 doi:10.20944/preprints202202.0084.v1

Spatial and functional mapping of spliceosome and centrosome proximity revealed by Proteogenomics profiling.

Lorella Del Grosso[1,2,] Luigi Cerulo[1,2], Andrea Remo[3], Brunella Franco[4,5] Johan Busselez[6] and Massimo Pancione[2]*

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

[2]Department of Sciences and Technologies, University of Sannio, Benevento, Italy;
[3]Pathology Unit, Mater Salutis Hospital AULSS9, "Scaligera", 37122 Verona, Italy
[4]Telethon Institute of Genetics and Medicine (TIGEM), Via Campi Flegrei, 34, 80078, Pozzuoli, Naples, Italy.
[5]Medical Genetics, Department of Translational Medicine, University of Naples "Federico II", Via Sergio Pansini, 80131, Naples, Italy
[6]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

*Correspondence should be addressed to:

Dr. Massimo Pancione, e-mail: massimo.pancione@unisannio.it Department of Sciences and Technologies, University of Sannio, 82100, Benevento, Italy; Phone: +39 0824 305116; Fax +39 0824 305147

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Abstract:

Ribonucleoproteins (RNP) condensates often contain intrinsically disordered proteins (IDPs) able to confer an exceptional multifunctionality necessary for gene expression, a process that defines cell types and enables cellular adaptation in metazoans. Centrosomes, the microtubule-organizing centers of animal cells, are particularly enriched in disordered proteins but the role of RNPs surrounding the centrosome and the ciliary basal body remains largely unknown. By refining and integrating the existing protein-protein interaction network, spatial proteomics and transcriptomic data, we here report the subcellular and genomic proximity between the spliceosome, a huge nuclear RNP complex that removes introns from a transcribed pre-mRNA, and centrosome/cilia components. We present a comprehensive map of pre-RNA processing factors and other RNA-binding proteins playing a role in the regulation of splicing but localized to the centrosome and cilia. Protein-protein interactions studies reveal that a large number of spliceosome components interact with both centrosome linker and centriolar satellites elements, necessary for cellular division and ciliogenesis. RNAseq data from mouse Embryonic Stem Cells (mESC) and human tissues revealed a co-transcriptional coordination program of splicing and centrosome-related genes with relevance to tissue-specific neurosensory disorders and cancer types. Additionally, we found that centrosome and spliceosome genes form linearly and spatially colocalized genomic loci (*CEP250*, *RBBM39*, *DHX35*, and *CTNNBL1*) conserved in human and mouse genome, then explaining similarities in co-modification and subcellular distribution. Our results suggest that centrosome and cilia constitute cytoplasmic sites for the exchange of molecular machinery with nucleus, storage of RNA splicing and spliceosome condensates previously unrecognized. These complexes in response to external signals, could play an integral part in ciliogenesis and nuclear division to establish and maintain cellular identity in metazoans.

Key words: ribonucleoproteins, splicing, centrosome, CEP250,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

1. Introduction

Ribonucleoprotein (RNP) particles or condensates are RNA-protein assemblies that lack surrounding membranes [1]. Eukaryotic cells contain a variety of RNP granules localized in the nucleus, in the cytoplasm and on membranes. These aggregates function to spatiotemporally organize various biomolecular processes ranging from RNA metabolism and gene regulation linked to memory, development and diseases [1,2]**.** Phase ransitions or liquid–liquid phase separation of bimolecular condensates in living cells can give rise to a restricted set of distinct intracellular compartments, in which evolutionary processes such as the selection and replication of biomolecules can take place [3-5]. Centrosomes are typical biomolecular condensates mainly composed of two microtubule-based barrel-shaped centrioles and a surrounding amorphous network of proteins or pericentriolar material (PCM) able to build and position the mitotic spindle and cilia in metazoans [6,7]. Centriolar satellites are membraneless granules that localize and move around centrosomes and cilia determining a dynamic centrosome- and satellite-associated pools more complex than once thought. Over the past few years, the function of the centrosome as microtubule organizing center and coordinator of the mitotic spindle has been questioned because centrioles are absent in up to half of all known eukaryotic species, and various mechanisms for acentrosomal microtubule nucleation have been described (e.g., plants cells and vertebrate oocytes) [7]. Interestingly, centrosomal proteins are significantly enriched in disordered and coiled-coil regions, more phosphorylated and longer than control proteins of the same organism [8]. The gain of disordered regions in centrosomal proteins is correlated with the increase in cell-types number (phenotypic complexity) raising the intriguing question that the resident proteins my function outside centrosomes and cilia [9]. In animal cells, the two centrioles are kept in close proximity by a machinery of cohesion named “centrosome linker” that persists from G1 until mitotic entry when duplicated centrosomes are disjoined and separated to form the spindle poles [10]. In early mitosis exit, the binding of C-NAP1 (CEP250) to CEP135 at the proximal end of the two centrioles is recognized as one of the first steps of centrosome cohesion assembly [10]. Surprisingly little or nothing is known on the functional importance of such subcellular structure typical of higher eukaryotes [11-13]. Additionally, recent observations have shown that the centrosome and RNP particles including ribosome and Eukaryotic initiation factors (eIFs) necessary for protein biosynthesis are functionally interconnected [13-15]. For example, the centrosomal protein OFD1,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

interacts with components of the Preinitiation complex of translation and modulates the translation of specific mRNA targets in the kidney [14]**.** The large PCM protein pericentrin (PCNT) is also delivered co-translationally to centrosomes during early mitosis [15]. In addition, mutations in RNP components cause pathological phenotypes such as Retinitis pigmentosa or Microcephaly linked to centriole defects [16,17]. Historically it is well-known that RNase treatment impairs the nucleation activity of centrosomes but the underlying molecular mechanisms remain still unknown [18,19]. Consistently, two studies identified a series of RNAs in the centrosome of surf clam (Spisula) oocytes, "centrosomal RNA" which were functionally correlated with nucleolinus, a RNA-rich compartment in the nucleolus [18,19]. However, over the past years, the role of RNP at centrosomes has remained enigmatic and often matter of debate and controversial. Pre-mRNA splicing is catalyzed by the spliceosome, a ribonucleoprotein complex comprised of five small nuclear **(**snRNPs) and numerous proteins [13,16]. Recent studies suggest that many of the metazoan-specific spliceosomal proteins, absent in yeast, might have roles in other molecular machines/biochemical pathways. We here show that spliceosome components constitute a structural and functionally relevant repertoire of centrosome-related molecular networks which defects can collectively explain the pathophysiology of apparently unrelated genetic disorders.

Methods

Datasets

We extracted the existing Protein-Protein interaction network, Proteomic, transcriptomic and genetic data from multiple bioinformatics tools comprising (N=21) databases. The detailed features of which are reported in **(Table S1)** [20-40]**.**

Reconstruction of protein-protein interactome and spatial proteomics

Protein-Protein interaction data of experimentally verified centrosome linker proteins were first analyzed by downloading raw data from The BioGRID database [20]. The potential interacting proteins were mined comparing the proteomic data from isolated centrosomes and integrating the effective relevance of Protein-Protein physical interactions with proteomic data to establish the relationships between key interacting proteins [21-27]. To mitigate the intrinsic inaccuracy and technical variability of protein interactions, the data were verified comparing independent datasets considering as significant interactions having a high Mascot Score >95% or Confidence Score > 0.9. In isolated centrosomes, we also evaluated the temporal change of ribonucleic proteins through the cell cycle progression [23]. In addition, potential interactors were correlated to their spatial subcellular localization using the map of the human proteome. In the light of potential physical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

interactions, candidate proteins were re-examined in the human proteome atlas to verify or not their presence at centrosome/cilia/centriolar satellites complex [21]. Proteins not clearly localized to the centrosome were not considered. The subcellular distributions of proteins encoded by genes of interest was visualized by immunofluorescence and immunohistochemical labelling by downloading high resolution images from stained cell lines and tissues. The Protein score and peptide matrix were hierarchically clustered by Spearman rank correlation (both baits and interactors) [22]. The Dot Plot, Heatmap and network graphical representation of interacting partners were generated through two open source bioinformatics software platform SAINT (significance analysis of interactome) for visualizing the strength of molecular interaction (http://prohitstools.mshri.on.ca/) and http://www.webgestalt.org/ [22,41]. Venn diagram was used to visually define the similarities and differences between datasets and overlapping traits were further analyzed. The functional profiling of potential interactors was explored by DAVID (the Database for Annotation, Visualization and Integrated Discovery) and g:Profiler [42]. Significant GO-Terms (Benjamini-Hochberg adjusted P-value< 0.05 for DAVID; g:SCS threshold< 0.05 for g:Profiler) were isolated. Only GO-Terms that were significant according to both functional annotation tools are reported.

Disorder and Secondary Protein Structure Predictions

Disorder predictions in centrosome proteins were obtained with the DISOPRED2 and FoldIndex programs as previously reported [8]. We tested that the two algorithms have a large overlap with each other and produce qualitatively equivalent results. We considered the longest isoform of each gene and their sequences were extracted from the Ensembl database. Phylogenetic distribution of centrosome genes in animals and fungi was performed by using reciprocal pairwise sequence-based (BLASTP and phmmer) and domain-based (hmmsearch) methods [43]

Cellular Fitness

Fitness scores for gene targets were collected from the Cancer Dependency Map dataset. (Table S1). The dataset consists of CRISPR-Cas9 whole-genome screens to identify dependencies in cancer cells or immortalized cell lines from different human tissues. The results are depicted as a negative fitness effect (the loss of cell viability in the absence of a test gene) or positive fitness effect (no loss of cell viability in the absence of a test gene), with the outcome presented as ‘fitness score’ for each gene in an indicated cell line. Score <0 is considered a statistically significant effect; values are scaled Bayes factors calculated using BAGEL [35]

Gene expression analysis from Human tissues and cell lines

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

The Genotype-Tissue Expression (GTEx) biobank and its associated bioinformatics portal https://www.gtexportal.org/home/ was used to analyze gene expression data from human tissue samples [29]. Gene expression data were collected from 54 non-diseased tissue sites and normalized RNA-seq data are shown as Transcripts Per Million (TPM). Deep RNA-sequencing data were extracted from the Cell Atlas in which the results are reported as normalized NX values. A NX value of 1.0 is defined as a threshold for expression of the corresponding protein. [29] We analyzed separately each gene-category (i.e. Ribosome, Splicing and centrosome) in a representative panel of immortalized and transformed cell lines. To determine the dynamic regulation of gene-category, we analyze the transcriptome data derived from MCF10A cells, an immortalized non-transformed human epithelial cell line, underlying the slow-cycling state [28]. The Relative expression of genes was expressed as log2FoldChange related to control across two different conditions, i.e., spontaneous quiescence vs quiescence induced with CDK4/6 inhibitor. The CDK4/6 are key cyclin-dependent kinases that promote G1 to S phase cell cycle progression and cap-dependent translation during mitosis–G1 (ref).

Single cell transcriptomic data analysis from mouse Embryonic Stem Cells (mESC)

To investigate the relationship between centrosomes and ribonucleic encoding genes during development, we applied the reconstruction of gene regulatory networks (GRNs) analysis using the single cell transcriptional profiles from mouse Embryonic Stem Cells (mESC) [40]. The dataset is made up of three expression matrices, each of which represents a phase cell cycle (G1 S G2-M), whose columns represent the single cell (96) while lines represent the relative expression levels of the genes (38390). We included into the analysis a list of genes sourced from the National Center for Biotechnology Information (NCBI) involved in the formation/regulation of the centrosome (659), transcription factors (2761), ribosomal (159), splicing (1305) and translation initiation (663) genes. The gene categories were extrapolated from the RPG (Ribosomal Protein Gene) database (http://ribosome.med.miyazaki-u.ac.jp) and National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). For the GRNs analysis, we used the inference algorithm SINCERITIES (SINgle CEll Regularized Inference using TIme-stamped Expression profileS) which allows to reconstruct genetic interactions from single cell transcriptional profiles and infers directed gene-gene relationships [45]. GNR inference is formulated taking into account the linear regressions and temporal changes of gene expression distributions. We used the Kolmogorov–Smirnov distance to quantify the distance between two cumulative distribution functions of gene expressions from subsequent time points. The Kolmogorov–Smirnov test (K–S test or KS test) is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). In this case we used the Kolmogorov-Smirnov test to verify the distributive distance of gene expression in successive time points. Sincerities uses Spearman's rank partial correlation analysis. In our study we set the parameter SIGN 0: this parameter does not select if the sign / mode of the gene regulation is inferred.

Ridge Regression, also called Tikhonov regularization, is a regularized version of Linear Regression: by adding a regularization term, commonly referred to as alpha, to the cost function, the learning algorithm is forced to keep the weight as low as possible. Ridge Regression adds a penalty factor to the cost function. This determines the loss of importance of the value of a feature, which, depending on the penalty, can be more or less accentuated. The strength of the penalty is tunable controlled, that is, by a hyperparameter that must be set. Speaking of regularization in general, there are two types of penalties:

L1 (absolute size) penalizes the absolute value of the model coefficients

L2 (squared size) penalizes the square of the value of the model coefficients.

Ridge Regression uses the L2 penalty.

The algorithm input is a list containing the following information:

- ✓ A list of matrices of length (n), where n is the number of capture time points. Each dataframe containing the observed expression levels of the m genes in single s_k cells.
- ✓ S by m matrix, where S is the total number of single cells (i.e., S=s_1+s_2+...+s_n) and m the number of genes.
- ✓ Vector of length n containing the cell capture time points or (time-stamps)

The algorithm is based on the premise 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. Sincerities returns the matrix of adjacencies, that is m by m matrix containing the weights of regulatory edges and DISTANCE_matrix, that is n-1 by m matrix containing the (normalized) distribution distance (DD) computed during the network inference, using linear regression:

$$DD_{j,l+1} = \alpha_{1,j}\widehat{DD_{1,l}} + \alpha_{2,l}\widehat{DD}_{2,l} + \ldots + \alpha_{m,j}\widehat{DD}_{m,l}$$

From the adjacency matrix generated by the Sincerities algorithm, a list was created using the functions of the “Igraph” package. The list of adjacencies, in descending order based on "weight" of the border, is made up of three columns: the first indicates the regulatory genes, the second the target genes and the third indicates the "weight" of the border connecting the two genes. The list generated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

from the Sincerities tool contained 3903843 gene interactions. The graphic representation of the regulatory activity of genes was performed with Cytoscape or SAINT (significance analysis of interactome) [22,41,46]

Disease modeling and mini intron containing genes analysis

The involvement of interacting genes in common disease phenotypes was investigated with WebGestalt [41] using as discriminant factor the enrichment ratio of cluster gene with a False Discovery Rate (FDR) adjusted p-value of 0.05. Transcriptome profiles from patient-derived (PRPF31-mutated) retinal organoids and Fibroblasts were used to investigate the impact of Mis-splicing of genes implicated in centrosome function and ciliogenesis [41]. To further understand the effect of splicing inhibition on centrosome genes expression, multiple whole transcriptome sequencing datasets obtained from cells treated with different splicing inhibitors (sudemycin, spliceostatin A targeting SF3B1) or (indisulam targeting RBM39) were reexamined [37-39]. The therapeutics response portal (https://portals.broadinstitute.org/ctrp/) which links genetic, lineage, and cellular features of cancer cell lines to small-molecule sensitivity was used to find molecules that target the gene clusters of interest. The platform allows to correlate the drug sensitivity to both gene expression levels and copy number variation [36]. The Cancer Genome Atlas (TCGA) database (https ://cance rgeno me.nih.gov/), which integrates gene expression data and clinical data (survival analysis), was used to analyze gene expression profiles of functionally distinct gene categories. TCGA is a landmark cancer genomics program that has molecularly characterized more than 20,000 primary cancers and matched normal samples spanning 33 cancer types. Tumor/normal differential expression analysis were explored using the analysis of variance (ANOVA). The graphical representation of gene expression data and the derived survival curves were imported from GEPIA, a bioinformatics web server for analyzing RNA sequencing expression data across the TCGA. The significance of genes in determining the overall survival was analyzed using the Kaplan–Meier curve. We considered as statistically relevant a *P* value less than 0.05. UALCAN, an interactive web resource for analyzing cancer OMICS and Proteomic data was used to validate the correlation between gene expression data and pathological parameters [31-33]. The cBioPortal database (https ://www.cbiop ortal .org/) which integrates cancer mRNA expression, genomic data, somatic cell mutation and DNA copy number changes allows us to confirm and verify differences in gene expression profiles and changes at genome level. To visualize Stoichiometric relationships of interacting proteins at mRNA and protein level, we used the DIA-expert software for the NCI-60 cell lines available in CellMiner (discover.nci.nih.gov/cellminercdb) [24]. Most eukaryotes contain two

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

types of spliceosomes; a) the canonical spliceosome, also called the major spliceosome which splices major introns, b) the minor spliceosome destined to the splicing of a small subset of introns called minor introns, that have divergent consensus sequences. These minor introns are found in genes that are predominately made up of major introns, and thus the expression of these minor intron-containing genes (MIGs) requires the coordinated action of both the major and the minor spliceosome [24]. To verify if critical interacting centrosome proteins identified in our study can be classified as MIGs and thus subjected to the action of minor spliceosome, we used the database https://midb.pnb.uconn.edu which allows to define the role of minor spliceosome in the regulation of MIG expression [34].

Statistical analysis

The statistical analyses were carried out using Prism version 4.02 (GraphPad Software, Inc), GeneSpring R/Bioconductor v.12.5 and R based package.

2. Results

2.1 Ribonucleoproteins interact with C-NAP1 (CEP250) and centrosome linker proteins

While investigating the protein-protein interaction (PPi) network of experimentally validated centrosome linker proteins (CLPs), we observed that C-NAP1 (CEP250) showed a significantly higher number of interactors compared to other CLPs **(Figure 1A,B).** The majority of C-NAP1 interactors included RNA-binding proteins, constituents of Ribosome, Heterogeneous nuclear ribonucleoproteins (HNRNPs) and RNA helicases **(Figure 1C).** Gene ontology (GO) analysis of the C-NAP1-interactome confirmed that top enriched terms were structural constituents of ribosome/translational initiation (27%) followed by cell-cell adherents junction (11%) and "mRNA splicing" components (9%) **(Figure 1D, Table S2).** Notably, a close connection with splicing components was also identified for other CLPs **(Figure S1A).** Focusing our attention on the interaction between Rootletin (CROCC) and pre-mRNA splicing component MAGOH revealed that the top ranking neighbors of the network were posttranscriptional regulation of gene expression, mRNA 3'-end processing and mRNA transport from the nucleus (Adjusted P-value < 0.001) **(Figure S1B, C)**. Additionally, we also found that a series of RNU genes (RNA, U1 Small Nuclear 2-4) classes and a poorly defined Rootletin interactor (CROCCP2) colocalized on chromosome band 1p36.13. Notably, *CROCCP2* and *CROCC* showed a consistent co-expression and co-regulation in a variety of human cell lines **(Figure S1C,D).** Thus, we asked whether the fraction of predicted disordered proteins across the known CLPs correlated with their ability to interact with a variety of proteins. First, we found that C-NAP1 displayed the largest fraction of predicted disordered residues (**Figure 1E).** In addition, investigation of a larger subset of centrosomal proteins localized at the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

pericentrosomal region, distal appendages and centrioles, showed that the increase in intrinsically disorder domains correlated with a larger interactome **(Figure 1E,F).** Intrinsically disordered proteins are also characterized by high evolvability and it is well-known that Archaea and Bacteria have, on average, lower disorder content than Eukaryota [9]. We thus analyzed the gain and loss of centrosome linker genes across unicellular and multicellular eukaryotic organisms using PCM markers (pericentrin and y-tubulin) for comparison. In contrast to PCM components that were extremely conserved across species, centrosome linker genes were limited to higher eukaryotes supporting previous reports [9] **(Figure 1G).** Therefore, the increase of disordered regions in centrosome linker proteins during the evolution is connected to abundance and variety of protein-protein interactions functionally unrelated to centrosome.

3. *mRNA splicing components selectively interact with centrosomes and satellites*

To further characterize the effective interactions occurring between centrosome and ribonucleoproteins, we integrated the data from Human proteomic and Interactome databases (Human Protein Atlas collection). We first included into PPi analysis, 548 proteins localizing to centrosome, 100 Ribosome/Eukaryotic initiation factors (eIFs) and 79 known splicing factors**.** We build a global map which confirmed a close proximity between the centrosome and ribonucleic proteome **(Figure 2A).** In order to define specific interactions, we included in the analysis a relevant subset of centrosomal partners (comprising centrioles, pericentriolar material, satellites, spindle and cilia) **(Figure 2B).** Reflecting the connections between centrosomes function and translational assembly, proteins known to work in centrosome maturation and nucleation (TUBG1) or centriole biogenesis (CEP57 and CEP76) or 3M complex factors (CUL7) also interacted with 60S and 40S core ribosome proteins indicating an extensive interaction landscape **(Figure 2C)**. Consistently with literature, the centriolar satellite (*OFD1*) and the distal appendages (*NINL*) markers enriched an mRNA translation subnetwork with participation of eIFs (i.e., EIF4ENIF1 and EIF6) and Cyclin-dependent kinase 2 (CDK2) **(Figure 2C)**. Indeed, GO analysis of the EIF4ENIF1 interactome revealed an enrichment of centrosome-related pathways and microtubule organization activity 16% and 18%, respectively **(Figure S2A)**. Additional analysis in form of interacting network confirmed that ciliogenic factors (e.g., OFD1 and FBF1), particularly mapped in closer association to cytosolic eIFs compared to Ribosome constituents **(Figure 3A)**. To establish the strength of the observed interactions, we investigated the relative abundances of eIFs and core Ribosome interacting proteins using as reference 9 centrosomal proteins including C-NAP1 [27] (T**able S1)**. Our analysis showed that several eIFs and core Ribosome proteins interacted with a restrict subset of centrosome components **(Figure 2C and Figure S2C).** For example, while RPL23, a component of the 60S subunit was broadly distributed across different centrosome proteins, others such as RPL19 or RPL26L1 were more selective for C-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

NAP1 and TUBG1, respectively **(Figure 2C)**. By extending the analysis to mRNA splicing factors we found a consistent number of potential interactions mostly with pre-RNA processing factors (PRPFs), RNA helicases, Small Nuclear Ribonucleoproteins and components of exon junction complex (EJC) (MAGOH). Mutations in several of these PRPFs (PRPF4 and PRPF31) and other components, for which cilia/centrosome association is less clear (MAGOH, SNRPD3), cause autosomal dominant forms of Microcephaly and retinitis pigmentosa (RP), a comparatively common inherited retinal blindness often associated with ciliary defects **(Figure 2D)** [10,16,30,47]. Unexpectedly, quantifying the effective abundance of interacting proteins involved in splicing, we found that C-NAP1 (CEP-250) exhibited intense and selective interactions with a large variety of proteins including pre-mRNA-processing factor (i.e. PRPF4); Serine And Arginine Rich Splicing Factors (i.e SFRS3, SFRS7), Spliceosome assembly RNA helicases (i.e. DDX24) and ubiquitously expressed HNRNPs (i.e. HNRNPM) **(Figure 2D and Figure S2D).** Overall, the data indicated that C-NAP1 could selectively interact with splicing components. To further determine that our approach was reliable, we collected the interactome data derived from the centrosome/cilia complex **(Gupta GD et al. 2015, Table S1).** GO analysis involving the globality of interactions (6000) showed that RNA binding and regulation of mRNA export from the nucleus were among the top 10 Enriched categories **(Figure S3A).** We next analyzed 10 centrosomal proteins reasoning that High-confidence interactors (n=418, FDR ~1%) validated by co-immunoprecipitation (coIP)/MS would provide more accurate results on the effective binding partners **(Figure 3A).** Notably, we found absence of Ribosome proteins and a predominant enrichment for mRNA splicing via spliceosome distributed broadly in ciliated and non-ciliated conditions **(Figure S3B).** Measuring the abundance of interacting partners, we found that CEP135, a well-established C-NAP1 interactor and SSX2IP, a crucial ciliogenic and mitotic spindle factor, were part of an extensive interaction landscape with 26 splicing factors **(Figure 3A).** Based on their distribution, centrosome-interactors enriched sub-compartment and functions involved in inter-centriole cohesion and centriolar satellites **(Figure 3A).** The data thus indicate that splicing components may indeed represent bona fide centrosome interactors.

4. Subcellular and spatial localization od splicing factors to the centrosome and cilia components

In order to better define the subcellular localization and the intimal relation of these Ribonucleoproteins with centrosome, we analyzed proteomic data generated from isolated centrosomes coupled to intracellular localization labeling data deriving from the Human proteome Atlas. We first identified a subset of highly significant proteins (735 out of 1452) expressed in the isolated centrosomes from lamb thymocytes [26]. GO analysis revealed that the top highest

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

enrichment term was represented by ribonucleoproteins **(Figure S3C).** Focusing on the distribution and type of identified ribonucleoproteins, we found that a subset of core Ribosome proteins and mRNA splicing components were highly represented **(Figure S3D).** Notably some components including RPS11, SRSF7 and PRP19 corresponded to high-confident centrosome interactors showing thus a good concordance with quantitative PPi data **(Figure 2D,E and Figure S3D).** The centrosome structure and composition vary considerably during the cell cycle; For example, quiescent cells could differ considerably from the proliferating ones. To explore this question, we examined the abundance and distribution of ribonucleoproteins during S and M phases from isolated centrosomes in HeLa S3 cells **(Kimura, 2012, Table S1).** Despite the similar amounts of EIFs and 40S/60S Ribosome proteins at M and S phases, mRNA splicing proteins remarkably increased 5-fold in centrosomes during M phase compared to the S phase **(Figure 3B).** These results suggested that splicing factors could participate in both centrosome division at spindle pole and ciliogenesis possibly playing a key a role in centriole-based function and centriolar satellites [48]. To further refine the identified protein-protein spatial interactions, we explored the comprehensive set of proteins distributed to centrosome (n=548) integrating image-based map from the Human proteome Atlas and interaction network**.** First, we noted that similarly to other subcellular structures, centrosomes share a strikingly high number of multilocalizing proteins with the nucleus **(Figure S4A,B).** Next, we asked if the effective subcellular distribution of different ribonucleoproteins spatially correlated with image-based centrosomes staining. While EIFs and core Ribosome proteins represented less than 1% of the total centrosomal proteome, notably 13 proteins potentially involved in mRNA splicing functions were identified (**Figure 3C, Figure S4C,D and Figure S5A).** One of these factors, WDR83 exhibited an intense centrosome labelling strengthening its interaction network with multiple centrosome proteins (**Figure 3A, 3D).** In addition, we found that a number of splicing proteins localized at interphase centrosomes including UPF1, DHX35, RBM39 or others, such as DDX53, marked centrosomes at each spindle pole but not in interphase **(Figure S5A-C)**. To understand the functional significance of our observations, we examined the cell viability or cellular fitness upon genetic perturbation or loss (Achilles project) of each gene of the interacting network (n=53) using as model the telomerase-immortalized RPE1 cell line hTERT-RPE1, a popular cellular model to study primary cilia. Core Ribosome and EIFs encoding genes were used for comparison. Almost all centrosomal genes screened did not result in loss of cell fitness. As expected, and strong dependency upon deletion of genes encoding for Ribosome, EIFs and splicing components was observed. Notably, with the exception of EIF4ENIF1, DDX11 and DHX35 the majority of splicing components interacting with or localizing at centrosome resulted in loss of cellular viability **(Figure S5C).** Therefore, the identified splicing factors could be critical executors of centrosome-related functions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

5. Genes encoding for splicing and centrosome constituents are co-regulated in a tissue-specific manner.

Next, we asked if the detected spatial and physical interplay between splicing and centrosomes could also have relevance at mRNA expression level. We use a Genotype-Tissue Expression (GTEx) platform to survey their differential expression across 54 non-diseased Human tissues **(Table S1).** While genes encoding for ribosome proteins were highly expressed and slightly varied among different tissues, splicing, EIFs and centrosomal genes were coordinately down-regulated in a tissue-specific manner in Heart, Pancreas, Brain and especially in Liver **(Figure S6A).** For some markers, we also found a good concordance between gene and protein expression comparing Liver vs Testis by immunohistochemistry (IHC) data derived from the Human proteome atlas **(Figure S6B).** Liver and heart have recently been proposed to work without a functioning minor spliceosome which likely operate on the so-called mini-intron containing genes (MIGs) promoting tissue-dependent retention and alternative splicing of minor introns [34]. Indeed, we found that ciliogenic factors belonging to MIGs (i.e. OFD1, TCTN3, CEP170) formed a protein-protein network with mRNA splicing regulators and mRNA transport from the nucleus indicative of a possible co-regulation trough cotranslational assembly and alternative splicing **(Figure S6C).** To ensure our observations are not simply explained by a species-specific effect, we collected single cell (scRNA-seq) data from mouse embryonic stem cells (mESCs) and analyzed them using the Sincerities algorithm [34,45]. We clustered 2786 genes into functional categories including centrosome, ribosomal, EIFs and splicing. By this approach, we identified a consistent number of annotated transcripts corresponding to 2420653 regulatory interactions. We focused on the top-ranked 2000 with greater bow weight. The analysis revealed that several centrosomal/cilia genes were highly interconnected with splicing genes supporting the existence of a robust and co-regulated transcriptional network also in murine embryonal cells**.** Globally, FBF1 and ODF1 were the most prominent regulatory genes. In fact, ~50% of interactors was constituted by genes encoding for splicing components **(Figure 4A).** Notably, Ofd1 was highly and selectively interconnected with a variety of splicing genes some of which found to be localized in human centrosomes **(Figure 4B).** These results are in line with recent studies showing that FBF1 and OFD1 could have a key role in modulating the translation of specific transcripts nearby centrosomes or integrating different centrosomal functions such as ciliogenesis and cell-cycle progression [14, 49]. To further clarify the possible genetic co-regulation, we used a public RNAseq dataset from human immortalized cells induced in spontaneous or CDK4/6 inhibitor-induced quiescence [28]. Also in this experimental setting, splicing and EIFs components show a comparable

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

and coordinated low gene expression profile in response to spontaneous or forced quiescence with the CDK inhibitor **(Figure 4C).** Altogether, these data suggest a close regulatory interdependence between centrosome and splicing encoding at transcriptional level.

6. Centrosome and splicing loci tend to be co-altered in non random fashion in human diseases

Alterations in spliceosome components cause often phenotypic abnormalities associated with centriole defects [28]. Our network of genes encoding for centrosome/cilia and ribonucleoprotein indeed indicated a number of overlapping diseases mostly related to neurodevelopment and neurosensory disorders **(Figure 4D).** For example, about 15% of the genetic causes of Retinitis Pigmentosa (RP) involves PRPFs encoding genes and recently also alterations in *CEP250* have been shown to result in RP in humans and in mouse models [16,47]. We therefore analyzed if aberrant splicing derived from mutations in *PRPF31* affected centrosomal genes [30]. Notably, we found that *CEP250* along with a subset of 17 centrosomal genes represents a subnetwork subjected to mis-splicing regardless of the cell type **(Figure 4E).** Next, we interrogated mRNA expression profiles across TCGA data using as reference non-neoplastic tissues. Some diseases, such as low-grade glioma (LGG) Glioblastoma multiforme (GBM), thymoma (THYM) exhibit concordant and high expression levels of 60S and 40S ribosomal proteins expressing genes. In contrast, we observed an elevated degree of variability for splicing and centrosome/cilia genes **(Figure 5A).** In contrast, liver hepatocellular carcinoma (LIHC) and cholangiocarcinoma (CHOL) exhibited a coordinated over-expression of these gene clusters then correlating with a shorter overall survival **(Figure 5B and Figure S7A,B).** We also found a high degree of concordance with independent IHC data. Compared to a variety of different cancers, LIHCs exhibited an intense staining for PRPF3 and OFD1 **(Figure 5B).** Notably, *CEP250* was the only linker gene to be commonly overexpressed in a variety of cancers and correlated with high CA20 value, a score used to measure the degree of 20 centrosome amplification-associated genes **(Figure 5A)** [50]. Moreover, we also found that the top-neighbor interactors "*XPO1* and *SSX2IP"* of the network were highly overexpressed mirroring the behavior of gene signature [22] **(Figure S7C).** Interestingly, the IHC data revealed that XPO1 was delocalized in the cytoplasm of cancer cells related to non-neoplastic hepatocytes indicating also defects in nucleo-cytoplasmic transport **(Figure 5C).** In order to better understand the origin of these relationships, we investigated the global set of genes co-expressed with *CEP250* using the RNA-seq data from the cancer cell line encyclopedia (CCLE) and NCI60 **(Table S1).** Focusing the attention on the Top 50 co-expressed genes, we found that a subset of mRNA splicing genes (CTNNBL1**,** RBM39, DHX35, RALY**)** and EIFs, in particular EIF6, were highly correlated **(Figure 5D).** Notably, the identified cluster represent a genetic module localized on human chromosome 20 and mouse Chromosome 2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

indicative of evolutionary proximity **(Figure 6A).** Consistently, whole chromosomal duplication events in Breast Ductal carcinoma and a subset of metastatic prostate cancers were strongly correlated with overexpression of this module **(Figure S8A,B).** These combinations could thus explain the strong co-modification observed across independent biological samples. These results reveal thus that apparently unrelated genes can form linearly colocalized and spatially colocalized loci which tend to have similarities in terms of co-expression, co-modification, and subcellular distribution in specific tissues.

7. **Splicing inhibitors elicit a selective repression of splicing and centrosome genes**

Having observed a nonrandom gene coregulation, we screened (https://portals.broadinstitute.org/ctrp/), a publicly available online tool that allows the extensive genetic characterizations of cancer dependencies with small molecules. Unexpectedly, we found that the molecule referred to as Indisulam, a splicing modulator that induces the proteasomal degradation of RBM39 [37] targeted similarly the gene module (CTNNBL1**,** RBM39, DHX35 and CEP250) strengthen the hypothesis that these genes interact with each other in a network module **(Figure 6B).** To further analyze the effect of splicing inhibition, we took advantage from recent RNA-seq data obtained from selective RBM39 inhibition and identification of RBM39 interacting RNAs in acute myeloid leukemia (AML) [37]. In RBM39-depleted cells, we found that about 2% of centrosome genes undergo aberrant splicing with a prevalence of exon skipping affecting 53 centrosome encoding genes including CEP250 and DHX35 **(Figure 6C).** Among the direct RBM39 mRNA-binding targets (Top 500 of 9560; score >4.5), we observed a relevant number of mRNAs encoding for centrosome-related genes (17/548) and notably GO analysis revealed that the Cytoskeleton and Microtubule binding was one of the most enriched terms **(Figure S8C).** To further examine the effects of splicing inhibition on centrosome-related genes, we analyzed transcriptomic data from Rh18 and HeLa cells following exposure to sudemycin D1 and spliceostatin A, respectively [38,39]. These compounds selectively bind to the essential spliceosome component SF3b, a subcomplex of the U2 snRNP, to inhibit pre-mRNA splicing. Overall, both the inhibitors displayed a concordant profile resulting in up-regulation of protein translation (encoding mostly for ribosome proteins) and splicing genes. Interestingly, DHX35, or different HNRNPs encoding genes such as HNRNPU and HNRNPM were commonly downregulated and underwent alternative splicing events following treatment with both drugs **(Figure 6C, D).** These changes were accompanied by a significant downregulation of centrosomal/ciliogenic genes as compared to those undergoing up-regulation (70-80% versus 30%-20%). However, a remarkable fraction of centrosomal genes decreased in the cytoplasm (54) and nucleus (74) following spliceostatin A treatment and only 7 genes were downregulated in both fractions **(Figure 6C).** For example, PCNT was downregulated in both cytoplasm and nucleus

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

following treatment with spliceostatin A in keeping with the evidence that *PCNT* mRNA is enriched and translated near centrosomes in the cytoplasm. Mis-splicing and in particular exon skipping was identified in many centrosomal genes following treatment with sudemycin D1 consistent with the literature [38,39] **(Figure S8D).** Focusing the attention on the genes undergoing differentially spliced exon junctions that potentially led out-of-frame transcripts upon sudemycin D1 treatment, we identified 58 centrosomal genes (3%) of 1739 total. Notably, the cluster undergoing aberrant splicing events included the colocalized gene pairs *CEP250*, *RBM39* and *CTNNBL1* involved in splicing and centrosome functions **(Figure 6D).** Overall, these results suggest that inhibition of splicing can indeed modulate the expression profiles of centrosome and cilia encoding genes.

8. Discussion

In the present article, we study the existing protein-protein interaction networks, spatial proteomics and trascriptomic data and present for the first time a comprehensive map of proteogenomic interactions between the spliceosomal and centrosomal components. Pre-mRNA splicing is carried out by a complex machinery called spliceosome composed by more than 200 proteins comprising core and regulatory elements. The spliceosome assembles on a pre-mRNA and undergoes numerous structural and compositional rearrangements during its assembly, activation, and catalytic activity. 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 [38,39]. Human spliceosomes contain numerous proteins absent in yeast, whose functions remain largely unknown [34, 38,39,51]. In line with this, many centrosome constituents tend to increase with the number of cell-types during evolution and the numerous disorder regions make them extremely flexible and able to interact with many partners. Consistently, our protein-protein interaction network study identified Ribosome and spliceosome components as the most abundant interactors of C-NAP1 (CEP250). The prominent high disorder of C-NAP1 may additionally provide a higher degree of plasticity and metazoan-specific protein-protein interactions to the machinery of cohesion, typical of centrosomes. Exploring high confident and stoichiometrically abundant interactions, allowed us to establish constant and selective associations between splicing factors and centrosome markers in which emerged centrosome cohesion components such as CEP135, C-NAP1, FBF1 and other crucial regulators involved in centriolar satellites (OFD1) function and ciliogenesis (SSX2IP). Image-based map of sub-cellular protein distribution and proteomes from isolated centrosomes revealed that a series of proteins involved in splicing were distributed to centrosomes and some of them refined the protein-protein interaction networks identified by restricting protein localization and interaction to the same organelle. The identified proteins belonged to distinct spliceosomal subcomplexes, including

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

components of early assembly complexes (SR proteins and mRNP) A and B including components of U1, U2, U4/U6 snRNP and Prp19 complex, and also potential components of catalytic C complexes. The presence of multiple subunits of the same intermediate complexes reinforced the idea of the spatial and functional interaction between these components **(Figure 7).** Notably, we found that the cellular spliceosome components were remarkably enhanced during M phase compared to S phase in isolated centrosomes indicating a high degree of spatiotemporally regulated functions related to this cellular compartment. Proteins involved in splicing and centrosomal functions are apparently unrelated. However, recent observations suggest that spliceosomal proteins, specifically a group of pre-RNA processing factors (PRPFs) that are expected to localize to nuclei, also localize to the ciliary basal body or the centrosome and promote ciliogenesis. However, whether localization of these ribonucleoproteins has any direct relevance for centrosome biology still needs to be determined. The analysis of transcriptomic data in mouse Embryonic Stem Cells (mESC), and across human tissues revealed a tissue-specific co-transcriptional coordination, a phenomenon observed also when perturbing cell-cycle transition in cultured human cells. Consistently, a recent study showed that Exon Junction Complexes (EJCs)-dependent mRNA trafficking towards centrosomes and basal bodies might contribute to proper mouse neural stem cell (mNSC) division and brain development [34, 38,39,51]. Transcriptomic profiles from PRPF31-mutated Retinitis Pigmentosa cells revealed that alteration in mRNA splicing programs heavily affected the genes implicated in ciliogenesis including *CEP250* [10,16,30, 47]. Similarly, the study across cancer samples showed a coordinated tissue-specific alteration of splicing and ciliogenic transcription program in selective tissues particularly in cancers originating in liver [34]. Although the disease mechanisms related to splicing and centrosome factors remain unclear, some tissues (particularly brain and liver) in which mRNA processing (PRPF3, PRPF4) and centrosome/ciliogenic genes (CROCC, CEP250) poorly expressed could be more susceptible to their deficiencies or overexpression. Interestingly, the fact that a subset of genes displaying a role in splicing and centrosomal function (CEP250, DHX35, CTNNBL1, RBM39 and RALY) are linearly and spatially colocalized on human chromosome 20 and mouse Chromosome 2 might explain, at least in part, why these different genes tend to have a similar co-expression, co-modification, and subcellular distribution pattern [53]. The strong co-modification of the gene cluster was also seen analyzing transcriptomic data derived from selective inhibition of splicing components. Today, the importance of the centrosome as microtubule organizing center and coordinator of the mitotic spindle is questioned and our knowledge of how centrosomes integrate spatial and temporal cues during interphase is limited. Therefore, why do spliceosome and centrosome proteins interact selectively and what is the purpose of those physical and spatial interactions? At present, we do not have a definitive

answer but several lines of evidence argue for a variety of functions centrosomal proteins may have outside centrioles and cilia. First, these spatial and physical relationships could arise from large and single condensates of ribonucleoparticles containing RNA–proteins and subsequently transported outside nucleus towards centrosomes. Consistent with a role of splicing complexes in the earliest steps in ciliogenesis, the EJC complex which forms at the junction of two exons is deposited by the nuclear-splicing machinery and accumulate around and at the base of cilia [34, 55-57]. These EJC proteins revealed an enrichment of untranslated or partially translated EJC-bound specific transcripts (e.g. NIN) necessary for ciliogenesis at centrosomes. A variety of kinases including NEK2 which disassembles C-NAP1 at the onset of mitosis have been reported to influence splicing [54]. We here propose that higher eukaryotes might require the proximity and co-regulation of splicing factors to regulate the expression of key centrosome transcripts to arrange specific condensates necessary for packaging and transport of RNA–protein complexes out of the nucleus. However, the precise role and mechanistic detail of how the spliceosome and EJC complex operate to the centrosome and cilia remains largely unknown. Thus, our study open an exciting area of future research on the dynamic compartmentalization of the RNP condensates in the cytosol in relation to the cell cycle and cilia-mediated signal transduction [55]. Many challenges remain in studying the native spliceosome and centrosome targets in a cell-free context at the level of chemical complexity (protein composition, variability and cellular modifications) and structural-dependent properties (i.e., RNA–protein interactions). In summary, we provide evidence that selective spliceosome components may be relevant executors of centrosome-related functions and that in collaboration with proteins that contain intrinsically disordered regions contribute to gene expression for cellular differentiation and development trajectories in metazoans.

Acknowledgments: This work was supported by the, FFABR grant n.4982, from the Italian Ministry of University and Research (MIUR) to M.P

References

1. Kaur, T.; Alshareedah, I.; Wang, W.; Ngo, J.; Moosa, M.M.; Banerjee, P.R. Molecular Crowding Tunes Material States of Ribonucleoprotein Condensates. **Biomolecules** 2019, **9**, 71. https://doi.org/10.3390/biom9020071
2. Alberti, S. The wisdom of crowds: Regulating cell function through condensed states of living matter. *J. Cell Sci.* 2017, *130*, 2789–2796. [Google Scholar] [CrossRef] [PubMed]
3. Tauber D, Tauber G, Parker R. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. Trends Biochem Sci. 2020 Sep;45(9):764-778.

4. Uversky, V.N.; Kuznetsova, I.M.; Turoverov, K.K.; Zaslavsky, B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.* 2015, *589*, 15–22.
5. 15. Uversky VN. Supramolecular Fuzziness of Intracellular Liquid Droplets: Liquid-Liquid Phase Transitions, Membrane-Less Organelles, and Intrinsic Disorder. Molecules. 2019 Sep 7;24(18):3265.
6. Woodruff JB, Ferreira Gomes B, Widlund PO, Mahamid J, Honigmann A, Hyman The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. AA. Cell. 2017 Jun 1;169(6):1066-1077.e10. doi: 10.1016/j.cell.2017.05.028.
7. 4. Rale, M.J. et al. (2018) Phase Transitioning the Centrosome into a Microtubule Nucleator. Biochemistry 57, 30–37
8. Dos Santos HG et al. Structure and non-structure of centrosomal proteins. PLoS One. 2013 May 9;8(5):e62633
9. Nido GS, Méndez R, Pascual-García A, Abia D, Bastolla U. Protein disorder in the centrosome correlates with complexity in cell types number. Mol Biosyst. 2012 Jan;8(1):353-67.
10. Remo A, Li X, Schiebel E, Pancione M. The Centrosome Linker and Its Role in Cancer and Genetic Disorders. Trends Mol Med. 2020 Apr;26(4):380-393.
11. Mazo, G. et al. Spatial Control of Primary Ciliogenesis by Subdistal Appendages Alters Sensation-Associated Properties of Cilia. Dev. Cell 39, 424–437 (2016).
12. Hata, S. et al. (2019) The balance between KIFC3 and EG5 tetrameric kinesins controls the onset of mitotic spindle assembly. Nat. Cell Biol. 21, 1138–1151
13. Ito KK, Watanabe K, Kitagawa D. The Emerging Role of ncRNAs and RNA-Binding Proteins in Mitotic Apparatus Formation. Noncoding RNA. 2020 Mar 20;6(1):13.
14. Iaconis, D., Monti, M., Renda, M. *et al.* The centrosomal OFD1 protein interacts with the translation machinery and regulates the synthesis of specific targets. *Sci Rep* **7,** 1224 (2017).
15. Sepulveda, G. et al. Co-translational protein targeting facilitates centrosomal recruitment of PCNT during centrosome maturation in vertebrates. Elife 7, (2018) e34959
16. Griffin C, Saint-Jeannet JP. Spliceosomopathies: Diseases and mechanisms. Dev Dyn. 2020 Jun 7.
17. Paolini, N.A. et al. A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. Am. J. Hum. Genet. 100, 506–522 (2017).
18. Alliegro, M.C.; Alliegro, M.A.; Palazzo, R.E. Centrosome-associated RNA in surf clam oocytes. Proc. Natl. Acad. Sci. USA USA 2006, 103, 9034–9038. [CrossRef]
19. Alliegro, M.A.; Henry, J.J.; Alliegro, M.C. Rediscovery of the nucleolinus, a dynamic RNA-rich organelle associated with the nucleolus, spindle, and centrosomes. Proc. Natl. Acad. Sci. USA 2010, 107, 13718–13723.
20. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. 2021 Jan;30(1):187-200. doi: 10.1002/pro.3978. Epub 2020 Nov 23. PMID: 33070389; PMCID: PMC7737760.
21. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Sci. 2018 Jan;27(1):233-244. doi: 10.1002/pro.3307. Epub 2017 Oct 10. PMID: 28940711; PMCID: PMC5734309.
22. Gupta GD, Coyaud É, Gonçalves J, Mojarad BA, Liu Y, Wu Q, Gheiratmand L, Comartin D, Tkach JM, Cheung SW, Bashkurov M, Hasegan M, Knight JD, Lin ZY, Schueler M, Hildebrandt F, Moffat J, Gingras AC, Raught B, Pelletier L. A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. Cell. 2015 Dec 3;163(6):1484-99. doi: 10.1016/j.cell.2015.10.065. PMID: 26638075; PMCID: PMC5089374.
23. Kimura H, Miki Y, Nakanishi A. Centrosomes at M phase act as a scaffold for the accumulation of intracellular ubiquitinated proteins. Cell Cycle. 2014;13(12):1928-37. doi: 10.4161/cc.28896. Epub 2014 Apr 17. PMID: 24743317; PMCID: PMC4111756.
24. Guo T, Luna A, Rajapakse VN, Koh CC, Wu Z, Liu W, Sun Y, Gao H, Menden MP, Xu C, Calzone L, Martignetti L, Auwerx C, Buljan M, Banaei-Esfahani A, Ori A, Iskar M, Gillet L, Bi R, Zhang J, Zhang H, Yu C, Zhong Q, Varma S, Schmitt U, Qiu P, Zhang Q, Zhu Y, Wild PJ, Garnett MJ, Bork P, Beck M, Liu K, Saez-Rodriguez J, Elloumi F, Reinhold WC, Sander C, Pommier Y, Aebersold R. Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. iScience. 2019 Nov 22;21:664-680. doi: 10.1016/j.isci.2019.10.059. Epub 2019 Oct 31. PMID: 31733513; PMCID: PMC6889472.
25. Jakobsen L, Vanselow K, Skogs M, Toyoda Y, Lundberg E, Poser I, Falkenby LG, Bennetzen M, Westendorf J, Nigg EA, Uhlen M, Hyman AA, Andersen JS. Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. EMBO J. 2011 Apr 20;30(8):1520-35. doi: 10.1038/emboj.2011.63. Epub 2011 Mar 11. PMID: 21399614; PMCID: PMC3102290.
26. Busselez J, Chichón FJ, Rodríguez MJ, Alpízar A, Gharbi SI, Franch M, Melero R, Paradela A, Carrascosa JL, Carazo JM. Cryo-Electron Tomography and Proteomics studies of centrosomes from differentiated quiescent thymocytes. Sci Rep. 2019 May 10;9(1):7187. doi: 10.1038/s41598-019-43338-9. PMID: 31076588; PMCID: PMC6510768.

27. Fogeron ML, Müller H, Schade S, Dreher F, Lehmann V, Kühnel A, Scholz AK, Kashofer K, Zerck A, Fauler B, Lurz R, Herwig R, Zatloukal K, Lehrach H, Gobom J, Nordhoff E, Lange BM. LGALS3BP regulates centriole biogenesis and centrosome hypertrophy in cancer cells. Nat Commun. 2013;4:1531. doi: 10.1038/ncomms2517. PMID: 23443559.
28. Min M, Spencer SL. Spontaneously slow-cycling subpopulations of human cells originate from activation of stress-response pathways. PLoS Biol. 2019 Mar 13;17(3):e3000178. doi: 10.1371/journal.pbio.3000178. PMID: 30865623; PMCID: PMC6433297.
29. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580-5. doi: 10.1038/ng.2653. PMID: 23715323; PMCID: PMC4010069.
30. Buskin A, Zhu L, Chichagova V, Basu B, Mozaffari-Jovin S, Dolan D, Droop A, Collin J, Bronstein R, Mehrotra S, Farkas M, Hilgen G, White K, Pan KT, Treumann A, Hallam D, Bialas K, Chung G, Mellough C, Ding Y, Krasnogor N, Przyborski S, Zwolinski S, Al-Aama J, Alharthi S, Xu Y, Wheway G, Szymanska K, McKibbin M, Inglehearn CF, Elliott DJ, Lindsay S, Ali RR, Steel DH, Armstrong L, Sernagor E, Urlaub H, Pierce E, Lührmann R, Grellscheid SN, Johnson CA, Lako M. Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. Nat Commun. 2018 Oct 12;9(1):4234. doi: 10.1038/s41467-018-06448-y. PMID: 30315276; PMCID: PMC6185938.
31. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017 Jul 3;45(W1):W98-W102. doi: 10.1093/nar/gkx247. PMID: 28407145; PMCID: PMC5570223.
32. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017 Aug;19(8):649-658. doi: 10.1016/j.neo.2017.05.002. Epub 2017 Jul 18. PMID: 28732212; PMCID: PMC5516091.
33. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4. doi: 10.1158/2159-8290.CD-12-0095. Erratum in: Cancer Discov. 2012 Oct;2(10):960. PMID: 22588877; PMCID: PMC3956037.
34. Olthof, A.M., Hyatt, K.C. & Kanadia, R.N. Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics* **20,** 686 (2019). https://doi.org/10.1186/s12864-019-6046-x
35. Dempster, J.M., Pacini, C., Pantel, S. *et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nat Commun* **10,** 5817 (2019). https://doi.org/10.1038/s41467-019-13805-y
36. Basu, Bodycombe, Cheah, *et al.*, An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules" *Cell*, **154**, 1151-1161 (2013)
37. Wang E, Lu SX, Pastore A, Chen X, Imig J, Chun-Wei Lee S, Hockemeyer K, Ghebrechristos YE, Yoshimi A, Inoue D, Ki M, Cho H, Bitner L, Kloetgen A, Lin KT, Uehara T, Owa T, Tibes R, Krainer AR, Abdel-Wahab O, Aifantis I. Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. Cancer Cell. 2019 Mar 18;35(3):369-384.e7. doi: 10.1016/j.ccell.2019.01.010. Epub 2019 Feb 21. PMID: 30799057; PMCID: PMC6424627.
38. Yoshimoto R, Kaida D, Furuno M, Burroughs AM, Noma S, Suzuki H, Kawamura Y, Hayashizaki Y, Mayeda A, Yoshida M. 2017. Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA 23: 47–57.
39. Wu G, Fan L, Edmonson MN, Shaw T, Boggs K, Easton J, Rusch MC, Webb TR, Zhang J, Potter PM. Inhibition of SF3B1 by molecules targeting the spliceosome results in massive aberrant exon skipping. RNA. 2018 Aug;24(8):1056-1066. doi: 10.1261/rna.065383.117. Epub 2018 May 29. Erratum in: RNA. 2018 Dec;24(12):1886. PMID: 29844105; PMCID: PMC6049506.
40. Buettner, F., Natarajan, K., Casale, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **33,** 155–160 (2015). https://doi.org/10.1038/nbt.3102
41. Yuxing Liao, Jing Wang, Eric J Jaehnig, Zhiao Shi, Bing Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, Nucleic Acids Research**,** Volume 47, Issue W1, 02 July 2019, Pages W199–W205.
42. Uku Raudvere, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, Jaak Vilo: g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update) Nucleic Acids Research 2019; doi:10.1093/nar/gkz369 [PDF].
43. Finn RD, Clements J, Arndt W, et al. HMMER web server: 2015 update. Nucleic Acids Research. 2015 Jul;43(W1):W30-8. DOI: 10.1093/nar/gkv397.

44. Álvarez-Fernández M, Malumbres M. Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition. Cancer Cell. 2020 Apr 13;37(4):514-529. doi: 10.1016/j.ccell.2020.03.010. PMID: 32289274
45. Papili Gao N, Ud-Dean SMM, Gandrillon O, Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics. 2018 Jan 15;34(2):258-266. doi: 10.1093/bioinformatics/btx575. PMID: 28968704; PMCID: PMC5860204.
46. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 2003 Nov; 13(11):2498-504
47. Huang XF, et al. Functional characterization of CEP250 variant identified in nonsyndromic retinitis pigmentosa. Hum Mutat 2019; 40: 1039-1045
48. Hofmann JC, Husedzinovic A, Gruss OJ. The function of spliceosome components in open mitosis. *Nucleus*. 2010;1(6):447-459.
49. Inoko A, Yano T, Miyamoto T, Matsuura S, Kiyono T, Goshima N, Inagaki M, Hayashi Y. Albatross/FBF1 contributes to both centriole duplication and centrosome separation. Genes Cells. 2018 Dec;23(12):1023-1042. doi: 10.1111/gtc.12648. Epub 2018 Nov 13. PMID: 30318703.
50. Ogden A, Rida PC, Aneja R. Prognostic value of CA20, a score based on centrosome amplification-associated genes, in breast tumors. *Sci Rep*. 2017;7(1):262. Published 2017 Mar 21. doi:10.1038/s41598-017-00363-w
51. Schmidt,C., Gronborg,M., Deckert,J., Bessonov,S., Conrad,T., Luhrmann,R. and Urlaub,H. Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. (2014) *RNA*, 20, 406–420.
52. Kwon, O.S., Mishra, R., Safieddine, A. *et al.* Exon junction complex dependent mRNA localization is linked to centrosome organization during ciliogenesis. *Nat Commun* **12,** 1351 (2021).
53. Bertram K, Agafonov DE, Dybkov O, Haselbach D, Leelaram MN, Will CL, Urlaub H, Kastner B, Lührmann R, Stark H. Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. Cell. 2017 Aug 10;170(4):701-713.e11. doi: 10.1016/j.cell.2017.07.011. Epub 2017 Aug 3. PMID: 28781166
54. Naro C, Barbagallo F, Chieffi P, Bourgeois CF, Paronetto MP, Sette C. The centrosomal kinase NEK2 is a novel splicing factor kinase involved in cell survival. Nucleic Acids Res. 2014 Mar;42(5):3218-27. doi: 10.1093/nar/gkt1307. Epub 2013 Dec 24. PMID: 24369428; PMCID: PMC3950702.
55. Joan A. Steitz, Gideon Dreyfuss, Adrian R. Krainer, Angus I. Lamond, A. Gregory Matera, Richard A. Padgett Where in the cell is the minor spliceosome? Proceedings of the National Academy of Sciences Jun 2008, 105 (25) 8485-8486; DOI: 10.1073/pnas.0804024105
56. Johnson CA, Malicki JJ. The Nuclear Arsenal of Cilia. Dev Cell. 2019 Apr 22;49(2):161-170. doi: 10.1016/j.devcel.2019.03.009. PMID: 31014478.
57. Amato R, Morleo M, Giaquinto L, di Bernardo D, Franco B. A network-based approach to dissect the cilia/centrosome complex interactome. BMC Genomics. 2014 Aug 7;15(1):658. doi: 10.1186/1471-2164-15-658. PMID: 25102769; PMCID: PMC4137083.

FIGURES AND LEGENDES

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 1

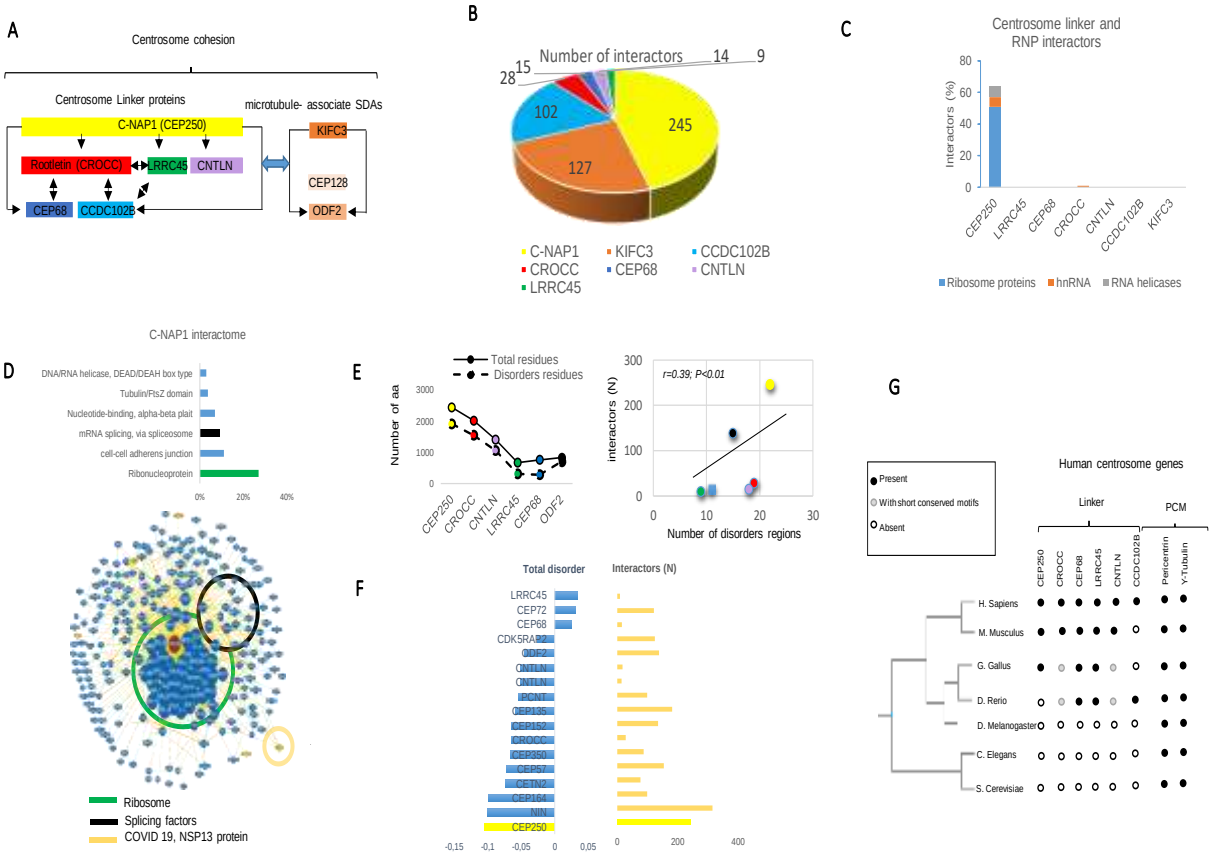

Fig.1 Protein-protein interaction network and centrosome cohesion proteins. **A)** Scheme of experimentally validated proteins involved in centrosome cohesion. **B)** Size of human interactome for the indicated centrosome-linker proteins derived from BioGRID database. **C)** Distribution of interacting ribonucleoproteic factors among centrosome linker proteins. **D)** Up, the graph shows the major pathways derived from C-NAP1 interactome (see Table S2); Down, Protein-Protein interaction network for C-NAP1 (BioGRID database). **E)** The degree of protein disorder in centrosome linker correlates with protein size and interactome profile; Down, correlation between protein disorder and the size of interactome for a larger subset of centrosome proteins. **G)** The genes encoding for centrosome linker proteins tend to be absent in lower Eukaryotes and non-mammalian systems. PCM markers are shown for comparison. Abbreviations: SDA, subdistal appendages.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 2

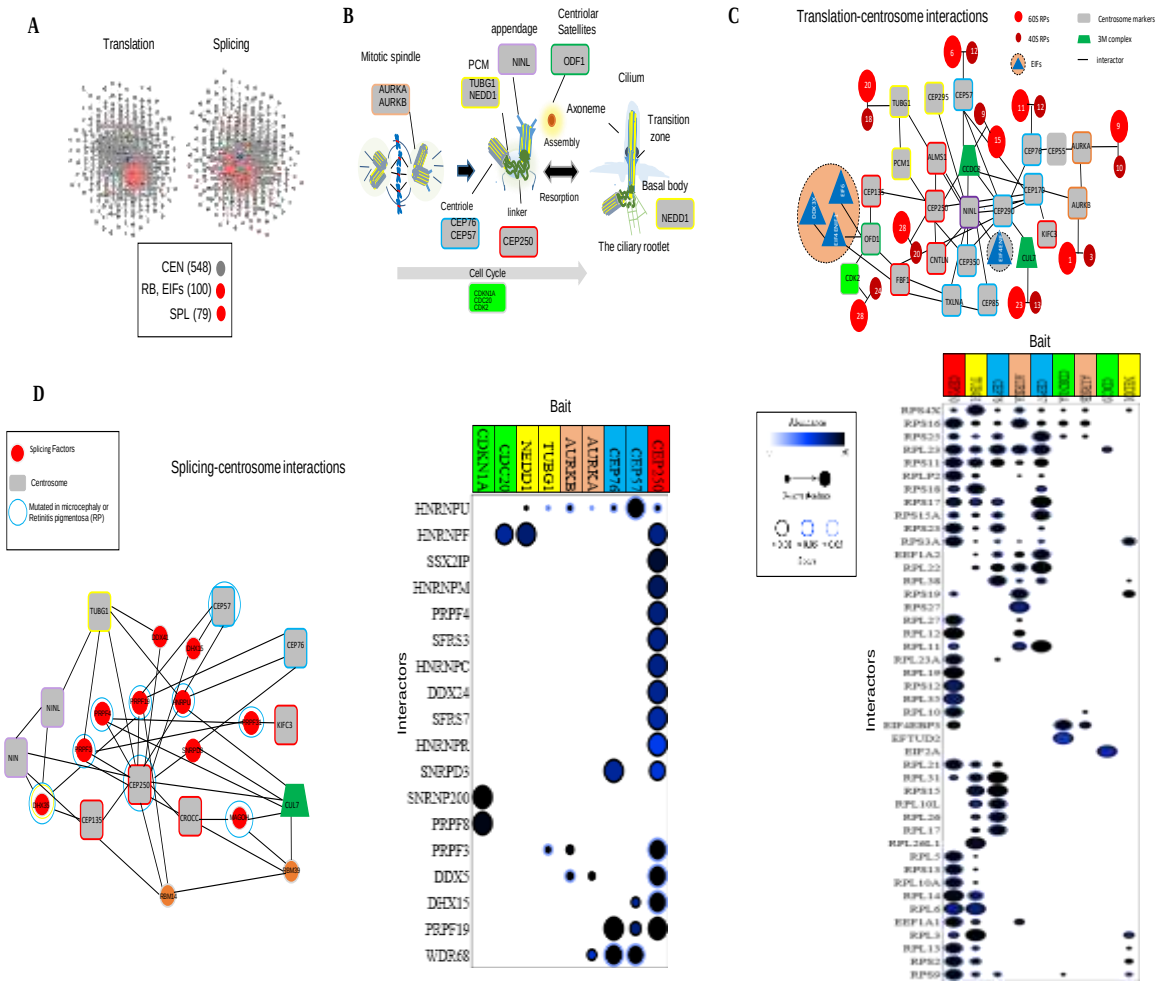

Fig.2 Splicing proteins interact selectively with centrosome components **A)** Spatial proximity between centrosome, translation and splicing components using protein-protein interaction network. **B)** Schematic representation of the mammalian centrosome and distribution of centrosome-related proteins. **C)** Up, Protein-Protein interaction network and distribution of Ribosome and EIFs at centrosome. Down, Dot Plot view of protein-protein interactions. **D)** Protein-Protein interaction network of splicing and centrosome proteins. Dot Plot view of protein-protein interactions between centrosome markers and different splicing proteins. Dot shading (blue-black gradient) indicates total number of spectral counts detected for each prey protein. Dot size indicates relative abundance of prey protein in each analysis. Confidence levels for each interaction according to SAINT (significance analysis of interactome). False discovery rate (FDR) are indicated by dot border.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

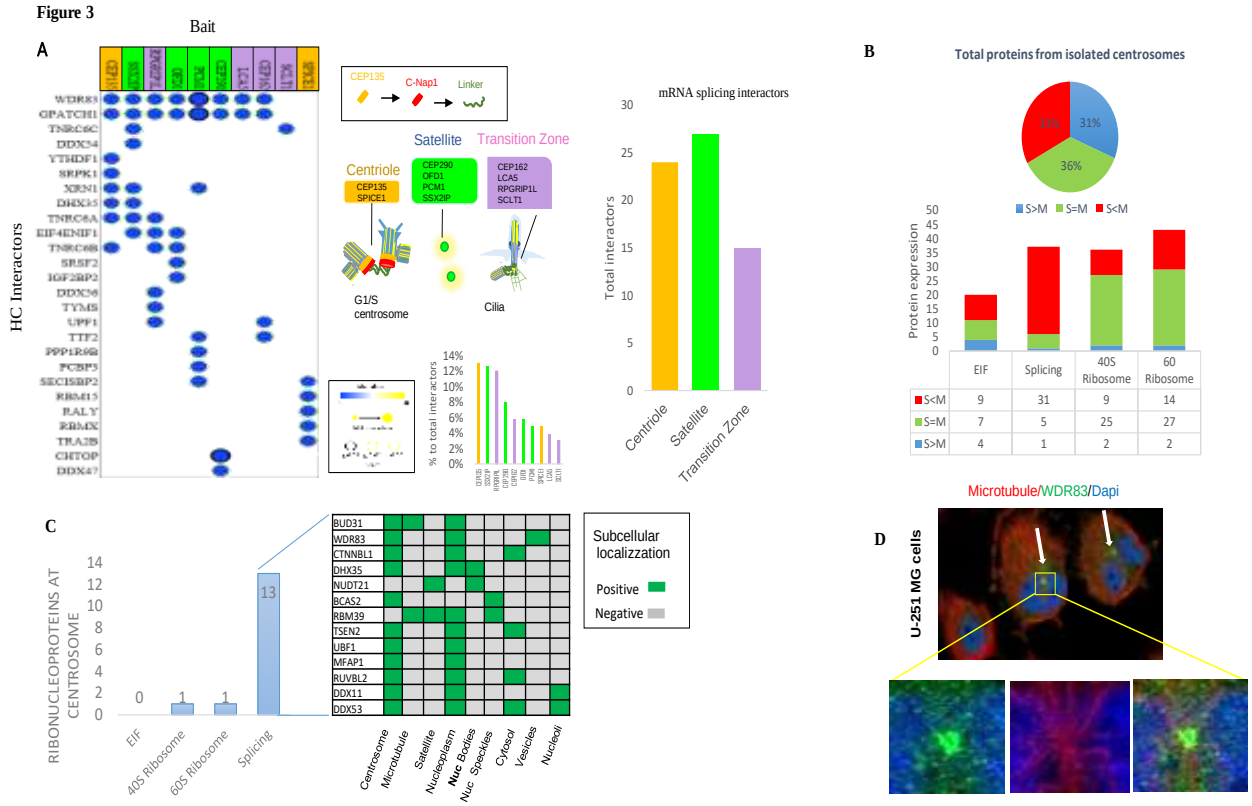

Fig.3 Protein-protein interaction network refines subcellular distribution of splicing indicating a role in ciliogenesis and centriole-based function A) Left, Dot Plot view of High-confident Protein-Protein interactions derived from FLAG-IP Mass spectrometry data **(Ref).** SAINT (significance analysis of interactome) was used to analyze the relative abundance/presence of interactors. False discovery rate (FDR) are indicated by dot border. Right, the graphs indicate the distribution of interactors relative to centrosome markers represented in the scheme. **B)** Quantification of different types of Ribonucleoproteins during S and M phases derived from isolated centrosomes in HeLa S3 cells. Up, the pie chart indicates the abundance of proteins detected in each condition (Ref). **C)** Left, number of different Ribonucleoproteins that have been experimentally detected in the centrosome by the Human Protein Atlas (Ref). Right, The splicing factors that localize to the centrosome are also localized "frequently" in the nucleoplasm as expected. **D)** Centrosomes from U251-MG cells are intensely stained for WDR83, immunofluorescence images extracted by "Human proteome atlas".

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 4

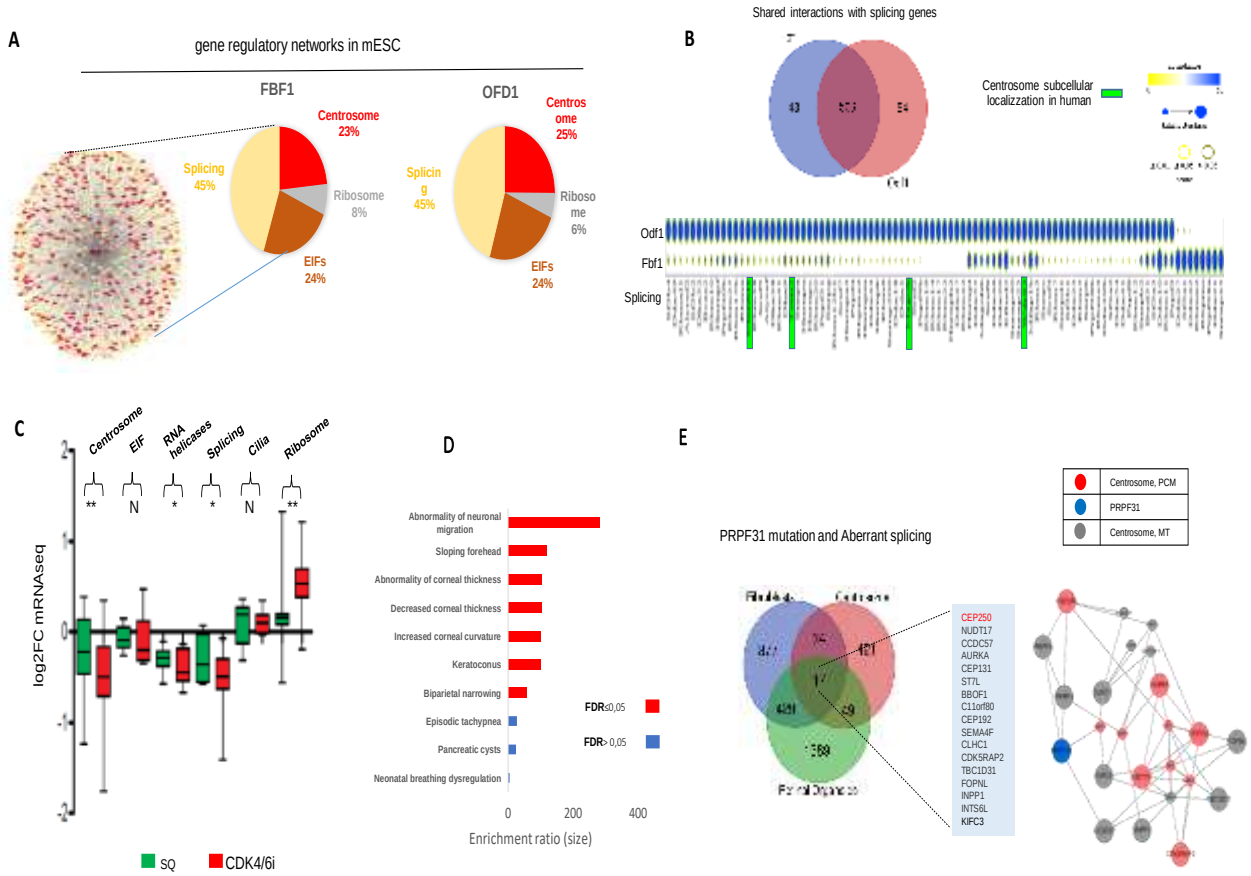

Fig.4 **A gene regulatory network characterizes splicing and centrosome encoding genes A)** Gene network analysis of Fbf1 and Odf1 in relation to the indicated categories derived from mouse embryonic stem cells (Ref). **B)** Up, the Venn diagram shows the shared interactions of splicing encoding genes between Fbf1 and Ofb1; Down, The Dot plot shows the pattern of gene regulatory network for a subset of splicing genes comparing Fbf1 vs Odf1 profile. SAINT (significance analysis of interactome) was used to analyze the "weight" of interactors according to SINCERITIES algorithm (Ref). False discovery rate (FDR) are indicated by dot border. **C)** Gene expression variation in spontaneous versus CDK inhibitor-induced slow-cycling cells (quiescence) for the indicated categories (ref) . For each category more than twenty genes were included. *P ≤ 0.05; **P ≤ 0.01; Student's t-test**. D)** The indicated enrichment categories were obtained using the phenotype-Human-Phenotype Ontology analysis integrating a list of 66 highly interconnected splicing and centrosome genes. **E)** *PRPF31* mutation causes aberrant splicing and retinitis pigmentosa (ref). The Venn diagram shows the shared centrosome encoding genes undergoing aberrant splicing in human fibroblasts and retinal organoids. Protein-Protein interaction network of centrosome genes targeted by PRPF31.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 5

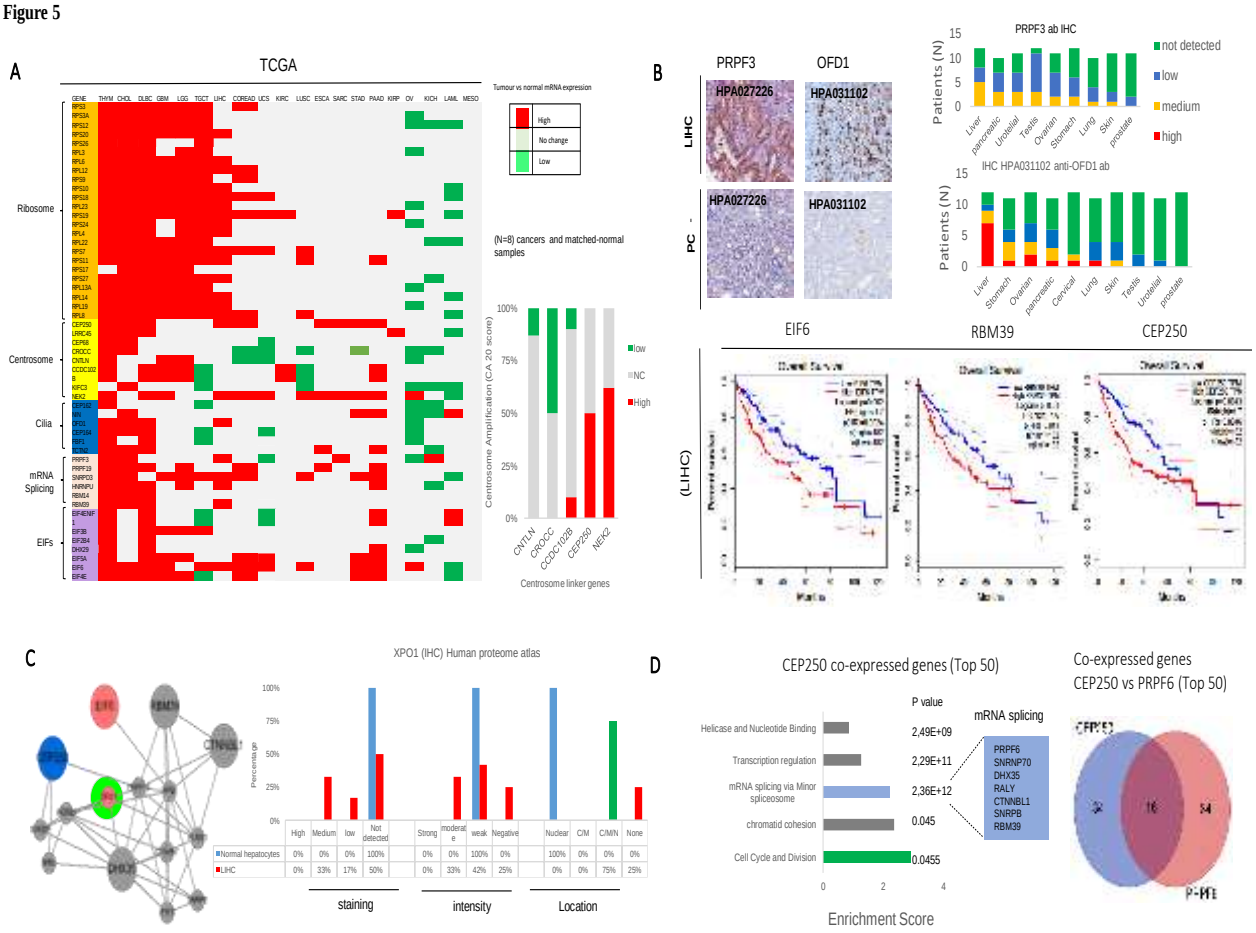

Fig.5. The centrosome and splicing transcriptional program are co-altered in cancers originating in Liver. A) Gene expression profile for the indicated categories across n=20 different cancer types; High expression in tumour versus normal is in "red" ($p<0,05$); low expression in tumour vs normal is in "green" ($p<0,05$); no changes between tumour vs normal are in grey ($p>0,05$). Right, The graph shows that *CEP250* and its binding partner *NEK2* are commonly over-expressed in cancer and correlate with centrosome amplification calculated by (CA 20 score). N=8 different cancers types for which tumour and matched-normal samples were available (Ref). **B)** Left, Immunohistochemistry images for PRPF3 and ODF1 in the indicated cancer tissues extracted by human proteome atlas. Right, PRPF3 and ODF1 immunostaining profile in Liver (LIHC) compared to any other cancer type. Low, High mRNA expression of the indicated genes correlates with shorter overall survival by Kaplan Meier analysis. **C)** Left, Protein-Protein interaction network reveals the mRNA transporter *XPO1* as the top ranking neighbours gene between CEP250 and its partner splicing regulators. Right, IHC data extracted from human proteome atlas to indicate the subcellular distribution of *XPO1* in normal hepatocytes versus LIHC. C, Cytoplasm; M; Membrane; N; Nucleus. **D)** Left, Shared co-expressed genes between PRPF6 and CEP250 taking into account the top 50. **Abbreviations:** CHOL: cholangiocarcinoma; COREAD: colon and rectum adenocarcinoma; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; ESCA: oesophageal carcinoma; GBM: glioblastoma multiforme; KICH: kidney chromophobe; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LAML: acute myeloid leukemia; LGG: low-grade glioma; LIHC: liver hepatocellular carcinoma; LUSC: lung squamous cell carcinoma; MESO: mesothelioma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; SARC: sarcoma; STAD: stomach adenocarcinoma; TGCT: testicular germ cell tumours; UCS: uterine carcinosarcoma;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 6

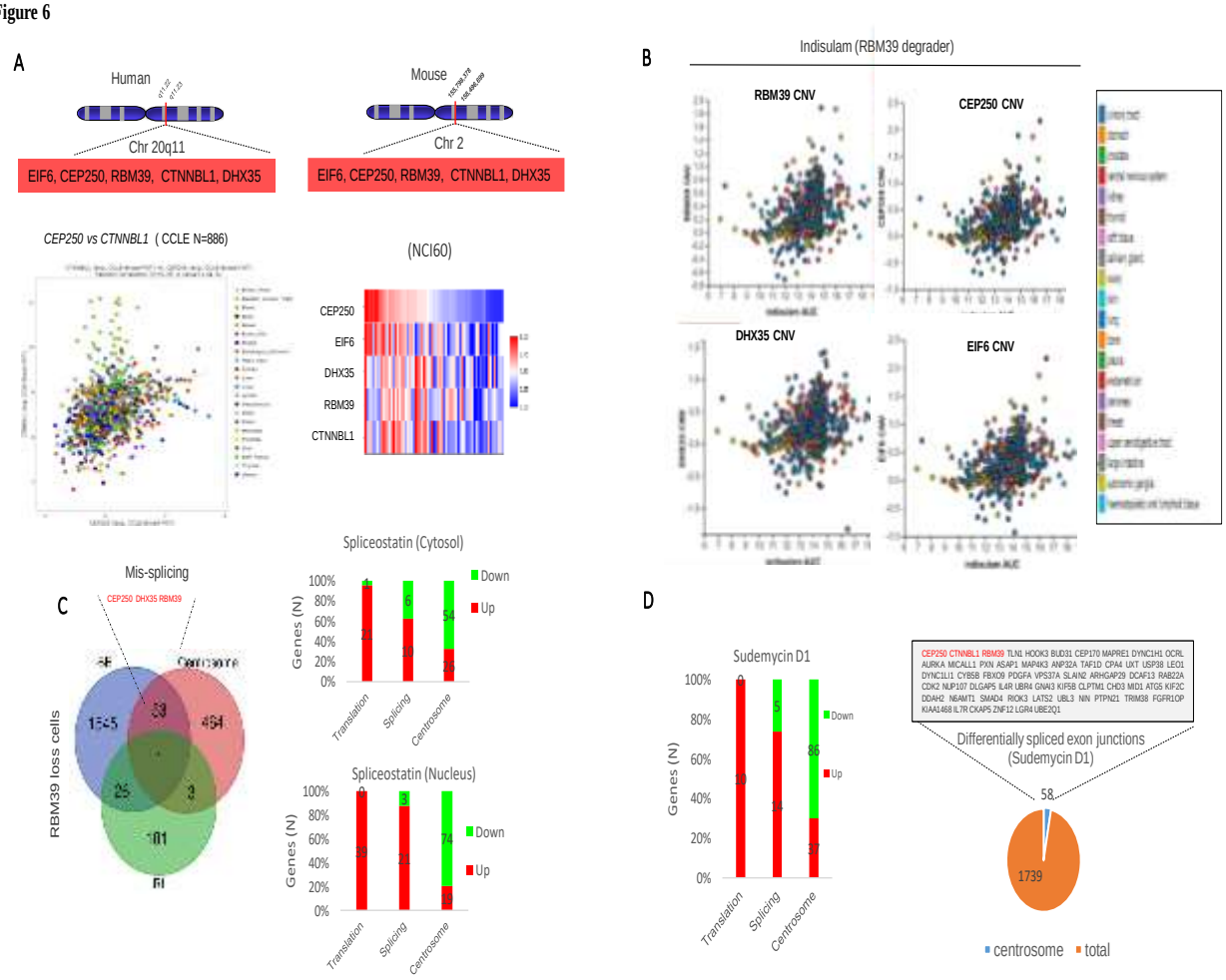

Fig.6. Intrachromosomal colocalization strengthens co-expression and co-modification of splicing and centrosome genes clusters A) Up, Linear and spatial co-localization of the indicated gene cluster on human chromosome 20 and mouse Chromosome 2; Down left, *CTNNBL1* and *CEP250* mRNA expression levels are highly correlated across CCLE database. Down right, Heatmap show a co-expression of the gene module across NCI60 database used as alternative sources. **B)** Indisulam a protein degrader drugs against RBM39 similarly co-modificates the gene cluster in relation to copy number variation across cancer cell lines; data extracted from the cancer therapeutics response portal. **C)** Left, Loss of RBM39 from (ref..) determines skipping exon dysfunction in 53 centrosome genes including CEP250, DHX35 and RBM39 itself. Right, The graphs show the changes in gene expression profiles for the indicated categories upon Spliceostatin treatment (Ref). **D)** Left, gene expression profiles for the indicated categories upon Sudemycin D1 treatment. Right, Aberrant spliced exon junctions in involve about more than 10% of centrosome encoding genes. The genes *CEP250, CTNNBL1 RBM39* belonging to the colocalized cluster are indicated in red.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 February 2022 doi:10.20944/preprints202202.0084.v1

Figure 7

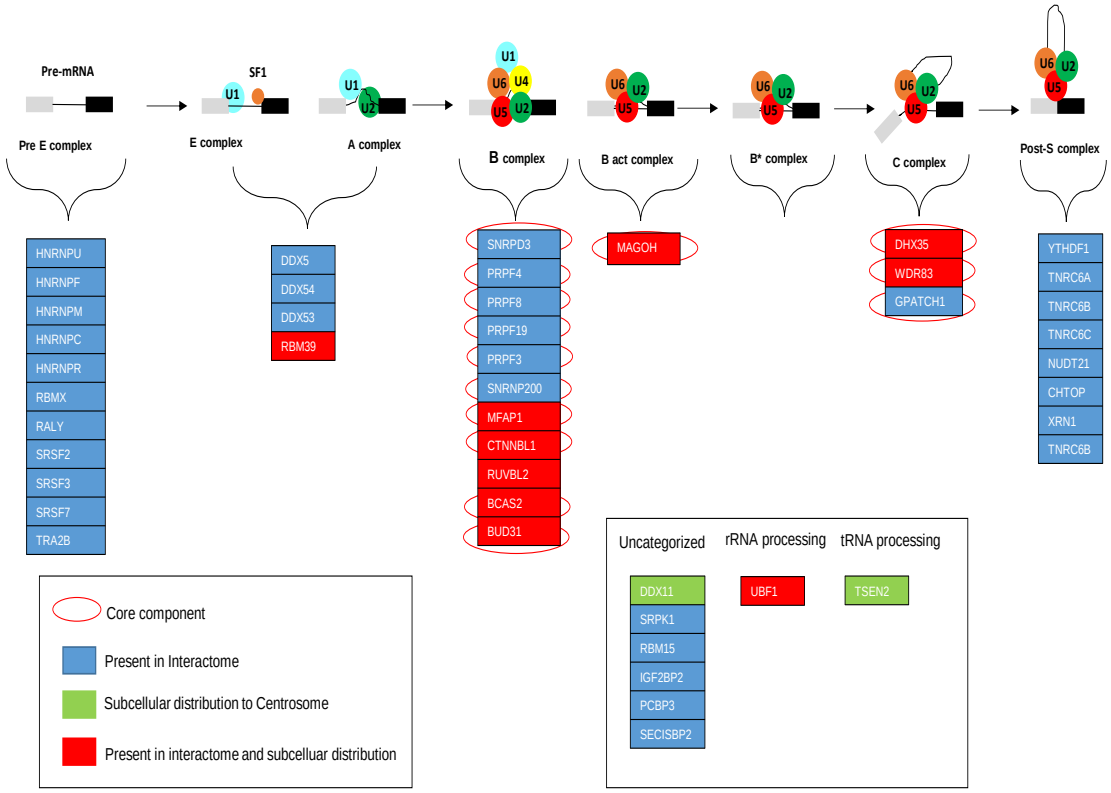

Fig.7. Distribution and functional features of Spliceosome components to the centrosome Schematic drawing showing the splicing proteins with subcellular localization to the centrosome and interacting splicing partners with centrosome proteins. They are categorized according to core and non-core component of the spliceosome. The spliceosomal sub-complexes are shown in the order of their recruitment and activity to the spliceosome.