rescue with wild type SRSF1, relative to the shControl ±SD.

F. *The recruitment of SRSF1 to the U2AF2 gene depends on IWS1 phosphorylation.* ChIP assays showing the binding of SRSF1 to the *U2AF2* and *GAPDH* genes in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells, transduced with a lentiviral V5-SRSF1 construct. The bars show the mean fold enrichment in SRSF1 binding (anti-V5 IP, vs IgG control IP) to the indicated regions of the *U2AF2* gene, in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%). The V5-SRSF1 expression is shown in Figure S6B.

G. *The recruitment of SRSF1 to the U2AF2 RNA depends on IWS1 phosphorylation.* RIP assays showing the binding of SRSF1 to the *U2AF2* RNA in the NCI-H522 cells in Fig. 4F. The bars show the mean fold enrichment in SRSF1 binding in the indicated regions of the *U2AF2* RNA (anti-V5 IP, vs IgG control IP) ±SD. Data were normalized relative to the input (2%). All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test)

H. *Model of the regulation of U2AF2 alternative RNA splicing by AKT3-mediated IWS1 phosphorylation at Ser720/Thr721.* The phosphorylation of IWS1 recruits SETD2 to the Spt6/IWS1/Aly complex in the CTD of RNA Pol II. SetD2 in the complex, trimethylates histone H3 at K36 during transcriptional elongation. p52/LEDGF and its RNA-binding partner SRSF1, bind histone H3K36me3. Binding of SRSF1 to exon 2 promotes inclusion of this exon in the mature *U2AF2* transcript.

**Figure 5. IWS1 phosphorylation controls the CDCA5/ERK phosphorylation feedback loop, through U2AF2 alternative RNA splicing.**

**A.** Domain organization of the U2AF65 protein, showing the Serine/Arginine Rich (SR) domain, the U2AF-ligand motif (ULM), two RNA recognition motifs (RRMs) and the U2AF Homology Motif (UHM). Exon 2 encodes the N-terminal RS domain. U2AF65 and U2AF35 interact through the U2AF65 UHM domain. The numbers on top identify the amino acids at the boundaries of the indicated motifs.

**B.** *The novel U2AF65β spliced variant, lacking exon 2, does not interact with Prp19.* HEK-293T cells were transduced with lentiviral constructs V5-U2AF65α, or V5-U2AF65β. Anti V5-U2AF65 immunoprecipitates from these cells were probed with anti-Prp19 or anti-V5-tag antibodies. (Lower panel). Western blots of the input lysates, probed with the indicated antibodies are also shown.

**C.** *The U2AF65-Prp19 interaction depends on IWS1-phosphorylation.* (Upper panel) Anti-U2AF65 immunoprecipitated from lysates of shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells were probed with anti-U2AF65, anti-Prp19, and anti-U2AF35 antibodies. Anti-IgG mouse isotype control immunoprecipitates were probed with the anti-Prp19 antibody. Western blots of the input lysates, probed with the indicated antibodies are also shown.

**D.** *IWS1-phosphorylation controls CDCA5 RNA splicing.* (Upper panel). The ratio of spliced to unspliced *CDCA5* RNA transcripts was examined by quantitative RT-PCR, using total RNA isolated from shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells and the listed oligonucleotide primers (E1-E2/E1-I1 and E2-E3/E2-I2). The same was done with the transcripts of a control gene (*GUSB*) (Oligonucleotide primers E5-E6/I5-E6). (Middle and lower panels) RIP assays in the same cells show that whereas the binding of Prp19 to the *CDCA5* RNA depends on the phosphorylation of IWS1, the binding of

U2AF65 does not. The bars show the mean fold enrichment in U2AF65 and Prp19 binding in the indicated regions of the *CDCA5* pre-mRNA (anti-U2AF65 or anti-Prp19 IP, vs IgG control IP) ± SD. Data were normalized relative to the input (2%). The primer location within *CDCA5* and *GUSB* can be found in Figure S7B.

**E.** *U2AF2 alternative RNA splicing, downstream of IWS1 phosphorylation controls CDCA5 splicing.* (Upper panel). The ratio of spliced to unspliced CDCA5 RNA transcripts was examined in shControl NCI-H522 cells that were either not rescued, or rescued with U2AF65α, or U2AF65β, and in shIWS1 NCI-H522 cells rescued with the same U2AF65α, or U2AF65β-encoding constructs. Quantitative RT-PCR was carried out and presented, as in Fig 5D. The GUSB gene was used again as the control. (Middle and lower panels) RIP assays in the same cells show that whereas U2AF65α rescues the binding of Prp19 to the CDCA5 RNA in shIWS1 cells, U2AF65β does not. Moreover, U2AF65α and U2AF65β, bound equally to the *CDCA5* RNA. The presentation of the data was the same as in Fig 5D.

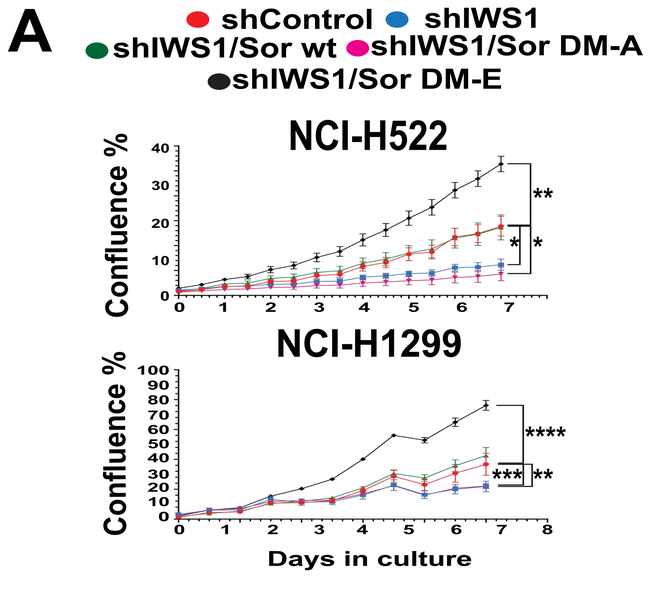
**F.** *The abundance of mature CDCA5 mRNA in the cytoplasm is under the control of the IWS1-phosphorylation-dependent expression of U2AF65α.* The abundance of the cytosolic *CDCA5* mRNA was measured by quantitative RT-PCR in the cytosolic fraction of shIWS1, shIWS1/WT-R, shIWS1/MT-R, shIWS1/U2AF65α-R and shIWS1/U2AF5β-R NCI-H522 (upper panel) and NCI-H1299 (lower panel) cells. Bars show the mean cytosolic *CDCA5* mRNA expression normalized to *GAPDH* ± SD in NCI-H522 and NCI-H1299 shIWS1, shIWS1/WT-R, shIWS1/MT-R, shIWS1/U2AF65α-R and shIWS1/U2AF5β-R cells compared to shControl. The accuracy of the fractionation was validated with the experiment in Figure S7E.

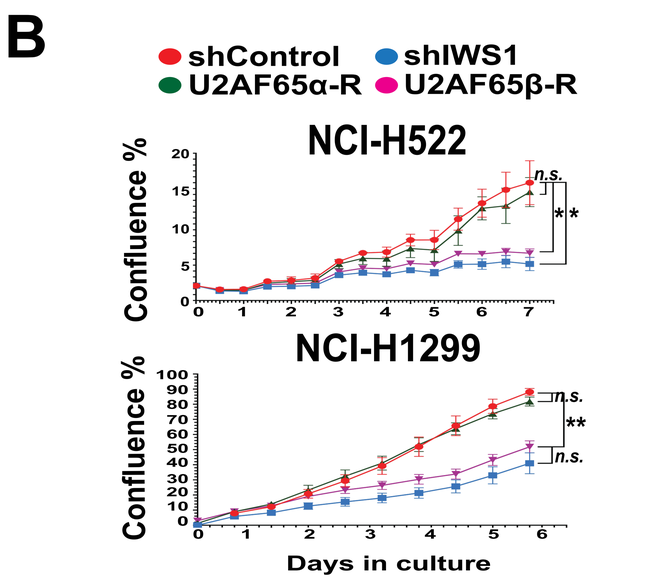
**G.** *CDCA5 and p-ERK, form a positive feedback loop, which is activated by IWS1 phosphorylation and U2AF2 alternative RNA splicing and promotes the expression of CDK1 and Cyclin B1.* Lysates of shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R

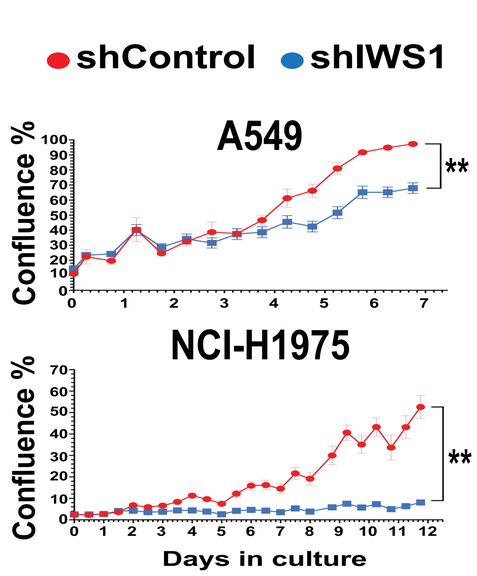
NCI-H522 cells, along with shControl/U2AF65α, shControl/U2AF65β, shIWS1/U2AF65α-R, shIWS1/U2AF5β-R and shIWS1/CDCA5-WT-R, shIWS1/CDCA5-DM-A-R and shIWS1/CDCA5-DM-D-R NCI-H522 cells, were probed with the indicated antibodies. (DM-A is the S79/S209 CDCA5 mutant and DM-E is the S79/S209EE CDCA5 mutant).

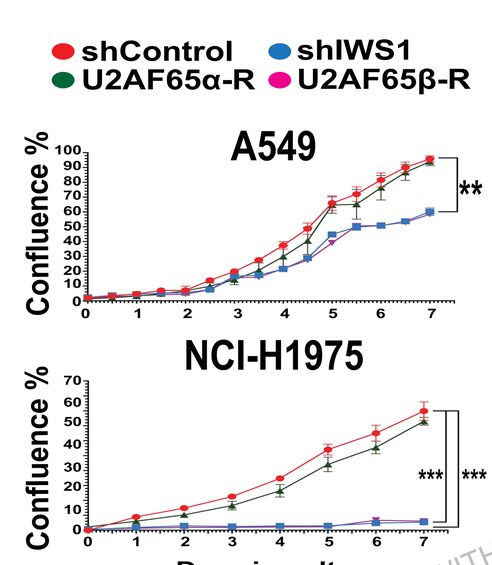
**H.** *IWS1 phosphorylation controls the phosphorylation of ERK in cells harboring EGFR and KRAS mutations.* Lysates of shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R *KRAS* (upper left panels) and *EGFR* mutant cell lines were probed with the indicated antibodies. ERK phosphorylation was reduced in all shIWS1 and shIWS1/MT-R cells, including the ones with *KRAS* and *EGFR* mutations. RT-PCR, using RNA isolated from these cells and *U2AF2* exon 1 and exon 3 oligonucleotide primers, confirmed the exclusion of exon 2 from the mature *U2AF2* mRNA, in all the shIWS1-transduced cells. (Lower left panel) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the E2/E3 ratio in all the shIWS1, shIWS1/WT-R, and shIWS1/MT-R cells, relative to the shControl ± SD. (Lower right panel) Quantification of the reduction of ERK-phosphorylation (Y202/T204), induced by shIWS1 in the indicated cell lines. Bars show the percent reduction, normalized to tubulin. The comparison between the non-*EGFR* mutant (NCI-H522, NCI-H1299, A549 and NCI-H460) and *EGFR* mutant cells (NCI-H1975, PC-9, NCI-H1650) was performed with one-way ANOVA statistical test.

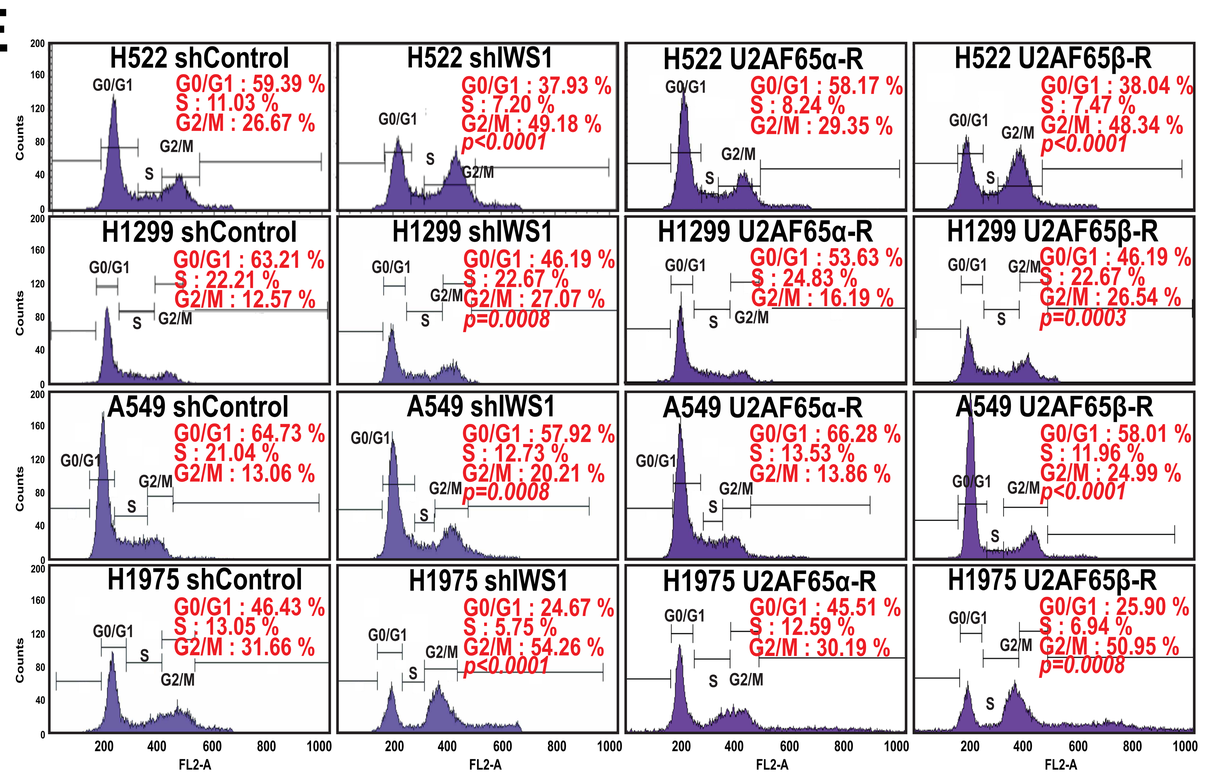
All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

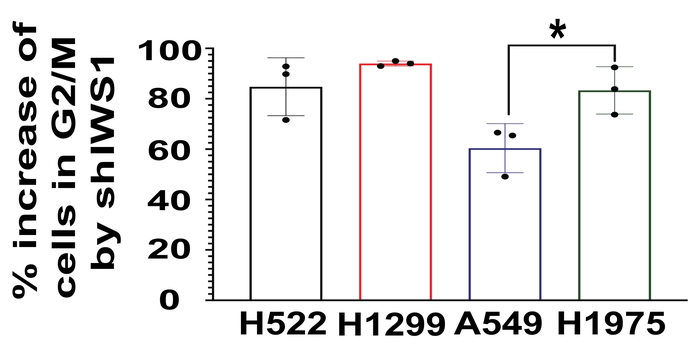

A
shControl
shIWS1
shIWS1/Sor wt
shIWS1/Sor DM-A
shIWS1/Sor DM-E
NCI-H522
NCI-H1299
Confluence %
Days in culture

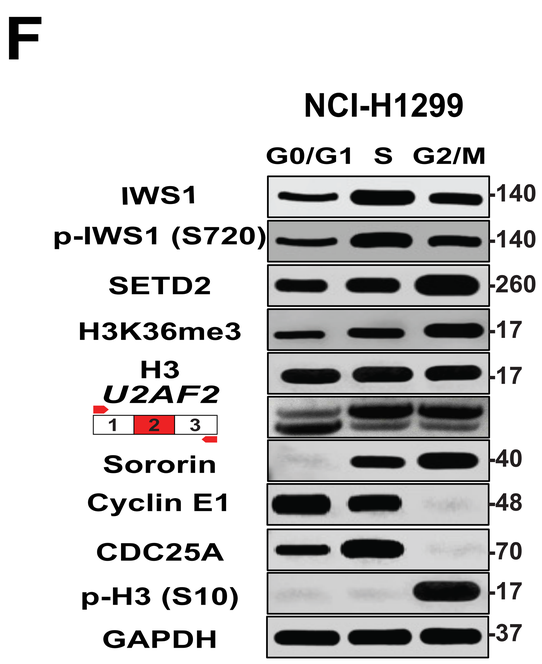

B
shControl
shIWS1
U2AF65α-R
U2AF65β-R
NCI-H522
NCI-H1299
Confluence %
Days in culture
n.s.

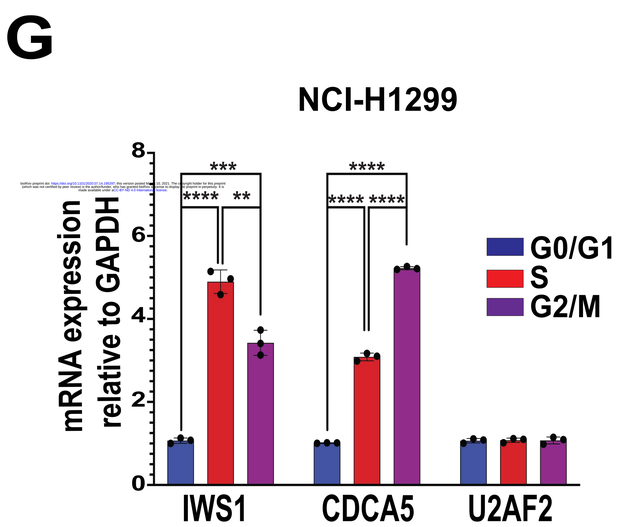

C
shControl
shIWS1
A549
NCI-H1975
Confluence %
Days in culture

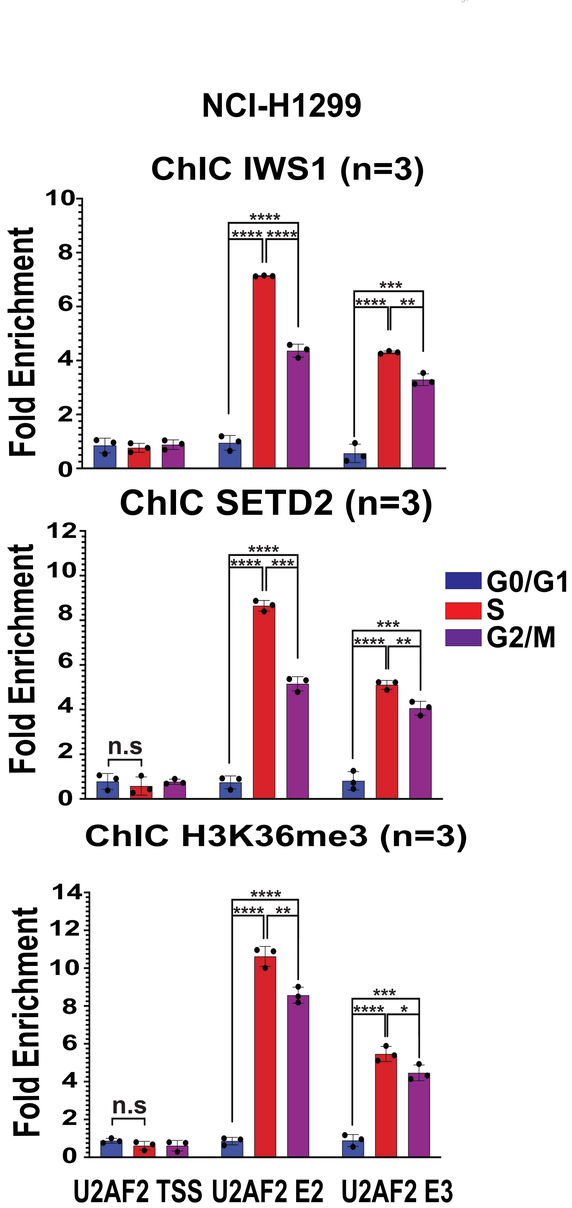

D
shControl
shIWS1
U2AF65α-R
U2AF65β-R
A549
NCI-H1975
Confluence %
Days in culture

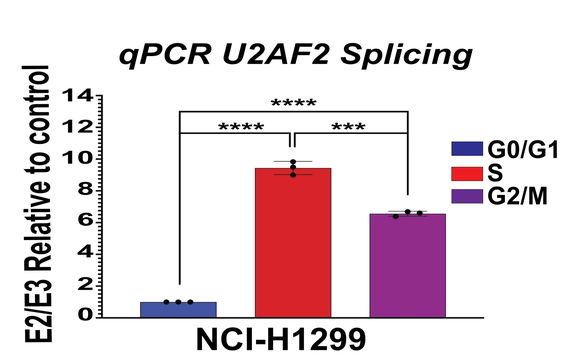

E
H522 shControl
G0/G1 : 59.39 %
S : 11.03 %
G2/M : 26.67 %
H522 shIWS1
G0/G1 : 37.93 %
S : 7.20 %
G2/M : 49.18 %
p<0.0001
H522 U2AF65α-R
G0/G1 : 58.17 %
S : 8.24 %
G2/M : 29.35 %
H522 U2AF65β-R
G0/G1 : 38.04 %
S : 7.47 %
G2/M : 48.34 %
p<0.0001
H1299 shControl
G0/G1 : 63.21 %
S : 22.21 %
G2/M : 12.57 %
H1299 shIWS1
G0/G1 : 46.19 %
S : 22.67 %
G2/M : 27.07 %
p=0.0008
H1299 U2AF65α-R
G0/G1 : 53.63 %
S : 24.83 %
G2/M : 16.19 %
H1299 U2AF65β-R
G0/G1 : 46.19 %
S : 22.67 %
G2/M : 26.54 %
p=0.0003
A549 shControl
G0/G1 : 64.73 %
S : 21.04 %
G2/M : 13.06 %
A549 shIWS1
G0/G1 : 57.92 %
S : 12.73 %
G2/M : 20.21 %
p=0.0008
A549 U2AF65α-R
G0/G1 : 66.28 %
S : 13.53 %
G2/M : 13.86 %
A549 U2AF65β-R
G0/G1 : 58.01 %
S : 11.96 %
G2/M : 24.99 %
p<0.0001
H1975 shControl
G0/G1 : 46.43 %
S : 13.05 %
G2/M : 31.66 %
H1975 shIWS1
G0/G1 : 24.67 %
S : 5.75 %
G2/M : 54.26 %
p<0.0001
H1975 U2AF65α-R
G0/G1 : 45.51 %
S : 12.59 %
G2/M : 30.19 %
H1975 U2AF65β-R
G0/G1 : 25.90 %
S : 6.94 %
G2/M : 50.95 %
p=0.0008
Counts
FL2-A


% increase of cells in G2/M by shIWS1
H522 H1299 A549 H1975


F
NCI-H1299
G0/G1 S G2/M
IWS1
p-IWS1 (S720)
SETD2
H3K36me3
H3
U2AF2
Sororin
Cyclin E1
CDC25A
p-H3 (S10)
GAPDH


G
NCI-H1299
mRNA expression relative to GAPDH
G0/G1
S
G2/M
IWS1
CDCA5
U2AF2


H
NCI-H1299
ChIC IWS1 (n=3)
ChIC SETD2 (n=3)
ChIC H3K36me3 (n=3)
Fold Enrichment
G0/G1
S
G2/M
n.s
U2AF2 TSS
U2AF2 E2
U2AF2 E3


I
qPCR U2AF2 Splicing
E2/E3 Relative to control
G0/G1
S
G2/M
NCI-H1299

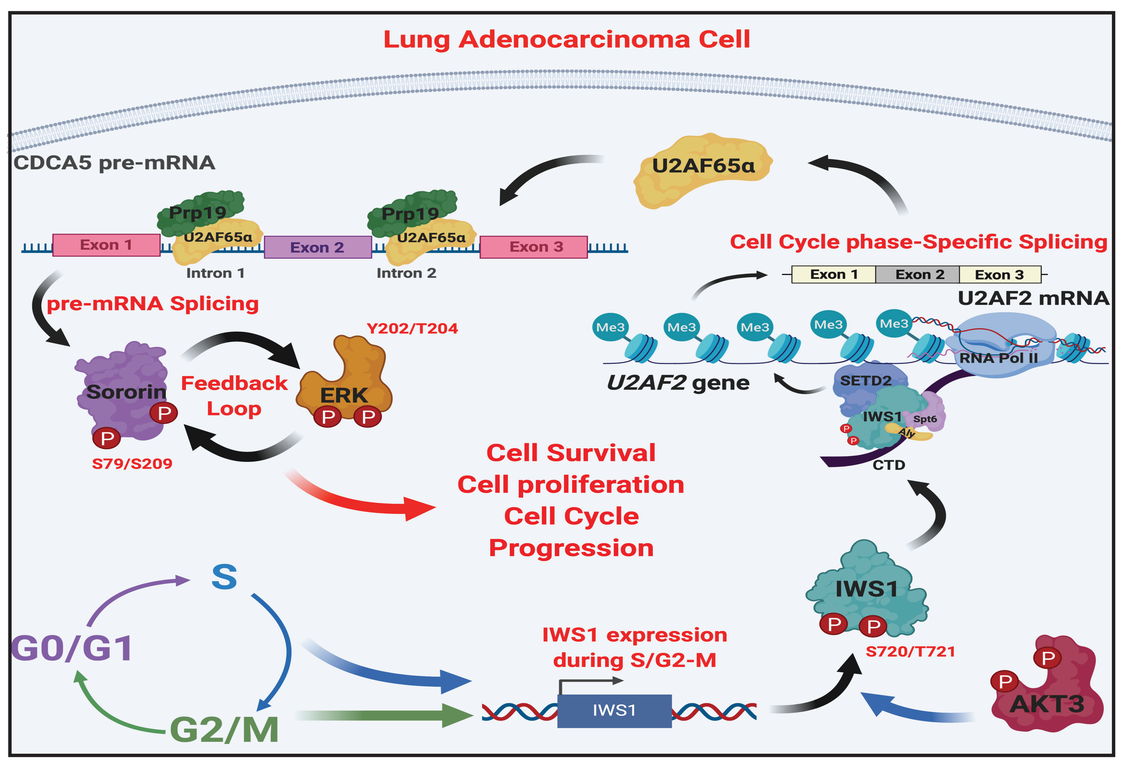

J
Lung Adenocarcinoma Cell
CDCA5 pre-mRNA
U2AF65α
Prp19
U2AF65α
Exon 1
Exon 2
Exon 3
Intron 1
Intron 2
pre-mRNA Splicing
Sororin
Feedback Loop
ERK
Y202/T204
S79/S209
Cell Survival
Cell proliferation
Cell Cycle Progression
Cell Cycle phase-Specific Splicing
U2AF2 mRNA
U2AF2 gene
RNA Pol II
SETD2
IWS1
Spt6
CTD
S
G0/G1
G2/M
IWS1 expression during S/G2-M
S720/T721
AKT3

**Figure 6. IWS1 phosphorylation promotes cell proliferation by controlling the Sororin/ERK phosphorylation feedback loop, through *U2AF2* RNA splicing.**

**A.** Growth curves of the indicated NCI-H522 (upper) and NCI-H1299 (lower) cells in media supplemented with 10% FBS. Cell proliferation was measured every 6 hours using an Incucyte live cell imager. Images of live cells in three independent cultures for each cell type were taken and analyzed using the native Incucyte confluence masking software. Results were expressed as confluence percentages ± SD. For simplicity, 12 and not 6 hour time points are shown. P values were calculated for the endpoint measurements, using the one-side unpaired t-test. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

**B.** *U2AF2 alternative RNA splicing affects cell proliferation.* Growth curves of the indicated NCI-H522 (upper) and NCI-H1299 (lower) cells in media supplemented with 10% FBS. The acquisition, representation and statistical analysis were performed similar to Fig. 6A.

**C.** *IWS1 phosphorylation promotes cell proliferation more robustly in the EGFR mutant lung adenocarcinoma cell line NCI-H1975, than in the KRAS mutant cell line A549.* Growth curves of shControl and shIWS1 A549 (*KRAS* mutant) and NCI-H1975 (*EGFR* mutant) cells in media supplemented with 10% FBS. Cell proliferation was measured as in A.

**D.** Growth curves of the indicated A549 (upper) and NCI-H1975 (lower) cells in media supplemented with 10% FBS. The acquisition, representation and statistical analysis were performed similar to Fig. 6A.

**E.** *IWS1 phosphorylation regulates progression through the G2/M phase of the cell cycle.* Cell cycle profiles of the indicated propidium iodide (PI)-stained cell lines, transduced with shControl, shIWS1 (Left upper panels) or rescued by U2AF65α oρ U2AF65β lentiviral constructs (Right upper panels). The cells were plated in equal numbers and they were harvested from semi-confluent cultures 48 hours later. Figure shows one representative, out of three biological replicates for each cell line. Numbers in red show the mean

percentage of cells in different phases of the cell cycle, measured based on the three biological replicates. The percent of cells in G2/M in shIWS1 cells was compared with the percent of cells in G2/M in shControl cells and the p value of the difference was calculated with the one-side unpaired t-test. (Lower panel) Bars show the shIWS1-induced percent change of the percentage of cells in G2/M in the indicated cell lines. The error bars show the SD of the percent change in 3 biological replicates. The percent change is smallest in the *KRAS* mutant cell line A549. Supplemental Table S4 lists the data for all the replicates and all the conditions. The statistical analysis between the NCI-H1975 and A549 was performed using one-sided unpaired t-test.

**F.** *IWS1 phosphorylation and U2AF2 alternative RNA splicing fluctuate during progression through the cell cycle.* Protein extracts derived from Exponentially growing NCI-H1299 cell cultures were separated into fractions enriched for cells in G0/G1, S and G2/M. Lysates of these cell fractions were probed with the indicated antibodies. RT-PCR, using RNA from the same cell fractions, and *U2AF2* exons 1 and 3 oligonucleotide primers, shows that *U2AF2* exon 2 inclusion in the mature *U2AF2* mRNA increases in cells in S and G2/M.

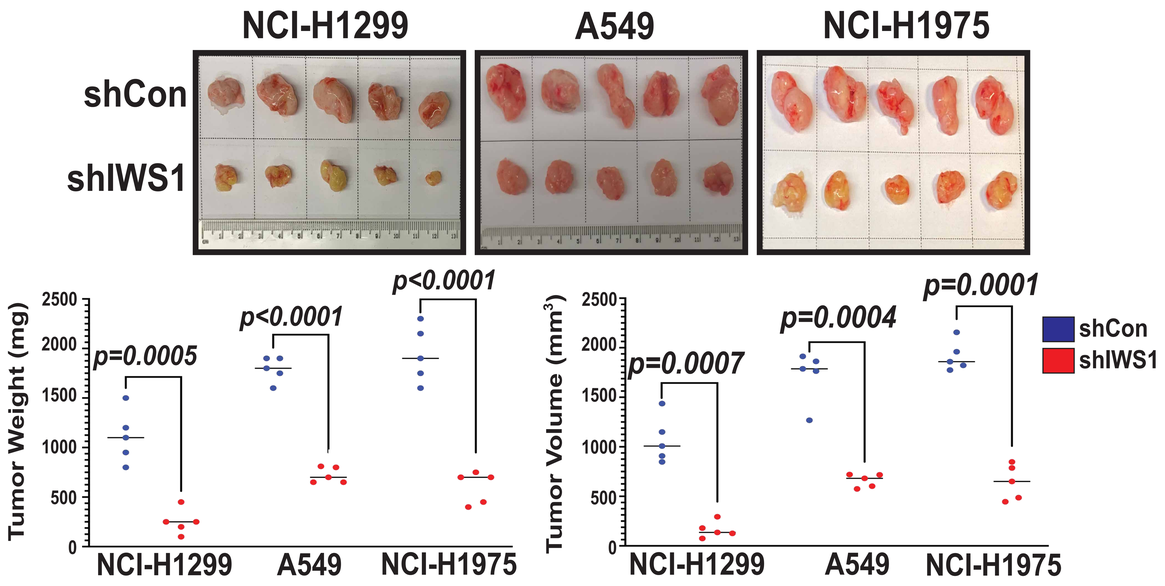
**G.** The abundance of the *IWS1*, *CDCA5* and *U2AF2* mRNAs in the G0/G1, S and G2/M phases of the cell cycle in NCI-H1299 cells was measured by quantitative RT-PCR and is presented relative to the abundance of GAPDH ± SD.

**H.** *IWS1 phosphorylation regulates the U2AF2 alternative RNA splicing in a cell-cycle specific manner*. ChIC assays showing the binding of IWS1 and SETD2 (upper and middle panels respectively) and the abundance of H3K36me3 (lower panel) in the *U2AF2* gene, in NCI-H1299 cells in the G0/G1, S and G2/M phases of the cell cycle. The bars show the mean fold enrichment in IWS1 and SETD2 binding, and in H3K36me3 abundance, in the indicated regions of the *U2AF2* gene ±SD.

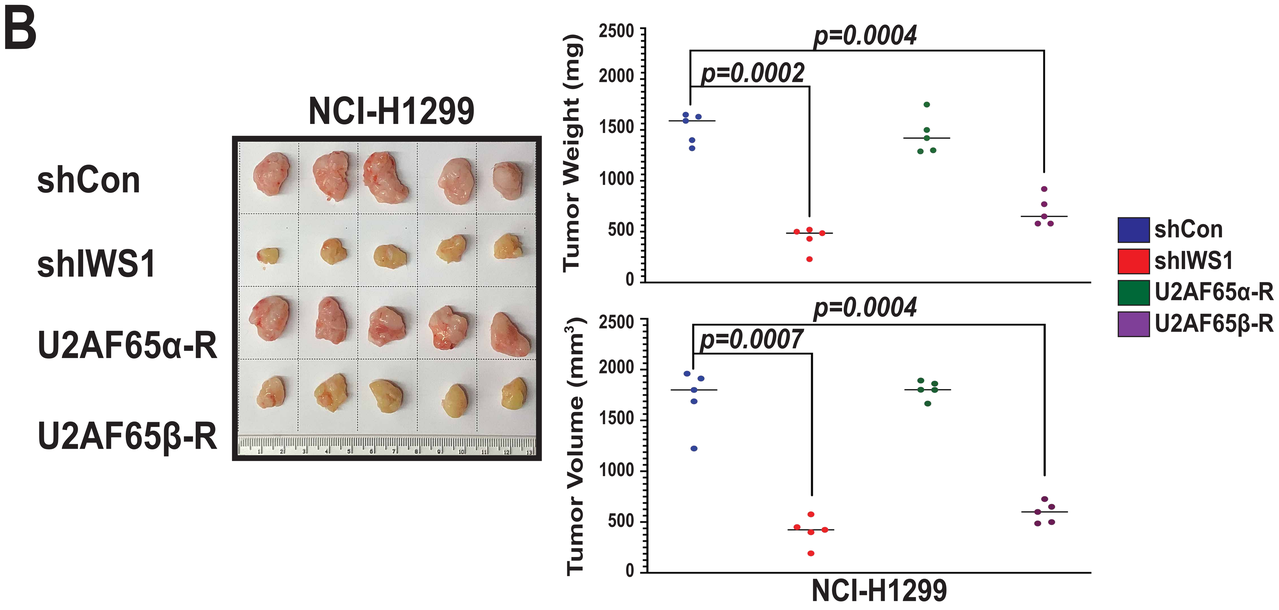
**I.** Quantitative RT-PCR showing the *E2/E3 U2AF2* ratio in NCI-H1299 cells. Bars show the mean E2/E3 *U2AF2* ratio in NCI-H1299 cellular fractions enriched for cells in S and G2/M phase relative to G0/G1 enriched fractions. In 6G to 6I, the error bars show the SD of at least 3 biological replicates. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test)

**J.** *Model of the regulation of the Sororin/ERK loop and cell cycle progression via the U2AF2 alternative RNA splicing, through the cell cycle.* IWS1 expression occurs specifically during the S and G2/M phase of the cell cycle. Following this cell cycle specific expression, IWS1 is phosphorylated by Akt3 at S720/T721 and orchestrates the assembly of epigenetic complexes on *U2AF2* gene, which translate the SETD2-mediated H3K36me3 signal into shifts of the alternative RNA splicing pattern of U2AF2, in a cell cycle specific manner. Subsequently, the U2AF65α isoform along with Prp19, facilitate the proper splicing of *CDCA5* pre-mRNA, leading to accumulation of Sororin during S and G2/M phase. Finally, Sororin forms a positive feedback loop with ERK phosphorylation. Activation of this loop plays an important role in the maintenance of ERK phosphorylation, and in the progression through the G2/M phase of the cell cycle in lung adenocarcinoma.

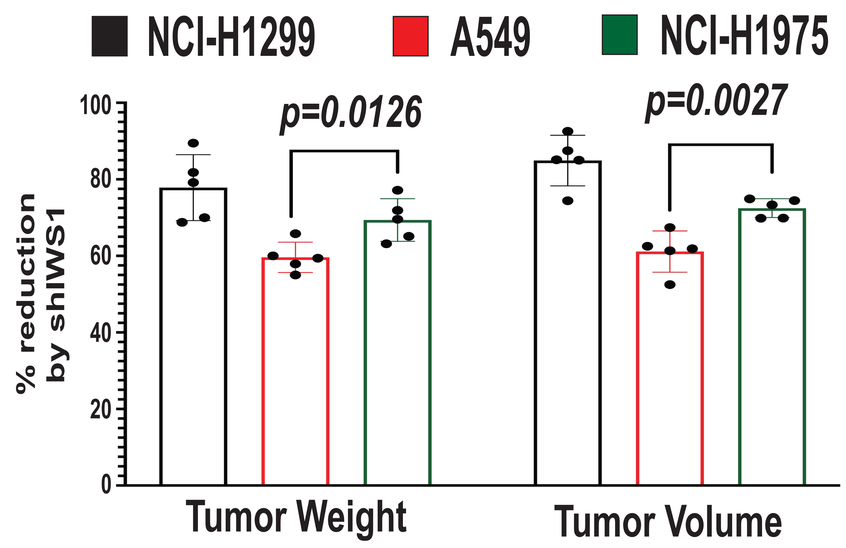
A



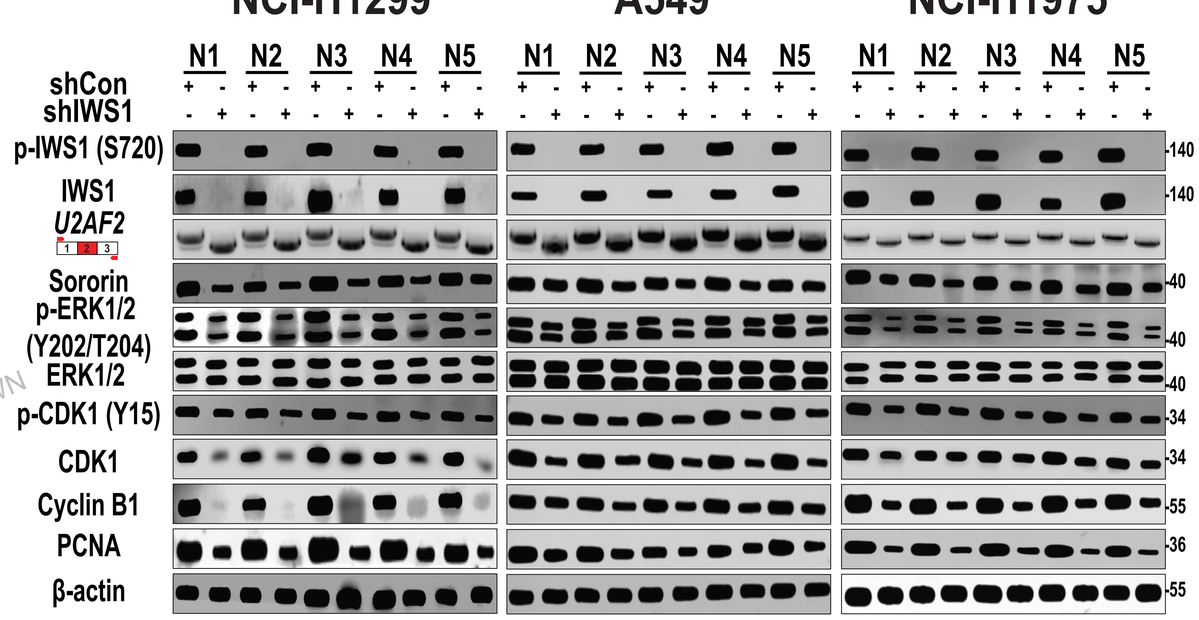

B




C

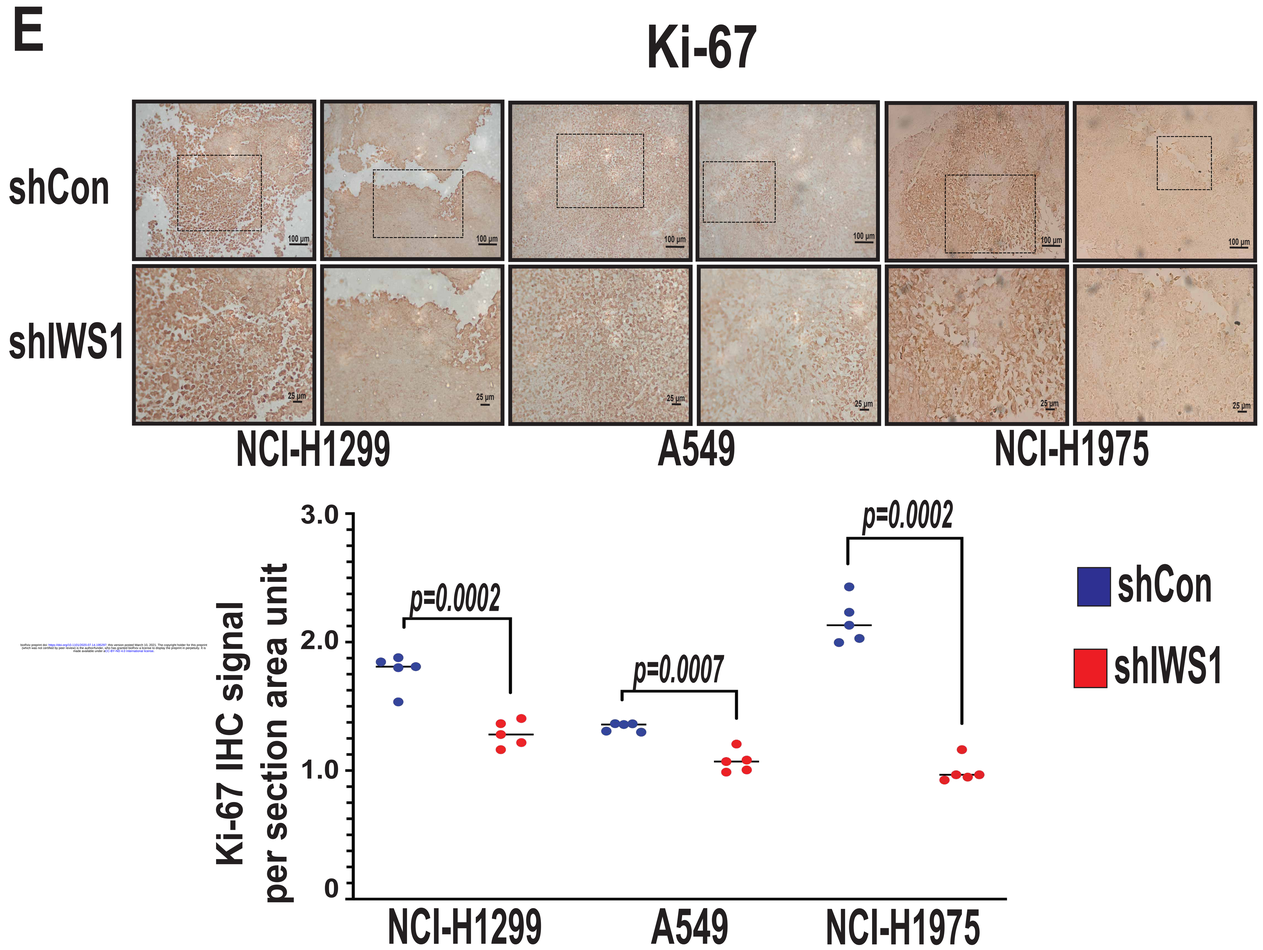



D






E




F

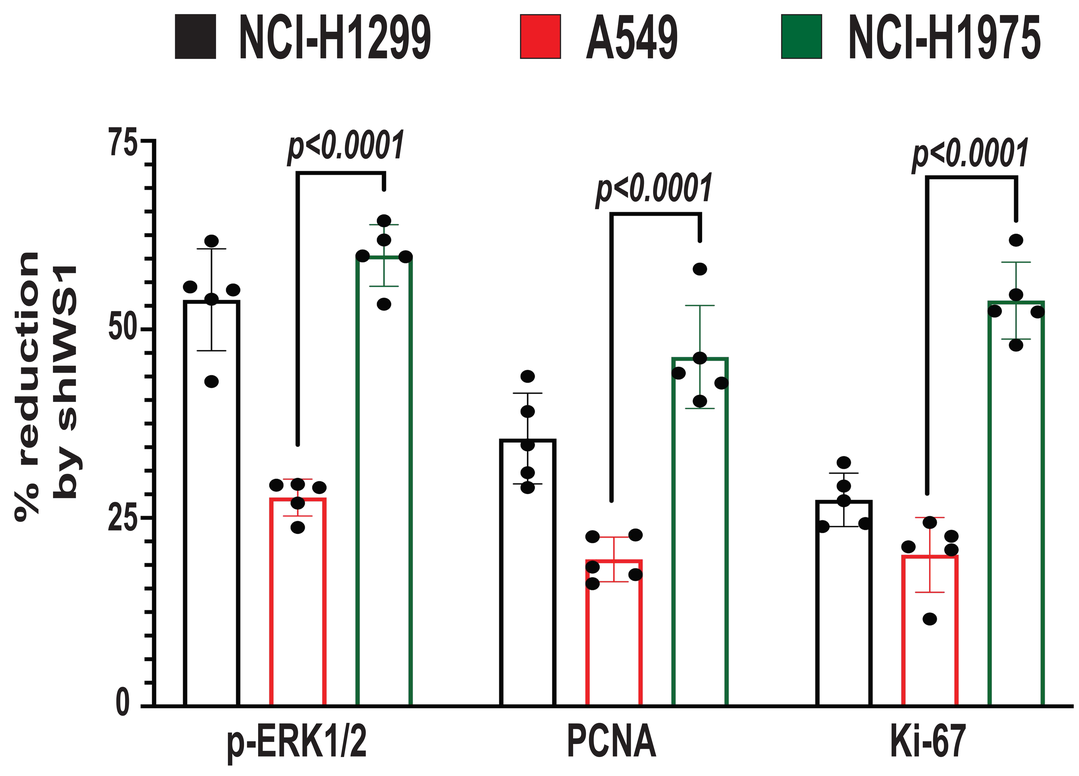


**Figure 7. IWS1 phosphorylation controls tumor growth *in vivo*, by regulating the U2AF2/Sororin/ERK axis.**

**A.** *IWS1 phosphorylation at Ser720/Thr721 promotes tumor growth in xenograft lung adenocarcinoma model in NSG mice*. NSG mice were injected subcutaneously with shControl or shIWS1 NCI H1299, A549 or NCI-H1975 cells. (N=5 mice/group). (Left panels) Images of the induced tumors, harvested at 4 weeks (NCI-H1299, NCI-H1975 cells) and 6 weeks (A549 cells) from the time of inoculation (Right panels) Scatter plots showing the tumor weight (up) and volume (down) of the harvested tumors. The horizontal lines indicate mean tumor weight or volume. Statistical analyses were done using the paired t-test.

**B.** *Inclusion of U2AF2 E2 promotes tumor growth in xenograft lung adenocarcinoma model in NSG mice.* NSG mice were injected subcutaneously with shControl, shIWS1, shIWS1/U2AF65α or shIWS1/U2AF65β NCI-H1299 cells. (N=5 mice/group). (Left panels) Images of the induced tumors, harvested at 4 weeks from the time of inoculation (Right panels) Scatter plots showing the tumor weight (up) and volume (down) of the harvested tumors. The horizontal lines indicate mean tumor weight or volume. Statistical analyses were done using the paired t-test.

**C.** From the weight and volume of the tumors induced by shIWS1-transduced cells and shControl-transduced cells in Figure 7A, we calculated the shIWS1-induced percent reduction of both tumor weight and volume. Bars show the mean percent reduction. The error bars show the SD of the percent reduction in each group. Statistical analyses were done using the one-sided unpaired t-test.

**D.** *IWS1 phosphorylation controls the Sororin/ERK phosphorylation axis in vivo.* Cell lysates derived from NCI-H1299, A549 and NCI-H1975 (shControl and shIWS1) mouse xenografts, were probed with the indicated antibodies. RT-PCR, using RNA isolated from

these xenografts and *U2AF2* exon 1 and exon 3 oligonucleotide primers shows that the knockdown of IWS1 in the xenografts results in *U2AF2* exon 2 exclusion, as expected.

**E.** (Upper panels) Ki-67 staining of tumor xenografts of shControl and shIWS1-transduced NCI-H1299 (left), A549 (middle) and NCI-H1975 (right) cells. Formalin-fixed, paraffin-embedded tumor samples were stained with a Ki-67 antibody. Secondary antibody was HRP-labelled. The boxes delineate the area of higher magnification shown in the image below. Scale bar in the right corner of each image. (Lower panel) Scatter plots showing the Ki-67 IHC signal relative to the section area in shControl and shIWS1 NCI-H1299, A549 and NCI-H1975 tumors. The horizontal line shows the mean Ki-67 signal in the indicated groups of xenografts. Statistical analyses were performed, using the paired t-test.

**F.** *KRAS mutant tumors are more resistant to the loss of IWS1 in vivo.* Percentage of reduction of the described marker induced by shIWS1 in the tumors derived from the NCI-H1299, A549 and NCI-H1975 cell lines. Bars show the mean reduction of the markers, derived from the Western blots data in Figure 7D (p-ERK and PCNA) and IHC data in Figure 7E (Ki-67). The error bars are SD of the values in each mice group. The statistics were performed with one-sided unpaired t-test.

**Figure 8. The p-IWS1/U2AF2 pathway is active in human lung adenocarcinomas and impacts tumor grade, stage, metastatic potential and treatment relapse in patients with *EGFR* mutant, but not *KRAS* mutant tumors.**



**A.** Lysates from 30 LUAD samples, paired with NAT and 10 unpaired LUAD samples were probed with the indicated antibodies. (Fifth line from the top). RT-PCR, using RNA isolated from these tumors, and *U2AF2* exon 1 and exon 3 oligonucleotide primers, shows that phosphor-IWS1 correlates with *U2AF2* exon 2 inclusion, as expected. (Sixth line) *U2AF2* E2/E3 exon ratio in the tumors in the upper panel, relative to the average of the 30 normal lung samples.

**B.** *IWS1 and phosphor-IWS1 correlate with U2AF2 exon 2 inclusion and the activity of the Sororin/ERK axis, most robustly in EGFR mutant lung adenocarcinomas.* Heatmaps of the correlation coefficients between the indicated components of the IWS1 phosphorylation pathway, in the entire cohort (left), in *KRAS* mutant (middle) and in *EGFR* mutant tumors (right). The correlation coefficients were calculated using simple linear regression. The values and the statistical confidence of all the comparisons can be found in Supplementary Table S5.

**C.** Violin plots showing the abundance of IWS1 phosphorylation (left) and the *U2AF2* E2/E3 ratio (right) in stage I and Stage II/III tumors. Data shown for all tumors in A and B, and selectively for *EGFR* or *KRAS* mutant tumors. The horizontal black lines indicate mean values for p-IWS1 levels and *U2AF2* E2/E3 ratios. Statistical analyses were performed using the one-sided unpaired t-test.

**D.** Kaplan-Meier Curves showing the impact of the abundance of phosphor-IWS1 (left) and *U2AF2* E2/E3 ratio (right) on patient survival in the LUAD cohort in Fig. 8A. Separate curves are shown again for all the patients, or for the patients with *KRAS* or *EGFR* mutant

tumors. The effect of phosphor-IWS1 levels and *U2AF2* exon 2 inclusion on patient survival is again significant in *EGFR* mutant, but not in *KRAS* mutant tumors. Statistical analyses were performed using the log rank test and Cox's proportional hazards model.

**E.** Heat Maps showing the correlation of IWS1 or SRSF1 with components of the IWS1/U2AF2 pathway in the TCGA LUAD database. The RNA-seq expression values from each individual patient were expressed as log2 normalized counts and they were used to calculate the correlation coefficients.

**F.** *The IWS1 phosphorylation axis correlated with relapse in lung adenocarcinoma patients harboring EGFR mutations.* Heatmaps showing the expression of IWS1, CDCA5, CDK1 and CCNB1 in relapsed and in non-relapsed tumors. Microarray analyses of lung adenocarcinomas in the entire cohort (left), *KRAS* mutant (middle) or *EGFR* mutant (right) cohort in the GSE13213 dataset. For quantification of gene expression we used the z-scores of the microarray signals.

**G.** *IWS1 expression and the U2AF2 exon 2 inclusion and FGFR2 exon IIIb exclusion pathways correlate strongly with metastasis in lung adenocarcinomas.* (Left panel) Heat Maps showing the expression of IWS1 and its downstream targets in the IWS1/U2AF2 pathway in resected brain metastatic lesions and primary lung adenocarcinomas. RNA-seq data from the GSE141685 dataset and primary lung adenocarcinomas derived from TCGA LUAD database. The TCGA LUAD samples were picked randomly, independent of their clinical stage or mutational status. For quantification of gene expression we used the z-scores of the normalized RNA-Seq reads. The primary tumors from the TCGA LUAD database are listed in Table S6. (Right Upper panel) Force-directed layout/t-SNE of all metastatic tumor cells isolated from patient-derived xenograft mice (Laughney et al., 2020[79]) colored by source, derived from scRNA-Seq data from the GSE123903. (Right lower panel) Force-directed layout (as in the upper panel) of all xenograft tumor cells colored by *z*-normalized *IWS1* and *CDCA5* expression.

**H.** *The IWS1 phosphorylation axis defines poor survival in lung adenocarcinoma patients harboring EGFR mutations.* Kaplan-Meier Curves showing the impact of IWS1 expression (left), *U2AF2* E2/E3 ratio (middle) and *FGFR2* IIIb/IIIc ratio (right) on patient survival in patients in the TCGA LUAD database. Separate curves are shown for all the patients, or for the patients with K*RAS* or *EGFR* mutant tumors. The effect of IWS1 levels and *U2AF2* exon 2 inclusion in patient survival is significant in *EGFR* mutant, but not in *KRAS* mutant tumors. Statistical analyses were performed using the log rank test and Cox's proportional hazards model.

**Figure S1 (relative to Figure 1). IWS1 expression and/or phosphorylation regulate alternative mRNA splicing**

**A.** Venn diagram shows the genes that are differentially expressed in shControl versus shIWS1 or shIWS1/WT-R versus shIWS1/MT-R NCI-H522 cells (q<0.01) and the overlap between the two sets.

**B.** Volcano plots of all the differentially expressed genes in shControl versus shIWS1(upper panel) and shIWS1/WT-R versus shIWS1/MT-R NCI-H522 cells (lower panel). Genes upregulated or downregulated with a p value <0.01 are shown in red and blue, respectively.

**C.** Heatmaps of the top 100 differentially-expressed genes (lowest FDR) in shIWS1 relative to shControl and in shIWS1/MT-R relative to shIWS1/WT-R NCI-H522 cells. Genes present in both sets and regulated to the same direction are shown in red. The abundance of IWS1 was significantly lower in shIWS1 cells, relative to the shControl cells, as expected (red box).

**D.** Gene Set Enrichment Analysis (GSEA) of genes differentially expressed in shControl versus shIWS1 (upper panels) and in shIWS1/WT-R versus shIWS1/MT-R (lower panels) NCI-H522 cells demonstrates significant enrichment in genes involved in RNA processing. (NES–normalized enrichment score, FDR–false discovery ratio)

**E.** *IWS1 expression and phosphorylation regulate the alternative splicing of the U2AF2 exon 2 and FGFR2 exon IIIb.* Quantitative RT-PCR showing the *E2/E3* and *IIIb/IIIc* ratio in the *U2AF2* (upper panel) and *FGFR2* (lower panel) mRNA transcripts in shControl, shIWS1, shIWS1/WT-R, shIWS1/MT-R and shIWS1/DE MT-R NCI-H522 and NCI-H1299 cells. Bars show the E2/E3 or IIIb/IIIc ratio (mean ± SD) in all the cell lines, normalized relative to the shControl.

**F.** Quantitative RT-PCR, showing the abundance of the total *U2AF2* mRNA transcripts (upper panel) or the U2AF2 E3/U2AF2 total ratio in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells. Bars show the total *U2AF2* mRNA (mean ± SD) relative to *GAPDH* and the U2AF2 E3/U2AF2 Total (mean ± SD) relative to shControl.

**G.** **H. I** and **J**. IWS1 expression and phosphorylation regulate the alternative RNA splicing of *SLC12A2* (exon 21), *IFT88* (exon 8), *STXBP1* (exon 18) and *C1qTNF6* (exon 3). In all cases, IWS1 promotes exon inclusion. Upper panels show the electrophoresed products of the corresponding RT-PCR reactions. These products were quantified and the ratios of the alternatively spliced to non-alternatively spliced adjacent exons were normalized relative to the ratio in the shControl cells. Bars show the relative exon ratios (mean ± SD) in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells. Assays were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

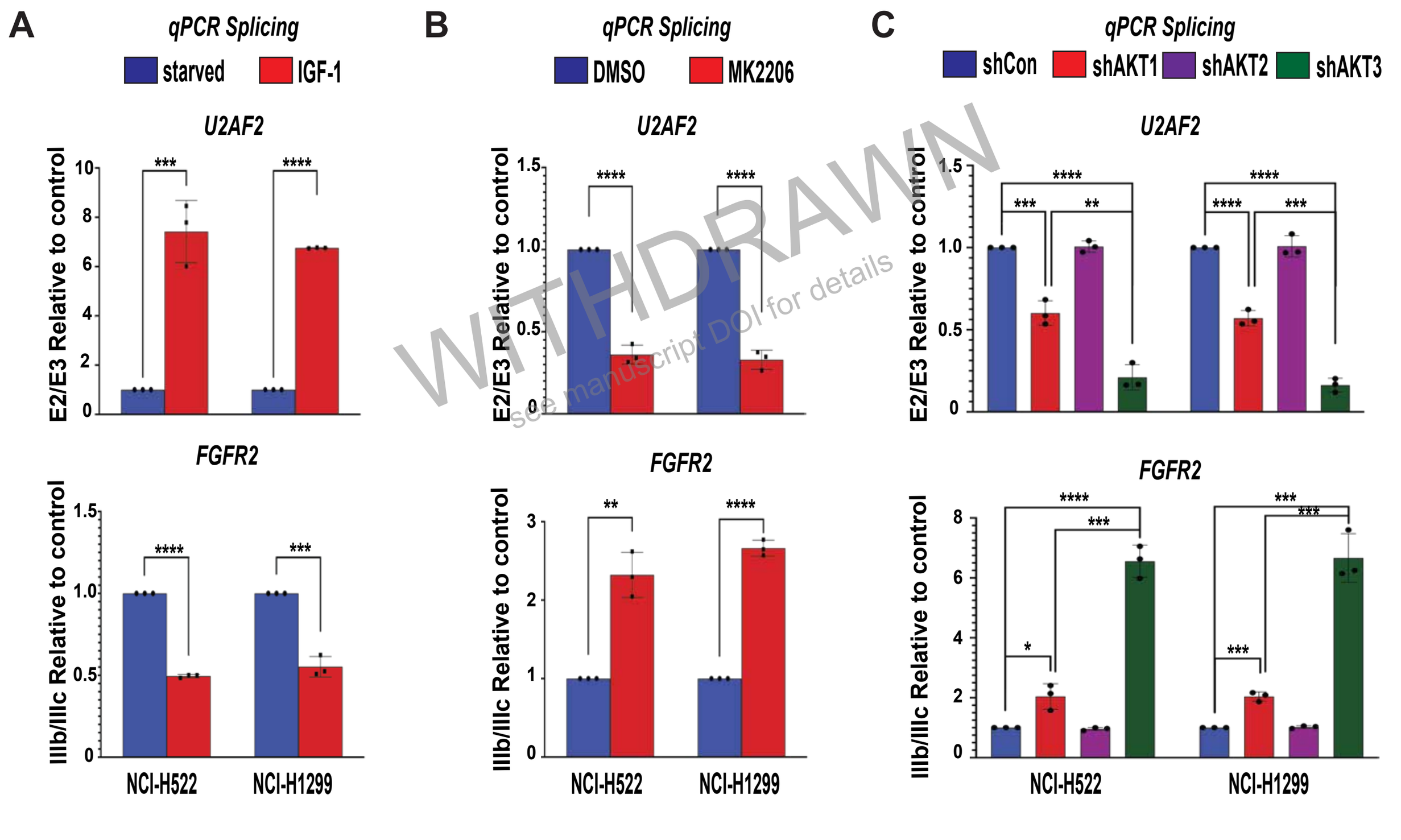

A
qPCR Splicing
starved
IGF-1
U2AF2
E2/E3 Relative to control
FGFR2
IIIb/IIIc Relative to control
NCI-H522
NCI-H1299
B
qPCR Splicing
DMSO
MK2206
U2AF2
E2/E3 Relative to control
FGFR2
IIIb/IIIc Relative to control
NCI-H522
NCI-H1299
C
qPCR Splicing
shCon
shAKT1
shAKT2
shAKT3
U2AF2
E2/E3 Relative to control
FGFR2
IIIb/IIIc Relative to control
NCI-H522
NCI-H1299
WITHDRAWN
see manuscript DOI for details

**Figure S2 (relative to Figure 2).IWS1 phosphorylation-dependent mRNA splicing of U2AF2 is regulated by serum and IGF-1-induced signals transduced by AKT3**

**A. B. and C**. (Upper panels) The *U2AF2* mRNA E2/E3 ratio in the cells in Figure 2 was measured by quantitative RT-PCR. Bars show the E2/E3 ratio in IGF-1, MK2206 and shAKT1, shAKT2 and shAKT3 cells relative to the control cells. (Lower panels) The *FGFR2* mRNA IIIb/IIIc ratio in the same cells was also measured by quantitative PCR. Bars show the IIIb/IIIc ratio in IGF-1, MK2206 and shAKT1, shAKT2 and shAKT3 cells relative to the control cells. Error bars indicate SD. All experiments in this figure were done on three biological replicates, in triplicate. n.s: non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

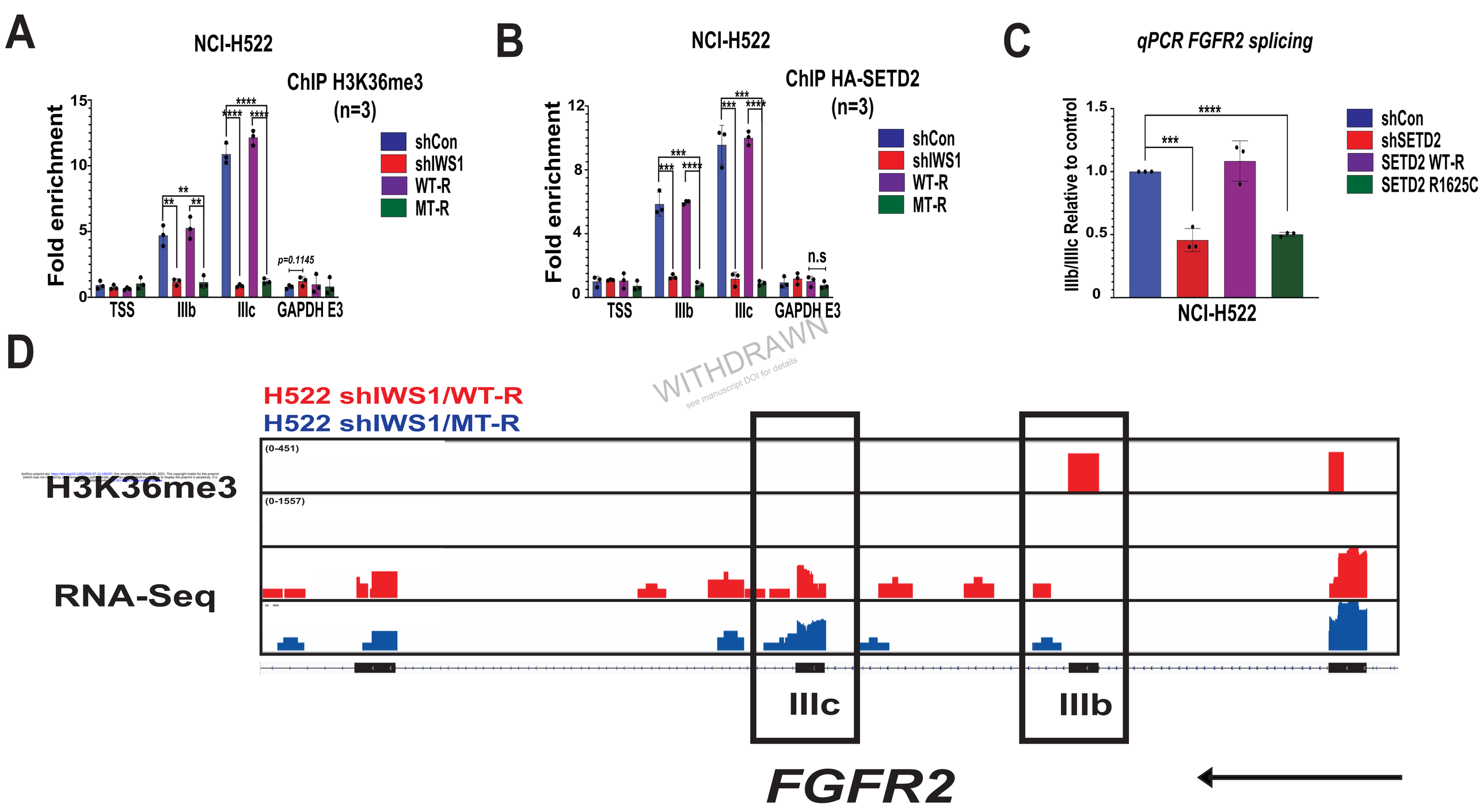


**Figure S3 (relative to Figure 3). *FGFR2* Exon IIIb exclusion, induced by IWS1 phosphorylation at Ser720/Thr721, depends on H3K36 trimethylation by SETD2.**

**A.** ChIP assays showing the abundance of H3K36me3 on the *FGFR2* and *GAPDH* genes in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells. Bars show the mean fold enrichment in H3K36me3 (anti-H3K36me3 IP, vs IgG control IP) in the indicated regions of the *FGFR2* gene, in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%).

**B.** ChIP assays showing the binding of HA-SETD2 to the indicated regions of the *FGFR2* and *GAPDH* genes, in shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522, transduced with a lentiviral HA-SETD2 construct. The bars show the mean fold enrichment in SETD2 binding (anti-HA IP, vs IgG control IP) in the indicated regions of the *FGFR2* gene, in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R cells ±SD. Data were normalized relative to the input (2%)

**C.** *The IWS1-regulated alternative RNA splicing of FGFR2, depends on enzymatically active SETD2*. The ratio of the *FGFR2* exons IIIb and IIIc was determined by quantitative RT-PCR, using the mRNA isolated from the cells from cells in Figure 3D. Again, bars show the *FGFR2* IIIb/IIIc ratio relative to the shControl. Error bars indicate SD. All experiments in this figure were done on three biological replicates, in triplicate. n.s: non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

**D.** Snapshots of the integrative genomic viewer showing the distribution of histone H3K36me3 marks, within the *FGFR2* gene, obtained from a ChIP-Seq analysis, along with the distribution of RNA reads obtained from the RNA-seq analysis of shIWS1/WT-R and shIWS1/MT-R NCI-H522 cells. Scale represents reads per million (RPM). Snapshots of peaks detected in both biological replicates are shown. The black box outlines *FGFR2* exons IIIb and IIIc and the adjacent regions. Arrow shows the direction of transcription.

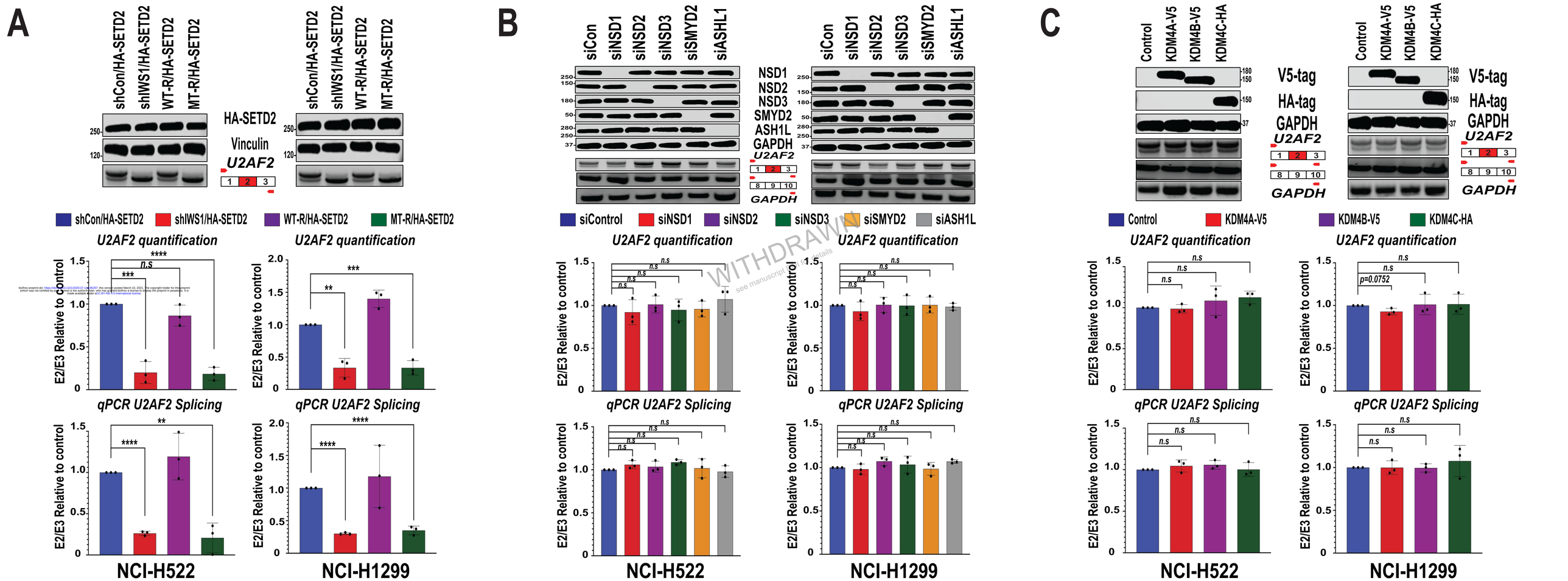


**Figure S4 (relative to Figure 3). *U2AF2* Exon 2 inclusion, induced by IWS1 phosphorylation at Ser720/Thr721, depends on H3K36 trimethylation by SETD2.**

**A.** *SETD2 does not rescue the shIWS1 and shIWS1/MT-R U2AF2 alternative splicing phenotype.* (Upper panels-First two lines) Lysates of shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 and NCI-H1299 cells, transduced with a lentiviral construct of wild type HA-SETD2, were probed with the indicated antibodies. (Upper panels-third line) RT-PCR reactions, using mRNA derived from the same cells, and oligonucleotide primers mapping in *U2AF2* exons 1 and 3 show that SETD2 does not rescue the shIWS1-induced exclusion of exon 2 from the *U2AF2* mRNA. (Middle panel) *U2AF2 E2/E3* ratio in HA-SETD2-transduced shIWS1, shIWS1/WT-R and shIWS1/MT-R cells, relative to HA-SETD2-transduced shControl cells. Data based on quantification of the RT-PCR products above. (Lower panel) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show again the E2/E3 ratio in HA-SETD2-transduced shIWS1, shIWS1/WT-R and shIWS1/MT-R cells, relative to HA-SETD2-transduced shControl cells ± SD.

**B.** (Upper panels. first 6 lines) The listed H3K36 methyltransferase were efficiently knocked down in NCI-H522 and NCI-H1299 cells, by siRNA transfection, as determined by probing western blots of transfected cell lysates with the indicated antibodies. (Upper panels, lines 7 and 9) RT-PCR, using RNA isolated from these cells, and *U2AF2* exon 1 and exon 3 or exon 8 and 10 oligonucleotide primers shows that none of these methyltransferases control the splicing pattern of the *U2AF2* mRNA. (Middle panels) *U2AF2 E2/E3* ratio in the siRNA-transfected cells, relative to the siControl cells. Data based on quantification of the RT-PCR products above. (Lower panels) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the *E2/E3* ratio in NCI-H522 and NCI-H1299 cells, after the knockdown of the indicated methyltransferases, relative to the siControl cells ± SD.

**C.** (Upper Panels, first 3 lines) The listed H3K36 demethylases were overexpressed in NCI-H522 and NCI-H1299 cells, using viral constructs (pLX304 constructs for KDM4A and KDM4B and a pBabe-puro construct for KDM4C). Expression was determined by probing western blots of transduced cell lysates with the indicated antibodies. (Upper panels 4$^{th}$ and 5$^{th}$ lines) RT-PCR, using RNA isolated from these cells, and *U2AF2* exon 1 and exon 3 or exon 8 and 10 oligonucleotide primers shows that none of these demethylases control the splicing pattern of the *U2AF2* mRNA. (Middle panels) *U2AF2 E2/E3* ratio in the transduced cells, relative to the vector-transduced (pLX304 EV plus pBabe-puro EV) cells. Data based on quantification of the RT-PCR data above. (Lower panels) The RT-PCR results were confirmed by quantitative RT-PCR. Bar graphs show the *E2/E3* ratio in NCI-H522 and NCI-H1299 cells, after transduction with the indicated demethylases, relative to the Control cells ± SD. All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test)

A

shControl shMRG15

qPCR FGFR2 Splicing

IIIb/IIIc Relative to control

5.0 2.5 0

** **

shControl shPTB

4.5 3.0 1.5 0

**** **

NCI-H522 NCI-H1299

B

shControl + - + -

shPTBP2 - + - +

60- 50- PTBP2

110- α-actinin

U2AF2

1 2 3

GAPDH

shControl shPTBP2

qPCR U2AF2 Splicing

E2/E3 Relative to control

1.5 1.0 0.5 0

n.s n.s

qPCR FGFR2 Splicing

IIIb/IIIc Relative to control

4.5 3.0 1.5 0

*** ****

NCI-H522 NCI-H1299

C

siCon siPHF1 siBRPF1 siMSH6 siGLYR1 siLEDGF

50- PHF1

180- BRPF1

150- MSH6

50- GLYR1

37- LEDGF

75- β-actin

50-

U2AF2

1 2 3

8 9 10

GAPDH

siControl siPHF1 siBRPF1 siMSH6 siGLYR1 siLEDGF

U2AF2 quantification

E2/E3 Relative to control

1.5 1.0 0.5 0

**** n.s n.s n.s n.s

qPCR U2AF2 Splicing

E2/E3 Relative to control

1.5 1.0 0.5 0

**** n.s n.s n.s n.s

NCI-H1299

D

shControl shLEDGF p75/WT-R p52/WT-R p52 A51P-R

V5-p75 p75- -75

V5-p52 p52- -52

LEDGF

-55 a-tubulin

U2AF2

1 2 3

8 9 10

GAPDH

shControl shLEDGF p75 WT-R

p52 WT-R p52 A51P-R

U2AF2 quantification

E2/E3 Relative to control

1.5 1.0 0.5 0

**** n.s **** ****

qPCR U2AF2 Splicing

E2/E3 Relative to control

1.5 1.0 0.5 0

**** **** ***

NCI-H1299

E

shControl shLEDGF

qPCR FGFR2 Splicing

IIIb/IIIc Relative to control

1.5 1.0 0.5 0

n.s n.s

NCI-H522 NCI-H1299

F

NCI-H1299 ChIP V5-p52/LEDGF (n=3)

Fold enrichment

12 10 8 6 4 2 0

*** n.s ** ** ** ** ** n.s n.s n.s

TSS E2 E3 GAPDH E3

shCon shIWS1 WT-R MT-R

G

shControl + - - - shControl + - - -

shIWS1 - + + + shIWS1 - + + +

WT-R - - + - WT-R - - + -

MT-R - - - + MT-R - - - +

V5-p52 + + + + V5-p52 + + + +

52- V5-p52 52-

55- a-tubulin 55-

U2AF2

1 2 3

shCon/V5-p52 shIWS1/V5-p52 WT-R/V5-p52 MT-R/V5-p52

U2AF2 quantification

E2/E3 Relative to control

1.5 1.0 0.5 0

*** **** *** ****

qPCR U2AF2 Splicing

E2/E3 Relative to control

1.5 1.0 0.5 0

** *** *** ****

NCI-H522 NCI-H1299

**Figure S5 (relative to Figure 4). The regulation of the alternative splicing of the *U2AF2* exon 2 by *IWS1* phosphorylation, depends on the p52 isoform of the H3K36me3 reader *LEDGF* and its splicing partner *SRSF1*.**

**A.** *The H3K36me3 reader MRG15 and its binding partner PTB, control the FGFR2 alternative RNA splicing.* The ratio of the *FGFR2* exons IIIb and IIIc was determined by quantitative RT-PCR, using the mRNA isolated from the NCI-H522 (upper panel) and NCI-H1299 (lower panel) cells in Figure 4A. Again, bars show the *FGFR2* IIIb/IIIc ratio relative to the shControl. Error bars indicate SD.

**B.** *The paralog of PTB, PTBP2 plays no role in the regulation of U2AF2 alternative RNA splicing.* (Upper panel) Western blots of lysates of NCI-H522 and NCI-H1299 cells transduced with shControl and shPTBP2 were probed with the indicated antibodies. (Middle and lower panel) Quantitative RT-PCR, using RNA isolated from these cells shows that the knockdown of PTBP2 does not affect the alternative splicing of *U2AF2,* but affects the *FGFR2* splicing pattern. Bar graphs show the *U2AF2* E2/E3 and *FGFR2* IIIb/IIIc ratio in shPTBP2 NCI-H522 and NCI-H1299 cells, relative to the shControl ± SD.

**C.** The knockdown of the indicated histone H3K36me3 readers in NCI-H1299 cells confirmed that only the knockdown of LEDGF results in the exclusion of exon 2 from the mature *U2AF2* mRNA transcript. This is a repeat of the experiment in figure 4C, which was done in NCI-H522 cells. The results confirmed that LEDGF controls the mRNA splicing of *U2AF2* in multiple cell lines.

**D.** The expression of LEDGF-p75-WT, LEDGF-p52-WT and LEDGF-p52A51P (a LEDGF mutant, which has lost the ability to bind H3K36me3) in shLEDGF-transduced NCI-H1299 cells, confirmed that only LEDGF-p52-WT can rescue the shLEDGF-induced exclusion of exon 2 from the mature *U2AF2* mRNA transcript. This is a repeat of the experiment in figure 4C, which was done in NCI-H522 cells.

**E.** *The H3K36me3 reader LEDGF, plays no role in the regulation of FGFR2 alternative RNA splicing.* The ratio of the *FGFR2* exons IIIb and IIIc was determined by quantitative RT-PCR, using the mRNA isolated from the NCI-H522 and NCI-H1299 cells in Figure 4C. Bars show the *FGFR2* IIIb/IIIc ratio in shLEDGF relative to the shControl. Error bars indicate SD.

**F.** ChIP assays addressing the binding of V5-LEDGF-p52 in TSS, exon 2 and exon 3 of U2AF2, in shControl, shIWS1, shIWS1/WT-R and shIWS1/MT-R NCI-H1299 cells, confirmed the binding of this LEDGF isoform in both, exon 2 and exon 3. This is a repeat of the experiment in figure 4D, which was done in NCI-H522 cells. The results confirmed that the p52 isoform of wild type LEDGF controls the mRNA splicing of *U2AF2,* by binding the U2AF2 exons 2 and 3, in multiple cell lines.

**G.** *V5-LEDGF-p52 does not rescue the shIWS1 and shIWS1/MT-R U2AF2 alternative splicing phenotype.* (Upper panel-First two lines) shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 (left panel) and NCI-H1299 (right panel) cells were transduced with a lentiviral construct of V5-p52/LEDGF. The expression of V5-p52-LEDGF was confirmed by probing lysates of these cells with an anti-V5 or an anti-α-tubulin (loading control) antibody. (Upper panel-Third line) RT-PCR, using RNA isolated from these cells and *U2AF2* exons 1 and 3 oligonucleotide primers shows that V5-p52-LEDGF does not rescue the shIWS1 and shIWS1/MT-R *U2AF2* alternative splicing phenotype. (Middle panel) Bars show the mean *E2/E3 U2AF2* RNA ratio in NCI-H522 and NCI-H1299 shIWS1, shIWS1/WT rescue and shIWS1/MT rescue cells, in the experiment in the top panel, normalized to the shControl cells ± SD (one-side unpaired t-test) (Lower panel) The RT-PCR results were confirmed by quantitative RT-PCR. All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 (one-side unpaired t-test).

**Figure S6 (relative to Figure 4). The regulation of the alternative splicing of the *U2AF2* exon 2 by *IWS1* phosphorylation, depends on the p52 isoform of the H3K36me3 reader *LEDGF* and its splicing partner *SRSF1*.**

**A.** *The alternative RNA splicing of the U2AF2 exon 2 depends on SRSF1.* (Upper panel) The experiment in figure 4E, which was carried out in NCI-H522 cells, was repeated in NCI-H1299 cells, with identical results. (Lower panel) As a confirmation of the independence of *FGFR2* RNA splicing to SRSF1, quantitative RT-PCR from NCI-H522 cells in Figure 4E and NCI-H1299 cells above, was used to determine the *FGFR2* IIIb/IIIc ratio. Again, bars show the *FGFR*2 IIIb/IIIc ratio in shSRSF1 relative to the shControl. Error bars indicate SD.

**B.** *SRSF1 does not rescue the shIWS1 and shIWS1/MT-R U2AF2 alternative splicing phenotype.* (Upper panel-First two lines) shControl, shIWS1 shIWS1/WT-R and shIWS1/MT-R NCI-H522 (left panel) and NCI-H1299 (right panel) cells were transduced with a lentiviral construct of V5-SRSF1. The expression of V5-SRSF1 was confirmed by probing lysates of these cells with an anti-V5 or an anti-β-actin (loading control) antibody. (Upper panel-Third line) RT-PCR, using RNA isolated from these cells and *U2AF2* exons 1 and 3 oligonucleotide primers shows that V5-SRSF1 does not rescue the shIWS1 and shIWS1/MT-R *U2AF2* alternative splicing phenotype. (Middle panel) Bars show the mean *E2/E3 U2AF2* RNA ratio in NCI-H522 and NCI-H1299 shIWS1, shIWS1/WT-R and shIWS1/MT rescue cells, in the experiment in the top panel, normalized to the shControl cells ± SD (one-side unpaired t-test) (Lower panel) The RT-PCR results were confirmed by quantitative RT-PCR.

**C.** The experiments in NCI-H522 cells, in figure 4F, were repeated in NCI-H1299 cells with identical results.

**D.** The experiments in NCI-H522 cells, in figure 4G, were repeated in NCI-H1299 cells with identical results. All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test).

**E.** (Upper panel) The 7-nt SRSF1 binding motif, involved in the regulation of alternative RNA splicing by SRSF1 (Anczuków et al., 2015[52]) (Middle and Lower panels) Boxes show the location of the SRSF1 binding motifs in *U2AF2* exons 2 and 3 and their nucleotide sequence. The height of the boxes is proportional to the -log(*p-value*) and it is presented relative to the height of the box mapping between nucleotides 184 and 192. The binding scores (z-scores) and p values are calculated using the RBPmap pipeline (http://rbpmap.technion.ac.il/index.html).

**F., G., H., and I.** The validated splicing targets of IWS1 phosphorylation in Fig. S1 (*SLC12A2, IFT88, STXBP1 and C1qTNF6*) were subjected to SRSF1 binding sites analysis as above.

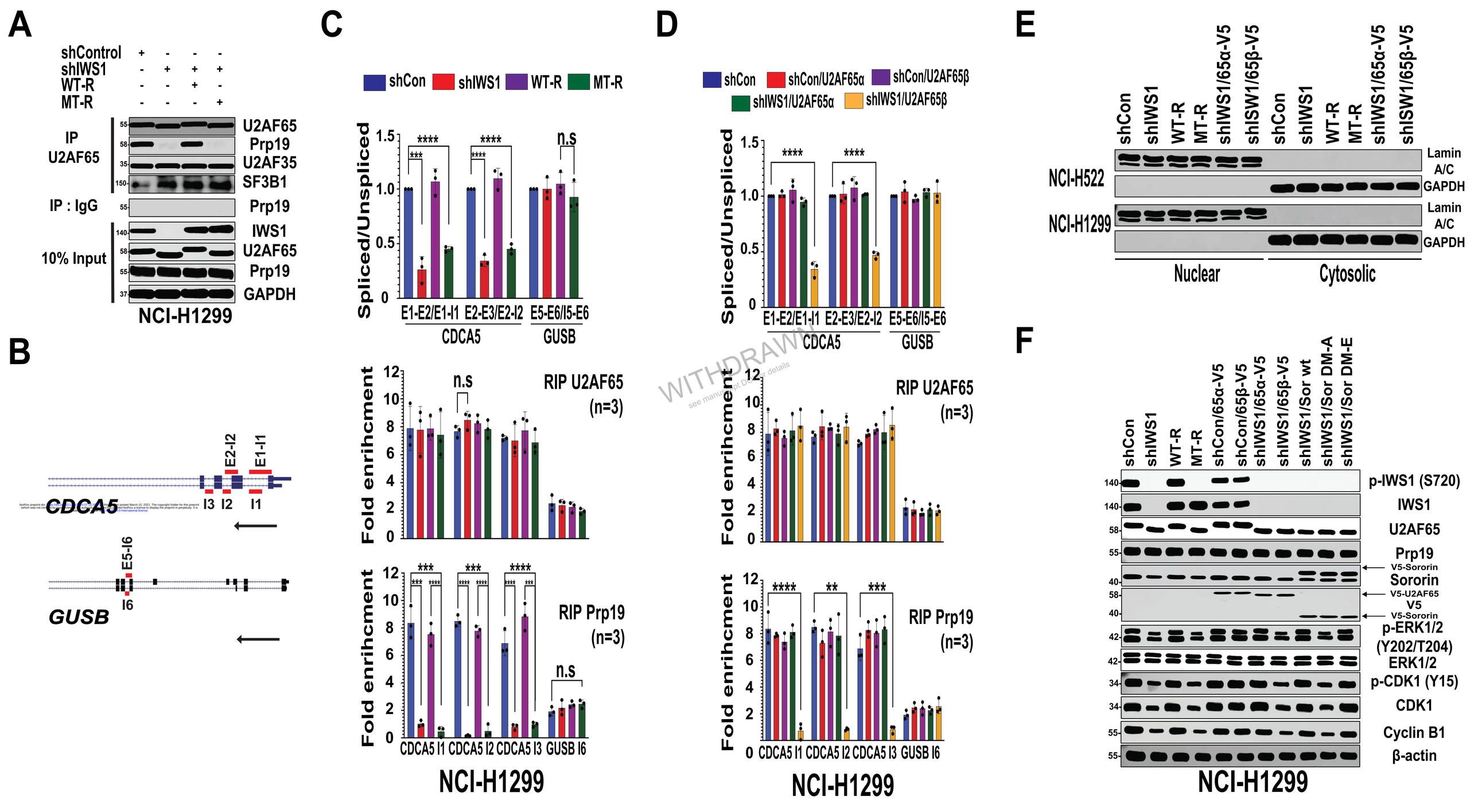

A
shControl + - - -
shIWS1 - + + +
WT-R - - + -
MT-R - - - +
IP U2AF65
U2AF65
Prp19
U2AF35
SF3B1
IP : IgG
Prp19
10% Input
IWS1
U2AF65
Prp19
GAPDH
NCI-H1299
B
E2-I2
E1-I1
CDCA5
I3 I2 I1
E5-I6
I6
GUSB
C
shCon shIWS1 WT-R MT-R
Spliced/Unspliced
E1-E2/E1-I1 E2-E3/E2-I2 E5-E6/I5-E6
CDCA5
GUSB
n.s
RIP U2AF65 (n=3)
Fold enrihcment
RIP Prp19 (n=3)
CDCA5 I1 CDCA5 I2 CDCA5 I3 GUSB I6
NCI-H1299
D
shCon shCon/U2AF65α shCon/U2AF65β
shIWS1/U2AF65α shIWS1/U2AF65β
WITHDRAWN
see manuscript DOI for details
E
shCon shIWS1 WT-R MT-R shIWS1/65α-V5 shISW1/65β-V5
NCI-H522
NCI-H1299
Lamin A/C
GAPDH
Nuclear
Cytosolic
F
shCon shIWS1 WT-R MT-R shCon/65α-V5 shCon/65β-V5 shIWS1/65α-V5 shIWS1/65β-V5 shIWS1/Sor wt shIWS1/Sor DM-A shIWS1/Sor DM-E
p-IWS1 (S720)
IWS1
U2AF65
Prp19
V5-Sororin
Sororin
V5-U2AF65
V5
V5-Sororin
p-ERK1/2 (Y202/T204)
ERK1/2
p-CDK1 (Y15)
CDK1
Cyclin B1
β-actin
NCI-H1299

**Figure S7 (relative to Figure 5). IWS1 phosphorylation controls the CDCA5/ERK phosphorylation feedback loop, through U2AF2 alternative RNA splicing.**

**A.** The experiment in NCI-H522 cells, in figure 5C, was repeated in NCI-H1299 cells with identical results.

**B.** UCSC browser snapshot showing the exon position of the relevant portions of the human *CDCA5* and *GUSB* genes. The map position of the sequences amplified by PCR is indicated with red lines. The arrows indicate the direction of transcription.

**C.** The quantitative RT-PCR and RIP experiments in NCI-H522 cells, in figure 5D, were repeated in NCI-H1299 cells with identical results.

**D.** The quantitative RT-PCR and RIP experiments in NCI-H522 cells, in figure 5E, were repeated in NCI-H1299 cells with identical results.

**E.** The cytosolic and nuclear compartment of the indicated NCI-H522 (upper two lines) and NCI-H1299 cells (lower two lines) were probed with the indicated antibodies, to validate the fractionation.

**F.** The experiment in NCI-H522 cells, in figure 5G, was repeated in NCI-H1299 cells with identical results. All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test).

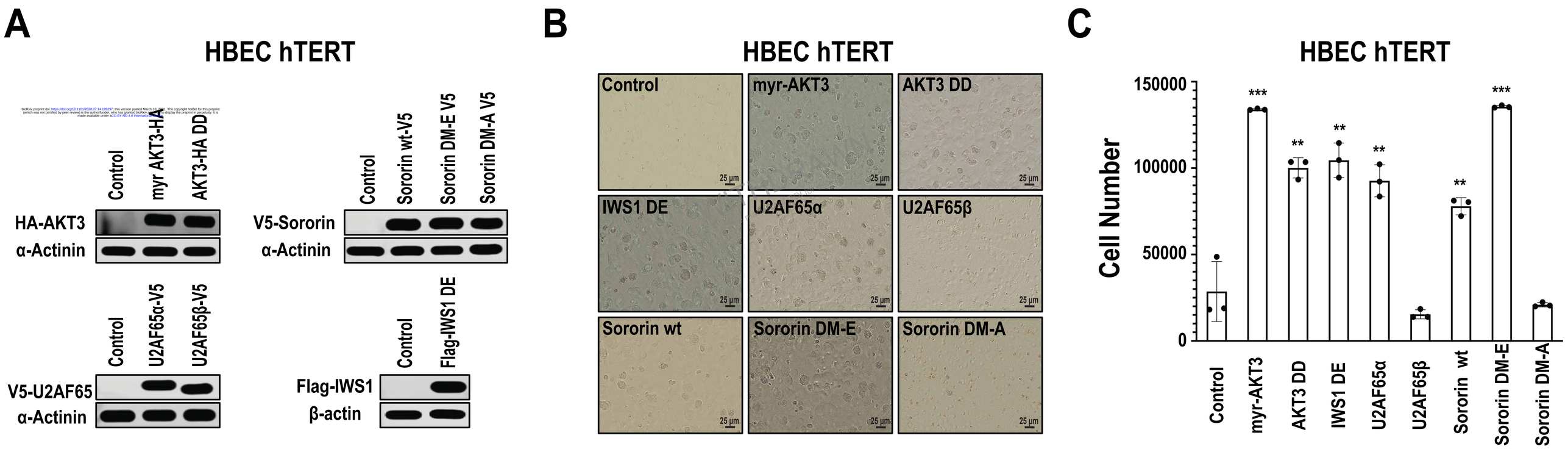


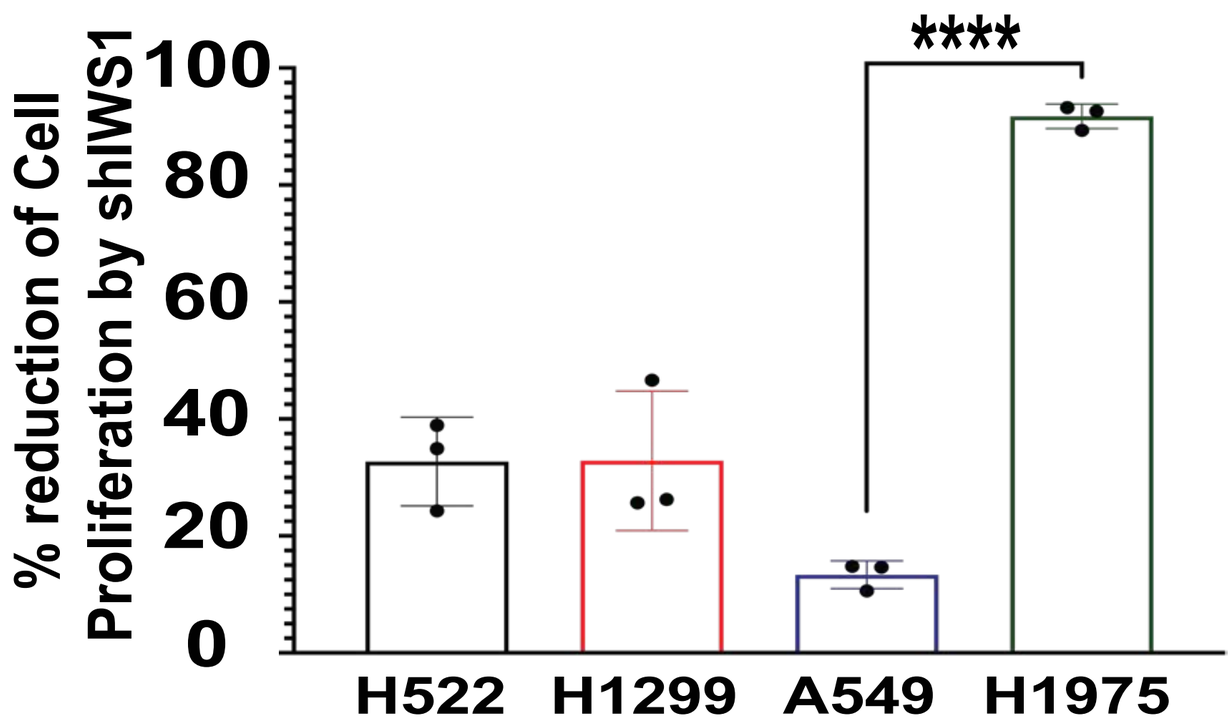
**Figure S8 (relative to Figure 5). IWS1 phosphorylation by AKT3 promotes cellular transformation through *U2AF2* alternative RNA splicing and Sororin expression.**

**A.** Western blots of lysates of HBEC hTERT cells, transduced with the indicated constructs, were probed with the indicated antibodies.

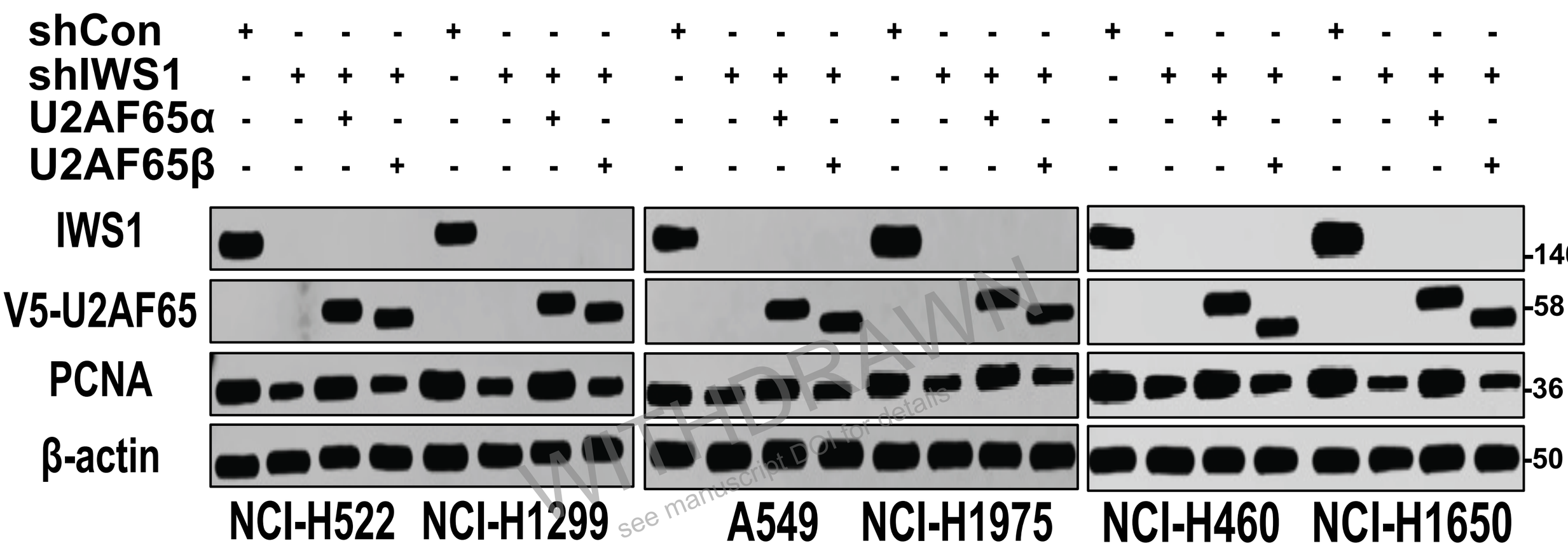
***B.*** *The AKT/IWS1 axis promotes cellular transformation*. HBEC hTERT cells in Figure S8A were subjected to cell transformation assay for 7 days. At that time point the cells were imaged at Incucyte live cell imager. Scale bar in the right corner of each image.

**C.** After imaging, the cells were incubated for 4 hours in WST solution as outlined in Cell Transformation Assay Kit (Abcam) and the percentage of transformed cells was measured as the absorbance in 450nm. All assays in this figure were done in triplicate, on three biological replicates. n.s : non-significant *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (one-side unpaired t-test).

A




B






C



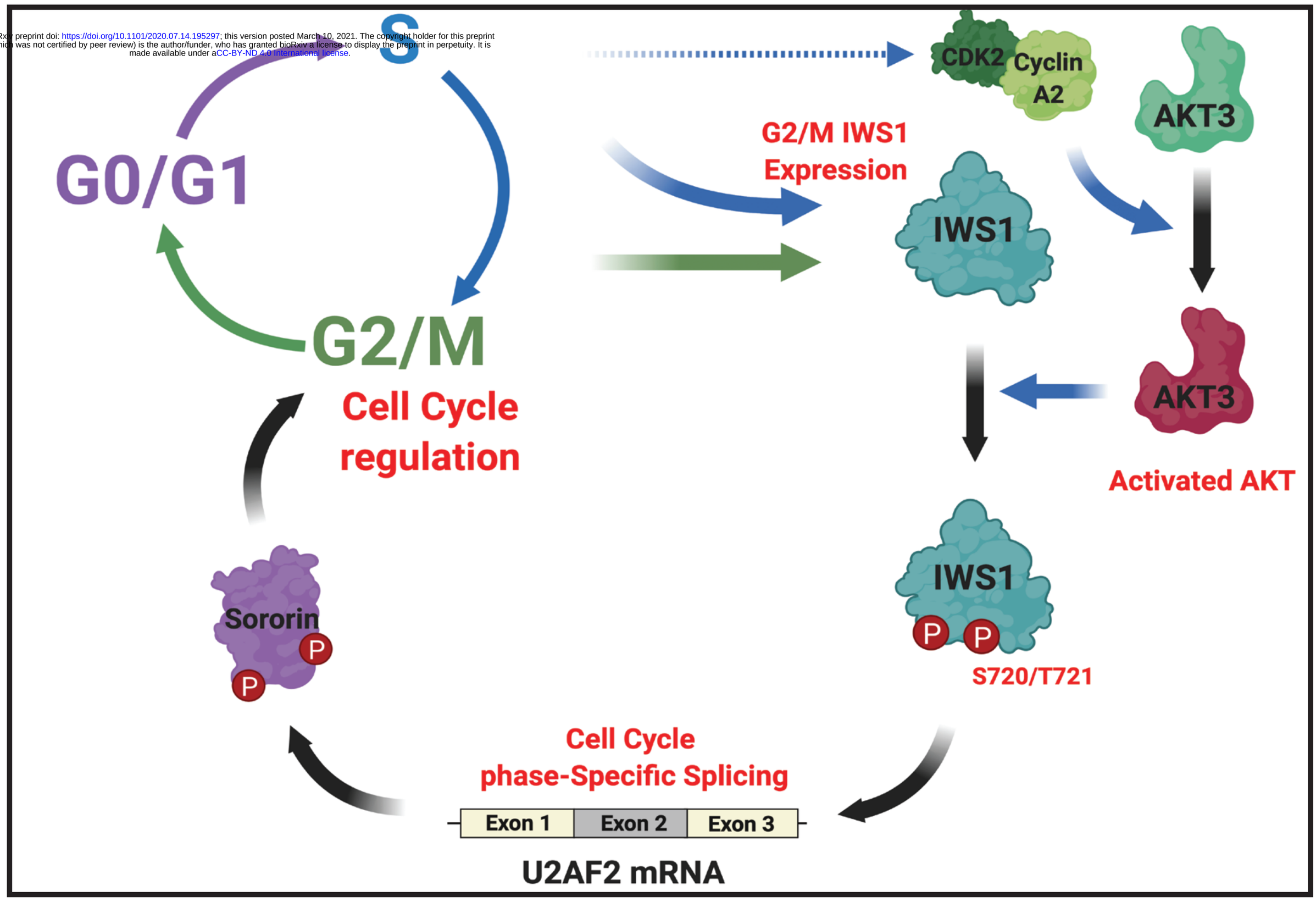


**Figure S9 (relative to Figure 6). IWS1 phosphorylation promotes cell proliferation by controlling the Sororin/ERK phosphorylation feedback loop, through *U2AF2* RNA splicing.**

**A.** Percent reduction of cell proliferation, induced by shIWS1 in the indicated cell lines. Bars show the mean reduction of cell proliferation, derived from the end-point data in Figure 6A and 6C. The error bars show the SD of at least 3 biological replicates. The statistics were performed using one-sided unpaired t-test (NCI-H522 vs A549 : p = 0.0225, NCI-H1299 vs A549 : 0.0493, A549 vs NCI-H1975 : p = 0.0001, NCI-H522 vs NCI-H1975 : p = 0.0174, NCI-H1299 vs NCI-H1975 : p = 0.0011).

**B.** Lysates derived from the indicated cells were probed with the indicated antibodies. In agreement with the data in Fig. 6A, 6C PCNA was downregulated by shIWS1 in both cell lines, and its downregulation was rescued by U2AF65α, but not U2AF65β.

**C.** *Model of the regulation of U2AF2 alternative RNA splicing, downstream of IWS1 phosphorylation, through the cell cycle.* IWS1 is expressed during S and G2/M (this report) and published data indicate that AKT is also activated as the cells enter S phase, via phosphorylation by CDK2/Cyclin A2 (Liu et al, 2014[74]). This induces a cell cycle specific alternative RNA splicing, resulting in inclusion of exon 2 in *U2AF2* mRNA. Subsequently, this isoform regulates the splicing and expression of Sororin, which in turn regulates the progression through the G2/M phase of the cell cycle.

A

NCI-H1299

N1 N2 N3 N4 N5

shCon + - + - + - + - + - + - + - + - + - + - + -
shIWS1 - + + + - + + + - + + + - + + + - + + +
U2AF65α-R - - + - - - + - - - + - - - + - - - + -
U2AF65β-R - - - + - - - + - - - + - - - + - - - +

IWS1
V5-U2AF65
GAPDH

140
58
37

B

shCon shIWS1

qPCR U2AF2 Splicing

E2/E3 Relative to control

qPCR U2AF2 Total

mRNA expression relative to GAPDH

n.s

N1 N2 N3 N4 N5

NCI-H1299 A549 NCI-H1975

C

NCI-H1299

shControl

shIWS1

100 µm

25 µm

p-IWS1 (S720)

Sororin

p-ERK 1/2 (Y202/T204)

p-CDK1 (Y15)

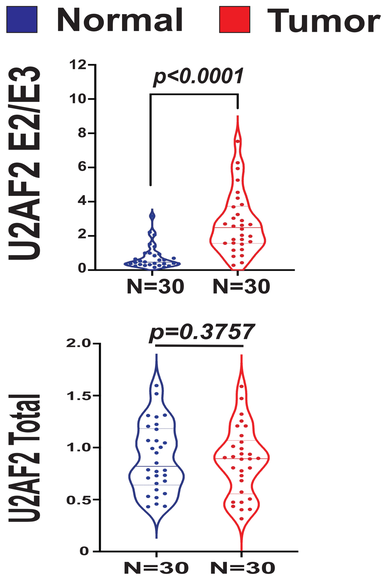
**Figure S10 (Relative to Figure 7). IWS1 phosphorylation controls tumor growth *in vivo*, by regulating the U2AF2/Sororin/ERK axis.**

**A.** Cell lysates derived from NCI-H1299 shControl, shIWS1, shIWS1/U2AF65α and shIWS1/U2AF65β mouse xenografts, were probed with the indicated antibodies.

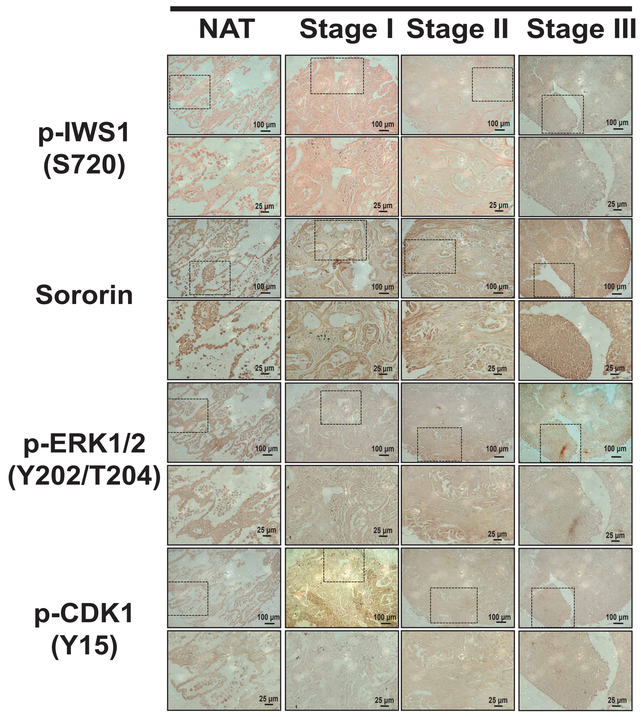
**B.** *IWS1 phosphorylation controls U2AF2 alternative RNA splicing in vivo.* Quantitative RT-PCR showing the *E2/E3 U2AF2* ratio (upper panel) or total *U2AF2* in lysates derived from tumors in Fig 7C. Bars show *U2AF2* E2/E3 ratio or the mean *U2AF2* levels, relative to GAPDH, in tumors derived from NCI-H1299 (left), A549 (middle) and NCI-H1975 (right) shControl and shIWS1 cells. The error bars are SD of at least 3 technical replicates. **p<0.001, ***p<0.001. (paired t-test), n.s : non-significant (paired t-test)

**C.** *IWS1 controls the Sororin/ERK phosphorylation axis in vivo.* (Upper panels) Phospho-IWS1 (Ser720), Sororin phosphor-ERK (Y202/T204) and phospho-CDK1 (Y15) staining, of NCI-H1299 shControl (upper) and shIWS1 (lower) tumor xenografts. Formalin-fixed, paraffin-embedded tumor samples from the experiment in Figure 7 were stained with the indicated antibodies. Secondary antibody was HRP-labelled. The staining of individual tumors with different antibodies was done, using sequential tumor sections, so that we could determine whether the expression and/or phosphorylation of proteins of interest spatially overlapped. Boxes delineate the area of higher magnification, shown in the image below. Scale in the right lower corner of each image.

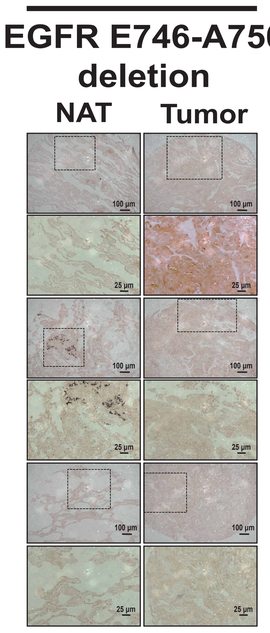
A



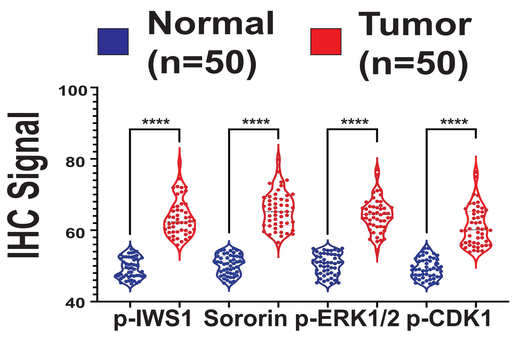

B



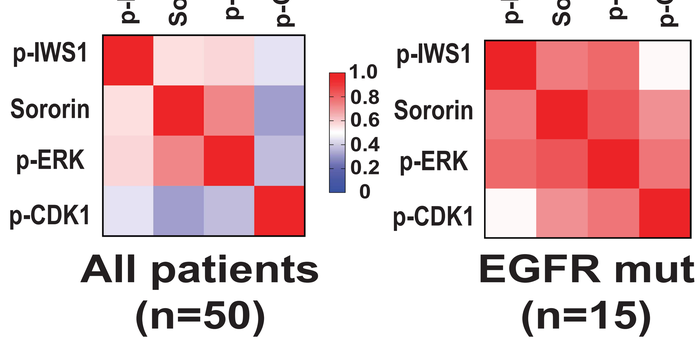

C



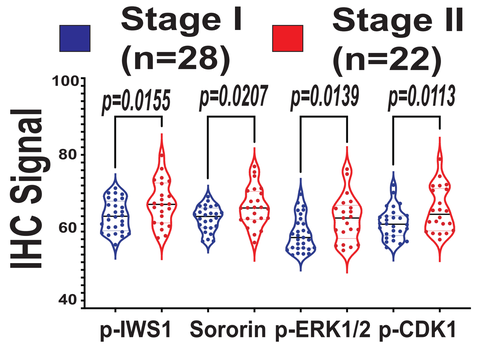

D



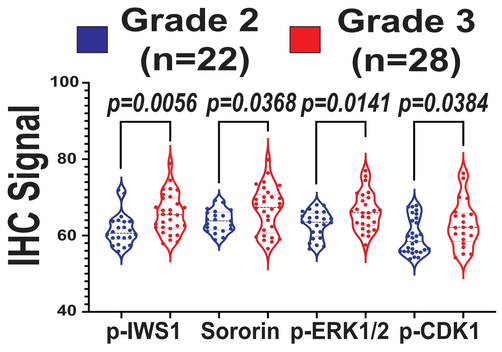

E



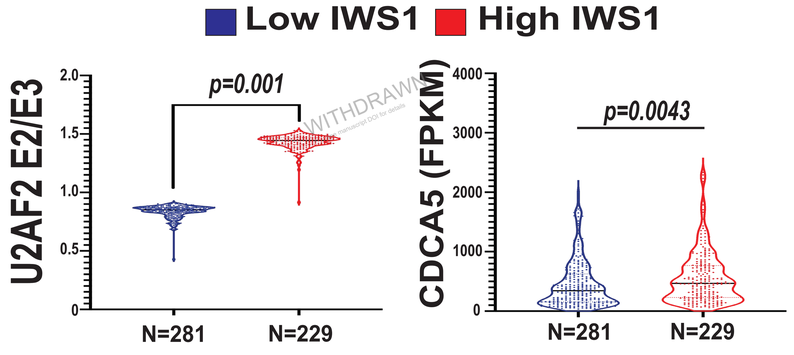

F



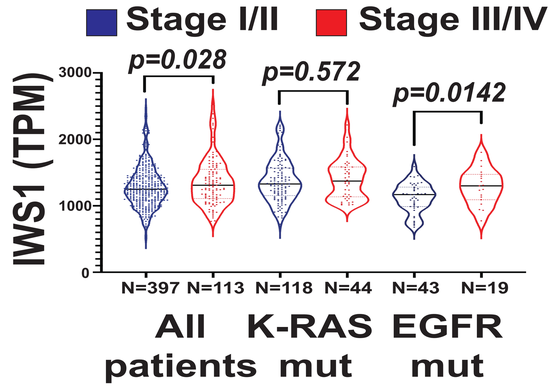

G



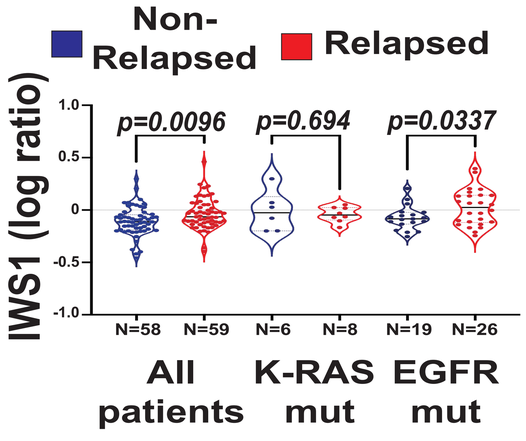

H



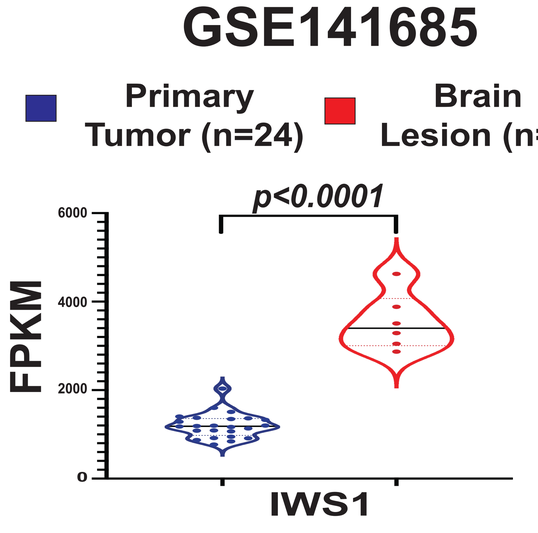

I



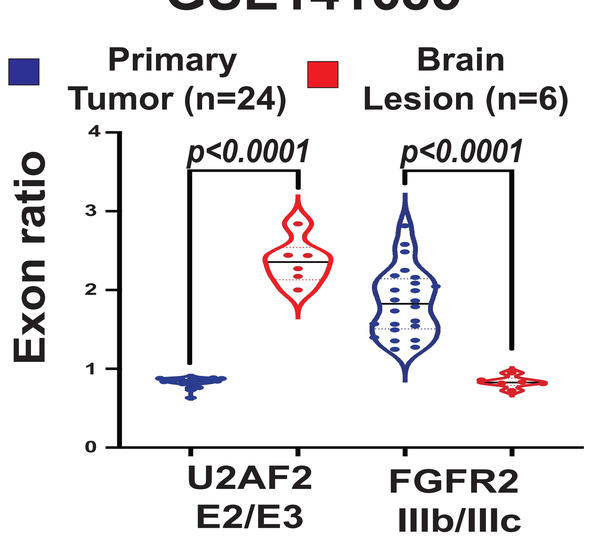

J



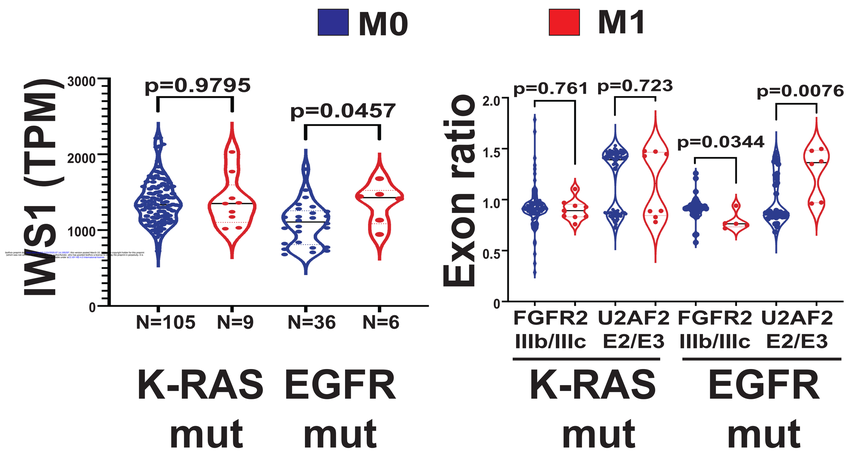

K



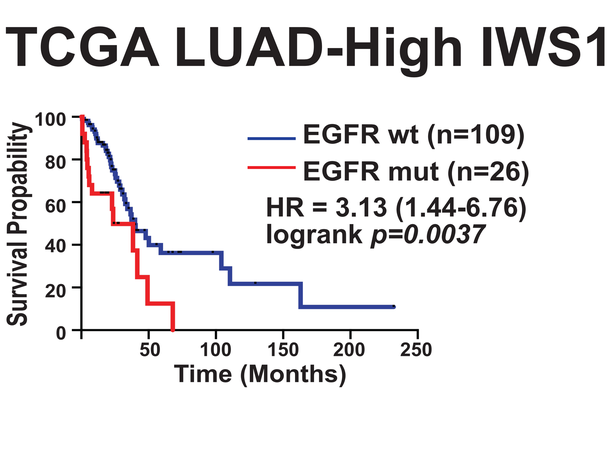

L




M




N




O

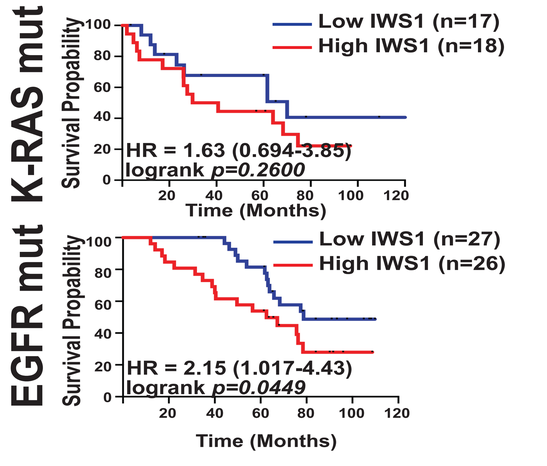


**Figure S11 (relative to Figure 8). The effect of the Akt3/p-IWS1/U2AF2 axis in lung adenocarcinoma**

**A.** *The overall abundance of the U2AF2 mRNA in lung adenocarcinomas and the adjacent normal tissue is similar, but the E2/E3 ratio is higher in the tumors.* Quantitative RT-PCR showing the *U2AF2 E2/E3* ratio (top) and the total *U2AF2* mRNA (bottom) in lysates derived from the 30 LUAD samples and the paired normal adjacent tissue in Fig 8A. The black lines in the violin plots show the mean *U2AF2 E2/E3* ratio relative to the matching normal samples, or *U2AF2* levels, relative to *GAPDH*. Statistical analyses were done using the paired t-test.

**B.** *IWS1 phosphorylation correlates with U2AF2 alternative RNA splicing and the Sororin/ERK axis in lung adenocarcinoma patients***.** Sequential sections of a commercially available tissue microarray (TMA) of 50 LUAD with paired NAT, were probed with the indicated antibodies. Figure shows the staining of a representative NAT sample and representative samples of Stage I, II and III lung adenocarcinomas (one of each). Boxes delineate the area of higher magnification, shown in the image below. Scale in the right lower corner of each image.

**C.** The TMA sections used before were stained with the specific monoclonal antibody against EGFR E746-A750 deletion. Figure shows three representative lung adenocarcinomas with EGFR deletion (E746-A750 deletion) and the corresponding NATs. Boxes delineate the area of higher magnification, shown in the image below. Scale in the right lower corner of each image.

**D.** *The overall activity of the p-IWS1/CDCA5 is higher in human lung adenocarcinomas than in the adjacent normal tissue.* Violin plots showing the abundance of IWS1 phosphorylation, along with the abundance of Sororin, phospho-ERK and phospho-CDK1, in the LUAD and NAT samples in the tissue microarray in B and C. The black line indicates

the mean p-IWS1, Sororin, p-ERK and p-CDK1 levels. Statistical analyses were done using the paired t-test.

**E.** *IWS1 phosphorylation exhibits a more robust correlation with Sororin, phospho-ERK and phospho-CDK1 in lung adenocarcinomas with EGFR mutations.* Heatmaps showing the correlation coefficient between the indicated components of the IWS1 phosphorylation pathway in the entire TMA cohort (left) and in the *EGFR* mutant cohort (right). The correlations were calculated using simple linear regression. The statistics and the p values for all the comparisons, can be found in Supplementary Table S5.

**F.** *The activity of the p-IWS1/CDCA5 pathway correlates with tumor stage in human lung adenocarcinomas.* Violin plots comparing the phosphorylation of IWS1 and the abundance of Sororin, phospho-ERK and phospho-CDK1 in stage I and Stage II/III lung adenocarcinomas in the group of 50 tumors sampled in the commercial TMA described above. Statistical analyses were performed using the one-sided unpaired t-test.

**G.** *The activity of the p-IWS1/CDCA5 pathway correlates with tumor grade in human lung adenocarcinomas.* Violin plots comparing the phosphorylation of IWS1 and the abundance of Sororin, phospho-ERK and phospho-CDK1 in histological grade 2 and grade 3 lung adenocarcinomas in the group of 50 tumors in the commercial TMA described above. Statistical analyses were performed using the one-sided unpaired t-test.

**H.** The expression of IWS1 correlates with Exon 2 inclusion in the *U2AF2* mRNA transcripts (upper panel) and with the expression of *CDCA5* (lower panel) in the TCGA LUAD dataset. The black line indicates the mean expression of *U2AF2* E2/E3 ratio and *CDCA5*. Statistical analyses were performed using the one-sided unpaired t-test.

**I.** Violin plots showing the abundance of IWS1 in stage I/II and in Stage III/IV tumors of patients in the TCGA LUAD database. Data shown for all tumors, and selectively for *EGFR* or *KRAS* mutant tumors. The horizontal lines indicate mean values for IWS1 levels. Statistical analyses were performed using the one-sided unpaired t-test.

**J.** Violin plots showing the abundance of IWS1 in relapsed and non-relapsed tumors of patients in the GSE13213 dataset. Data shown for all tumors, and selectively for *EGFR* or *KRAS* mutant tumors. The horizontal lines indicate mean values for IWS1. Statistical analyses were performed using the one-sided unpaired t-test.

**K.** (Left panel) Violin plots showing the expression of IWS1 in transcripts per million (TPM) in *KRAS* and *EGFR* mutant, primary and metastatic lung adenocarcinomas in the TCGA LUAD database. IWS1 expression is higher in tumors from patients with metastatic disease (M1) only if the tumors harbour *EGFR* mutations. (Right panel) Exon E2/E3 ratio in the *U2AF2* mRNA and exon IIIb/IIIc ratio in the *FGFR2* mRNA in *KRAS* and *EGFR* mutant lung adenocarcinomas. E2/E3 ratio in the *U2AF2* mRNA is higher and the E8/E9 ratio in the *FGFR2* mRNA is lower in tumors from patients with metastatic disease (M1), only if the tumors harbour *EGFR* mutations. The black lines indicate the mean IWS1 expression, and the mean *U2FA2* E2/E3 and *FGFR2* IIIb/IIIc ratios. Statistical analyses were performed using the one-sided unpaired t-test.

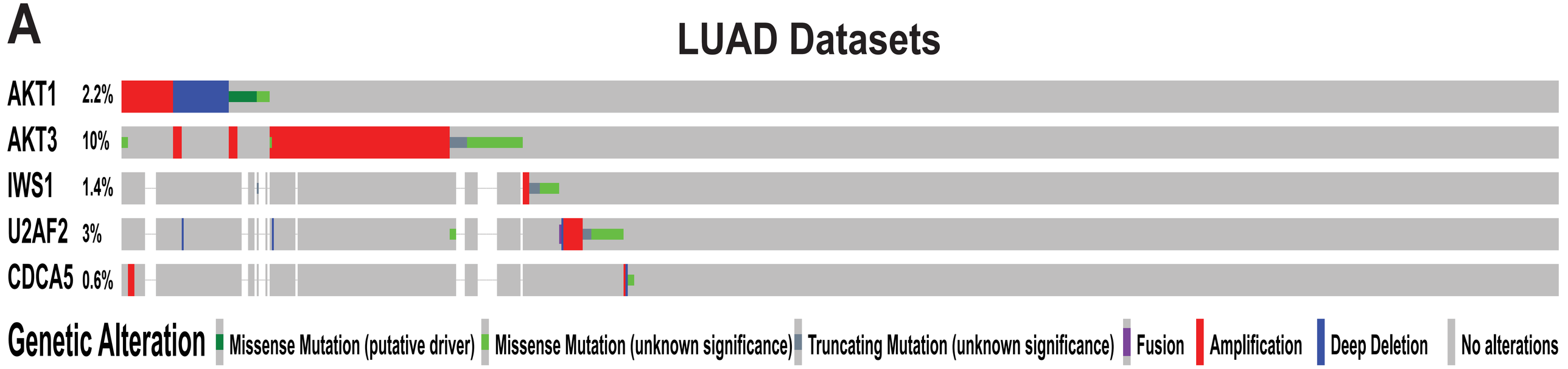
**L.** The expression of *IWS1* (upper panel) and the *E2/E3* ratio in mature *U2AF2* mRNA transcripts (lower panel-left) are higher in brain metastases of lung adenocarcinomas than in primary tumors, while the *IIIb/IIIc* ratio in mature *FGFR2* transcripts (lower panel-right) is lower in brain metastases relative to primary tumors. The black lines indicate the mean expression of IWS1 and the mean *U2AF2 E2/E3* and *FGFR2 IIIb/IIIc* ratio. Violin plots were based on the same data shown in Figure 8G. Statistical analyses were performed using the one-sided unpaired t-test.

**M.** The *E2/E3* ratio in mature *U2AF2* mRNA transcripts are higher in the same patient group as Fig S11L. The black lines indicate the mean *U2AF2 E2/E3* and *FGFR2 IIIb/IIIc* ratio. Violin plots were based on the same data shown in Figure 8G. Statistical analyses were performed using the one-sided unpaired t-test.

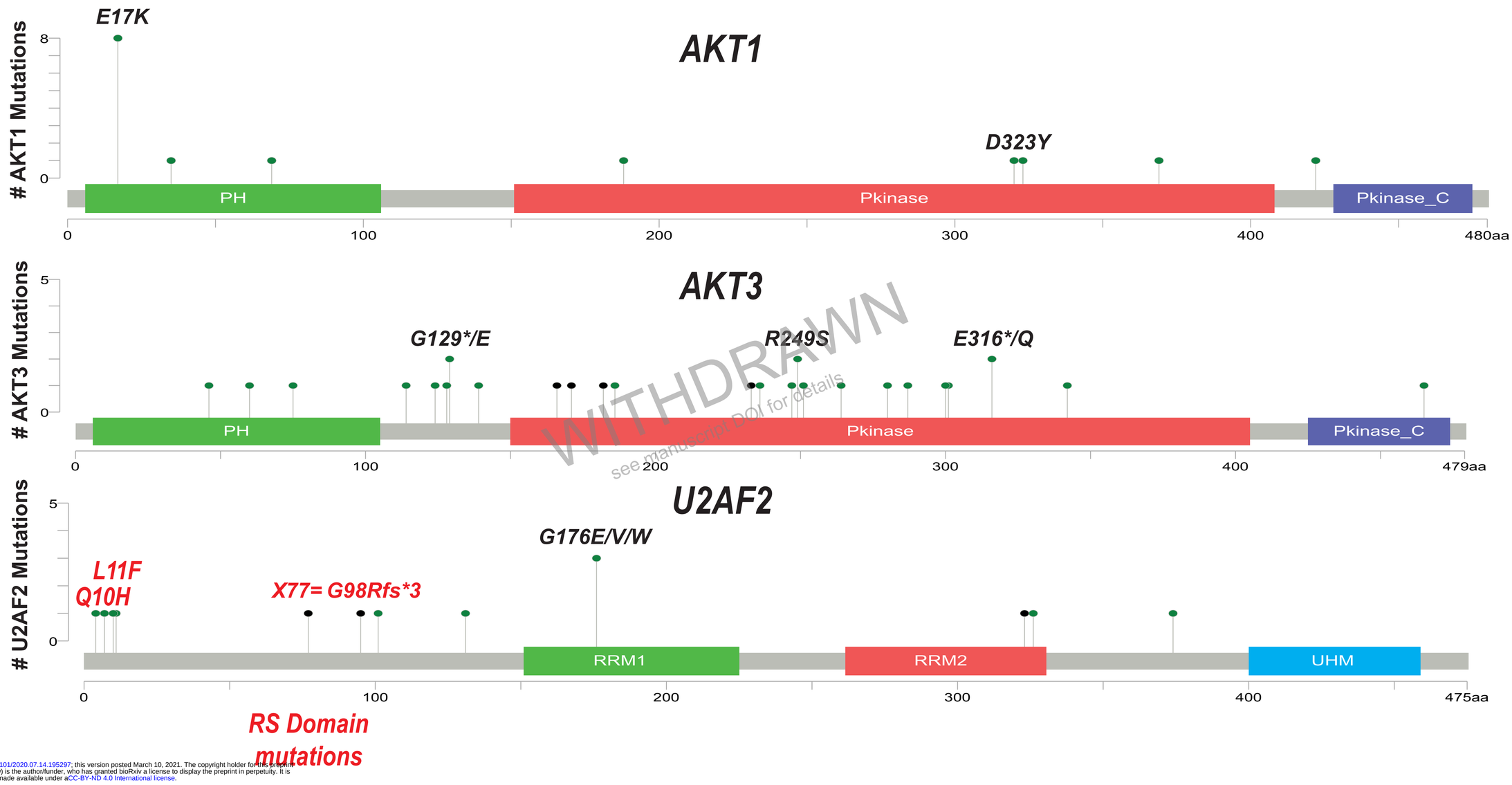
**N.** *EGFR mutations predict poor survival in patients expressing high levels of IWS1.* Kaplan-Meier curves showing the impact of *EGFR* mutations on survival, in patients with lung adenocarcinomas expressing high levels of IWS1. Statistical analyses were performed using the log rank test and Cox's proportional hazards model.

**O.** *The abundance of IWS1 has a stronger negative impact on the survival of patients with lung adenocarcinomas harbouring EGFR, than those with lung adenocarcinomas harbouring KRAS mutations.* Kaplan-Meier curves showing the impact of the abundance of IWS1 on the survival of patients with lung adenocarcinomas harbouring *KRAS* mutations (upper panel) or *EGFR* mutations (lower panel) in the GSE13213/GSE26969 dataset. Statistical analyses were performed using the log rank test and Cox's proportional hazards model.

A




B






C

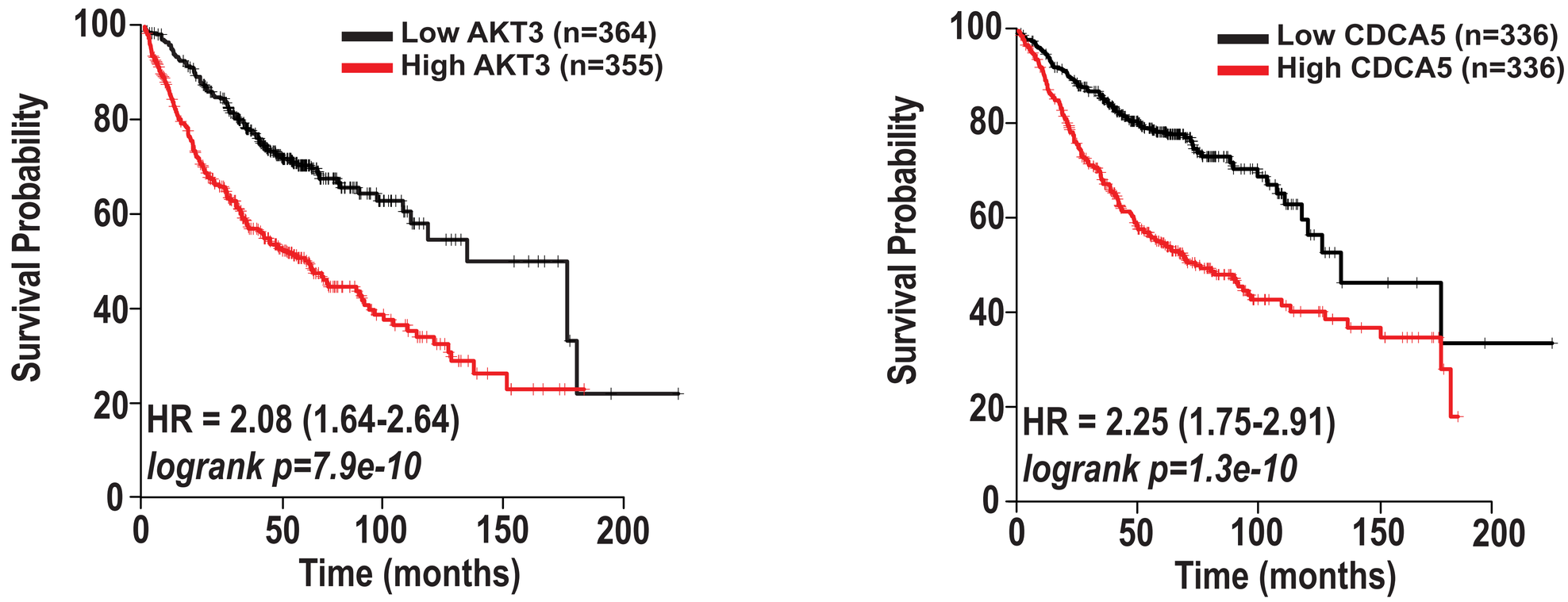


**Figure S12 (relative to Figure 8). The effect of the Akt3/p-IWS1/U2AF2 axis in lung adenocarcinoma**

***A.*** *The AKT/IWS1/U2AF2 pathway is genetically altered in patients with Lung Adenocarcinoma.* Alterations of the AKT/IWS1 pathway were visualized by cBioPortal for Cancer Genomics, using all the available LUAD datasets (TSP- Ding et al., 2008[81], Broad-Imielinksi et al., 2012[82], MSKCC-Rivzi et al., 2015[83], Jordan et al., 2017[84], OncoSG-Chen et al., 2020[85], TCGA Firehose Legacy). Mutation, deletion, amplification, fusion and other alterations are shown in different colors.

***B.*** The distribution of the mutations of the above patients in *AKT1*, *AKT3* and *U2AF2* genes. The most common mutations in these patients are shown in black. Regarding *U2AF2,* the missense and splice mutations of the RS domain can be found in red. The complete list of the mutations in this pathway can be found in Supplementary Table S7.

**C.** Kaplan-Meier Curves showing the impact of *AKT3* expression (left) and *CDCA5* (right) on patient survival in patients in the TCGA LUAD database. Statistical analyses were performed using the log rank test and Cox's proportional hazards model.