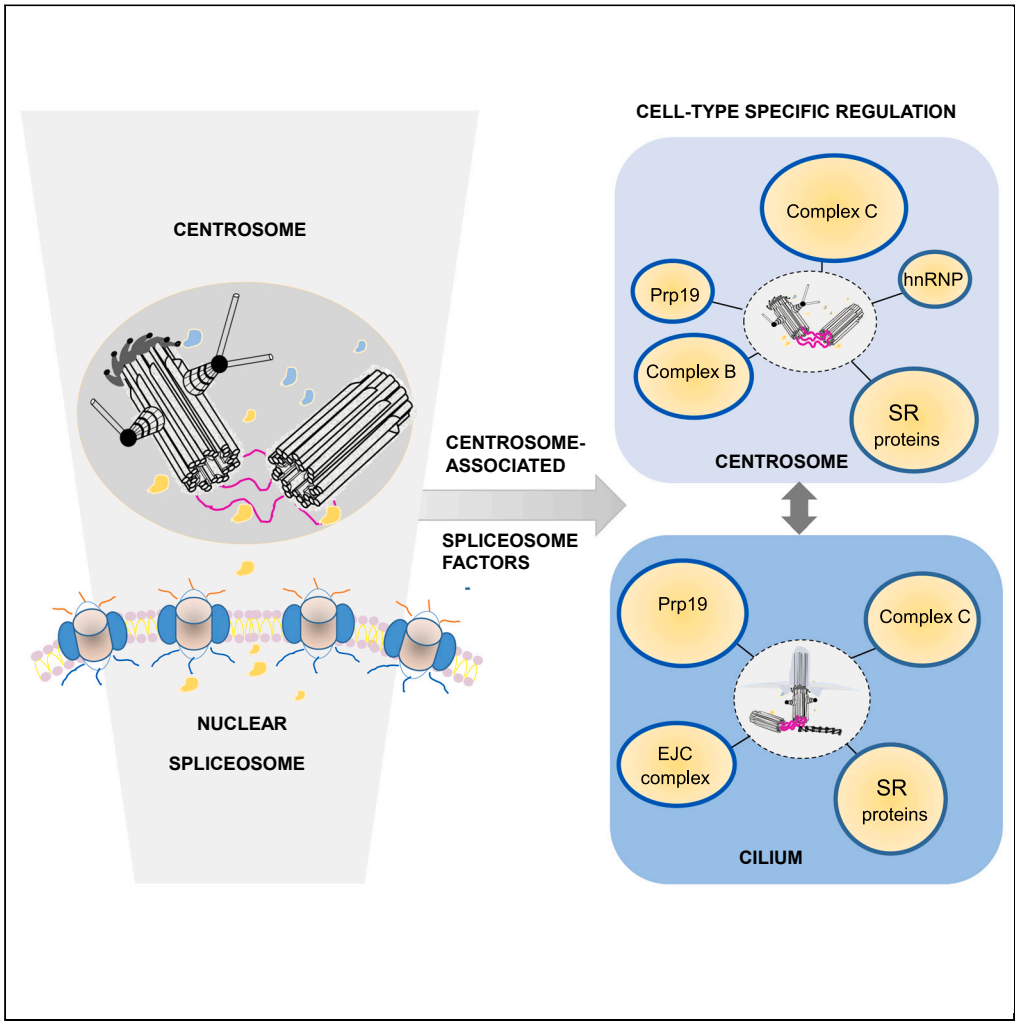
iScience

CellPress
OPEN ACCESS

Article

Single-cell proteo-genomic reveals a comprehensive map of centrosome-associated spliceosome components



Luigi Cerulo, Nunziana Pezzella, Francesca Pia Caruso, ..., Mirco Galiè, Brunella Franco, Massimo Pancione

franco@tigem.it (B.F.)
massimo.pancione@unisannio.it (M.P.)

Highlights

Spliceosome components interact with centrosome substructures

Centrosome-associated spliceosome components target OFD1 and CEP250

Centrosome-associated spliceosome components are cell-type specific

Cholangiocarcinoma is a target of centrosome-associated splicing alterations

Cerulo et al., iScience 26, 106602
May 19, 2023
https://doi.org/10.1016/j.isci.2023.106602

iScience

CellPress
OPEN ACCESS

Article

Single-cell proteo-genomic reveals a comprehensive map of centrosome-associated spliceosome components

Luigi Cerulo,[1,2] Nunziana Pezzella,[3,4] Francesca Pia Caruso,[1,2] Paola Parente,[5] Andrea Remo,[6] Guido Giordano,[7] Nicola Forte,[8] Johan Busselez,[9] Federico Boschi,[10] Mirco Galiè,[11] Brunella Franco,[3,4,12,*] and Massimo Pancione[2,13,14,*]

SUMMARY

Ribonucleoprotein (RNP) condensates are crucial for controlling RNA metabolism and splicing events in animal cells. We used spatial proteomics and transcriptomic to elucidate RNP interaction networks at the centrosome, the main microtubule-organizing center in animal cells. We found a number of cell-type specific centrosome-associated spliceosome interactions localized in subcellular structures involved in nuclear division and ciliogenesis. A component of the nuclear spliceosome BUD31 was validated as an interactor of the centriolar satellite protein OFD1. Analysis of normal and disease cohorts identified the cholangiocarcinoma as target of centrosome-associated spliceosome alterations. Multiplexed single-cell fluorescent microscopy for the centriole linker CEP250 and spliceosome components including BCAS2, BUD31, SRSF2 and DHX35 recapitulated bioinformatic predictions on the centrosome-associated spliceosome components tissue-type specific composition. Collectively, centrosomes and cilia act as anchor for cell-type specific spliceosome components, and provide a helpful reference for explore cytoplasmic condensates functions in defining cell identity and in the origin of rare diseases.

INTRODUCTION

Ribonucleoprotein (RNP) particles are membrane-less structures composed of RNA and proteins.[1] Eukaryotic cells contain a variety of RNP structures, such as stress granules and P-bodies, localized in the nucleus, in the cytoplasm, and on membranes. These aggregates are vital in controlling RNA metabolism and splice events, which are crucial for cellular development and differentiation.[2–4] Centrosomes are biomolecular condensates known to be the major microtubule-organizing centers (MTOC) of animal cells. Centrosomes play a key role both in cell division and in the formation of sensory and motile cilia. A pair of barrel-shaped microtubule-based centrioles form the core of each centrosome, which is surrounded by an elegantly organized protein network termed pericentriolar material (PCM) and centriolar satellites.[5,6] The older, more mature centriole, also known as the "mother" centriole, plays a crucial role in ciliogenesis by acting as a nucleating basal body to organize the axonemes of cilia.[5,6] Centrioles are absent in up to half of all known eukaryotic species, and a variety of mechanisms for a centrosomal microtubule nucleation have been described (for example, in plants cells and vertebrate oocytes), raising doubts about the centrosome's role as a microtubule organizing center.[5] It is interesting to note that centrosomal proteins are enriched in disordered and coiled-coil regions compared to non centrosomal proteins from the same organism.[7,8] Through RNA–RNA and RNA–protein interactions, intrinsically disordered proteins are essential for the coordination and control of gene expression. This suggests the intriguing theory that centrosomes and cilia may contain RNA and RNA-processing proteins. In animal cells, a centrosome substructure termed the "centrosome linker" maintains the two centrioles in close proximity.[9] The binding of the centrosomal protein CEP250 (C-NAP1) to CEP135 at the proximal end of the two centrioles is known to be one of the initial steps of centrosome cohesion assembly.[9] Of interest, CEP250 is functionally interconnected with subdistal appendage markers (ODF2, NIN) and Golgi function.[10] These functional partnerships have been proposed to play a role in the spatial control of ciliogenesis.[9,11] However, little, if anything, is known about the functional significance of such subcellular structures.[9–12] Recent research has also revealed a functional

[1]Bioinformatics Laboratory, BIOGEM scrl, Ariano Irpino, Avellino, Italy

[2]Department of Sciences and Technologies, University of Sannio, Benevento, Italy

[3]Telethon Institute of Genetics and Medicine (TIGEM), Via Campi Flegrei, 34, Pozzuoli, 80078 Naples, Italy

[4]School for Advanced Studies, Genomics and Experimental Medicine Program, Naples, Italy

[5]Unit of Pathology, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[6]Pathology Unit, Mater Salutis Hospital AULSS9, "Scaligera", 37122 Verona, Italy

[7]Unit of Medical Oncology and Biomolecular Therapy, Department of Medical and Surgical Sciences, University of Foggia, Policlinico Riuniti, 71122 Foggia, Italy

[8]Department of Clinical Pathology, Fatebenefratelli Hospital, 82100 Benevento, Italy

[9]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

[10]Department of Computer Science, University of Verona, Strada Le Grazie 8, Verona, Italy

[11]Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy

[12]Medical Genetics, Department of Translational Medicine, University of Naples "Federico II", Via Sergio Pansini, 80131 Naples, Italy

Continued

connection between the centrosome, RNP particles, ribosomes, and eukaryotic initiation factors (eIFs) required for protein synthesis.[12–14] For instance, the centrosomal protein OFD1 influences the translation of specific mRNA targets in the kidney by interacting with elements of the preinitiation complex of translation.[13] The large PCM protein pericentrin (PCNT) is also delivered co-translationally to centrosomes during early mitosis.[14] Notably, mutations in the splicing and core ribosomal elements result in genetic disorders such as retinitis pigmentosa and microcephaly, which are frequently associated with abnormalities in the centriole.[15,16] In 1970, the observation that centrioles are self-replicating organelles led to the proposal that centrosomes contain their own complement of nucleic acids.[17] Although it is well-established that DNA is not present in centrosomes, RNA is known to associate with centrosomes in diverse cell types, including early embryos (Drosophila, Xenopus, zebrafish), surf clams, and cultured mammalian cells.[17,18] These findings raise the possibility that RNA localization and other post-transcriptional regulatory mechanisms may contribute to centrosome activity and/or function.[19] Pre-mRNA splicing is catalyzed by the nuclear spliceosome, a huge, ribonucleoprotein complex comprising five small nuclear RNP (snRNPs) and numerous other proteins.[15,20] Many of the metazoan-specific spliceosomal proteins that are absent in yeast may have cytoplasmic roles, but this hypothesis has still to be tested. We show here that spliceosome components constitute a relevant repertoire of centrosome-cilia molecular networks, which is probably required to control phenotypic diversity in cells and tissues.

RESULTS

Ribonucleoproteins and centrosome cohesion components are tightly interconnected

When we looked at the protein-protein interaction (PPi) network of experimentally validated centrosome linker proteins (CLPs), we observed that CEP250 had a significantly higher number of interactors compared to other CLPs (Figures 1A and 1B). Notably, RNA-binding proteins including ribosome components, heterogeneous nuclear ribonucleoproteins (HNRNPs), and RNA helicases comprised the vast majority of CEP250 interactors (Figures 1C and S1A). Cell-cell adherent junctions (11%), structural constituents of ribosome/translational initiation (27%), and mRNA splicing components (9%) were the top enriched terms according to gene ontology (GO) analysis of the CEP250 interactome (Figure 1C and Table S1). It is noteworthy that the pre-mRNA splicing factor MAGOH and the centrosome linker protein Ciliary Rootlet Coiled Coil (CROCC) were part of another protein-protein interaction network. In fact, post-transcriptional regulators, splicing, and mRNA nuclear export molecules were the top ranked interaction partners of CROCC, adjusted p-*value 0.001* (Figure S1B).

We found that the Pseudogene 2 (CROCCP2) and several small nuclear RNAs (e.g., U1) involved in the processing of pre-messenger RNA colocalized with the CROCC gene on 1p36.13 (Figure S1C).

As a result, *CROCC* and *CROCCP2* transcripts exhibited a coordinated expression pattern in a range of samples, indicating the presence of a co-regulated expression network (Figure S1D). We next investigated whether the proportion of predicted disordered regions among CLPs was related to the interactome size. Among CLPs, we discovered that CEP250 had the highest percentage of predicted disordered domains and it was associated with a larger interactome (Figure 1D). By investigating additional centrosomal proteins, including those in the pericentrosomal region, distal appendages, and centrioles, we found evidence that the expansion of the protein interactome mirrored the increase in intrinsically disordered domains in distinct portions of the centrosome (Figure 1E). It is generally known that Archaea and Bacteria have a lower average content of disordered domains compared to Eukaryota.[8] Using PCM markers (pericentrin and γ-tubulin) for comparison, we examined the gain and loss of centrosomal genes in unicellular and multicellular eukaryotes, including animals. In contrast to PCM genes that were highly conserved across species, the gain of linker, centriolar satellites and appendage genes correlated with apparent biological complexity, in keeping with the literature (Figure 1F).[8] As a result, the CLP protein-domains architecture reflects the sizes of interactomes and perhaps the complexity of mammalian cells.

The nuclear mRNA processing machinery interacts with linker and satellites components

We next focused on a relevant subset of nine distinct centrosome-associated components including centrioles, pericentriolar material, spindle and ciliary proteins, to better elucidate RNP interaction networks (Figure 2A and Table S2).[21] We found that centrosome proteins primarily involved in centriole biogenesis and centrosome maturation interacted with the translational machinery, 60S and 40S core ribosome proteins (Figure 2B). Furthermore, the distal appendages (NINL) and ciliogenic factors (OFD1, FBF1) correlated with the translation initiation factors (EIF4ENIF1 and EIF6) and cyclin-dependent kinase 2 (CDK2),

[13]Department of Biochemistry and Molecular Biology, Faculty of Pharmacy, Complutense University Madrid, 28040 Madrid, Spain

[14]Lead contact

*Correspondence: franco@tigem.it (B.F.), massimo.pancione@unisannio.it (M.P.)

https://doi.org/10.1016/j.isci.2023.106602

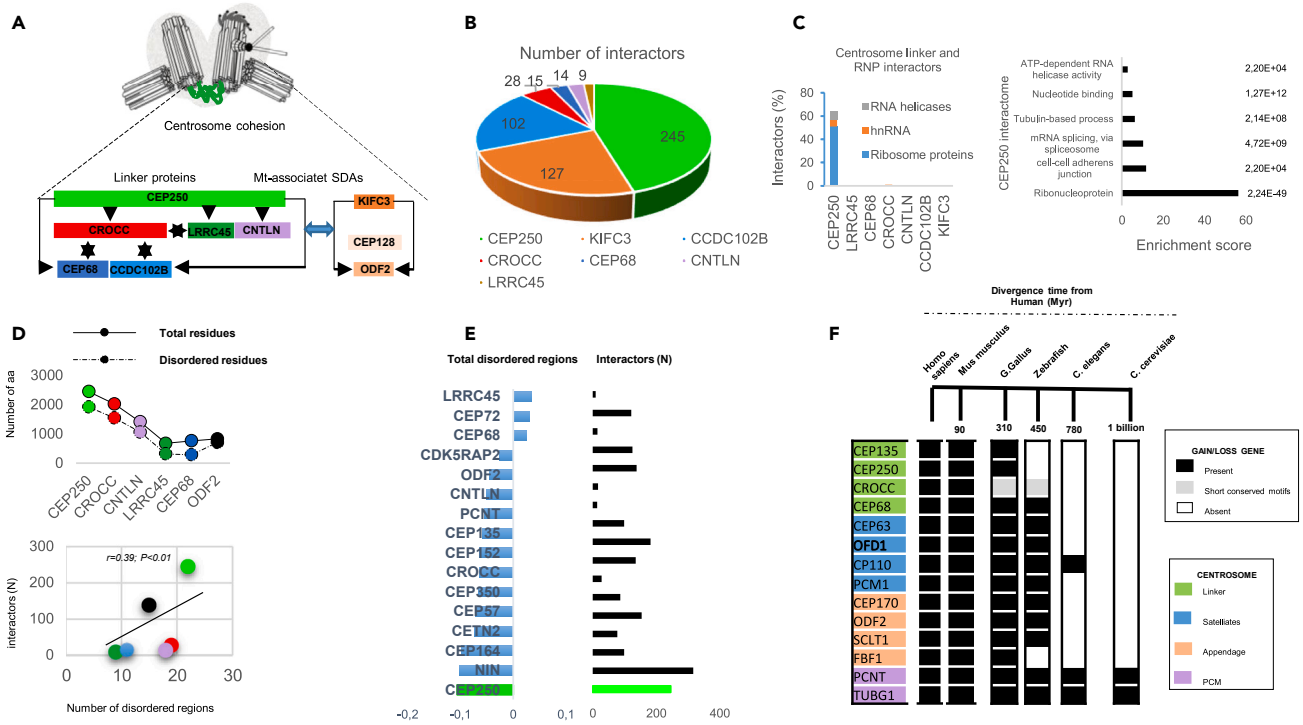


Figure 1. The centrosome linker proteome network reflects the interactome size and complexity of eukaryotic organisms

(A) Scheme of the protein interaction network involved in centrosome cohesion.

(B) The protein interaction network of the centrosome linker proteins using the "BioGRID database".

(C) Left, the class and number of ribonucleoproteins interacting with centrosome linker proteins are shown graphically. Right, top GO terms of the CEP250 interactome (see Table S2).

(D) Top, number of intrinsically disordered domains among centrosome linker proteins. Bottom, the higher number of intrinsically disordered domains in CEP250 protein correlates with the increased size of the interactome. Centrosome linker proteins are indicated by color circles.

(E) Intrinsically disordered domains are correlated with the interactome of additional centrosome proteins including components of the pericentriolar material (PCNT), centriole growth (CEP350), and centrosome appendages (NIN, CEP164).

(F) The levels of expression of centrosome linker, centriolar satellite, and centrosome appendage genes are absent or less represented in lower eukaryotes and non-mammalian organisms. PCM markers are shown for comparison.

indicating a close interaction landscape with the translation machinery, consistent with the literature (Figures S2A–S2C).[13]

When we quantified interactions of centrosome- and cilia-related proteins with spliceosome factors, we unexpectedly discovered that a wide range of pre-mRNA-processing factors (e.g., PRPF4), Serine and Arginine Rich Splicing Factors (SFRS3, SFRS7), RNA Helicases (DDX24), and ubiquitously expressed hnRNPs (e.g., HNRNPM), displayed a robust interaction predominantly with CEP250 (Figures 2B and S2D). CEP250-associated splicing factors showed a significant correlation with genetic disorders including retinitis pigmentosa and microcephaly that are associated with mutations in centrosome genes (Figure 2C).[15,22,23] Similarly, a large number of RNA-processing proteins including factors involved in mRNA splicing and regulation of translation at the centrosome were identified in an independent database of human neural cells (Figure 2D and Table S2).[24] To assess the validity of our observations, we next analyzed an independent centrosome proteome database where the Human Centrosome-Cilium interactome has been captured for a large subset of centrosome proteins (n = 58), resulting in more than 6,000 interactions (Table S2).[25] The global centrosome interactome was enriched for Gene Ontology (GO) terms related to RNA binding and the regulation of mRNA export, expanding the significance of specific RNA-processing proteins at this location (Figure S3A). We reasoned that extrapolating highly reliable and experimentally validated interactors (n = 418, FDR 1%) would result in more precise predictions of the real binding partners. The findings revealed no interactions with translation-related proteins. Notably, we discovered that crucial satellite/cilia basal body (SSX2IP and OFD1) and centrosome linker (CEP135) components established a robust protein interaction network with several spliceosome components alone or in complexes

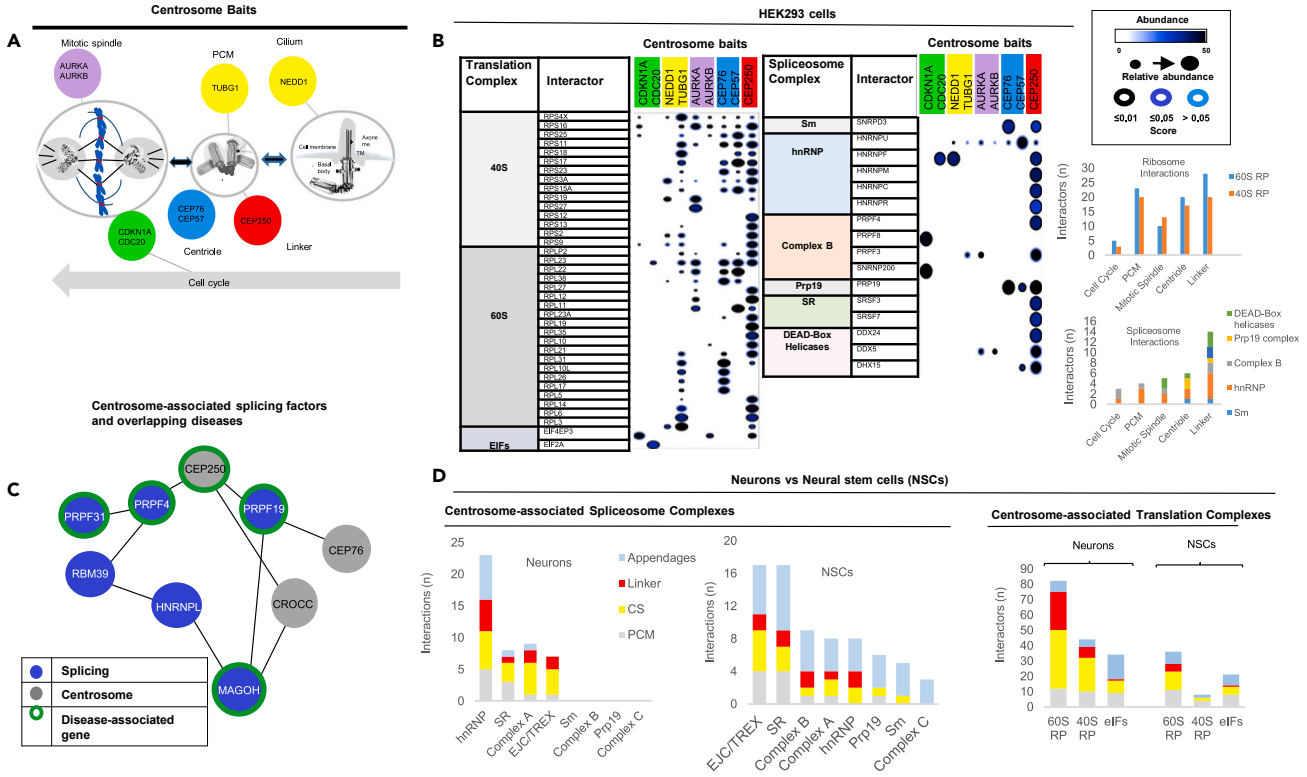


Figure 2. Ribonucleoprotein particles interact with centrosomal proteins

(A) Schematic representation of the centrosome with position and function of the nine bait proteins informed by Fogeron et al. (2013).

(B) Left, the protein interaction maps for the centrosome baits and the translation and splicing components were analyzed according to SAINT (significance analysis of interactome) and are shown by dot plots. Dot shading (blue-black gradient) is the spectral counts for each prey protein. Dot size indicates relative abundance of prey protein in each analysis. Confidence levels and False discovery rate (FDR) are indicated by dot border. Right, the number of interactors for translation/splicing components related to centrosome substructures are shown graphically.

(C) Mutations in both CEP250 and interacting splicing components are associated with inherited retinal blindness and microcephaly (green circles).

(D) Spliceosome and translation complexes identified in different centrosome regions in neurons and neural stem cells informed by O'Neill et al. (2022).

(Figures 3A, 3B, and S3B). Together, these data demonstrated the non-random interaction between splicing and centrosome substructures.

A variety of nuclear spliceosome components localize at the centrosome

To further assess the consistency of the protein interaction networks at the centrosome, proteomic data of centrosomes (>700 proteins) isolated from lamb thymocytes were analyzed.[26] As expected, the thymocytes centrosome proteome was enriched for Gene Ontology (GO) terms related to ribosome and mRNA splicing components. Notably, a set of RNPs including PRP19, RPS11 and SRSF7 showed concordance with human centrosome-associated proteins (Figures S3C and S3D). Proteomic data of centrosomes isolated from HeLa S3 cells revealed that, relative to ribosomes and eIFs factors, splicing proteins at centrosomes increased five-fold in M phase compared to S phase (Figure S4A).[27] The subcellular distribution of ribonucleoproteins localized at centrosomes and centriolar satellites (n = 564) was then mapped using imaging data from the human proteome atlas (Figures S4B and S4C).[28] Notably, the image-based approach allowed us to identify a large number of proteins associated with splicing and RNA export (n = 14) but not with translation-related proteins (Figures 3C and S5A–S5C). The splicing factors located at the centrosome predominantly included components of the Prp19 complex (BCAS2, BUD31), complex C (WDR83, DHX35) and the DEAD-Box-helicases family (Figures 3C and S5B). For instance, the intense labeling of WDR83 at the centrosome of U-251 MG cells derived from a human malignant Glioblastoma multiforme, showed a high concordance with the protein interaction networks in distinct portions of the centrosome (Figures 3A and 3C).

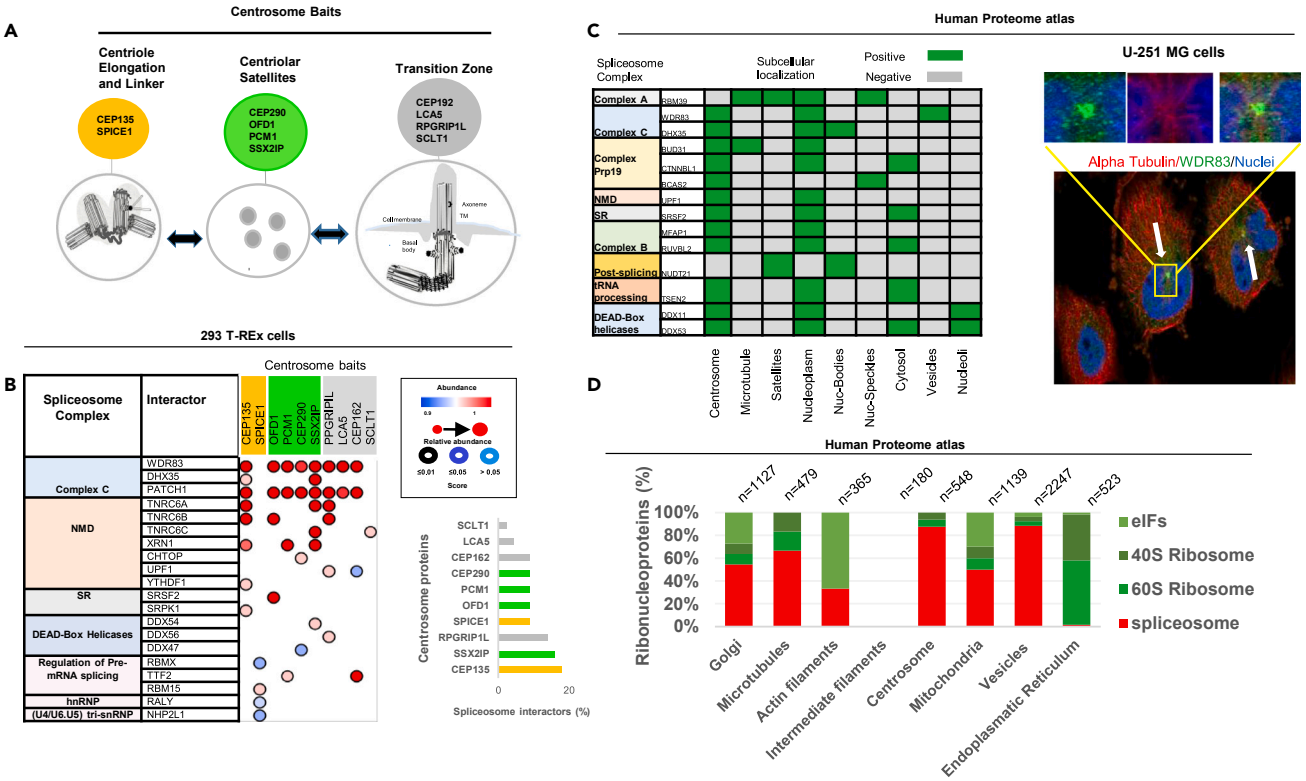


Figure 3. Spliceosome components are localized at centrosomes and centriolar satellites

(A) Schematic representation of the centrosome bait proteins according to Gupta et al. (2015). Centrosome proteins including components of the centriole, centriolar satellites and the cilium are indicated by different colors.

(B) Left, protein interaction data are shown by SAINT. Right, the number of splicing interactors for each centrosome bait is indicated graphically.

(C) Left, subcellular distribution of spliceosome components identified in the Human Protein Atlas database. Right, example of U251-MG cells stained with the splicing factor WDR83 localized at interphase centrosomes (white arrows) from the human proteome atlas. WDR83 (green), Alpha Tubulin (red) and nuclei (blue).

(D) Subcellular distribution of ribonucleoproteins, eIFs, ribosomes (40S and 60S) and splicing components identified in different intracellular structures from the Human Protein Atlas database.

The large majority of the discovered splicing factors were localized at interphase centrosomes (Figures S5A and S5B). In contrast, one of them (DDX53) was clearly confined to the spindle poles in mitosis but not in interphase centrosomes, supporting the notion that the variations of splicing proteins at the centrosome could be cell-type specific and related to the cell cycle (Figure S5C).[29] The analysis of additional subcellular organelles revealed that proteins associated with splicing localized predominantly at the centrosome and vesicles, supporting a possible nucleo-cytosolic shuttling of splicing factors at distinct intracellular sites (Figure 3D). Altogether, these findings suggested that centrosome-associated factors may have cell-type specific relevance at the centrosome.

A subset of centrosome-associated splicing factors are co-regulated in a cell-specific manner

To determine how centrosome and RNP-related transcripts are regulated, we investigated gene expression data across the human GTEx database.[30] As expected, ribosomal-related transcripts were stably expressed in a variety of tissues. Instead, a subset of splicing and centrosomal genes tended to be similarly downregulated in heart, brain, and liver tissues, indicating a tissue-specific co-regulation (Figures S6A–S6C). Of note, these tissues function without a minor spliceosome.[31] Some ciliogenic genes with minor introns such as TCTN3 and CEP170 that were downregulated in liver tissues established a protein interaction network mainly with the identified centrosome-associated splicing factors (CTNNBL1, BUD31, and BCAS2) (Figure S6D). To ensure our observations are not merely explained by a species-specific effect, we collected single cell RNA-seq data from mouse embryonic stem cells and developed a gene regulatory network using SINCERITIES.[32,33] By focusing our attention on the top 2,000 interactions with the highest score, we found

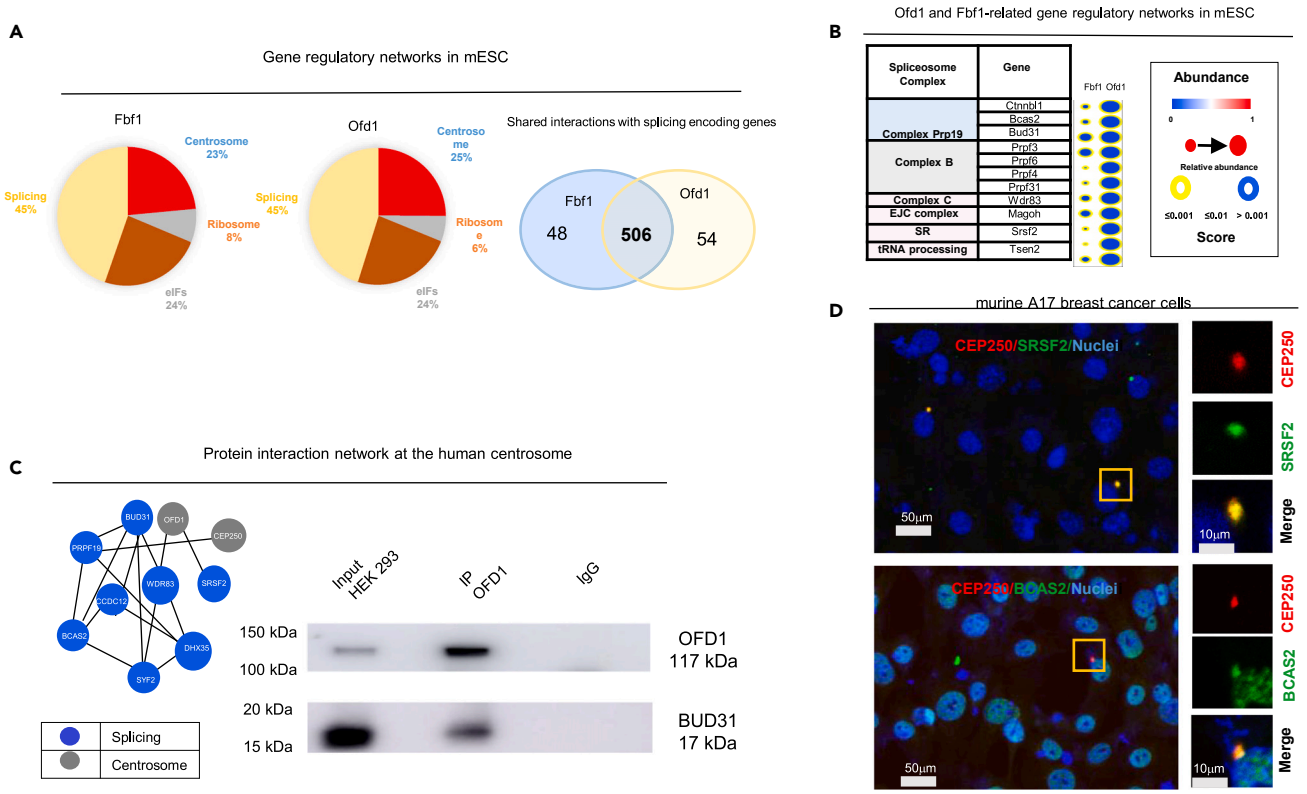


Figure 4. Splicing and centrosomal genes are regulated in a celltype-specific manner in humans and mice

(A) Left, gene regulatory network of Fbf1 and Odf1 in mouse embryonic stem cells (mESC) according to the SINCERITIES algorithm. mESC data from Buettner et al. (2015). Right, the overlapping and divergent interactions with splicing components for Fbf1 and Ofb1 are shown by a Venn diagram.

(B). Genetic interactions of different strengths between query (Fbf1 vs. Ofd1) and target splicing genes are shown by SAINT.

(C) Left, the murine genetic network is recapitulated in the human centrosome protein interaction network. The major hub of this network includes BUD31 (5 interactions). Right, the interaction between OFD1 and BUD31 was confirmed through co-immunoprecipitation of OFD1 and verified by WB in HEK293 cells. Immunoprecipitated BUD31 was detected by WB. IgG were used as loading control.

(D) A17 murine breast cancer stem cells stained with CEP250 (red), BCAS2 or SRSF2 (green). Nuclei are stained with Dapi in blue. Bar 50 µm. On the right of each panel: magnifications show the colocalization of CEP250/BCAS2 and CEP250/SRSF2 at centrosomes. Bar 10 µm.

that centrosome transcripts were highly related to splicing genes as compared to translation-related transcripts (Figure S7A). Notably, *Fbf1* and *Ofd1* were the most prominent centrosomal genes of the regulatory networks (Figures 4A and 4B).[13,34] Moreover, the same splicing factors including a new one SRSF2, were part of the OFD1 and CEP250-regulatory network at the human centrosome (Figure 4C). Major hubs of this network were BUD31 (5 interactions), BCAS2 (4 interactions), and DHX35 (4 interactions). Because the above data suggest a preponderance of Prp19 complex (BUD31) interactions, we validated these *in silico* predictions with western blot analysis after co-immunoprecipitation (co-IP) using as bait protein OFD1. Co-IP experiments validated the BUD31/OFD1 interaction in HEK293 cells but not in HeLa cells, supporting the cell-type specificity of the identified networks (Figure 4C). To validate the predicted centrosome-associated splicing networks, we used immunofluorescence combining CEP250 with the splicing markers BCAS2 and SRSF2 in A17 murine cancer stem cells. We found that nearly 60% of tumor cells had no CEP250 immunolabeling. Notably, CEP250-positive cells revealed a single and large signal adjacent to the nuclei as expected for centrosomes that colocalized with both splicing factors in A17 cells (Figure 4D). In addition to the centrosome, BCAS2 mainly localized to the nucleoplasm as expected for a splicing protein (Figure 4D). To further elucidate the observed interactions, the transcripts of a large number of splicing (n = 91) and centrosomal (548) genes were investigated in A17 and BB1, murine breast cancer cells characterized by mesenchymal and epithelial features, respectively (Figure S7B).[35] Gene-set analysis revealed that a cluster of centrosome-associated splicing genes (CTNNBL1, RBM39, DHX35) and centrosomal genes (OFD1, PCM1 and CEP250) were differentially expressed in BB1 as compared to A17 (Figure S7B). Comprehensive transcriptomic analysis of cancer cell lines from the CCLE and NCI60 panels[36,37] confirmed a

specific co-expression module that included CEP250, CTNNBL1, RBM39 and DHX35, linearly conserved in the human and mouse genome (Figures S7B and S7C). These results suggested the existence of cell-type-specific, tissue-specific and co-expression modules in a subset of centrosome-associated splicing genes.

Tissue-specific alterations in centrosome-associated splicing factors

To better understand the reciprocal relationship between splicing and centrosomes in human diseases, we interrogated gene expression profiles in the TCGA database using as a reference non-neoplastic tissue.[38,39] As expected, ribosome-associated transcripts exhibited higher expression patterns in a variety of blood, brain and gastrointestinal cancers. In contrast, the splicing and centrosome expression profiles were highly tissue-specific (Figure S7D). Co-expression networks were evident in liver cancers (LC), especially in a rare and aggressive variant of the bile duct termed cholangiocarcinoma (CHOL) (Figures S8A and S8B). For instance, *CEP250* expression was significantly higher in liver cancers (~ three times higher) than in normal liver (Figure S8C). In addition, *CEP250* resulted significantly overexpressed and correlated with centrosome amplification in a variety of cancers (Figure S8C).[40] To verify the predicted functional networks at centrosomes, we analyzed human FFPE samples of normal liver tissues and CHOL using colorectal cancers (CRCs) for comparison. Fallopian tube tissue was also used, because it is a standard for establishing a tissue-based expression of the ciliary rootlet coiled-coil protein (CROCC) and because it contains a variegated panel of ciliated and non-ciliated secretory cells.[41] As a first test of our assay, we verified CEP250 and a number of splicing proteins including BCAS2, BUD31, and DHX35 in normal liver and fallopian tube tissues by IHC (Figures 5A and 5B). In fallopian tissues, CEP250 showed selective and intense staining in ciliated cells at the ciliary rootlets. The proteins BCAS2, BUD31 and DHX35 also exhibited a moderate/high expression in ciliated cells (black arrows in Figure 5A). Staining localization in ciliated cells was consistent with splicing proteins being expressed also in the nucleus, which was particularly strong for BCAS2. Of interest, in normal liver samples, CEP250 showed no staining and all splicing proteins displayed a lack of nuclear localization. In contrast, CEP250 and all splicing factors exhibited a strong expression pattern in CHOL as compared to normal liver, confirming the bioinformatics predictions. For instance, the subcellular staining patterns of BUD31 and BCAS2 were mostly nuclear in CHOL tissues (Figure 4C). To verify the localization of the splicing factors at centrosomes within the tissues, we used a multiplexed immunofluorescence method to quantify co-staining at centrosomes combining CEP250 with each splicing marker.[42,43] In fallopian tissues, CEP250 colocalized variably with all splicing factors as discrete foci at cilia (Figures 4A and 4D). Of interest, BUD31 appeared as distinct and large foci at ciliary basal bodies and as a collection of spots localized in nuclei as illustrated in Figure 5A. This appearance of CEP250 and splicing factor colocalization was not detected in normal liver (Figure 5B). Importantly, in the CHOL tissues, CEP250 appeared as discrete bright foci as expected for centrosomes and colocalized with BUD31 and BCAS2 (Figures 5C and 5D). To further evaluate the potential of high-throughput quantification of the detected colocalization at centrosomes, we tested our method using CRC tissue microarrays. In CRCs, cancer cells contained single uniformly shaped CEP250 foci that colocalized with DHX35. This pattern of colocalization was not evident for CEP250-BCAS2 labeling (Figures S9A and S9B). In these experiments, using the 60S ribosomal protein RPL12 as a marker to identify ribosomes, the quantification of colocalization was significantly lower than those detected for the splicing proteins (Figures S9B and S9C). These experiments indicated a high degree of concordance with bioinformatic data based on gene transcript profiling and splicing protein networks detected at centrosomes.

Deregulation of splicing factors influences centrosome-related transcripts

Pathogenetic mutations in splicing genes often cause phenotypic abnormalities associated with centriole defects.[15] Analyzing RNA-seq data from retinal organoids and fibroblasts mutated in the pre-RNA processing factor PRPF31 identified *CEP250* as a common target of splicing alterations (Figure S10A).[22,23] In addition, by exploring an online tool (https://portals.broadinstitute.org/ctrp/)[44] we found that CEP250 and DHX35 were similarly targeted by indisulam, a drug known to induce degradation of the splicing factor RBM39 via the proteasome (Figure S10B). In an independent database, we confirmed that a large number of RBM39 mRNA targets comprised a cytoskeleton and microtubule related-pathway (Figure S11A).[45] We next analyzed RNA-seq data in cells exposed to sudemycin D1 (SD1) or spliceostatin A (STA), two inhibitors of the splicing factor SF3B1 to further understand the effect of splicing inhibition on centrosome genes.[46,47] More than 70% of centrosomal transcripts tended to be downregulated by both inhibitors (Figure S11B). When we compared the differentially expressed genes in nucleus vs. cytoplasm following STA treatment, we identified 54 centrosomal genes that were downregulated in the cytoplasm, suggesting a possible post-transcriptional control of centrosomal gene expression (Figure S11C). Finally, when we considered

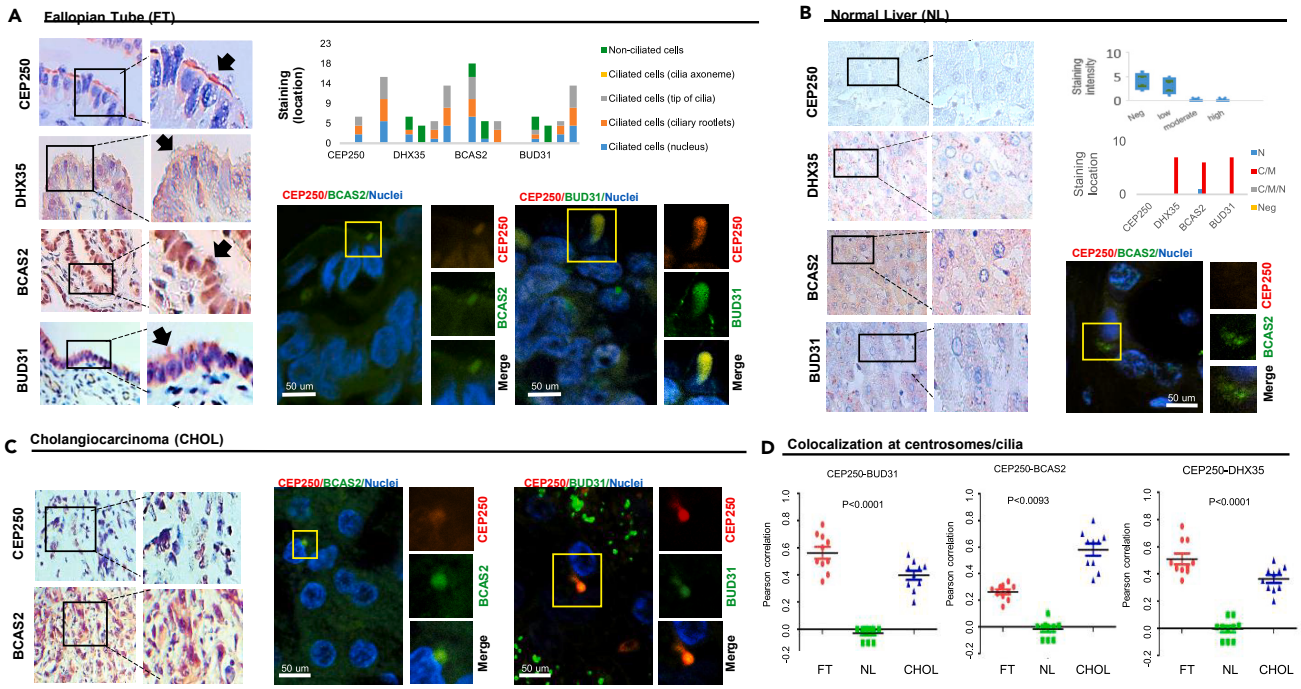


Figure 5. Centrosome-associated splicing factors colocalize with CEP250 in fallopian tubes and cholangiocarcinoma

(A) Left panel, IHC of CEP250, DHX35, BCAS2 and BUD31 in human fallopian tube (FT) tissues. Cilia are indicated by black arrows. Bar 50 μm. Right panel, top, subcellular distribution (Nuclear (N), cytoplasm (C) and membrane (M)) of each marker in ciliated and non-ciliated cells in 10 FT samples is shown graphically. Bottom, FT tissues stained with CEP250 (red), BCAS2 or BUD31 (green). Nuclei in blue. On the right of each panel: magnifications show the colocalization of CEP250/BCAS2 and CEP250/BUD31 at cilia. Bar 10 μm.

(B) Left panel, IHC in human normal liver samples with markers used in A. Bar 50 μm. Right panel, top, staining intensity for all markers is reported as a boxplot. In the middle, the subcellular distribution (Nuclear (N), cytoplasm (C) and membrane (M)) for each marker in 7 samples is shown graphically. Bottom panel, FFPE tissue sample stained with CEP250 (red) and BCAS2 (green). Nuclei in blue. On the right: magnifications show that CEP250 and BCAS2 do not colocalize. Bar 10 μm.

(C) Left panel, representative IHC images of CEP250 and BCAS2 staining in cholangiocarcinoma tissues. Right panel, cholangiocarcinoma sections stained with BCAS2 or BUD31 (green), CEP250 (red), and DNA (blue). On the right magnifications show that the colocalization of CEP250/BCAS2 and CEP250/BUD31 foci marks centrosomes in several cells. Bar 10 μm.

(D) Quantification of the colocalization at centrosomes/cilia in the indicated tissues estimated according to Pearson's correlation. From −1.0 to 1.0; 0 indicates no significant correlation and −1.0 indicates negative correlation. At least 100 cells in ten different high-magnifications fields (60x) tissue sections were examined. Statistical significance was evaluated using unpaired t tests. *p values* are reported in each graph.

junctions that potentially led out-of-frame transcripts following SD1 treatment, 58 aberrant transcripts including centrosome *(CEP250)* and splicing-related transcripts (*RBM39* and *CTNNBL1*) clustered in a genetic module were identified (Figure S11D). These data suggested a possible post-transcriptional control of the splicing on centrosome-related transcripts.

DISCUSSION

In this study we present a comprehensive map of the proteogenomic interactions between spliceosomal and centrosomal components. Pre-mRNA splicing is carried out by a complex machinery called the spliceosome composed of more than 200 proteins comprising core and regulatory elements. Globally, the process of splicing in eukaryotic cells is tightly coupled to the transcription machinery, and ultimately leads to the packaging of mature mRNAs into large ribonucleoparticles composed of numerous RNA-binding proteins.[48] Human spliceosomes contain numerous proteins that are absent in yeast, whose functions remain largely unknown.[20] We first found that the protein-domain architecture of the centrosome linker, which is critical for ciliogenesis and mitosis, reflects the interactome sizes and the complexity of eukaryotic organisms. The numerous disordered regions identified could confer extreme flexibility and enable centrosomal proteins to interact with many partners. Consistent with this, our protein-protein interaction network study identified ribosome and spliceosome components as the most abundant interactors of CEP250. The

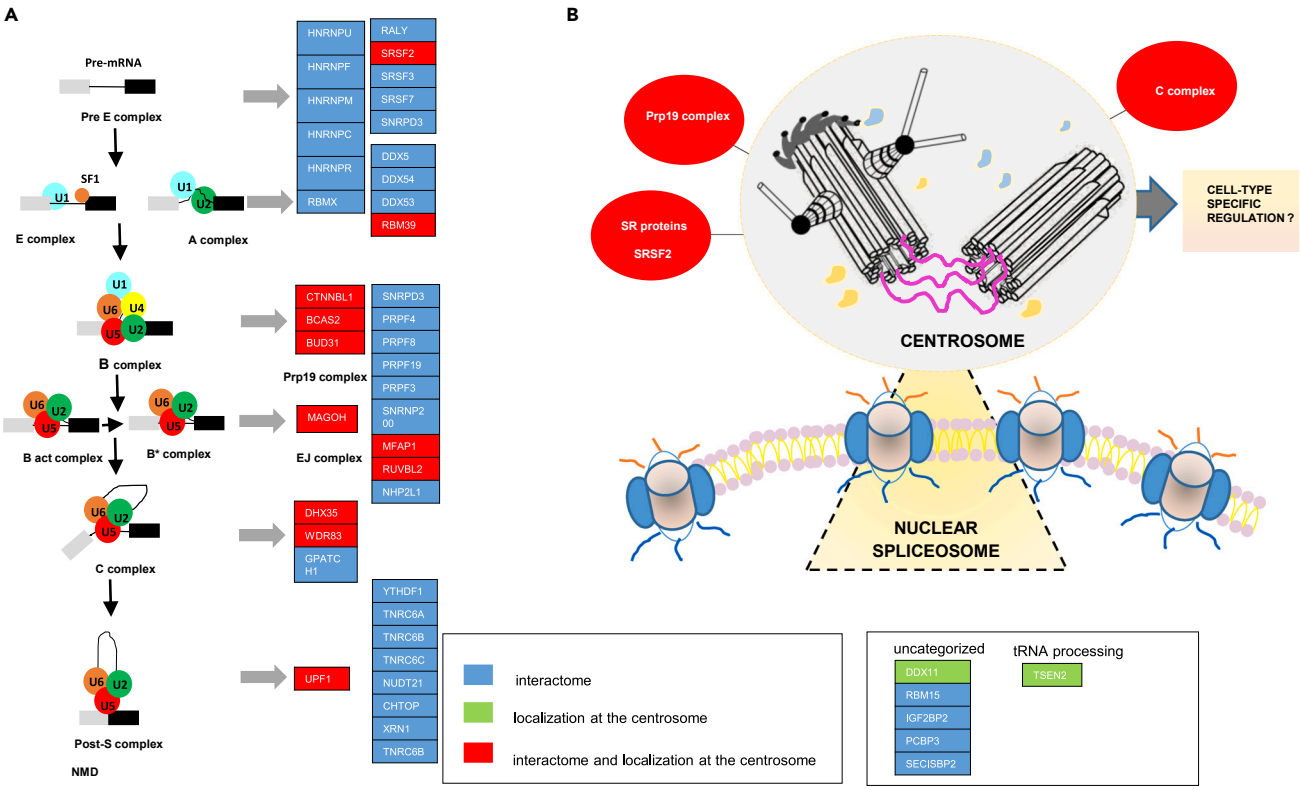


Figure 6. Spliceosome components interact or reside at the centrosome or cilium

(A) Schematic drawing showing the splicing proteins playing a role in the nuclear spliceosome that interact or localize at the centrosome. The function of splicing proteins and the temporal order of sub-complexes recruitment is given.

(B) Schematic representation of the centrosome and the associated splicing components validated in this study.

abundance of intrinsically disordered regions in CEP250 may provide a higher degree of plasticity and promote multivalent interactions at the centrosome.

Spatial proteomics analysis from independent databases confirmed that spliceosome factors expected to localize to the nucleus also localized to the centrosome or centriolar satellites, refining the protein interaction network data. The identified centrosome-associated RNA processing factors were assigned to distinct spliceosome sub-complexes: SR proteins, complexes A and B, Prp19 complex, and potential components of the catalytic complex C (Figure 6A). The presence of multiple subunits of a given complex reinforces the idea of the spatial and functional relevance of these components at the centrosome (Figure 6B). Transcriptomic data in mouse Embryonic Stem Cells (mESC) and human tissues revealed cell-type–specific gene expression networks, a phenomenon corroborated in murine mammary mesenchymal and epithelial cancer cells. Notably, BUD31, a component of the Prp19 complex that functions in splicing was validated as an interactor of the centriolar satellite protein OFD1.

We found that a large subset of splicing and centrosomal genes was expressed at the lowest levels in brain, heart, and liver tissues. Of interest, these tissues manage without a functioning minor spliceosome.[31] Our identification of a genetic module that included CEP250 and centrosome-associated splicing genes in human and mouse explained the similar co-expression and subcellular distribution pattern. A consistent tissue-specific co-expression network was identified in liver cancers, mainly in cholangiocarcinoma, a rare and lethal form of tumor that occurs in the bile ducts. Indeed, we verified that CEP250 and candidate centrosome-associated splicing proteins (BUD31, BCAS2 and DHX35) were consistently expressed and colocalized in the ciliary body of fallopian tube tissues. In contrast, these proteins were poorly expressed or displayed a lack of nuclear localization in normal liver. In cholangiocarcinoma tissues, we observed a strong overexpression of CEP250 and splicing proteins that colocalized as discrete cytoplasmic foci, as expected for centrosomes. It is interesting to note that our validation on human tissue samples recapitulated

bioinformatic predictions based on the subcellular localization profiles in different cancer cells. However, more work is needed, including comparisons between RNA and protein expression levels in cells and tissues, to understand tissue specificity in more detail. Recent studies showed evidence that centrosome-associated splicing factors may be cell-type specific in neurons and neural stem cells, suggesting a dual role of the centrosomal proteome in the maintenance of pluripotency and cell-cycle progression.[24,48] Why splicing factors reside at centrosomes is a question under active investigation. The cytoplasmic functions of splicing factors at centrosomes may include roles in "transcriptome homeostasis", specifically the surveillance of unspliced pre-mRNAs and the control of mRNA stability and translation.[48] In this scenario, a critical issue will be to test if centrosomes and cilia also contain a dedicated pool of spliceosomal U snRNAs.[49] We propose that centrosome-associated splicing factors could facilitate the transport and assembly of RNP particles contributing to cell division and ciliogenesis in a cell-type specific manner. Many challenges remain in studying the native RNP aggregates surrounding centrosomes in a cell-free context. Thus, our study opens an exciting area of future research on the cytosolic compartmentalization of the RNP condensates in the context of cellular differentiation and development trajectories in metazoans.

Limitations of the study

The current data just suggested a close interaction between spliceosome and centrosome-related proteins, the cause-and-effect relationship between them need to be illustrated. Another limitation of our research is the lack of functional experiments to explore the dynamic molecular changes of splicing factors localized at the centrosome. The reason why splicing proteins reside at centrosomes has not been addressed in our study. Our results must further be corroborated as more spatial multi-omics data in the same single cells across development, aging and disease become available.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Culture of murine breast cancer cell lines
- METHOD DETAILS
  - Datasets
  - Reconstruction of protein-protein interactomes and spatial proteomics
  - Disordered and secondary protein structure predictions
  - Gene expression analysis from human tissues
  - Mouse embryonic stem cell (mESC) single cell transcriptomic data
  - Disease modeling and mini-intron containing gene analysis
  - Normal and cancer tissue specimens
  - Splicing and centrosome protein quantification by immunohistochemistry (IHC)
  - Multiplexed immunofluorescence analysis and confocal microscopy
  - Co-immunoprecipitation (co-IP) experiments and Immunoblot analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106602.

ACKNOWLEDGMENTS

We are very grateful to Dr. Pietro Cirigliano (Nikon Instruments S.p.A) for support and invaluable technical assistance in the use of confocal microscopy and acquisitions of images. This work was supported by the (FFABR grant n.4982/2018), from the Italian Ministry of University and Research (MIUR) to M.P and by the Associazione Italiana per la Ricerca sul Cancro (AIRC; Grant IG 26414/2022 to BF).

AUTHOR CONTRIBUTIONS

F.P.C., L.C., and M.P. developed bioinformatics analysis. P.P., G.G., and A.R. provided human FFPE samples. N.F. and M.P. developed and acquired IHC and IF analysis on human tissues samples. J.B., F.B., and M.G. provided proteomic and transcriptomic analysis on centrosomes isolated from thymocytes and murine cancer cells. N.P. performed co-immunoprecipitation experiments in cultured cells supervised by B.F., M.P., and B.F. acquired the financial support for the project leading to this publication. M.P. drafted the manuscript with inputs from B.F. and L.C. M.P. conceived and supervised the study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 25, 2022
Revised: January 16, 2023
Accepted: March 29, 2023
Published: April 10, 2023

REFERENCES

1. Kaur, T., Alshareedah, I., Wang, W., Ngo, J., Moosa, M.M., and Banerjee, P.R. (2019). Molecular crowding tunes material states of ribonucleoprotein condensates. Biomolecules *9*, 71.

2. Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. J. Cell Sci. *130*, 2789–2796.

3. Tauber, D., Tauber, G., and Parker, R. (2020). Mechanisms and regulation of RNA condensation in RNP granule formation. Trends Biochem. Sci. *45*, 764–778.

4. Uversky, V.N. (2019). Supramolecular fuzziness of intracellular liquid droplets: liquid–liquid phase transitions, membraneless organelles, and intrinsic disorder. Molecules *24*, 3265.

5. Rale, M.J., Kadzik, R.S., and Petry, S. (2018). Phase transitioning the centrosome into a microtubule nucleator. Biochemistry *57*, 30–37.

6. Woodruff, J.B., Ferreira Gomes, B., Widlund, P.O., Mahamid, J., Honigmann, A., and Hyman, A.A. (2017). The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. Cell *169*, 1066–1077.e10.

7. Dos Santos, H.G., Abia, D., Janowski, R., Mortuza, G., Bertero, M.G., Boutin, M., Guarín, N., Méndez-Giraldez, R., Nuñez, A., Pedrero, J.G., et al. (2013). Structure and non-structure of centrosomal proteins. PLoS One *8*, e62633.

8. Nido, G.S., Méndez, R., Pascual-García, A., Abia, D., and Bastolla, U. (2012). Protein disorder in the centrosome correlates with complexity in cell types number. Mol. Biosyst. *8*, 353–367.

9. Remo, A., Li, X., Schiebel, E., and Pancione, M. (2020). The centrosome linker and its role in cancer and genetic disorders. Trends Mol. Med. *26*, 380–393.

10. Mazo, G., Soplop, N., Wang, W.-J., Uryu, K., and Tsou, M.-F.B. (2016). Spatial control of primary ciliogenesis by subdistal appendages alters sensation-associated properties of cilia. Dev. Cell *39*, 424–437.

11. Hata, S., Pastor Peidro, A., Panic, M., Liu, P., Atorino, E., Funaya, C., Jäkle, U., Pereira, G., and Schiebel, E. (2019). The balance between KIFC3 and EG5 tetrameric kinesins controls the onset of mitotic spindle assembly. Nat. Cell Biol. *21*, 1138–1151.

12. Ito, K.K., Watanabe, K., and Kitagawa, D. (2020). The emerging role of ncRNAs and RNA-binding proteins in mitotic apparatus formation. Noncoding. RNA *6*, 13.

13. Iaconis, D., Monti, M., Renda, M., van Koppen, A., Tammaro, R., Chiaravalli, M., Cozzolino, F., Pignata, P., Crina, C., Pucci, P., et al. (2017). The centrosomal OFD1 protein interacts with the translation machinery and regulates the synthesis of specific targets. Sci. Rep. *7*, 1224.

14. Sepulveda, G., Antkowiak, M., Brust-Mascher, I., Mahe, K., Ou, T., Castro, N.M., Christensen, L.N., Cheung, L., Jiang, X., Yoon, D., et al. (2018). Co-translational protein targeting facilitates centrosomal recruitment of PCNT during centrosome maturation in vertebrates. Elife *7*, e34959.

15. Griffin, C., and Saint-Jeannet, J. (2020). Spliceosomopathies: diseases and mechanisms. Dev. Dyn. *249*, 1038–1046.

16. Paolini, N.A., Attwood, M., Sondalle, S.B., Vieira, C.M.D.S., van Adrichem, A.M., di Summa, F.M., O'Donohue, M.F., O'Donohue, M.-F., Rachuri, S., Briggs, J.W., et al. (2017). A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. Am. J. Hum. Genet. *100*, 506–522.

17. Alliegro, M.A., Henry, J.J., and Alliegro, M.C. (2010). Rediscovery of the nucleolinus, a dynamic RNA-rich organelle associated with the nucleolus, spindle, and centrosomes. Proc. Natl. Acad. Sci. USA *107*, 13718–13723.

18. Alliegro, M.C., Alliegro, M.A., and Palazzo, R.E. (2006). Centrosome-associated RNA in surf clam oocytes. Proc. Natl. Acad. Sci. USA *103*, 9034–9038.

19. Amato, R., Morleo, M., Giaquinto, L., di Bernardo, D., and Franco, B. (2014). A network-based approach to dissect the cilia/centrosome complex interactome. BMC Genom. *15*, 658.

20. Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell *170*, 701–713.e11.

21. Fogeron, M.-L., Müller, H., Schade, S., Dreher, F., Lehmann, V., Kühnel, A., Scholz, A.-K., Kashofer, K., Zerck, A., Fauler, B., et al. (2013). LGALS3BP regulates centriole biogenesis and centrosome hypertrophy in cancer cells. Nat. Commun. *4*, 1531.

22. Buskin, A., Zhu, L., Chichagova, V., Basu, B., Mozaffari-Jovin, S., Dolan, D., Droop, A., Collin, J., Bronstein, R., Mehrotra, S., et al. (2018). Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. Nat. Commun. *9*, 4234.

23. Huang, X.F., Xiang, L., Fang, X.L., Liu, W.Q., Zhuang, Y.Y., Chen, Z.J., Shen, R.J., Cheng, W., Han, R.Y., Zheng, S.S., et al. (2019). Functional characterization of CEP250 variant identified in nonsyndromic retinitis pigmentosa. Hum. Mutat. *40*, 1039–1045.

24. O'Neill, A.C., Uzbas, F., Antognolli, G., Merino, F., Draganova, K., Jäck, A., Zhang, S., Pedini, G., Schessner, J.P., Cramer, K., et al. (2022). Spatial centrosome proteome of human neural cells uncovers disease-relevant heterogeneity. Science *376*, eabf9088.

25. Gupta, G.D., Coyaud, É., Gonçalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J.M., Cheung, S.W.T., et al. (2015). A dynamic protein interaction landscape of the human centrosome-cilium interface. Cell *163*, 1484–1499.

26. Busselez, J., Chichón, F.J., Rodríguez, M.J., Alpízar, A., Gharbi, S.I., Franch, M., Melero, R., Paradela, A., Carrascosa, J.L., and Carazo, J.-M. (2019). Cryo-Electron Tomography and Proteomics studies of centrosomes from differentiated quiescent thymocytes. Sci. Rep. *9*, 7187.

27. Kimura, H., Miki, Y., and Nakanishi, A. (2014b). Centrosomes at M phase act as a scaffold for the accumulation of intracellular ubiquitinated proteins. Cell Cycle *13*, 1928–1937.

28. Thul, P.J., and Lindskog, C. (2018). The human protein atlas: a spatial map of the human proteome. Protein Sci. *27*, 233–244.

29. Hofmann, J.C., Husedzinovic, A., and Gruss, O.J. (2010). The function of spliceosome components in open mitosis. Nucleus *1*, 447–459.

30. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The genotype-tissue expression (GTEx) project. Nat. Genet. *45*, 580–585.

31. Olthof, A.M., Hyatt, K.C., and Kanadia, R.N. (2019). Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. BMC Genom. *20*, 686.

32. Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

33. Papili Gao, N., Ud-Dean, S.M.M., Gandrillon, O., and Gunawan, R. (2018). SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics *34*, 258–266.

34. Inoko, A., Yano, T., Miyamoto, T., Matsuura, S., Kiyono, T., Goshima, N., Inagaki, M., and Hayashi, Y. (2018). Albatross/FBF1 contributes to both centriole duplication and centrosome separation. Gene Cell. *23*, 1023–1042.

35. Galiè, M., Sorrentino, C., Montani, M., Micossi, L., Di Carlo, E., D'Antuono, T., Calderan, L., Marzola, P., Benati, D., Merigo, F., et al. (2005). Mammary carcinoma provides highly tumourigenic and invasive reactive stromal cells. Carcinogenesis *26*, 1868–1878.

36. 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., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

37. Guo, T., Luna, A., Rajapakse, V.N., Koh, C.C., Wu, Z., Liu, W., Sun, Y., Gao, H., Menden, M.P., Xu, C., et al. (2019). Quantitative proteome landscape of the NCI-60 cancer cell lines. iScience *21*, 664–680.

38. Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. *45*, W98–W102.

39. Chandrashekar, D.S., Bashel, B., Balasubramanya, S.A.H., Creighton, C.J., Ponce-Rodriguez, I., Chakravarthi, B.V.S.K., and Varambally, S. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia *19*, 649–658.

40. Ogden, A., Rida, P.C.G., and Aneja, R. (2017). Prognostic value of CA20, a score based on centrosome amplification-associated genes, in breast tumors. Sci. Rep. *7*, 262.

41. Remo, A., Manfrin, E., Parcesepe, P., Ferrarini, A., Han, H.S., Mickys, U., Laudanna, C., Simbolo, M., Malanga, D., Oliveira, D.M., et al. (2018a). Centrosome linker–induced tetraploid segregation errors link rhabdoid phenotypes and lethal colorectal cancers. Mol. Cancer Res. *16*, 1385–1395.

42. Wang, M., Knudsen, B.S., Nagle, R.B., Rogers, G.C., and Cress, A.E. (2019b). A method of quantifying centrosomes at the single-cell level in human normal and cancer tissue. Mol. Biol. Cell *30*, 811–819.

43. Zinchuk, V., and Grossenbacher-Zinchuk, O. (2011). Quantitative colocalization analysis of confocal fluorescence microscopy images. Curr. Protoc. Cell Biol. *Chapter 4*. Unit 4.19.

44. Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

45. Wang, E., Lu, S.X., Pastore, A., Chen, X., Imig, J., Chun-Wei Lee, S., Hockemeyer, K., Ghebrechristos, Y.E., Yoshimi, A., Inoue, D., et al. (2019a). Targeting an RNA-binding protein network in acute myeloid leukemia. Cancer Cell *35*, 369–384.e7.

46. Wu, G., Fan, L., Edmonson, M.N., Shaw, T., Boggs, K., Easton, J., Rusch, M.C., Webb, T.R., Zhang, J., and Potter, P.M. (2018). Inhibition of SF3B1 by molecules targeting the spliceosome results in massive aberrant exon skipping. RNA *24*, 1056–1066.

47. Yoshimoto, R., Kaida, D., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Kawamura, Y., Hayashizaki, Y., Mayeda, A., and Yoshida, M. (2017). Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA *23*, 47–57.

48. Busselez, J., Uzbekov, R.E., Franco, B., and Pancione, M. (2023). New insights into the centrosome-associated spliceosome components as regulators of ciliogenesis and tissue identity. Wiley Interdiscip. Rev. RNA, e1776.

49. Steitz, J.A., Dreyfuss, G., Krainer, A.R., Lamond, A.I., Matera, A.G., and Padgett, R.A. (2008). Where in the cell is the minor spliceosome? Proc. Natl. Acad. Sci. USA *105*, 8485–8486.

50. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. *30*, 187–200.

51. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205.

52. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-CEP250	Proteintech Europe	Cat#66814-1-Ig; RRID: AB_2882157
Rabbit polyclonal anti-BUD31	Cusabio	Cat# CSB-PA01749A0Rb; RRID: AB_2936399
Rabbit polyclonal anti-BCAS2	Abclonal	Cat#A4398; RRID: AB_2765658
Rabbit polyclonal anti-DHX35	Cusabio	Cat#CSB-PA887985LA01HU; RRID: AB_2936400
Rabbit polyclonal anti-SRSF2	Abclonal	Cat#A3635; RRID: AB_2765195
Rabbit polyclonal anti-OFD1	Sigma-Aldrich	Cat#HPA031103: RRID: AB_10602188
Rabbit polyclonal anti-RPL12	Sigma-Aldrich	Cat#HPA003403; RRID: AB_1856432
Goat Anti-Mouse IgG H&L (Alexa Fluor® 647)	Abcam	Cat#ab150115; RRID: AB_2687948
Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488)	Abcam	Cat#ab150077; RRID: AB_2630356
Goat IgG anti-Rabbit IgG (H+L)-Alexa Fluor 488	Jackson ImmunoResearch	Cat#111545144; RRID: AB_2338052
Alexa Fluor® 647 AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	Cat#115605003: RRID: AB_2338902
Biological samples		
Healthy Human Fallopian Tube tissue	Mater Salutis Hospital tissue bank, Verona, Italy.	N/A
Human colorectal cancer tissue microarrays	DST, unisannio tissue bank, Benevento Italy.	N/A
Healthy Human liver tissue	Casa Sollievo della Sofferenza Hospital tissue bank, Foggia, Italy	N/A
Human liver cancer tissue	Casa Sollievo della Sofferenza Hospital tissue bank, Foggia, Italy	N/A
Chemicals, peptides, and recombinant proteins		
Universal HIER antigen retrieval reagent	Abcam	Cat#ab208572
Hoechst 33342 Staining Dye Solution	Abcam	Cat#ab228551
Anti-Fade Fluorescence Mounting Medium	Abcam	Cat#ab104135
Paraphormaldeyde	Bio-Optika	Cat#O5-K01015
Phosphatase inhibitors	Roche	Cat#04906837001
Polyvinylidene difluoride (PVDF) membrane	Millipore	Cat#IPVH00010
Protease inhibitors	Sigma-Aldrich	Cat#P8340
Immobilon Crescendo Western HRP substrate	Millipore	Cat#WBLUR0500
SuperSignal™ West Femto Maximum Sensitivity Substrate	Thermoscientific	Cat#34095
Restore™ Western Blot Stripping Buffer	Thermoscientific	Cat#21059
Critical commercial assays		
RNeasy kit	Qiagen	Cat#74004
Deposited data		
Centrosome proteome from HeLa S3	Kimura et al.[27]	www.landesbioscience.com/journals/cc/article/28896
Centrosome proteome from the young lamb thymus,.	Busselez et al.[26]	PXD003928
Protein-protein interaction of centrosome proteins	Fogeron et al.[21]	thebiogrid.org/165943/publication/lgals3bp
Spatial centrosome proteome of human neural stem cells and neurons	O'Neill et al.[24]	https://doi.org/10.1126/science.abf9088

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spatial centrosome proteome of the Human Centrosome-Cilium Interface	Gupta et al.[25]	https://prohits-web.lunenfeld.ca
RNA-seq data from PRPF31-mutated cells	Buskin et al.[22]	PRJEB22885
RNA-seq data from RBM39ko cells	Wang et al.[42]	GSE114558
Transcript profiling from Spliceostatin A-treated cells	Yoshimoto et al.[47]	GSE72156
Transcript profiling of Sudemycin-treated cells	Wu et al.[46]	GSE102539
Single cell-RNA expression profile of mouse embryonic stem cells (mESM)	Buettner et al.[32]	E-MTAB-2805
A17 and BB1 microarrays	This paper	See Table S4
Original western blot images	This paper	See Figure S12
Experimental models: Cell lines		
Murine breast cancer cell A17	Laboratory of Mirco Galiè	N/A
Murine breast cancer cell BB1	Laboratory of Mirco Galiè	N/A
Human HEK293 cell	ATCC	CRL-1573
Human Hela cell	ATCC CCL	CCL-2
Software and algorithms		
ImageJ	ImageJ	https://imagej.nih.gov/ij/
Multichip Average (RMA) algorithm	?	?
GraphPad Prism8	GraphPad	https://www.graphpad.com
Gepia2 database	GEPIA	GEPIA 2 (cancer-pku.cn)
BioGRID database	BioGRID	thebiogrid.org
Human Proteome Atlas database	Human Proteome Atlas	https://www.proteinatlas.org
Significance Analysis ofINTeractome (SAINT)	SAINT	https://prohits-viz.org/
WEBGestalt Software	WEBGestalt	http://www.webgestalt.org/
DAVID Software	DAVID	http://david.ncifcrf.gov
g:Profiler Software	g:Profiler	https://biit.cs.ut.ee/gprofiler/gost
DISOPRED2 algorithm	DISOPRED2	http://bioinf.cs.ucl.ac.uk/
FoldIndex algorithm	FoldIndex	http://fold.proteopedia.org/
Genotype-Tissue Expression (GTEx) portal	GTEx	https://gtexportal.org
SINCERITIES algorithm	SINCERITIES	https://github.com/CABSEL/SINCERITIES
Cancer Therapeutics Response Portal (CTRP)	CTRP	https://portals.broadinstitute.org/ctrp/
UALCAN portal	UALCAN	https://ualcan.path.uab.edu
cBioPortal database	cBioPortal	https://www.cbioportal.org/
CellMiner portal	CellMiner	discover.nci.nih.gov/cellminercdb
Minor Intron DataBase (MIDB)	MIDB	https://midb.pnb.uconn.edu
Cancer Cell Line Encyclopedia (CCLE)	CCLE	https://sites.broadinstitute.org
NCI60 panel database	NCI60	https://dtp.cancer.gov/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Massimo Pancione (massimo.pancione@unisannio.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper did not report any original code;
- Original western blot images are shown in Figure S12 and are publicly available as of the date of publication. Individual microarray data of mRNAs isolated from A17 and BB1 cells of splicing and centrosome genes are provided in Table S3.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Culture of murine breast cancer cell lines

To evaluate whether centrosome-associated splicing components are cell-specific, we used two murine breast cancer cell lines: A17 expressing mesenchymal and basal-like stem cells features and BB1 expressing epithelial features. Both cancer cell models were established from the same spontaneous mammary carcinomas of FVB mice transgenic for the HER-2/neuT oncogene as reported.[35] Whole-genome microarray analysis was performed in replicated cellular extracts (n=5 for A17) and (n=4 for BB1) using the NimbleGen Gene Expression system. Briefly, total RNA was isolated from samples using the Qiagen RNeasy kit (Qiagen #74004), following the manufacturer's instructions. RNA was used for cDNA synthesis, followed by labeling of the cDNA with Cy3. The labeled cDNA samples were hybridized to the *Mus musculus* 12×135K Array (Roche NimbleGen), which represents 44,170 murine genes. Single color NimbleGen arrays were scanned using a GenePix 4400A Microarray Scanner. Data were extracted from scanned images using the NimbleScan software, and the Robust Multichip Average (RMA) algorithm was used to generate gene expression values (Table S3). Hybridization, scanning and normalization of the data were performed by the Functional Genomic Center of the University of Verona (Verona, Italy). Gene set enrichment analysis (GSEA) was used to associate the specific-cancer cell phenotype to a group of genes coding for splicing (91 genes) and centrosome (548 genes). For immunofluorescence 2x105 cells/well of A17 cells were fixed 10 min in, paraphormaldeyde 4% Bio-Optika #O5-K01015, and permeabilized 10 min with PBS+0.2% Triton-x-100 at RT. After blocking of a-specific binding sites with PBS+1%BSA at RT cells were incubated 1h at 37°C with a mix of primary antibodies depending on the type of colocalization assessed. The following antibodies were used: anti-SRSF2 Abclonal #A3635, anti-CEP250 Proteintech #66814, anti-BCAS2 Abclonal #A4398, all diluted 1:100 in PBS+BSA. After 3 washes in PBS, cells were incubated 1h at 37°C with a mix of secondary antibodies (goat anti-rabbit 488 Alexafluor dil. 1:1000; goat anti-mouse 647 Jackson 1:200) in PBS -1%BSA. Cells were imaged using a Confocal Microscope (Nikon AXR system).

METHOD DETAILS

Datasets

We collected and analyzed independent datasets containing information on protein-protein interactions, proteomics, transcriptomics, and geneticdata from cellular and tissue resources. The characteristics of the databases and bioinformatics tools used in this study are reported in (Table S2).

Reconstruction of protein-protein interactomes and spatial proteomics

Protein-protein interaction data of centrosome linker proteins were analyzed by downloading raw data from the BioGRID database.[50] In order to correct for the inherent inaccuracy and technical variability of protein interactions, a Mascot Score >95 percent and a Confidence Score >0.9 was used. To independently determine the role of RNA binding proteins at centrosomes, proteomic data were also analyzed from isolated centrosomes as previously described.[26–28] Protein-protein interaction networks at centrosomes were validated using independent datasets analyzed by spatial protein proximity labelling.[21,24,25] These methods using multiple baits from the same subcellular localization followed by mass spectrometry yield detailed maps of individual centrosomes. The number of ribonucleic proteins present in the S-phase (S > M), M-phase (S < M) or both the phase (S = M) during cell cycle in centrosome proteomes was analyzed as described.[27] The identified ribonucleic proteins were categorized as eIFs (n=20), 40S (n=36), 60S ribosome (n=43) and splicing factors (n=37) and assigned to the mentioned cell cycle stages.[27] To visualize the subcellular distributions of the human centrosomal proteins, a systematic analysis of the centrosome proteome (n=548 proteins) from the human proteome atlas was performed.[28] The subcellular location of each protein was determined based on the signal pattern for nucleus (DAPI, blue), microtubules (mouse anti-alpha-tubulin, red) and the protein of interest (rabbit antibody green).[28] The subcellular localization

patterns were compared with known spliceosome proteins (n=257) extracted from the Oncomine database. The spatial proteomes of Golgi (n=1127 proteins); mitochondria (n=1139 proteins); vesicles (n=2247 proteins); endoplasmic Reticulum (n=523 proteins), and cytoskeleton elements including microtubules (n=479 proteins), intermediate filaments (n=180 proteins) and actin filaments (n=365 proteins), were analyzed for comparison. Ribonucleoproteins localized at centrosomes were cataloged as eukaryotic initiation factors (eIFs), ribosome, and spliceosome proteins. For the protein-protein interaction maps, both baits and interactors were hierarchically clustered, using the Spearman rank correlation. To visualize the strength of the molecular interaction, the tool significance analysis of interactome (SAINT) was used to create dotplots, heatmaps, and graphical representations of interacting partners.[25] The software platform WEBGestalt at (http://www.webgestalt.org/) was used to independently verify and visualize important protein-protein interaction networks at centrosomes.[51] Overlapping characteristics and dataset discrepancies were analyzed using Venn diagrams. To investigate the molecular pathways and functional profiles of protein-protein interaction at the centrosome, DAVID (Database for Annotation, Visualization, and Integrated Discovery) and g:Profiler were used.[52] Significant pathways and gene ontology (GO)-Terms were identified using a Benjamini-Hochberg adjusted P-value less than 0.05. GO-Terms that scored significant in two different bioinformatics tools were included.

Disordered and secondary protein structure predictions

Disordered predictions in centrosome proteins were obtained with the DISOPRED2 and FoldIndex algorithms as previously reported.[7] Tests were conducted to see if the two algorithms produced results that were both quantitatively and qualitatively comparable. The longest isoform of each gene was obtained from the Ensembl database at https://www.ensembl.org/index.html. Phylogenetic analysis of centrosomal genes was performed by using reciprocal pairwise sequence-based (BLASTP and phmmer) and domain-based (hmmsearch) approaches.

Gene expression analysis from human tissues

Gene expression profiles from human tissue samples were analyzed using the Genotype-Tissue Expression (GTEx) biobank at the https://www.gtexportal.org/home/. Normalized RNA-seq data expressed as Transcripts Per Million (TPM) were collected from 54 non-disease tissue samples.[30] Analysis of the expression levels of a given set of genes of interest (eIFs, ribosomes, splicing factors, and centrosome) was performed by including at least 10 genes from each category.

Mouse embryonic stem cell (mESC) single cell transcriptomic data

To investigate the relationship between the centrosome and ribonucleic encoding genes during development, we constructed gene regulatory networks (GRNs) from single cell transcriptional profiles in mouse Embryonic Stem Cells (mESC).[32] The dataset comprises three expression matrices of 38,390 genes. Each matrix represents a cell cycle phase (G1, S, G2-M) with 96 single cells. We reduced the original set of genes to genes involved in the formation/regulation of the centrosome (659), transcription factors (2761), ribosomal (159), splicing (1305), and translation initiation (663) genes. Gene categories were extrapolated from the Ribosomal Protein Gene (RPG) database (http://ribosome.med.miyazaki-u.ac.jp) and the National Center for Biotechnology Information (NCBI) dataset (www.ncbi.nlm.nih.gov). To infer gene regulatory networks from single cell transcriptional profiles the SINCERITIES (SINgle CEll Regularized Inference using TIme-stamped Expression profileS) algorithm was used.[33] Briefly, the algorithm is based on the hypothesis that changes in the expression of a gene at one-time point allows us to predict changes in the gene expression distributions of the corresponding target genes at the next time point. The algorithm returns the matrix of adjacencies, which is an m-by-m matrix containing the weights of regulatory edges, and a distance matrix, which is an n-1-by-m matrix containing the (normalized) distribution distance (DD) computed during the network inference, using the linear regression. We generated a list of 3,903,843 gene-gene interactions ordered by the relative importance that was further analyzed with SAINT as indicated above.

Disease modeling and mini-intron containing gene analysis

The enrichment of interacting genes in common disease phenotypes was tested with the WebGestalt tool, using a False Discovery Rate (FDR) less than 0.05. To evaluate how PRPF31 mutations affect mis-splicing in centrosome-related genes, transcriptome profiles from patient-derived retinal organoids and fibroblasts harboring mutations in PRPF31 were analyzed.[22] Furthermore, transcriptome profiles derived from cells treated with different types of splicing inhibitors (sudemycin and spliceostatin A targeting SF3B1 and

indisulam targeting RBM39) were analyzed to determine the effect of splicing inhibition on centrosome-related genes (Table S2). The therapeutics response portal at https://portals.broadinstitute.org/ctrp/, which links genetic, lineage, and cellular aspects of cancer cell lines to small-molecule sensitivity, was used to investigate drugs that target gene clusters of interest.[44] The Cancer Genome Atlas (TCGA) database at https://cancergenome.nih.gov/was used to evaluate the transcript levels of centrosome and RNA-binding protein genes. The differential expression of tumor/normal tissues was investigated using ANOVA (analysis of variance). The gene expression profiles in cancer cells were investigated in two independent databases, the Cancer Cell Line Encyclopedia (CCLE) and the NCI60 panel of 947 and 60 cancer cell lines, respectively (Table S2). Explorative immunohistochemistry (IHC) for centrosome (CROCC) and pre-mRNA processing factors (SNRPD3) in non-pathological and cancer tissues were collected from the Human Proteome Atlas.[28] GEPIA, a bioinformatics web application for analyzing RNA sequencing expression data across cancers and normal tissues, was used to import the graphical representation of gene expression data. A P-value less than 0.05 was considered as statistically significant. The correlation between gene expression data was verified independently using UALCAN, an interactive web platform for assessing cancer OMICS and proteomic data (Table S2). The cBioPortal database, available at https://www.cbioportal.org/, was used to confirm variations in gene expression profiles and alterations in genomic data and DNA copy number changes. The stoichiometric relationships between transcripts and proteins were investigated with CellMiner at discover.nci.nih.gov/cellminercdb.[37] To determine if the identified centrosome and cilia interactors could be categorized as minor intron-containing genes (MIGs), the tool available at https://midb.pnb.uconn.edu was used.[31]

Normal and cancer tissue specimens

The use of human tissue samples was approved by the ethical committee for research of the University of Verona (protocol 42160) and abides by the Declaration of Helsinki principles. FFPE (Formalin-fixed Paraffin-embedded) samples were collected from the archives of Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, FG, Italy and Mater Salutis Hospital AULSS9, "Scaligera", 37122 Verona, Italy. The human normal tissues included liver (n=7 cases), colon (n=3 cases), and fallopian tubes (n=10 cases). The pathology sections included two different forms of human cancers; cholangiocarcinoma (n=10 cases) and colorectal (CRC) tissue microarrays (n=35 cases). The luminal epithelium of the fallopian tube is composed of highly polarized secretory and ciliated cells, therefore, it has been chosen as a positive control as reported.[41]

Splicing and centrosome protein quantification by immunohistochemistry (IHC)

IHC was performed with the BenchMark staining instrument (Ventana discovery XT), in which all steps from deparaffinization to counterstaining are performed inside the instrument. Following paraffin embedding, thin sections (4 μm thick) were mounted onto positively charged slides and dried for 24h. The deparaffinization step was done in the Ventana Discovery XT platform using EZ prep solution. Heat-induced epitope retrieval was performed by heating the slides in Tris-EDTA buffer pH 7.8 at 95°C for 40 min. The block of endogenous peroxides was done using the inhibitor CM for 4min. Incubation with the primary antibody was done for 30 min. Slides were then incubated with OmniMap anti-Rb HRP for 15 min and OmniMap anti-Ms HRP for 15 min, respectively. The antigen-antibody complexes were then revealed with DAB CM and H_2O_2 CM for 8 min. Finally, slides were counterstain with Hematoxylin for 8 min and coverslipped in permanent mounting media. Two pathologists (A.R and N.F) independently evaluated the percentages of cells that stained positive for each marker. The IHC profile was evaluated in triplicate sections and categorized on the fraction of stained cells as "high >75%", "medium 25-75%", "low <25%" or "not detected 0%". Staining localization at the nucleus, cytoplasm, or membranes was also collected. The expression pattern of CEP250 at the centrosome was stratified as reported.[41] The working dilutions and primary antibodies employed for IHC are reported in (Table S4).

Multiplexed immunofluorescence analysis and confocal microscopy

A multiplex method was employed to quantify centrosomes at the level of single cells in human tissue samples, as reported.[42] FFPE blocks were sectioned at 5-μm thickness and mounted on slides. Subsequently, slides were deparaffined by washing in xylene three times, 7 min each, followed by passing through 100%, 75%, and 50% ethanol and ddH2O for rehydration. Antigen retrieval was performed using the Universal HIER antigen retrieval reagent, 10X, diluted 1:10 (ab208572, abcam) and heated at 97°C using a decloaking chamber for 20 min. Slides were washed in washing buffer (PBS and 0.1% Tween 20 at 25°C) following by blocking buffer (5% normal goat serum, 0.1 M Tris-HCl, and 0.15 M NaCl, pH 7.6 at 25°C) for 30minat room temperature (RT). Primary antibodies were diluted in blocking buffer and incubated at 4°C overnight in a

humidified chamber as follows: Anti-CEP250 (Proteintech Europe, 66814-1); diluted 1:100; Anti-BUD31 (Cusabio Italy, CSB-PA01749A0Rb), diluted 1:200; Anti-BCAS2 Rabbit Polyclonal Antibody (CliniSciences Italy, A4398-20); diluted 1:200; Anti-DHX35 Rabbit Polyclonal Antibody (Cusabio Italy, CSB-PA887985LA01HU), diluted 1:100. Slides were washed three times in washing buffer and incubated with secondary antibody and Hoechst, 33342 Staining Dye Solution; diluted 1:10000 (ab228551, abcam) for 30 min to 1h at RT. Secondary antibodies were: Goat Anti-Mouse IgG H&L (Alexa Fluor® 647, red) (abcam, ab150115) and Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488, green) (ab150077) diluted 1:500 or 1:1000. Slides were washed three times in washing buffer, mounted using Anti-Fade Fluorescence Mounting Medium - Aqueous, Fluoroshield (ab104135, abcam) and stored at –20°C for future analysis. Unstained control samples were run to check and exclude autofluorescence. An anti-ribosomal protein L12 (RPL12) rabbit Polyclonal Antibody (DUOTECH Italy, HPA003403), diluted 1:200, was used to mark ribosomes for comparison to spliceosome components. Specimens were imaged using a Confocal Microscope (Nikon AXR system, 60× objective) equipped with Ti2-E Inverted Microscope 25 mm FOV High speed Resonant scanner and High-resolution Galvano scanner. High resolution images were acquired with NIS-Elements C Software (Nikon). Tissue-based protein colocalization at centrosomes/cilia was estimated according to Pearson's correlation coefficient as reported.[43] At least 100 cells in ten different high-magnifications fields (60x) of tissue sections were estimated.

Co-immunoprecipitation (co-IP) experiments and Immunoblot analysis

The IP were performed on HEK293 and Hela cell lysates at least three times as reported.[13] Briefly, the cells were homogenized in lysis buffer (Tris 50 mM pH 7.9, 1% triton X-100, 0.1% Tween20, 150 mM NaCl, 10% glycerol, 5 mM MgCl2) or in a specific buffer for Co-IP experiments (10 mM Tris-HCl, 1 mM EDTA, 10 mM $MgCl_2$, 5 mM EGTA, 50 μM β-Glycerophosphate,0.5% NP40, 20% Sodium deoxycholate). Lysates were treated with protease inhibitors from Sigma-Aldrich (P8340) and phosphatase inhibitors from Roche (PhosSTOP, 04906837001). For Co-IP experiments, lysates were incubated with specific antibodies and IgG, as control. Western blot (WB) studies were performed in triplicate. For Co-IP experiments, the ratio of IP proteins with respect to the input was 1:50. Blots were quantified by ImageJ. The control was settled as 1 and the fold change was calculated and reported below the panels as the mean ± standard error of the mean (SEM). Polyvinylidene difluoride (PVDF) membranes were used for Immunoblot (Millipore, US, Immobilon-P, IPVH00010) and ECL western blotting reagent (Thermoscientific, 32106) or Femto (Thermoscientific, 34095) were used for detection.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were carried out using Prism version 4.02 (GraphPad Software, Inc), GeneSpring R/Bioconductor v.12.5 and R default packages. The Bonferroni correction was used for multiple comparison tests.

ADDITIONAL RESOURCES

This paper did not create any additional resources.