UNIVERSITY OF NAPOLI FEDERICO II

Doctorate School in Molecular Medicine

Doctorate Program in
Genetics and Molecular Medicine
Coordinator: Prof. Lucio Nitsch
XXVIII Cycle

"Functional analysis of Dnajc17 gene function: in vivo and in vitro studies"

CANDIDATE: Dr. Angela Pascarella
MENTOR: Prof. Roberto Di Lauro

Napoli 2016

TABLE OF CONTENTS

ABSTRACT

Congenital hypothyroidism (CH) is the most frequent endocrinological congenital disorder often due to thyroid dysgenesis (TD). Mutations in genes involved in thyroid development such as Nxk2.1, Pax8, Foxe1, and Tshr cause TD in animal models, although in humans loss of function mutations in these genes have been identified only in few cases of CH with TD, suggesting the involvement of additional loci. A polygenic mouse model of CH with TD allowed our laboratory to identify Dnajc17, a member of the type III heat-shock protein-40 (Hsp40) family, as a candidate TD modifier gene. Our laboratory demonstrated that homozygous null mutation in Dnajc17 gene are early embryo lethal.

Moreover, as DNAJC17 is still poorly characterized, we started a detailed study in vitro. First, we studied its subcellular localization, showing that it localizes predominantly in the nucleus. Then, to uncover DNAJC17 function we used two different in vitro approaches: 1) analysis of DNAJC17 interactome by CO-IP and mass spectrometry, 2) analysis of gene expression profile induced by Dnajc17 knock-down, both performed in HeLa cells. Interactome studies show that DNAJC17 interacts with several proteins involved in splicing. Some of these proteins are essential for embryonic development, similarly to Dnajc17 (PPR19, PLRG-1, snRNP200, XAB2). SNRNP200 (U5) is a component of the U5 snRNP and U4/U6-U5 tri-snRNP complexes and XAB2 that is involved in transcription-coupled repair and transcription. PRP19 and PLRG-1 belong to Prp19/CDC5L complex. Depletion of this complex leads to a block in splicing before the first transesterfication step in splicing and lariat formation of the pre-mRNA.

Furthermore, the confirmation of DNAJC17 interaction with splicing machinery components comes from immunofluorescence demonstrating that DNAJC17 colocalize whith SC35, a spliceosome assembly factor localized in nuclear speckles. Consistently, gene expression profile of cells knocked-down for Dnajc17 displays alterations in the balance between alternatively spliced isoforms of many genes. We are currently characterizing differentially spliced genes.

To elucidate the functional role of Dnajc17 in thyroid we generated a conditional knock-out mouse in which Dnajc17 has been disrupted specifically in the thyroid.

1.BACKGROUND

1.1 Thyroid Gland development

The thyroid anlage is first identified in mouse embryos at embryonic day E8.5 in the ventral wall of the primitive pharynx, caudal to the region of the first branchial arch. This median thickening deepens, forms first a small pit (thyroid bud). At this time, primitive thyroid cells already have a distinct molecular signature, with co-expression of four transcription factors Hhex, Tift1, Pax8 and Foxe1.
At E9–E9.5 the endodermal cells of the thyroid anlage form the thyroid bud by proliferation and invasion of the surrounding mesenchyme. By E10, a small hole at the site of origin in the pharyngeal floor (the foramen cecum) is the remnant of the anlage, connected with the migrating thyroid primordium by a narrow channel (the thyroglossal duct). At E11.5 the thyroglossal duct disappears, and the thyroid primordium loses its connections with the floor of the pharynx and begins to expand laterally. Two days later the thyroid primordium reaches the trachea, which has extended ventrocaudally starting from the primitive laryngotracheal groove. By E15–E16 the thyroid lobes expand considerably, and the gland exhibits its definitive shape: two lobes connected by a narrow isthmus (De Felice and Di Lauro 2004)

1.2 Molecular aspects of thyroid development

The molecular basis of thyroid development begin with the investigation of a transcriptional factor, Nkx2.1, identified as responsible of Thyroglobulin (Tg) and thyroperoxidase (Tpo) expression. Nkx2.1 was found expressed in adult and functional thyroid cells and in their precursor.
Subsequently other three transcriptional factors was found expressed in thyroid precursor cells: Hhex, Pax8 and Foxe1.
This four transcriptional factors are also expressed in other tissues, but their simultaneous presence is restricted at only thyroid cells since their specification to the rest of their life.
Titf1/Nkx2.1 belongs to the Nkx2 family of transcriptional factors, was first identified as homeodomain-containing transcription factor able to bind to specific DNA sequences present in the Tg gene promoter.
Titf1/Nkx2.1 knockout mice are characterized by lack of thyroid and death at birth (Kimura et al. 1996). The thyroid primordium forms in its correct position but subsequently undergoes degeneration and eventually disappears (Kimura et al. 1999).
Pax8 (Paired box gene 8) is a transcriptional factor belongs to Pax family which nine members are characterized by a DNA binding domain named paired.
The expression of this gene during embryo development is detected in thyroid since its specification (E8.5).

Pax8$^{-/-}$ mice are born with expected Mendelian frequency, but they show growth retardation and die in 2-3 weeks. The thyroid gland of this mice is severely affected because neither follicles nor TFCs can be detected, and the rudimentary gland is composed almost completely of calcitonin-producing C cells (Mansouri et al. 1998).
Foxe1 was originally identified as a thyroid specific nuclear protein that recognizes a DNA sequence present on both Tg and TPO promoters under hormone stimulation (Civitareale et al. 1989, Santisteban et al. 1992). Expression of Foxe1 in the thyroid cell precursors is maintained during development and persists in adult TFCs (Zannini et al. 2001).
Homozygous Foxe1$^{-/-}$ mice are born at the expected ratio but die within 48 h. These mice display no thyroid in its normal location and an absence of thyroid hormones. Furthermore, the mice show a severe cleft palate, probably responsible for the perinatal death, and elevated TSH levels in the bloodstream. The lack of Foxe1 doesn't affects the thyroid primordium which is easily detected by the expression of Titf1/Nkx2-1 and Pax8. However, at E9.5 in Foxe1 null embryos, thyroid precursor cells are still on the floor of the pharynx and at later stages of development, mutant mice exhibit either a small thyroid remnant still attached to the pharyngeal floor or no thyroid gland at all (De Felice et al. 1998).
Hhex is a transcription factor expressed from primordium to adult thyroid. This factor is fundamental for thyroid development, indeed Hhex$^{-/-}$ mice at E9.5 show the absence of thyroid precursor, but the thyroid anlage is properly formed and expresses Titf1/Nkx2-1, Foxe1, and Pax8 (Martinez Barbera Et al. 2000, De Felice and Di Lauro 2004).

1.3 Congenital Hypothyroidism with Thyroid Dysgenesis and its Multigenic origin

Congenital hypothyroidism (CH) is the most frequent endocrinological congenital disorders with a reported incidence of 1:1,500 live newborns. In a large fraction of CH thyroid dysfunction is due to thyroid dysgenesis (TD), including developmental disorders such as athyreosis, ectopy, hemiagenesis and hypoplasia. Mutations in genes involved in thyroid development such as Nxk2.1, Pax8, Foxe1, and Tshr cause TD in animal models, and loss of function mutations in the same genes are associated with TD in patients. Indeed, TD can be a genetic and heritable affection. However, mutations in these genes are present only in few patients affected by CH with TD, suggesting that other loci could be involved in the genesis of TD (Castanet et al. 2005; De Felice and Di Lauro 2004).
To provide proof for a possible multigenic origin of TD a mouse model was generated.
The null mutations generated in the genes encoding the transcription factors Titf1 and Pax8 only display a TD phenotype when homozygous. No phenotype was observed when these genes are in heterozygosity, but mice heterozygous

for both Titf1- and Pax8-null alleles, denominated DHTP (double heterozygous for both Titf1- and Pax8-null mutations), present TD. The disease is characterized by a small thyroid gland, elevated levels of TSH, reduced thyroglobulin biosynthesis, and high occurrence of hemiagenesis.
TD condition is strain specific. TD is displayed when the DHTP mutations are in the C57BL/6 genetic background, and it is not observed in the 129/Sv strain (Amendola et al. 2005). The B6-specific TD phenotype of DHTP mice is associated with a reduced activation of a transgenic thyroglobulin (Tg) promoter, thus indicating that the modifier is acting on a transcriptional mechanism (Amendola et al. 2010).

1.4 Identification of candidate modifier genes

A genetic linkage analysis in a DHTP backcross population identified a potential locus linked to the B6-specific phenotypethe. This locus mapping on chromosome 2. On the basis of a single-nucleotide polymorphism (SNP) that causes a non synonymous amino acid change in a highly conserved protein domain between B6 and Sv strains, was identified Dnajc17, encoding for a member of the type III heat-shock protein-40 (Hsp40) family, as a candidate modifier gene for hypothyroidism. The B6 allele encodes for a protein containing a phenylalanine residue at position 273, whereas Sv mice contain a tyrosine in that position. The Sv129 variant is conserved in all species examined, from Drosophila to humans (Table 1).

Species	Alignment
Mus muris (B6 strain)	RD**F**ESLVMMRMRQAAERQQLIAQMQQED–EGRPT
Mus muris (Sv strain)	RD**Y**ESLVMMRMRQAAERQQLIAQMQQED–EGRPT
Rattus norvegicus	RD**Y**ESLVMMRMRQAAERQQLIAQMQQED–EGRPT
Homo sapiens	RD**Y**ESLVMMRMRQAAERQQLIARMQQEDQ–EGPP
Macaca mulatta	RD**Y**ESLVMMRMRQAAERQQLIARMQQEDQ–EGPPT
Monodelphis domestica	RD**Y**ESLVMMRMRQAAERQQLIEQMQREDE–EGLPT
Danio rerio	RD**Y**ESVVLMRLRQAEERKRLIEQMKREE–EDDTA–
Anopheles gambiae	RD**Y**ESLVLRQLRQAEERKRLIEQMMKEEAEAEGETTAQ
Drosophila melanogaster	TD**Y**EDLVMRKLRQAEERKRLIEQMMKDE–EGE–

Table 1. Alignment of the deduced amino acid sequence of the C-terminal region of DNAJC17 orthologous genes among different species. Tyrosine273 residue is in bold. Sequence data are according to the Ensembl data (Adapted from Amendola et al. 2010).

Various data suggest this hypothesis. Dnajc17 is high expressed in both fetal and adult thyroid, affects Tg transcription and is essential in mouse embryo development. Indeed The Dnajc17$^{+/-}$ mice appeared normal, while Dnajc17$^{-/-}$ pups were never observed postnatally, indicating that the Dnajc17 null mutation is recessive embryonic lethal. Only Dnajc17$^{-/-}$ two-cell embryos was found, but no homozygous embryos were recovered after the morula stage. These findings indicate that Dnajc17$^{-/-}$ embryos were developmentally arrested before implantation and indicate an essential role of this gene in early mouse development (Amendola et al. 2010).

Furthermore, Patel and collaborators identified a homozygous truncating mutation in DNAJC17 in a family with an apparently novel syndrome of retinitis pigmentosa and hypogammaglobulinemia (Patel N. et al. 2015).

1.5 DNAJC17 protein

DNAJC17 protein belong to Dnaj family, which canonical function is to stimulate the ATPase activity of molecular chaperone DNAJ/HSP70 in various cell function such protein folding, transport, degradation and heat shock response.
All the members of the DnaJ/Hsp40 family contain the J domain through which they bind to their partner Hsp70s. This domain is usually present at the N-terminal region of the proteins. The J domain is a 70-amino acid sequence consisting of four helices and a loop region between helices II and III that contains a highly conserved tripeptide of histidine, proline, and aspartic acid (the HPD motif). In addition to the J domain, many DnaJ/Hsp40 proteins contain other conserved regions, which are critical to their functions. Based on the difference in these regions, DnaJ proteins can be categorized into three groups (Fig. 1). Type I proteins with the J domain, the Gly/Phe-rich region, and the cysteine repeats. Type II proteins possess the J domain and the Gly/Phe-rich region, but lack the cysteine repeats. Type III proteins do not have any of these conserved regions other than the J domain. Although type I and II proteins are different in their conserved regions, it seems that both types function similarly and bind to non-native substrates. In contrast, the type III proteins may not bind to non-native polypeptides and thus should not function as molecular chaperones on their own (Qiu XB 2006).

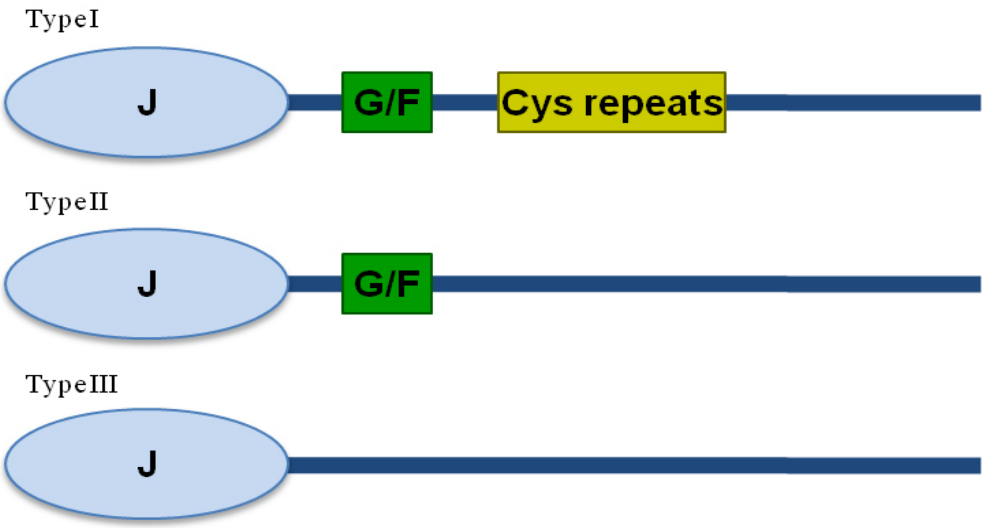

Figure 1. Classification of DnaJ/Hsp40 proteins. Depending on the presence of the Gly/Phe-rich region (G/F) and/or the cysteine repeats, a DnaJ/Hsp40 protein can be categorized as type I, II, or III.

DNAJC17 belongs to type III proteins and the three-dimensional structure prediction show that, the J domain is located near to the N-terminus, moreover there is an RNA recognition motif RRM located at C-terminus of protein (Fig. 2).

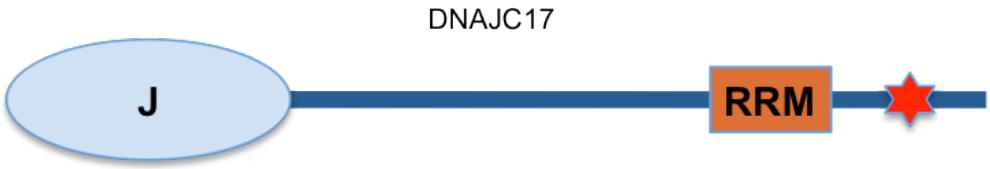

Figure 2. Schematic representation of DNAJC17 protein. The J domain is located near to the N-terminus and the RNA recognition motif (RRM) at C-terminus of protein. The star represent thy273 localization.

1.6 Mouse preimplantation development

The period of preimplantation development in mouse, extending from egg fertilization to implantation of the blastocyst in the uterus, is a key stage during which the first three major cell lineages of the embryo and its extraembryonic membranes are set aside. These three lineages contribute to distinct tissues in later development: the epiblast (EPI) gives rise to the fetus itself; the trophectoderm (TE) goes on to form the majority of the fetal contribution to the placenta; and the primitive endoderm (PE) becomes the parietal and visceral endoderm, which later contributes to the yolk sac (Cockburn and Rossant 2010).

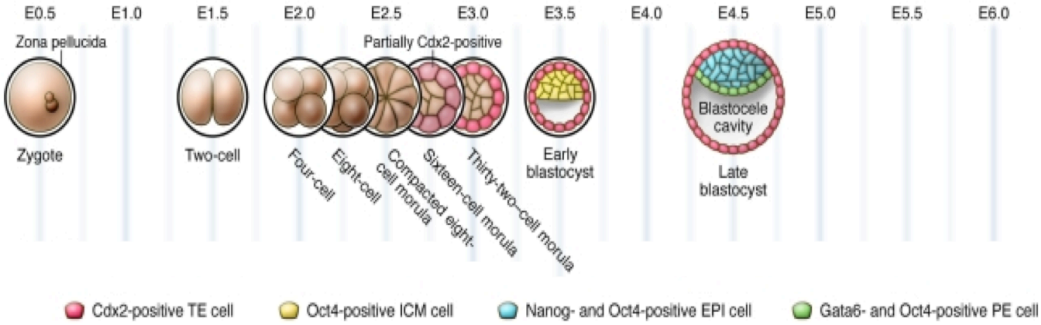

Figure 3. Stages of preimplantation development. In the mouse, the fertilized egg undergoes three rounds of cleavage, producing an eight-cell embryo that then undergoes compaction. From the eight-cell stage onward, cell divisions produce two populations of cells, those that occupy the inside of the embryo and those that are located on the outside. The blastocoel cavity begins to form inside the embryo beginning at the 32-cell stage and continues to expand as the embryo grows and matures into the late blastocyst stage. Cdx2 becomes upregulated in outside, future TE cells, starting at the 32-cell stage, while Oct4 expression becomes limited to the ICM in the early blastocyst stage. By the late blastocyst stage, while continuing to express Oct4 ubiquitously, the ICM contains a population of Nanog-positive EPI cells and a population of Gata6-positive PE cells (Adapted from, Cockburn and Rossant 2010).

The fertilized egg first undergoes a series of early cleavage divisions, producing increasing numbers of progressively smaller cells, known as blastomeres, without changing the overall size of the embryo (Fig. 3). Protein

synthesis in the mammalian zygote initially relies on a deposit of maternally loaded mRNA (Bachvarova 1985). Transcription of mRNA coded by the zygotic genome begins during the first few cleavage divisions. In mouse there is an initial burst of zygotic transcription at the end of the one-cell stage (Aoki et at. 1997; Hamatani et al. 2004).

The early cleavage divisions produce an eight-cell embryo (morula stage) that subsequently undergoes an increase in intercellular adhesion known as compaction, causing all cells to adopt a more flattened morphology (Fig. 3). This process of compaction is essential for later morphogenetic events and for the proper segregation of the three embryonic lineages. In the mouse, compaction is associated with the formation of adherens and, later, tight junctions between cells. Blastomeres do not show any signs of intracellular polarity until compaction but, concomitant with the increase in cell adhesion at this stage, all cells rapidly polarize along the axis perpendicular to cell contact such that outward facing (apical) regions become distinct from inward facing (basolateral) regions.

Once the eight-cell embryo has compacted and polarized, it undergoes two further rounds of cleavage, growing from eight cells to 16, and from 16 cells to 32. During these divisions, inheritance of the polarized state is influenced by the orientation of the cleavage plane of the blastomere. If a cell undergoes mitosis at an angle perpendicular to its axis of polarity (that is, parallel to its inside-outside axis), its two daughter cells will both be polar and will remain on the outside of the embryo. However, cells can also divide parallel to their axis of polarity, producing one polarized outside daughter cell and one apolar cell that is located on the inside of the embryo (Johnson and Ziomek 1981; Sutherland et al. 1990). In this way the preimplantation embryo, which was previously composed of a uniform population of cells, has now generated two separate groups of cells: apolar inside cells and polar outside cells.

From the 32-cell stage onward, these two cell populations have distinct developmental fates: cells on the outside of the embryo contribute to the TE lineage, while inside cells contribute to the inner cell mass (ICM), the group of cells that further diverges into the EPI and PE lineages.

Starting at the 32-cell stage, as the outside cells of the embryo are becoming fully committed to the TE lineage (Dyce et al. 1987; Pedersen et al. 1986), a fluid-filled cavity known as the blastocoel begins to form (Fig. 3). The presence of a blastocoel is essential for proper development of the ICM (Pedersen et al. 1986). Once it begins to form, maintenance of the blastocoel depends on the epithelial character of the TE.

With the formation of the blastocoel at E3.5, the mouse embryo is now considered a blastocyst. It continues to mature for an additional 24 hours and is ready to implant into the uterine wall by E4.5.

TE and ICM lineage segregation is controlled by a small group of transcription factors. Specifically, Cdx2 is required for TE development, while the pluripotency markers octamer 3/4 (Oct4), Nanog, and SRYbox containing gene 2 (Sox2) are involved in establishing the ICM fate. In the mouse, Cdx2 is

expressed at varying levels in all blastomeres starting at the eight-cell stage, but it becomes restricted to outside, future TE cells, prior to blastocyst formation (Fig. 3) (Dietrich and Hiirag 2007; Ralston and Rossant 2008; Cockburn and Rossant 2010).

1.7 Splicing mechanism

Most eukaryotic genes are expressed as precursor mRNA (pre-mRNA) that undergoes extensive modification and processing in the nucleus to convert it into mature mRNA for export to the cytoplasm, where it is subsequently translated into protein. An important processing step is the splicing, an essential step of gene expression in which noncoding sequences (introns) are removed and coding sequences (exons) are ligated together. Whereas some exons are constitutively spliced, many are alternatively spliced to generate variable forms of mRNA from a single pre-mRNA species. Alternative splicing is prevalent in higher eukaryotes and it enhances their complexity by increasing the number of unique proteins expressed from a single gene (Nilsen and Graveley 2010). Unraveling splicing at the molecular level is not only important for understanding gene expression, but it is also of medical relevance, as aberrant pre-mRNA splicing is the basis of many human diseases or contributes to their severity (Novoyatleva et al. 2006; Ward and Cooper 2010).

Nuclear pre-mRNA splicing is catalyzed by the spliceosome, a large complex ribonucleoprotein (RNP) complex. Both the conformation and composition of the spliceosome are highly dynamic.

Two unique spliceosomes coexist in most eukaryotes: the U2-dependent spliceosome, which catalyzes the removal of U2-type introns, and the less abundant U12-dependent spliceosome. this new spliceosome is structurally and functionally analogous to the well-characterized major-class splicing apparatus, it mediates the excision of a minor class of evolutionarily conserved introns that have non-canonical consensus sequences (reviewed by Patel and Steitz 2003; Will and Lührmann 2011).

Information provided by a pre-mRNA that contributes to defining an intron is limited to short, conserved sequences at the 5′splice site (ss), 3′ss and branch site (BS) (Burge et al. 1999; Will and Lührmann 2011). The BS is typically located 18-40 nucleotides upstream from the 3'ss and in higher eukaryotes is followed by a polypyrimidine tract (PPT) (Fig. 4A). Different splice site and branch site sequences are found in U2- versus U12-type introns (Burge et al. 1999). Additional, cis-acting pre-mRNA elements include exonic and intronic splicing enhancers (ESEs and ISEs) or silencers (ESSs and ISSs). They are typically short and diverse in sequence and modulate both constitutive and alternative splicing by binding regulatory proteins that either stimulate or repress the assembly of spliceosomal complexes at an adjacent splice site (Smith and Valcárcel 2000; Wang and Burge 2008; Will and Lührmann 2011).

Nuclear pre-mRNA introns are removed by two consecutive transesterification reactions (reviewed by Moore et al. 1993). First, the 2′OH group of the branch adenosine of the intron carries out a nucleophilic attack on the 5′ss. This results in cleavage at this site and ligation of the 5′end of the intron to the branch adenosine, forming a lariat structure. Second, the 3′ss is attacked by the 3′OH group of the 5′exon, leading to the ligation of the 5′and 3′exons (forming the mRNA), and release of the intron (Fig. 4A).

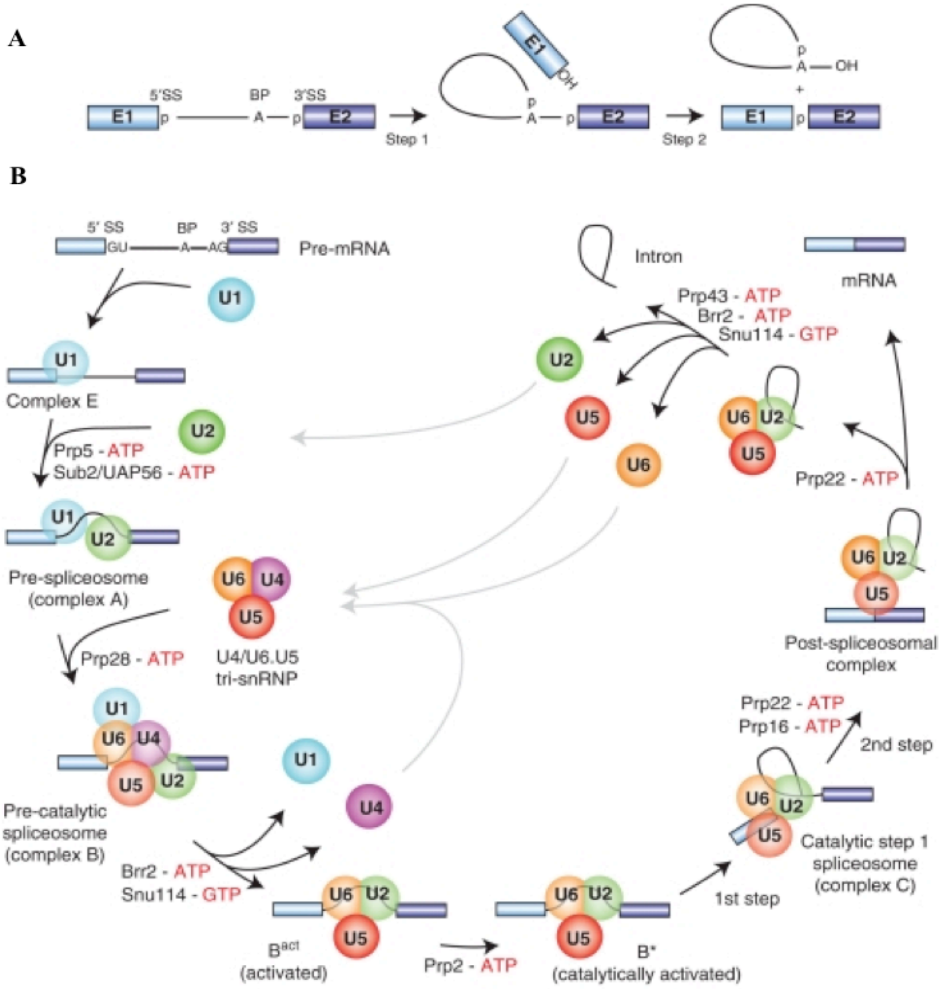

Figure 4. Pre-mRNA splicing. (A) Schematic representation of the two-step mechanism of pre-mRNA splicing. Boxes and solid lines represent the exons (E1, E2) and the intron, respectively. The branch site adenosine is indicated by the letter A and the phosphate groups (p) at the 5′ and 3′ splice sites, which are conserved in the splicing products, are also shown. (Adapted from Will and Lührmann 2011).

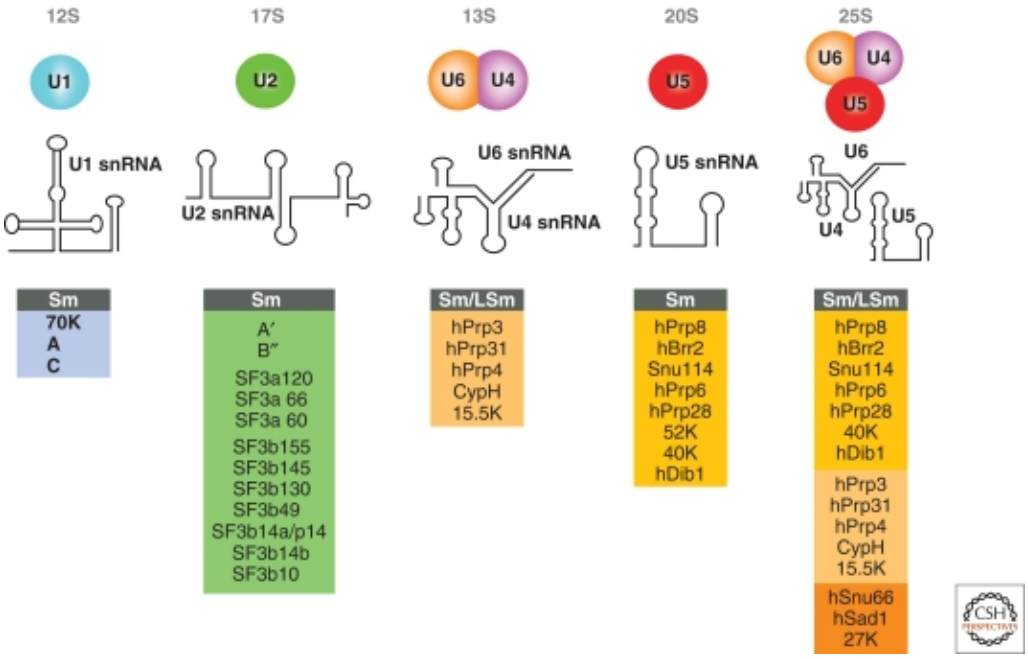

Figure 5. Protein composition and snRNA secondary structures of the major human spliceosomal snRNPs. All seven Sm proteins (B/B', D3, D2, D1, E, F, and G) or LSm proteins (Lsm2-8) are indicated by "Sm" or "LSm" at the top of the boxes showing the proteins associated with each snRNP. The U4/U6.U5 tri-snRNP contains two sets of Sm proteins and one set of LSm proteins (Adapted fromWill and Lührmann 2011).

Assembly and functioning of the spliceosome requires approximately 100 polypeptides and five small nuclear RNAs (snRNAs), not considering gene-specific RNA-binding factors. The U2-dependent spliceosome is assembled from the U1, U2, U5, and U4/U6 snRNPs and numerous non-snRNP proteins. Each snRNP consists of an snRNA (or two in the case of U4/U6), a common set of seven Sm proteins (**B/B′, D3, D2, D1, E, F, and G**) and a variable number of particle-specific proteins (Will and Lührmann 2006; Will and Lührmann 2011) (Fig. 5). Assembly of the Sm proteins onto snRNA is a critical step of the biogenesis of the snRNP particles. Following export from the nucleus to the cytoplasm, each of the U1, U2, U4 and U5 snRNAs associates with one set of the Sm proteins to generate the snRNP core particle (henceforth termed Sm core RNP). Formation of the Sm core RNP is essential for the metabolic stability of the snRNP, as well as for the subsequent hypermethylation of the 5′cap from m7G to 2,2,7-trimethylguanosine, import of the particle into the nucleus and the recruitment of at least some particle-specific proteins (Plessel et al. 1994). The spliceosomal snRNPs do not possesse a pre-formed active site and several of them are substantially remodeled in the course of splicing.

Spliceosome assembly occurs by the ordered interaction of the spliceosomal snRNPs and numerous other splicing factors (reviewed by Brow 2002; Matlin and Moore 2007; Staley and Woolford 2009). In the event that an intron does not exceed ~200–250 nts, the spliceosome initially assembles across the intron

(Fox-Walsh et al. 2005) (Fig. 4B). In the earliest cross-intron spliceosomal complex (i.e., the E complex), the U1 snRNP is recruited to the 5′ss and non-snRNP factors such as SF1/mBBP and U2AF interact with the BS and PPT, respectively. In a subsequent step, the U2 snRNP stably associates with the BS, forming the A complex (also denoted prespliceosome). The U4/U6.U5 tri-snRNP, which is pre-assembled from the U5 and U4/U6 snRNPs, is then recruited, generating the pre-catalytic B complex. Major rearrangements in RNA–RNA and RNA–protein interactions, leading to the destabilization of the U1 and U4 snRNPs, give rise to the activated spliceosome (i.e., the Bact complex). Subsequent catalytic activation by the DEAH-box RNA helicase Prp2, generates the B* complex, which catalyzes the first of the two steps of splicing. This yields the C complex, which in turn catalyzes the second step. The spliceosome then dissociates and, after additional remodeling, the released snRNPs take part in additional rounds of splicing (Will and Lührmann 2011).

1.8 Prp19/CDC5L complex

The protein composition of mammalian spliceosomes has been studied using complexes purified from HeLa nuclear extracts by a combination of gel filtration and affinity chromatography (Reed 1990; Bennett et al. 1992); mass spectrometry and database searches (Neubauer et al. 1998). A conserved subspliceosomal complex was identified and it is known as Prp19/CDC5L complex (Pso4-complex) in mammals (Ajuh et al. 2000; Makarova et al. 2002; Grote et al. 2010; Burns et al. 1999; Chen et al. 1999; McDonald et al. 1999). Depletion of this complex leads to a block in splicing before the first transesterfication step in splicing and lariat formation of the pre-mRNA (Makarova et al. 2002). Prp19/CDC5L complex consist of seven proteins: hPrp19, CDC5L, PRLG-1, AD002, SPF27, CTNNBL1 (β-catenin-like 1), and HSP73 (Makarova et al. 2004). CDC5L, PRLG-1, PRP19, and SPF27 form a stable subcomplex of the hPrp19/CDC5L complex (Grote et al. 2010).

In mice the homozygous null mutation in prp19 gene is early embryo lethal. Indeed, homozigous knockout embryos die at the blastocyst stage (Fortschegger et al. 2007). Also the homozygous null mutation in PLRG-1 gene is early embryo lethal (Kleinridders et al. 2009).

snRNP200 (U5) is required for spliceosome assembly and its depletion can block the second step of splicing (Umen and Guthrie 1995). The omozigous knockout of snRNP200 induce complete embryonic lethality before implantation (J:204739 International Knockout Mouse Consortium, MGI download of modified allele data from IKMC and creation of new knockout alleles. Database Download. 2014).

Another important protein in the splicing process is XAB2 (XPA-binding protein 2) is involved in the processes associated with cell cycle control and pre-mRNA splicing (Nakatsu et al. 2000; Dix et al. 1999; McDonald et al.

1999). XAB2-homozygous mutants could survive until the morula stage, but could not develop to the blastocyst stage (Yonemasu et al. 2005).

1.9 Nuclear Speckles

In mammalian cells the pre-mRNA splicing factors, including small nuclear ribonucleoprotein particles (snRNPs), spliceosome subunits, and other non-snRNP protein splicing factors, shows a punctate nuclear localization pattern that is usually termed "a speckled pattern" or named "SC35 domains" (Wansink et al. 1993). The speckles are located in the interchromatin regions of the nucleus. They usually appear as 20–50 irregularly shaped structures that vary in size.

To targeting some of SR pre-mRNA splicing proteins to nuclear speckles is necessary and sufficient the presence of arginine/serine-rich domain (RS domain) (Caceres et al. 1997; Hedley et al. 1995; Li et al. 1991). For speckle-targeting have also been implicated the threonine-proline repeats of SF3b1 (Eilbracht et al. 2001) and the forkhead-associated domain in NIPP1 (Jagiello et al. 2000)

Although they apparently contain few, if any, genes speckles are often observed close to highly active transcription sites. This suggests that they likely have a functional relationship with gene expression, and some specific genes have been reported to preferentially localize near speckles (Brown et al. 2008; Huang et al. 1991; Johnson et al. 2000; Moen et al. 2004; Smith et al. 1999; Xing et al. 1993; Xing et al. 1995),

However in the literature there are apparent discrepancies concerning the possible direct role of speckles as splicing sites.

Although many in the field do not view speckles as direct transcription/pre-mRNA splicing centers, others suggest that they may have a more direct role relating to the splicing and transport of pre-mRNA (reviewed in Hall et al. 2006; Melcak et al. 2000; Shopland et al. 2002; Wei et al. 1999).

Many pre-mRNA splicing factors, including snRNPs and SR proteins (Fu 1995), have been localized to nuclear speckles. In addition, several kinases (Clk/STY, hPRP4, and PSKHI) (Brede et al. 2002; Colwill et al. 1996; Ko et al. 2001; Kojima et al. 2001; Sacco-Bubulya et al. 2002) and phosphatases (PP1) (Trinkle-Mulcahy et al. 1999; Trinkle-Mulcahy et al. 2001) that phosphorylate/dephosphorylate components of the splicing machinery have also been localized to nuclear speckles.

The proteomic information, together with additional localization studies, has revealed that speckles contain many other proteins apart from pre-mRNA splicing factors. Of particular interest is the localization of transcription factors (Larsson et al. 1995; Mortillaro et al. 1996; Zeng et al. 1997), 3'-end RNA processing factors (Krause et al. 1994; Schul et al. 1998), eukaryotic translation initiation factor eIF4E (Dostie et al. 2000), eif4AIII, a protein involved in translation inhibition (Li et al. 1999), and structural proteins (Jagatheesan et al. 1999; Nakayasu and Ueda 1984; Sharma et al. 2010).

2. AIM OF STUDY

Dnajc17 was identified as a candidate of congenital hypothyroidism (CH) with thyroid dysgenesis (TD) and the homozygous null mutation in DnajC17 gene are early embryo lethal. Embryonic lethality makes it difficult to study the functions of Dnjc17 in the various tissues. Thus, to elucidate the functional role of DNAJC17 in thyroid I generated a conditional knock-out mouse in which Dnajc17 has been disrupted specifically in the thyroid.

Moreover, as DNAJC17 molecular function is unknown, I performed detailed characterization studies in vitro. To this purpose I used an inducible in vitro system cells exhibiting tetracycline-inducible expression of DNAJC17 fused to GFP protein (GFP-DNAJC17). This cell line allowed me to study the subcellular localization of Dnajc17, its interactors and its targets.

DNAJC17 cellular localization was studied by immunofluorescence assays.

The interactors of DNAJC17 were identified by co-IP followed by Mass Spectrometry (MS). Interactors were then validated by co-immunoprecipitation and subcellular co-localization of DNAJC17 and its interactors was analyzed by Immunofluorescence Assays.

As DNAJC17 has an RNA-recognition motif (RRM domain) and about one-half of its interactors identified by MS are involved in splicing, I verified if Dnajc17 plays a role in gene expression and in the alternative mRNA splicing. For this purpose I performed an RNA-seq analysis in Dnajc17-depleted HeLa cells. The whole transcriptome of DNAJC17kd cells was compared with that of DNAJC17wt cells to identify up-regulated and down-regulated genes. The RNA-Seq data were also analyzed to identify alternatively-spliced genes differentially expressed between $DNAJC17_{kd}$ versus $DNAJC17_{wt}$. I am currently confirming and characterizing differentially spliced genes.

To analyze homozygous dnajc17 knockout embryos I setup a protocol to collect embryos at the morula stage. Total RNA obtained by collected embryos will be used to perform a RNA-Seq experiment to identify the genes altered by the absence of Dnajc17 and presumably responsible for the early embryo lethality of knockout mice.

3. MATHERIALS AND METHODS

3.1 Establishment of Stable HeLa Cell Lines

We used cDNA from sv129 mice tissue amplified by PCR with specific primers for Dnajc17. The resulting cDNA was used to obtain a fusion cDNA where Dnajc17 is in frame with GFP. Resulting chimeric cDNA and the cDNA containing only GFP were cloned into pCDNA5/FRT/TO (Invitrogen) to obtain stable HeLa cell lines exhibiting tetracycline-inducible expression of DNAJC17 tagged with GFP at the N-terminus (GFP-DNAJC17 cells) or only GFP protein (GFP cells) from a specific genomic location. This is a 5.1 kb expression vector Contains a hybrid promoter consisting of the human cytomegalovirus immediate-early (CMV) promoter and tetracycline operator 2 (TetO2) sites for high-level tetracycline-regulated expression in a wide range of mammalian cells. The TetO sites are 3' of transcription initiation at pCMV (downstream). TetO2 sequences consist of 2 copies of the 19 nucleotide sequence: 5′-TCCCTATCAGTGATAGAGA-3′ separated by a 2 base pair spacer. It serves as binding sites for 4 Tet repressor molecules (comprising two Tet repressor homodimers). Contains a single FRT site immediately upstream of the hygromycin resistance gene for Flp recombinasemediated integration and selection of the pcDNA5/FRT/TO plasmid. Used in co-transfection into Flp-In T-REx HeLa host cells with pOG44 (Invitrogen) that contains the FLP gene encoding the Flp recombinase to mediate integration of the expression plasmid into the genome.. Stable transfectants were selected by adding hygromycin to the culture medium

3.2 Immunoblotting

Whole-cell lysates of HeLa cells, GFP-DNAJC17 and GFP cells were prepared in sample buffer and normalized for equal protein concentration by Pierce BCA Protein Assay Kit (Thermo Scientific). A 20 μg amount of protein samples was resolved on a precast NuPAGE 4-12% Bis-Tris Gel (Life Technologies) and transferred on a polyvinylidene difluoride (PVDF) membrane (Millipore). Nonspecific binding sites were blocked by incubation with 5% nonfat dry milk in TBS (TrisHCl 20 MM pH 7.6, NaCl 140 mM) 0.1% Tween 20. Immunodetection was performed by using rabbit polyclonal antibodies against DNAJC17, and monoclonal GFP antibody (Chromoteck). Immune complexes were detected by enhanced chemiluminescence as instructed by the manufacturer (Pierce).

3.3 Co-immunoprecipitation

GFP-DNAJC17 fusion protein or GFP protein expression were induced 16h according to manufacturer instructions (Invitrogen K1020).
4 confluent 15cm plates of GFP-DNAJC17 or GFP were harvested in PBS, Spin down for 5' at 1000g and resuspended on 1 mL of Lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 10mM EDTA, 2 mM MgCl2, 0.1% triton X100). After 30' incubation in ice, extracts were passed several times through a 20g 1and ½ needle. Protein extracts were clarified at 13000g for 10' and transferred in clean tubes. 10 µL of GFP trap (chromotek gta-20) were added to the protein extract and the tubes were incubated in rotation for 2h at 4°C. GFP trap beads were pulled down at 1000g for 5' and washed 5 times with washing buffer (50 mM Tris pH 7.4, 200 mM NaCl, 10mM EDTA, 2 mM MgCl2, 0.1% triton X100).
After the washes beads were boiled in 50 µL of 1X LDS (Invitrogen NP0008), 1X sample reducing agent (Invitrogen NP0004). 5 µL of eluate for each sample were separated on 4-12% Bis-tris acrylamide gel (Invitrogen NP0321BOX) and silver stained. The remaining 45µL were stored for mass spectrometry protein identification.

3.4 In gel digestion, mass spectrometry analysis and protein identification

Eluated were separated by SDS-PAGE on a 12% gel, which was then stained with colloidal Coomassie G-250 (Candiano et al. 2004). Gel image was acquired by using an Image Scanner III densitometer (GE Healthcare Life Sciences). Each lane corresponding to 0.2M, 0.3M and 0.4M samples was manually cut into 23 slices. Slices were excised from gels, triturated and washed with water. Proteins were in-gel reduced, S-alkylated and digested with trypsin as previously reported [D'Ambrosio et al. 2008]. Digest aliquots were removed and subjected to a desalting/concentration step on ZipTip mC18 disk (Millipore) using 5% formic acid/50% acetonitrile as eluent before nanoLC-ESI-LIT-MS/MS analysis.
Digests were analyzed by nanoLC-ESI-LIT-MS/MS using a LTQ XL mass spectrometer (Thermo Finnigan, San Jose, CA, USA) equipped with Proxeon nanospray source connected to an Easy-nanoLC (Proxeon, Odense, Denmark). Peptide mixtures were separated on an Easy C18 column (100 x 0.075 mm, 3 mm) (Proxeon, Odense, Denmark) by using a linear gradient of acetonitrile containing 0.1% formic acid in aqueous 0.1% formic acid; acetonitrile was ramped from 5% to 40% over 60 min, and from 40% to 80% in 10 min, at a flow rate of 300 nL/min. Spectra were acquired the range m/z 400-2000. Acquisition was controlled by a data-dependent product ion scanning procedure over the three most abundant ions, enabling dynamic exclusion (repeat count 1 and exclusion duration 1 min). The mass isolation window and collision energy were set to m/z 3 and 35%, respectively. Duplicate analysis of each sample was performed.

Proteome Discoverer 1.3 platform (ThermoFisher, San Jose, CA, USA) was employed to search raw mass data against an updated mouse Uniprot non-redundant sequence database (76058 protein sequences) using Mascot (Matrix Science, UK) algorithms by using a mass tolerance value of 2.0 Da for precursor ion and 0.8 Da for MS/MS fragments, trypsin as proteolytic enzyme, a missed cleavages maximum value of 2 and Cys carbamidomethylation as fixed modification, Met oxidation and pyro-Glu (N-term Q) variable modification. Candidates with more than 2 assigned unique peptides with an individual MASCOT score >25 were further considered for protein identification.

3.5 Immunofluorescence Assay

Cells were fixed with 3,7 % paraformaldehyde and permeabilized with PBS1X-Triton 0,1%. For detection of protein, we used rabbit polyclonal DNAJC17 antibody, PRP19 mouse polyclonal antibody (LifeSpan BioSciences), CDC5L mouse monoclonal anibody (Santa Cruz Biotechnology), PLRG-1 goat polyclonal antibody (Novus Biologicals), HELIC2 (snRNP 200) goat polyclonal antibody (Santa Cruz Biotechnology), SC-35 rabbit monoclonal antibody (Sigma). Secondary antibodies were mouse, human or goat IgG conjugated to a fluorochrome (Alexa Fluor). Antibody incubations were for 1 h at RT in PBS1X with BSA (2%) and Triton X-100 (0.1%). Coverslips were mounted with mounting medium for fluorescence with DAPI (Vectashield) or DRAQ5 (Cell Signaling). Confocal microscopy and overlay analysis were performed on a Zeiss LSM510 Confocal Laser Scanning Microscope with a x63 objective.

3.6 Image analysis

Colocalization analysis was performed using JACoP plug-in embedded in the visualization and analysis software ImageJ (Bolte and Cordelières, 2006). Analysis was performed on a similar-sized symmetrical region of interest (ROI) selected for each dye. Each colored image was split into respective red, green, and blue channels. The comparative degree of colocalization was calculated as mean Pearson's R coefficients and Coste's randomization P value on the red and green channels using the embedded colocalization analysis plug-in at default settings .

3.7 RNA interference

The HeLa cell line was grown in Dulbecco's modified Eagle medium (Euroclone) supplemented with 10% fetal calf serum (Hyclone). Cells were plated (18*10^4) in a 6 well plate in antibiotic-free complete medium and transfected the following day with 50nM Dnajc17 siRNA (ON-TARGETplus Human Dnajc17 siRNA, cat n° L-021141) and 50 nM siRNA

negative control (ON-TARGETplus Non-targeting Pool, cat n° # D-001810) in INTERFERin transfection reagent (Polyplus 409-10) following the manufacturer's protocol; three technical replicates were prepared for each condition.
Cells were harvested 96 hours after transfection and total RNA was extracted for RNA-Seq transcriptome analysis.

3.8 E1A splicing assay

GFP-DNAJC17, HeLa-GFP or HeLa trasfected with specific siRNA to knockdown Dnajc17, were transfected with E1A minigene (Yang et al. 1994) using Lipofectamine 2000 (Invitrogen), according to manufacturer's instruction. After transfection, at the medium was added or not 2mg/ml of tetracycline and the cells were collected after 16hrs for RNA extraction. 1 mg of total DNA-free RNA was used for cDNA generation by using M–MLV reverse transcriptase (Promega). The oligos RT-PCR was performed as previously described (Yang et al. 1994). The specific primers used for RT-PCR were ElA569, 5'-ATTATCTGCCACGGAGGTGT-3'and ElA1315, 5'-GGATAGCAGGCGCCATTTTA-3'.

3.9 RNA sequencing

Differential gene expression was analysed in DNAJC17 interfered HeLA cells by RNA sequencing (seq) approach. RNA was extracted from three independent samples for both knockdown and wild type conditions and sequenced on HiSeq1500 (Illumina) according to manufacturer's protocols. About 7x107 high quality sequence reads were produced for sample with a mean quality score ≥ 30 (99.9% base quality accuracy). The human Ensembl GRCh37 genome was used as reference for read mapping, performed by TopHat aligner (Trapnell et al. 2009). Properly mapped reads (about 5x107) were further processed using Cufflinks package (Trapnell et al. 2010). Selected reads were assembled in transcripts, which were compared among the six samples and with known annotations to generate a unique final transcriptome. Resulting 20825 genes were then tested for differential expression between $DNAJC17_{kd}$ vs $DNAJC17_{wt}$ and filtered for corrected p-value (FDR) ≤ 0.05 and absolute fold change ≥ 1.5 (884 significantly deregulated genes). In addition to the simple changes in gene expression, the RNA-Seq analysis allowed the identification of differentially spliced genes. All isoforms derived from the same pre-mRNA were investigated to calculate their relative abundance that represents the alternative splicing of the gene. This measurement was compared between the two conditions with Jensen-Shannon test to verify if in $DNAJC17_{kd}$ one or more splicing isoforms were more representative of the gene than in $DNAJC17_{wt}$. The differential splicing test resulted significant for 287 genes (FDR ≤ 0.05).

3.10 Generation of Dnajc17 conditional knockout mouse

The knock-in targeting vector to generate Dnajc17 conditional KO mice was purchased from EUCOMM CONSORTIUM http://www.eucomm.org (Fig. 6). The knock-in targeting vector contained a 4605 bp 5' homology arm, a FRT-flanked cassette driven by Dnajc17 promoter containing both a LacZ reporter gene and a neomycin resistance gene for positive selection (both separated by a LoxP sequence), 2 loxP sites flanking exon 2, 3 and 4 of Dnajc17 gene as well as a 4876 bp 3' homology arm.
Removal of the FRT cassette in the Dnajc17 knockin mice was achieved by constitutive FLP-mediated recombination. Subsequently, an additional round of breedings was performed in order to remove Flp transgene from the progeny. At this point, the resulting Dnajc17 loxP (floxed) mice were ready to be bred with Pax8-Cre mice (Bouchard et al. 2004). As a consequence of Cre-mediated recombination of exon 2, 3 and 4, the Dnajc17 knockout transcript lacks exon 2, 3 and 4 mainly in thyroid and kidney. Absence of a protein in Dnajc17 conditional knockout mice will be verified by western blot and immunohistochemistry.

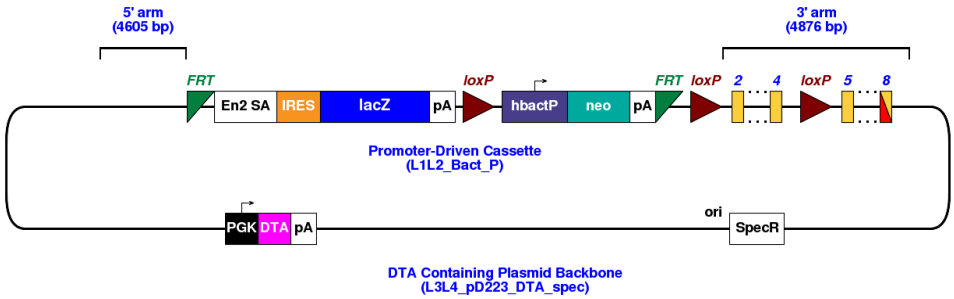

Figure 6. knock-in targeting vector.

Mouse genotyping was performed with genomic DNA samples isolated from tail biopsies that were taken from each animal at weaning. Standard PCR protocols were used to amplify the wildtype, knockin, floxed and knockout alleles of the Dnajc17 gene, as well as to detect the presence of Cre or Flp. The specific primers used in each case are indicated in table 2.

Cre	AACGCACTGATTTCGACC	CAACACCATTTTTTCTGACCC
Flp	CACCTAAGGTCCTGGTTCGTCAGT	CCAGATGCTTTCACCCTCACTTAG
Wildtype and floxed	GCTTGGCTAGTTCCACTA	GGACAATGTGAAAGCCTAAGC
knockin	GCTTGGCTAGTTCCACTA	CACAACGGGTTCTTCTGTTAGTCC

Table 2. Mouse genotyping primers.

4. RESULTS

4.1 Construction of DNAJC17 inducible-expression system

The specific function of Dnajc17 is still poorly characterized, for this reason we generated an in vitro system cells exhibiting tetracycline-inducible expression of DNAJC17 fused to GFP protein (GFP-DNAJC17). To obtain this cell line we used cDNA from sv129 mice tissue in Flp-In T-REx System (Flip In TRex, Invitrogen).

In this system, expression of Gfp-Dnajc17 gene is repressed in the absence of tetracycline and induced in the presence of tetracycline. In the absence of tetracycline, the Tet repressor forms an homodimer that binds with extremely high affinity to each Tet operator (TetO2) sequence in the promoter controlling expression of the Gfp-Dnajc17 gene repressing the gene transcription. Upon addition, tetracycline binds with high affinity to each Tet repressor homodimer and causes a conformational change in the repressor that renders it unable to bind to the Tet operator.

The Tet repressor:tetracycline complex then dissociates from the Tet operator allows induction of transcription from the Gfp-Dnajc17 gene (Fig. 7A). As control, we also generated stable HeLa cell lines exhibiting tetracycline-inducible expression of GFP (GFP cells line).

Western blot analysis of whole HeLa, GFP-DNAJC17 and GFP cells extracts (Fig 7B) and immunofluorescence (Fig. 7C) confirmed the expression of DNAJC17-GFP in a tetracycline dose-dependent manner. Immunofluorescence at the same time points confirmed the nuclear localization of DNAJC17-GFP while GFP alone localizes also in the cytoplasm (Fig. 7C). For all findings showed in this work we used 16h of tet treatment.

These data confirm not only the distinct nuclear localization of GFP-DNAJC17 fused protein but allow me to use an in inducible *in-vitro* system to study the function of Dnajc17 gene.

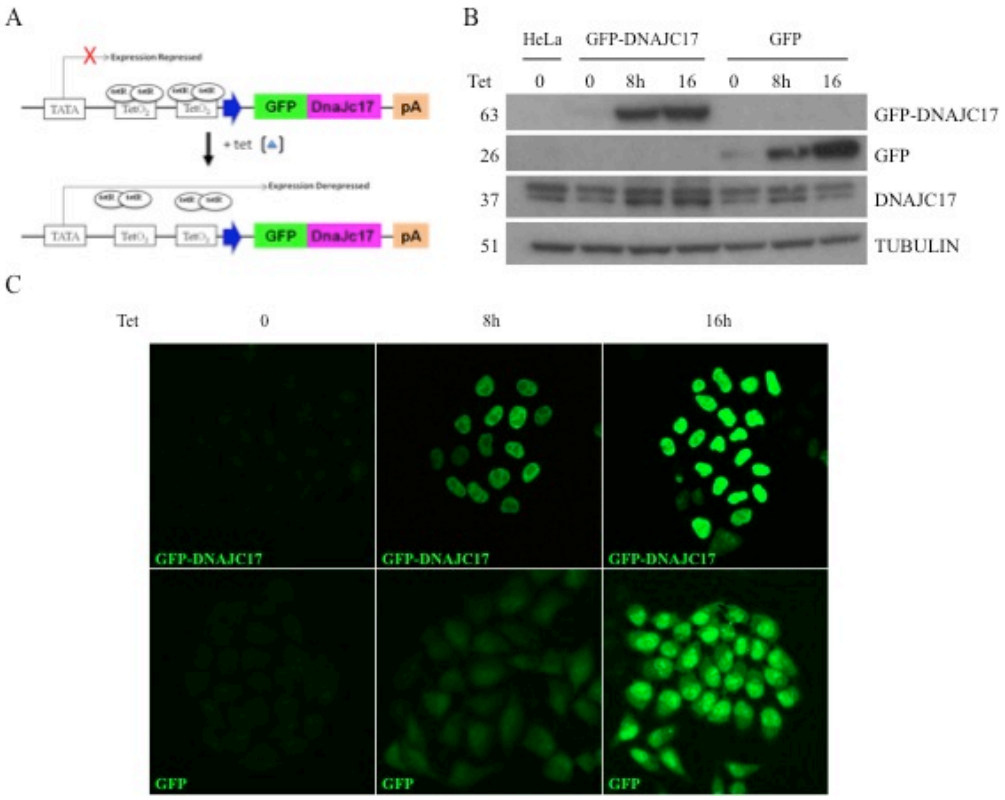

Figure 7. HeLa cell lines exhibiting tetracycline-inducible expression of DNAJC17 fused to GFP protein or GFP. The binding of tetracycline to the Tet repressor that active the promoter controlling expression of the GFP-DNAJC17 gene (A). Expression of GFP-DNAJC17 and GFP gene is repressed in the absence of tetracycline and induced in the presence of tetracycline. Western blot analysis of whole HeLa, GFP-DNAJC17 and GFP cells extracts (B) and immunofluorescence (C) confirmed the expression of GFP-DNAJC17 in a tetracycline dose-dependent manner. Immunofluorescence at the same time points confirmed the nuclear localization of DNAJC17GFP while GFP alone localize also in the cytoplasm (C).

4.2 DNAJC17 interactome includes several proteins involved in splicing

To identify DNAJC17 interacting proteins, we used GFP-DNAC17 cells and GFP cells as negative control in Co-IP assay. GFP-DNAJC17 fusion protein or GFP protein expression were induced 16h and whole lysate was immunoprecipitated using GFP trap beads (chromotek). Eluated were processed by mass spectrometry. 63 proteins were identified, and the apparent observed molecular mass and isoelectric point for all identified proteins matched very well with their calculated values. A Map of the Interactome Network shows the proteins interacting with DNAJC17 (Fig. 8).

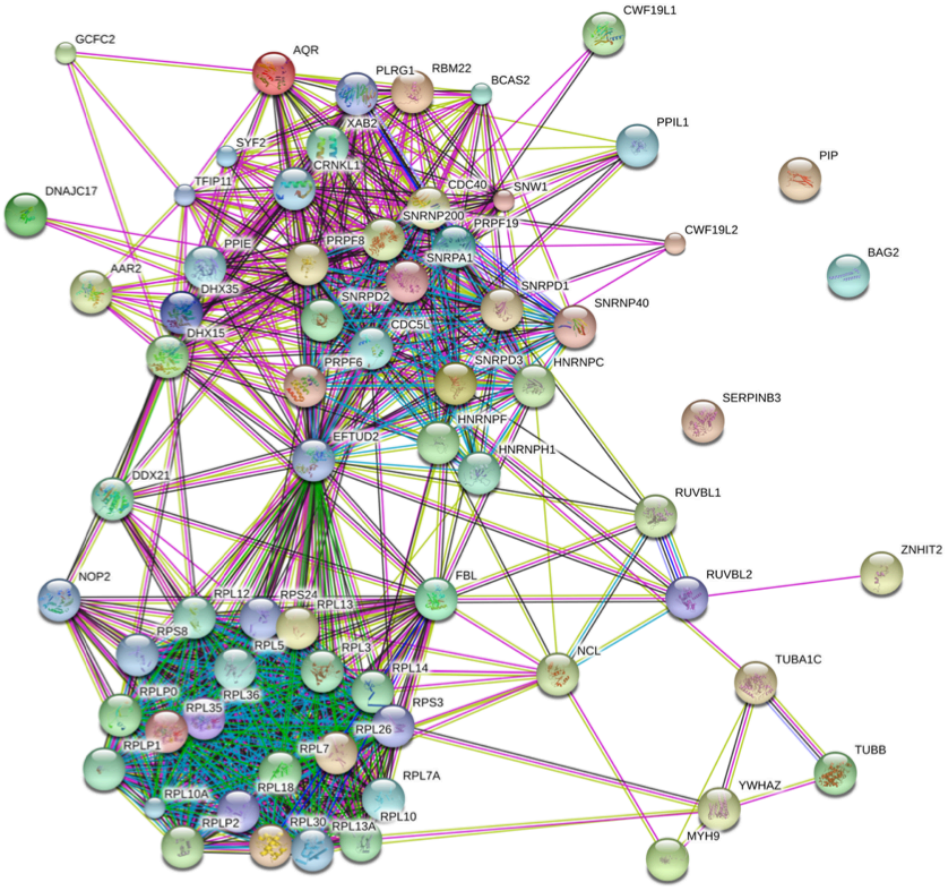

Figure 8. Map of the Interactome Network showing the proteins interating with GFP-DNAJC17 protein fusion (STRING software). 63 proteins was identified as interactors of DNAJC17. Most of these proteins could be classified into two general groups according to their cellular function: spliceosome and ribosome.

Most of these proteins could be classified into two general groups according to their cellular function: spliceosome and ribosome. The largest group of proteins consists of factors involved in splicing. A total of 24 proteins have been identified within this category. Among these 24 proteins we identified 4

proteins playing essential roles in early embryogenesis similarly to Dnajc17 (PPR19, PLRG-1, snRNP200, XAB2). (Tab. 3).

Accession	Gene names	Function	Localization	Knockout mice	References	Human diseases	References
O75643	SNRNP200	Prp19/CDC5L complex	Nucleus	Embryonic lethality before implantation	J:204739 International Knockout Mouse Consortium	Retinitis pigmentosa 33	Zhao et al. 2009
Q9UMS4	PRP19	Prp19/CDC5L complex	Nucleus	Embryonic lethality before implantation	Fortschegger et al., 2007		
A8MW61	PLRG-1	Prp19/CDC5L complex	Nucleus	Embryonic lethality before implantation	Kleinridders et al., 2009		
B4DSH1	CDC5L	Prp19/CDC5L complex	Nucleus			Bilateral multicystic renal dysplasia	Groenen et al. 1998
Q68CN2	XAB2	Splicing factor	Nucleus	Embryonic lethality before implantation	Yonemasu R et al. 2005		

Table 3. Proteins selected for validation. From mass spectrometry 4 proteins have been identified as splicing factors essential in embryogenesis like Dnajc17 (PRP19, PLRG-1, snRNP200, XAB2). CDC5L belong to the Prp19/CDC5L complex and was identified in a human disease: the bilateral multicystic renal dysplasia.

PLRG-1 and PRP19 are nuclear proteins component of a complex containing other splicing factors (Grote M et al., 2010; Ajuh et al., 2000; Burns et al., 1999; Chen et al., 1999; McDonald et al., 1999; Neubauer et al., 1998). This conserved subspliceosomal complex is known as Prp19/CDC5L complex in mammals (Ajuh et al., 2000; Makarova et al., 2002; Grote M et al., 2010). Depletion of this complex leads to a block in splicing before the first transesterfication step in splicing and lariat formation of the pre-mRNA (Makarova et al., 2002). PRP19 e PLRG-1 belong to the subcomplex of the hPrp19/CDC5L complex consisting of CDC5L, PRLG1, PRP19, and SPF27 (Grote M et al., 2010). Also CDC5L e SPF27 have been selected by mass spectrometry (Tab. 3). snRNP200 (U5) is required for spliceosome assembly and its depletion can block the second step of splicing (Umen JG, Guthrie C, 1995). XAB2 (XPA-binding protein 2) is involved in the processes associated with cell cycle control and pre-mRNA splicing (Nakatsu et al., 2000; Dix et al., 1999; McDonald et al., 1999).
As DNAJC17 main localization is nuclear, we decided to focus on interactors involved in splicing.

4.3 DNAJC17 interacts with several splicing proteins in Vivo

To test whether the interactions observed *in vitro* by IP assays could be observed also *in vivo*, immunoprecipitation on whole protein extracts and co-immunofluorescence of HeLa modified cells were performed. GFP-DNAJC17 fusion protein or GFP were induced 16h and GFP trap beads was used to immunoprecipitate the recombinant proteins from the total protein extract prepared from GFP-DNAJC17 cells and from GFP cells used as control.
Subsequently, the bound proteins were subjected to Western blot analysis after separation by SDS-PAGE. Western blot developed with a specific anti-target protein (Fig. 9) confirmed the presence of the tested interactors (PRP19, PLRG-1, CDC5L, snRNP200 and XAB2) in the co-immunoprecipitated proteins, only in the extract prepared from GFP-DNAJC17 cells treated with

tetracycline.

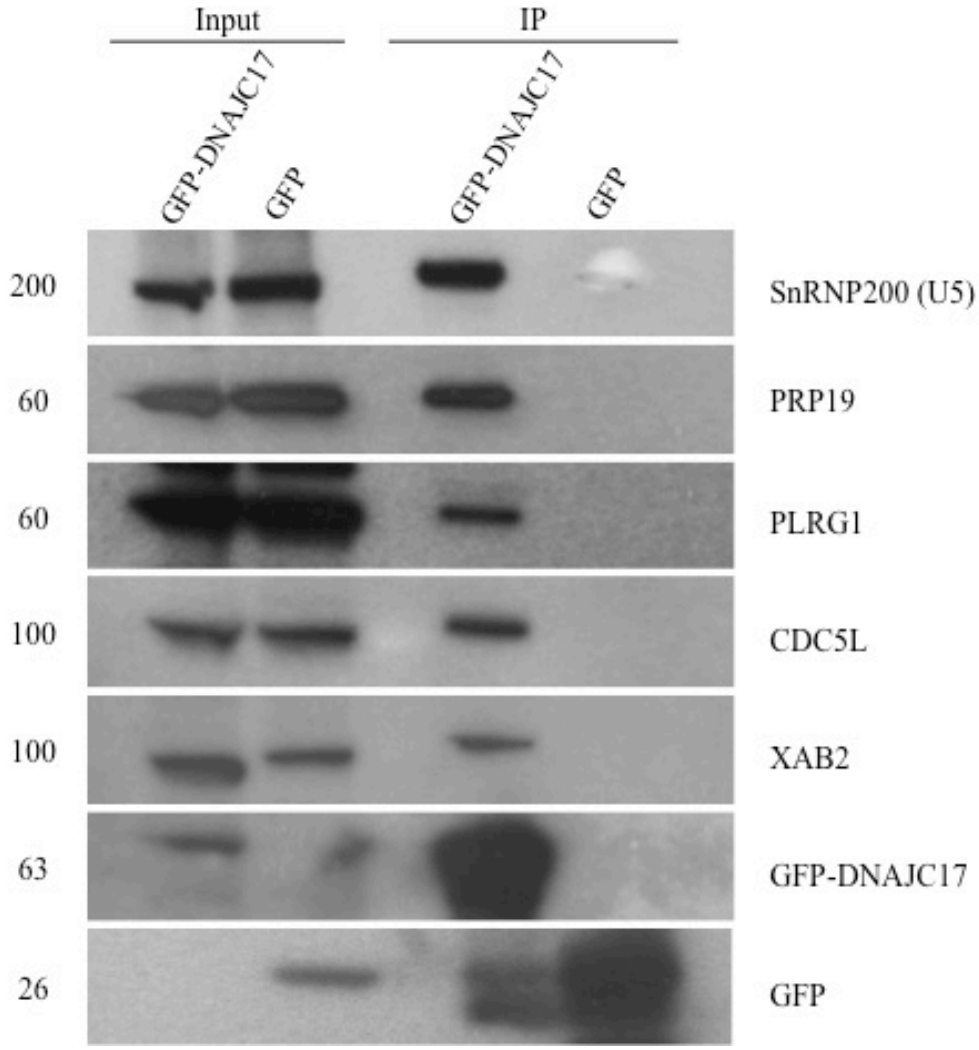

Figure 9. Identification of DNAJC17-interacting partners. GFP-DNAJC17 and GFP cells were treated with tetracicline for 16h and cell lysates were immunoprecipitated with GFP-Trap®_A affinity beads (Chromoteck) and bound proteins were eluted and analysed by Western blotting with the indicated antibodies.

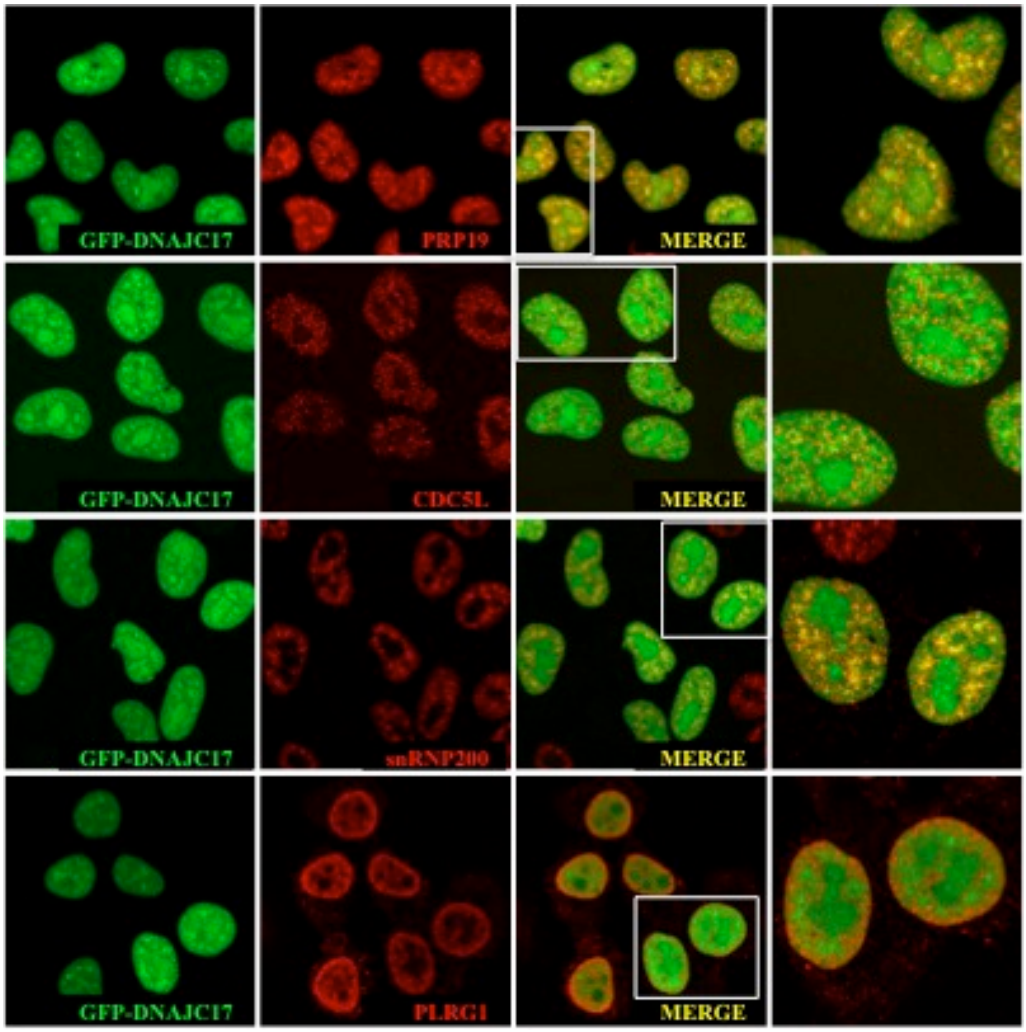

Figure 10. Colocalization images in GFP-DNAJC17 cells. GFP-DNAJC17 cells stably transfected with GFP-DNAJC17 were induced with tetracycline end after 16h fixed and permeabilized. Each panel is composed of 4 images: the first one is GFP-DNAJC17; the red showing the staining with anti-PRP19, anti-CDC5L, anti-snRNP200 (u5) or anti-PLRG1 (secondary antibody: Alexa Fluor 555); the third image is a merged picture of GFP-DNAJC17 and the red-stained PRP19, CDC5L, snRNP200 or PLRG1. The last column shows a detail of the merged image. Colocalization is represented with increasing intensity of yellow in the third and last images.

Furthermore, subcellular co-localization was observed by immunofluorescence (Fig. 10).

Confocal images were analyzed with the ImageJ software using the JACoP plugin calculating the Pearson's coefficient (Bolte and Cordelières, 2006) and following the Coste's approach (Bolte and Cordelières, 2006). The immunofluorescence assay shows that DNAJC17 colocalizes with PRP19 (mean Pearson's coefficient: 0,805 ± 0,057; Coste's randomization P value: 100%), PLRG-1 (mean Pearson's coefficient: 0,739 ± 0,050; Coste's randomization P value: 100%), CDC5L (mean Pearson's coefficient: manca da fare), snRNP200 (mean Pearson's coefficient: 0,651 ± 0,057; Coste's

randomization P value: 100%). These data confirm that DNAJC17 shares the same compartment of PRP19, PLRG-1, CDC5L and snRNP200 and suggest a possible role of DNAJC17 in modulating splicing activity.

4.4 DNAJC17 co-localizes with the splicing complex 35, a nuclear speckle molecular machinery marker

The nucleus contains various types of subnuclear structures, such as nucleoli, speckles particles, Cajal bodies, each having different nuclear activities. Based on the morphological appearance I speculated that DNAJC17 might be localized with the speckles that are nuclear domains enriched in pre-mRNA splicing factors, located in the interchromatin regions of the nucleoplasm of mammalian cells. An immunofluorescence analysis was performed on GFP-DNAJC17 cells, using SC35 as marker of nuclear speckles (Fig. 11). These findings establish that DNAJC17 co-localizes with SC35 (mean Pearson's coefficient: 0,648 ± 0,035; Coste's randomization P value: 100%). This data support the idea that DNAJC17 could be involved in maintaining folding and stability of proteins stored in the nuclear speckles dynamic complex and mainly involved in splicing machinery. Furthermore I also demonstrated by immunofluorescence on GFP-DNAJC17 cells that DNAJC17, SC35 and PRP19 colocalize (mean Pearson's coefficient: 0,722 ± 0,052; Coste's randomization P value: 100%) (Fig. 11). DNAJC17 colocalizes with SC35 spekles and SC35 colocalizes with PRP19 that is a nuclear proteins component Prp19/CDC5L complex (Pso4-complex) in mammals.

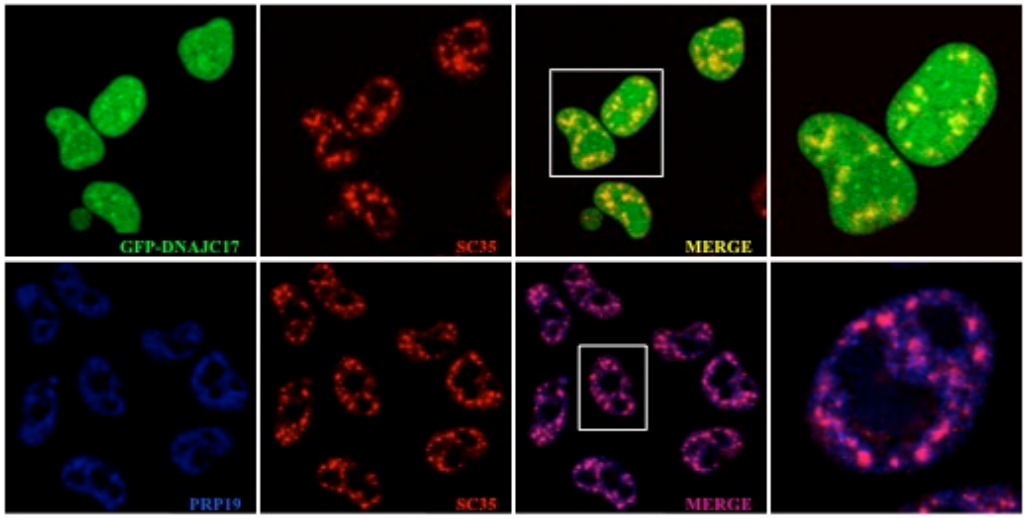

Figure 11. DNAJC17 and PRP19 colocalize with SC35 speckles. GFP-DNAJC17 cells stably transfected with GFP-DNAJC17 were induced with tetracycline end after 16h fixed and permeabilized. Cells were stained with anti-SC35 (upper image) or anti-PRP19 and anti-SC35 (lower image). The SC35 antibody was detected with Alexa Fluor 555, while the PRP19 antibody was detected with Alexa Fluor 647. Colocalization is represented with increasing intensity of yellow (upper) or pink (lower) in the third and last images.

4.5 E1A splicing assay

To test whether DNAJC17 modulates alternative splicing, GFP-DNAJC17 and GFP cells, as a control, were transfected with E1A minigene (Yang et al. 1994) in presence (+) or not (-) of tetracycline. RT-PCR analysis showed that overexpression of DNAJC17-GFP did not alter E1A alternative splicing (Fig. 12A).

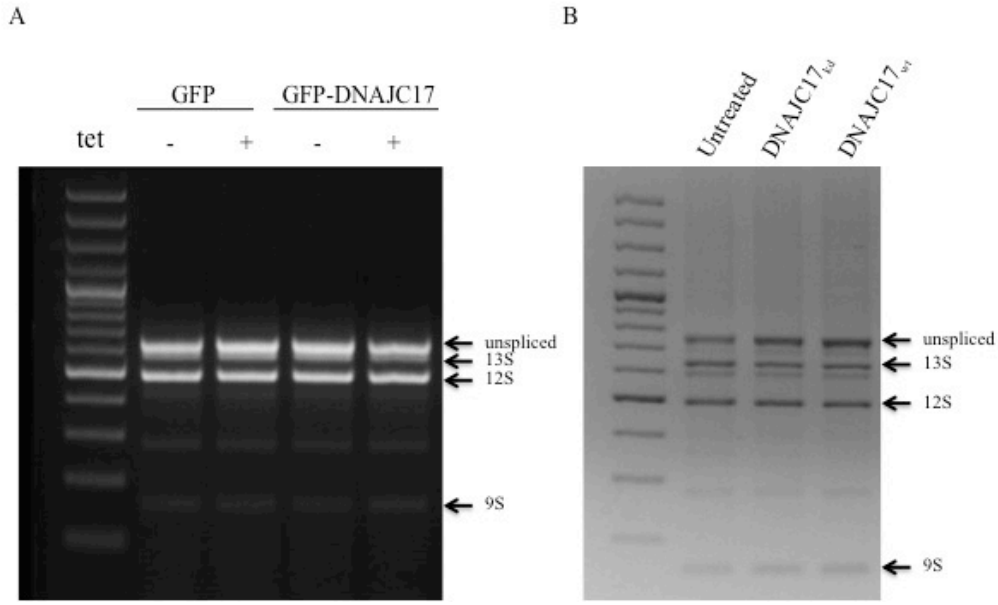

Figure 12 . In vivo splicing assays using GFP-DNAJC17, GFP and DNACJ17-depleted HeLa cells. GFP-DNAJC17 and GFP cells, as a control, were transfected with E1A minigene in presence (+) or not (-) of tetracicline. Total RNA was extracted and analyzed for ElA splicing by RT-PCR (A). HeLa cells were transfected with specific siRNA to knockdown DNAJC17 (DNAJC17kd) or with scrambles as control (DNAJC17wt). After 24h the cells were transfected with E1A minigene. Total RNA was extracted and analyzed for ElA splicing by RT-PCR.

This assay was performed also by RNA interference approach. HeLa cells were transfected with specific siRNA to knockdown DNAJC17 ($DNAJC17_{kd}$), or with scrambles as control ($DNAJC17_{wt}$), after 48h the cells were transfected with E1A minigene. PCR analysis following 96h after siRNA trasfection showed that silencing of Dnajc17 did not alter E1A alternative splicing (Fig. 12B). Although, this result suggest that DNAJC17 was not directly involved in in splicing regulation of E1A gene, its role in alternative splicing process can not be excluded.

4.6 RNA interference of Dnajc17

To test wether DNACJ17 could have an impact in the splicing on the whole transcriptome, I performed RNA-Seq analysis in HeLa cells depleted of DNACJ17 by RNA interference. To this aim, HeLa cells were transfected with specific siRNA to knockdown Dnajc17. The Western blot analysis following 96h of trasfection revealed the specific decline of protein for siRNA induced Dnajc17 knockdown ($DNAJC17_{kd}$) as compared to its wild type counterpart treated with scramble ($DNAJC17_{wt}$) (Fig. 13).

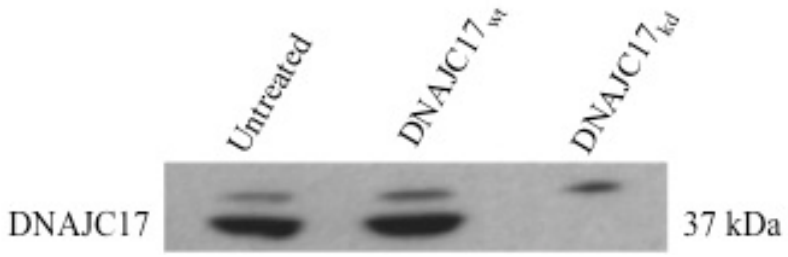

Figure 13. Westernblot analysis of RNA interference. HeLa cells were transduced with specific siRNA to knockdown DNAJC17. The Western blot analysis following 96h of transfection revealed the specific decline of protein for siRNA induced Dnajc17 knockdown ($DNAJC17_{kd}$) as compared to its wild type counterpart treated with scramble ($DNAJC17_{wt}$).

The effect of Dnajc17 silencing on global gene expression was investigated by high-throughput sequencing of RNA (RNA-Seq) and the whole transcriptome of $DNAJC17_{kd}$ cells was compared to that of $DNAJC17_{wt}$ cells . Three replicates for both conditions were sequenced producing about 7×10^7 reads for sample, that were mapped to reference genome (Homo sapiens, GRCh37) using TopHat aligner (Trapnell et al, 2009). Correctly mapped reeds were then processed by Cufflinks (Trapnell et al, 2010) to assembly genes and transcripts and to calculate their relative expression levels. A total of 20825 assembled genes were selected (removing those with too low aligned reads) and further analysed to test differential expression. Genes with corrected p-value ≤ 0.05 and absolute fold change ≥ 1.5 were differentially expressed in knockdown cells compared to wild type. The resulting 359 up-regulated and 525 down-regulated genes represent the significant variation of expression profile in response to Dnajc17 silencing.

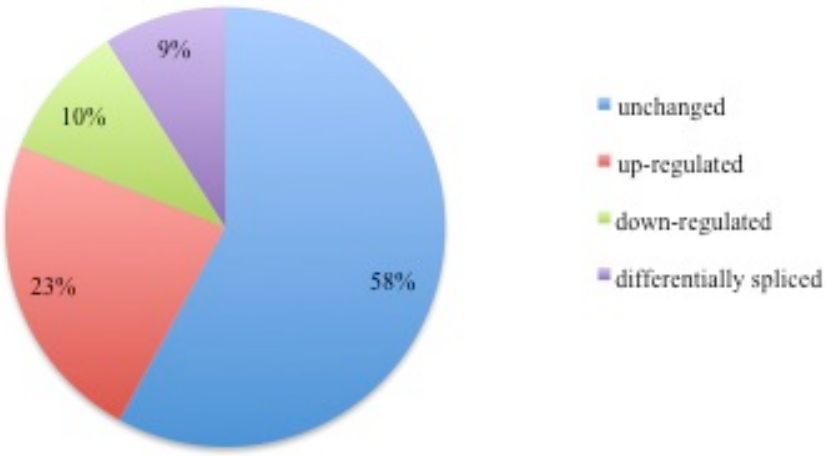

Figure 14. RNA-Seq analysis.

Furthermore, sequencing analysis of transcriptome allows to investigate also qualitative differences across the samples. The expression levels of the isoforms produced by alternative splicing of a single transcript were compared between $DNAJC17_{kd}$ and $DNAJC17_{wt}$. This analysis identified 287 significant differentially spliced genes, showing that Dnajc17 silencing affected the alternative splicing, and suggesting a critical role of DNAJC17 in primary transcripts processing (Fig. 14).

4.7 Generation of Dnajc17 conditional knockout mouse

To study the in vivo role of this Dnajc17, we generated a Dnajc17 conditional KO. The knock-in targeting vector contained a 4605 bp 5' homology arm, a FRT-flanked cassette driven by promoter containing both a LacZ reporter gene and a neomycin resistance gene for positive selection (both separated by a LoxP sequence), 2 loxP sites flanking exon 2, 3 and 4 of Dnajc17 gene as well as a 4876 bp 3' homology arm. Removal of the FRT cassette in the Dnajc17 knockin mice was achieved by constitutive FLP-mediated recombination. Subsequently, an additional round of breedings was performed in order to remove Flp transgene from the progeny. The resulting Dnajc17 loxP (floxed) mice appear normal and fertile.
At this point, the resulting Dnajc17 loxP (floxed) mice were ready to be bred with Pax8-Cre mice (Bouchard et al. 2004). As a consequence of Cre-mediated recombination of exon 2, 3 and 4, the Dnajc17 knockout transcript lacks exon 2, 3 and 4 mainly in thyroid and kidney. Absence of a DNAJC17 protein in Dnajc17 conditional knockout mice will be verified by western blot and immunohistochemistry.

5. DISCUSSION

Congenital hypothyroidism (CH) is the most frequent endocrinological congenital disorder often due to thyroid dysgenesis (TD). Mutations in genes involved in thyroid development such as Nxk2.1, Pax8, Foxe1, and Tshr cause TD in animal models, although in humans loss of function mutations in these genes have been identified only in few cases of CH with TD, suggesting the involvement of additional loci. A polygenic mouse model of CH with TD allowed our laboratory to identify Dnajc17 as a candidate TD modifier gene.

Furthermore, an homozygous truncating mutation in Dnajc17 has been identified in a family with an apparently novel syndrome of retinitis pigmentosa and hypogammaglobulinemia (Patel N. et al. 2015). Suggesting the important role of DNAJC17 in different cellular processes.

The homozygous null mutation in Dnajc17 gene are early embryo lethal. In particular, only Dnajc17$^{-/-}$ two-cell embryos were found, but no homozygous embryos were recovered after the morula stage. These findings indicate that Dnajc17$^{-/-}$ embryos were developmentally arrested before implantation and indicate an essential role of this gene in early mouse development (Amendola E. et al. 2010).

DNAJC17 protein is a member of the type III heat-shock protein-40 (Hsp40) family. Canonical function of Dnaj family is to stimulate the ATPase activity of molecular chaperone DNAJ/HSP70 in various cell function such protein folding, transport, degradation and heat shock response.

Type III proteins have only the J domain as conserved regions. These proteins may not bind to non-native polypeptides and thus should not function as molecular chaperones on their own (Qiu XB 2006).

DNAJC17 three-dimensional structure prediction show that, the J domain is located near to the N-terminus, moreover there is an RNA recognition motif RRM located at C-terminus of protein (Fig.2).

In this thesis work, I performed in vitro studies to shed light on the function of Dnajc17, which is still uncharacterized. In this work, DNAJC17 interactome analysis by CO-IP and mass spectrometry demonstrates that DNAJC17 interacts with 63 proteins. Most of these proteins could be classified into two general groups according to their cellular function: spliceosome and ribosome. The largest group of proteins consisting of factors involved in splicing. A total of 24 proteins have been identified within this category. Among these 24 proteins we identified 4 that form a stable subcomplex of the hPrp19/CDC5L complex (CDC5L, PRLG-1, PRP19, and SPF27). Depletion of this complex leads to a block in splicing before the first transesterfication step in splicing and lariat formation of the pre-mRNA (Grote et al. 2010; Makarova et al. 2002). Furthermore, CDC5L, PRLG1 and PRP19 are essential proteins in embryogenesis like Dnajc17. The interaction and colocalization of DNAJC17 with CDC5L, PRLG-1 and PRP19 was confirmed with CO-IP and immunofluorescence essay. The interaction with the core of the

hPrp19/CDC5L complex suggest a possible involment of DNAJC17 in the correct formation or function of the complex.
snRNP200 (U5) was another protein identified among DNAJC17 interactors. This protein is required for spliceosome assembly and its depletion can block the second step of splicing (Umen and Guthrie 1995). The homozygous knockout of snRNP200 induces complete embryonic lethality before implantation.
These data suggest the involvement of DNAJC17 in splicing mechanism.
Based on the interaction of DNAJC17 with several proteins involved in splicing and on the analysis of the morphological appearance of DNAJC17 localization, I speculated that DNAJC17 might be localized within the nuclear speckles. I confirmed this hypothesis by immunofluorescence assay that demonstrated the colocalization between DNAJC17 and SC35, a spliceosome assembly factor localized in nuclear speckles. Furthermore, DNAJC17 localizes in nucleus.
Taken together these data support the idea that DNAJC17 could be involved in maintaining folding and stability of proteins stored in the nuclear speckles dynamic complex and mainly involved in splicing machinery
However, the E1A minigene splicing assay performed either with overexpression or knockdown of DNAJC17 gave negative results. This does not excludes that DNAJC17 could play a role in the splicing process, as DNAJC17 might bind specific RNA sequences that are not present in E1A minigene, or it might regulates specific splicing factors that are not involved in E1A splicing regulation. An investigation beyond the scope of this study is required to demonstrate the direct role/involvement of DNAJC17 in alternative splicing.
The role of DNAJC17 in gene expression regulation was investigated by RNA interference approach. Sequencing analysis of transcriptome from DNAJC17-interfered cells shows 359 up-regulated and 525 down-regulated genes showing significant variation of expression level in response to Dnajc17 silencing.
Finally, sequencing analysis of expression levels of the isoforms produced by alternative splicing of a single transcript were compared between $DNAJC17_{kd}$ versus $DNAJC17_{wt}$. This analysis identified 287 significant differentially spliced genes, showing that Dnajc17 silencing affected alternative splicing.
These findings suggest a critical role of DNAJC17 in primary transcripts processing.
Furthermore, to elucidate the functional role of DNAJC17 in thyroid I generated a conditional knock-out mouse in which Dnajc17 has been disrupted specifically in the thyroid. I am currently studying the involvement of Dnajc17 in thyroid gland development and physiology.
Finally, to understand the involvement of DNAJC17 in embryonic development I am collecting embryos at the morula stage to analyze homozygous Dnajc17 KO embryos. After the collection, I plan to perform a RNA-Seq experiment to identify the genes altered by the absence of Dnajc17.

6. CONCLUSION

The work described in this thesis is the first attempt to unveil the molecular function of DNAJC17, a housekeeping gene still poorly characterized, The interest for this gene stems from previous studies that identified Dnajc17 as a potential modifier gene for CH, highlighting also its essential role in early embryogenesis (Amendola et al. 2010).
Here we characterized DNAJC17 interactome, showing that a large number of its interactors are components of the splicing machinery. We also analyzed the modifications of gene expression profiles induced by DNAJC17 knockdown, including many differentially spliced transcripts. Taken together, these results strongly suggest that DNAJC17 protein could play a role in the regulation of gene expression possibly by interacting with proteins involved in splicing.
Several experiments are still in progress, including: 1) detailed study of the transcripts spliced differentially in the absence of Dnajc17, 2) generation and characterization of conditional Dnajc17 knock-out mouse, 3) collection of Dnajc17 KO morulae for the analysis of gene expression profiles.
The completion of these studies both in vitro and in vivo, will be essential to elucidate both the role played by Dnajc17 in thyroid and in whole embryo development, and to clarify the molecular mechanisms underlying its biological function.

7. REFERENCES

Ajuh P, Kuster B, Panov K, Zomerdijk JC, Mann M, Lamond AI. *Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry*. EMBO J. 2000; 19:6569–6581. [PubMed: 11101529]

Amendola E, De Luca P, Macchia PE, Terracciano D, Rosica A, Chiappetta G, Kimura S, Mansouri A, Affuso A, Arra C, Macchia V, Di Lauro R, De Felice M. *A Mouse Model Demonstrates a Multigenic Origin of Congenital Hypothyroidism*. Endocrinology. 2005 Dec;146(12):5038-47. [Pubmed: 16150900]

Amendola E, Sanges R, Galvan A, Dathan N, Manenti G, Ferrandino G, Alvino FM, Di Palma T, Scarfò M, Zannini M, Dragani TA, De Felice M, Di Lauro R. A *locus on mouse chromosome 2 is involved in susceptibility to congenital hypothyroidism and contains an essential gene expressed in thyroid. Endocrinology. 2010 Apr;151(4):1948-58.* [Pubmed: 20160132]

Aoki F, Worrad DM, Schultz RM. *Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo*. Dev Biol. 1997 Jan 15;181(2):296-307. [Pubmed: 9013938]

Bachvarova R. *Gene expression during oogenesis and oocyte development in mammals*. Dev Biol (N Y 1985). 1985;1:453-524. [Pubmed: 2481471] (Abstract)

Bennett M, Michaud S, Kingston J, Reed R. *Protein components specifically associated with prespliceosome and spliceosome complexes*. Genes Dev. 1992 Oct;6(10):1986-2000. [Pubmed: 1398075]

Bolte S and Cordelières FP. A guided tour into subcellular colocalization analysis in light Microscopy. Journal of Microscopy. 2006; Vol. 224, pp. 213–232. [Pubmed: 17210054]

Bouchard M, Souabni A, Busslinger M. *Tissue-specific expression of cre recombinase from the Pax8 locus*. Genesis. 2004 Mar;38(3):105-9. [15048807]
Brede G, Solheim J, Prydz H. *PSKH1, a novel splice factor compartment-associated serine kinase*. Nucleic Acids Res. 2002 Dec 1;30(23):5301-9. [PubMed: 12466556]

Brow DA. *Allosteric cascade of spliceosome activation. Annu Rev*. Genet. 2002; 36: 333–360. [Pubmed: 12429696]

Brown JM, Green J, das Neves RP, Wallace HA, Smith AJ, Hughes J, Gray N, Taylor S, Wood WG, Higgs DR, et al. *Association between active genes occurs*

at nuclear speckles and is modulated by chromatin environment. J Cell Biol. 2008;182: 1083–1097. [PubMed: 18809724]

Burge CB, Tuschl T, Sharp PA. *Splicing of Precursors to mRNAs by the Spliceosomes*. In The RNA world Second edition (ed. Gesteland RF et al.), pp. 525–560; 1999. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Burns CG, Ohi R, Krainer AR, and Gould KL. *Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing*. Cell Biology. 1999; 96(24): 13789–13794. [Pubmed: 10570151]

Caceres JF, Misteli T, Screaton GR, Spector DL, Krainer AR. *Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity*. J Cell Biol. 1997; 138: 225–238. [PubMed: 9230067]

Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, Righetti PG. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis. 2004 May;25(9):1327-33. [PubMed: 15174055] (Abstract)

Castanet M, Sura-Trueba S, Chauty A, Carré A, de Roux N, Heath S, Léger J, Lyonnet S, Czernichow P, Polak M. *Linkage and mutational analysis of familial thyroid dysgenesis demonstrate genetic heterogeneity implicating novel genes*. Eur J Hum Genet. 2005 Feb;13(2):232-9. [Pubmed: 15547625]

Chen HR, Tsao TY, Chen CH, Tsai WY, Her LS, Hsu MM, Cheng SC. *Snt309p modulates interactions of Prp19p with its associated components to stabilize the Prp19p-associated complex essential for pre-mRNA splicing*. Proc Natl Acad Sci U S A. 1999; 96(10):5406-11. [Pubmed: 10318896]

Civitareale D, Lonigro R, Sinclair AJ, Di Lauro R. 1989 *A thyroidspecific nuclear protein essential for tissue-specific expression of the thyroglobulin promoter*. EMBO J., 1989; 8:2537–2542. [Pubmed: 2583123]

Cockburn K, Rossant J. *Making the blastocyst: lessons from the mouse*. J Clin Invest. 2010 Apr;120(4):995-1003. [Pubmed: 20364097]

Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI. *The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution*. EMBO J. 1996 Jan 15;15(2):265-75. [PubMed: 8617202]

D'Ambrosio C, Arena S, Salzano AM, Renzone G, Ledda L, Scaloni A. A proteomic characterization of water buffalo milk fractions describing PTM of

major species and the identification of minor components involved in nutrient delivery and defense against pathogens. Proteomics. 2008 Sep;8(17):3657-66. [Pubmed: 18668696] (Abstract)

De Felice M, Ovitt C, Biffali E, Rodriguez-Mallon A, Arra C, Anastassiadis K, Macchia PE, Mattei MG, Mariano A, Schöler H, Macchia V, Di Lauro R. *A mouse model for hereditary thyroid dysgenesis and cleft palate.* Nat Genet. 1998 Aug;19(4):395-8. [Pubmed: 9697704]

De Felice M. and Di Lauro R. *Thyroid development and its disorders: genetics and molecular mechanisms*. Endocr Rev, 2004. 25(5): p. 722-46. [Pubmed: 5466939]

Dix, I, Russell C, Yehuda, S B, Kupiec M, and Beggs J D. *The identification and characterization of a novel splicing protein, Isy1p, of Saccharomyces cerevisiae*. RNA. 1999 5(3):360-8. [PubMed: 10094305]

Dostie J, Lejbkowicz F, Sonenberg N. *Nuclear eukaryotic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles*. J Cell Biol. 2000 Jan 24;148(2):239-47. [PubMed: 10648556]

Dyce J, George M, Goodall H, Fleming TP. *Do trophectoderm and inner cell mass cells in the mouse blastocyst maintain discrete lineages?* Development. 1987 Aug;100(4):685-98. [Pubmed: 3443052]

Eilbracht J, Schmidt-Zachmann MS. *Identification of a sequence element directing a protein to nuclear speckles*. Proc Natl Acad Sci. 2001; 98: 3849–3854. [PubMed: 11274404]

Fortschegger K, Wagner B, Voglauer R, Katinger H, Sibilia M, and Grillari J. Early *Embryonic Lethality of Mice Lacking the Essential Protein SNEV*. Mol Cell Biol. Apr. 2007; 27(8):3123–3130. [PubMed: 17283042]

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. *The architecture of pre-mRNAs affects mechanisms of splice-site pairing*. Proc Natl Acad Sci U S A. 2005 Nov 8;102(45):16176-81. [Pubmed: 16260721]

Fu XD. *The superfamily of arginine/serine-rich splicing factors*. RNA. 1995 Sep;1(7):663-80. [PubMed: 7585252]

Groenen PM, Vanderlinden G, Devriendt K, Fryns JP, Van de Ven WJ. *Rearrangement of the human CDC5L gene by a t(6;19)(p21;q13.1) in a patient with multicystic renal dysplasia*. Genomics. 1998 Apr 15;49(2):218-29. [Pubmed: 9598309]

Grote M, Wolf E, Will CL, Lemm I, Agafonov DE, Schomburg A, Fischle W, Urlaub H, Lührmann R. *Molecular architecture of the human Prp19/Cdc5l complex.* Mol Cell Biol. 2010; 30:2105–2119. [PubMed: 20176811]

Hall LL, Smith KP, Byron M, Lawrence JB. *Molecular anatomy of a speckle.* Anat Rec A DiscovMol Cell Evol Biol. 2006; 288: 664–675. [PubMed: 16761280]

Hamatani T, Carter MG, Sharov AA, Ko MS. *Dynamics of global gene expression changes during mouse preimplantation development*. Dev Cell. 2004 Jan;6(1):117-31. [Pubmed: 14723852]

Hedley ML, Amrein H, Maniatis T. *An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine rich splicing factor*. Proc Natl Acad Sci USA. 1995; 92: 11524–11528. [PubMed: 8524796]

Huang S, Spector DL. *Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors*. Genes Dev. 1991; 5: 2288–2302. [PubMed: 1748285]

Jagatheesan G, Thanumalayan S, Muralikrishna B, Rangaraj N, Karande AA, Parnaik VK. *Colocalization of intranuclear lamin foci with RNA splicing factors*. J Cell Sci. 1999 Dec;112 (Pt 24):4651-61. [PubMed: 10574713]

Jagiello I, Van Eynde A, Vulsteke V, Beullens M, Boudrez A, Keppens S, Stalmans W, Bollen M. *Nuclear and subnuclear targeting sequences of the protein phosphatase-1 regulator NIPP1*. J Cell Sci. 2000; 21: 3761–3768. [PubMed: 11034904]

Johnson C, Primorac D, McKinstry M, McNeil J, Rowe D, Lawrence JB. Tracking COL1A1 *RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain*. J Cell Biol. 2000; 150: 417–432. [PubMed:10931857]

Johnson MH, Ziomek CA. *The foundation of two distinct cell lineages within the mouse morula*. Cell. 1981 Apr;24(1):71-80. [Pubmed: 7237545]

Keene JD. *Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome.* Proc Natl Acad Sci U S A. 2001 Jun 19;98(13):7018-24. [Pubmed: 11416181]

Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox C, Ward J, Gonzalez F. *The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary.* Genes Dev. 1996 Jan 1;10(1):60-9. [Pubmed: 8557195]

Kimura S, Ward JD, Minoo P Thyroid-specific enhancerbinding protein/transcription factor 1 is not required for the initial specification of the thyroid and lung primordia. Biochemie. 1999; 81:321–328.

Kimura S, Ward JD, Minoo P. *Thyroid-specific enhancerbinding protein/transcription factor 1 is not required for the initial specification of the thyroid and lung primordia.* Biochimie. 1999 Apr;81(4):321-7. [Pubmed: 10401665]

Kleinridders A, Pogoda H, Irlenbusch S, Smyth N, Koncz C, Hammerschmidt M, and Bru¨ning J.C. PLRG-1 Is an Essential Regulator of Cell Proliferation and Apoptosis. Mol Cell Biol. 29(11):3173-85. [PubMed: 17283042]

Ko TK, Kelly E, Pines J. *CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles*. J Cell Sci. 2001 Jul;114(Pt 14):2591-603. [PubMed: 11683387]

Krause S, Fakan S, Weis K, Wahle E. *Immunodetection of poly(A) binding protein II in the cell nucleus*. Exp Cell Res. 1994 Sep;214(1):75-82. [PubMed: 8082750]

Kojima T1, Zama T, Wada K, Onogi H, Hagiwara M. *Cloning of human PRP4 reveals interaction with Clk1*. J Biol Chem. 2001 Aug 24;276(34):32247-56. [PubMed: 11418604]

Larsson SH, Charlieu JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, Ross A, Cuzin F, van Heyningen V, Hastie ND. *Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing*. Cell. 1995 May 5;81(3):391-401. [PubMed: 7736591]

Li H, Bingham PM. *Arginine/Serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing*. Cell. 1991; 67: 335–342. [PubMed: 1655279]

Li Q, Imataka H, Morino S, Rogers GW Jr, Richter-Cook NJ, Merrick WC, Sonenberg N. *Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII*. Mol Cell Biol. 1999 Nov;19(11):7336-46. [PubMed: 10523622]

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Lührmann R. *Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome.* Science. 2002 Dec 13;298(5601):2205-8. [PubMed: 12411573]

Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Lührmann R. *A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing.* EMBO J. 2004 Jun 16;23(12):2381-91. [Pubmed: 15175653]

Mansouri A, Chowdhury K, Gruss P. *Follicular cells of the thyroid gland require Pax8 gene function.* Nat Genet, 1998 May;19(1):87-90. [Pubmed: 9590297]

Maris C, Dominguez C, Allain FH. *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* FEBS J. 2005 May;272(9):2118-31. [Pubmed: 15853797]

Martinez Barbera JP, Clements M, Thomas P, Rodriguez T, Meloy D, Kioussis D, Beddington RS. *The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation.* Development. 2000 Jun;127(11):2433-45. [Pubmed: 10804184]
M
atlin AJ, Moore MJ. Spliceosome assembly and composition. Adv Exp Med Biol. 2007;623:14-35. [Pubmed: 18380338]

McDonald WH, Ohi R, Smelkova N, Frendewey D, Gould KL. *Myb-Related Fission Yeast cdc5p Is a Component of a 40S snRNP-Containing Complex and Is Essential for Pre-mRNA Splicing*. Mol Cell Biol. 1999; 19(8): 5352–5362. [Pubmed: 10409726]

Melcak I, Cermanova S, Jirsova K, Koberna K, Malinsky J, Raska I. *Nuclear pre-mRNA compartmentalization: Trafficking of released transcripts to splicing factor reservoirs*. Mol Biol Cell. 2000; 11: 497–510. [PubMed: 10679009]

Moen PT Jr, Johnson CV, Byron M, Shopland LS, de la Serna IL, Imbalzano AN, Lawrence JB. *Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis*. Mol Biol Cell. 2004; 15:197–206. [PubMed: 14617810]

Moore MJ, Query CC, Sharp PA. *Splicing of precursors to mRNA by the spliceosome*. In RNA World (ed. Gesteland RF, Atkins JF), pp. 303–357; 1993. Cold Spring Harbor Labratory Press, Cold Spring Harbor, New York.

Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, Warren SL, Sharp PA, Berezney R. *A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix*. Proc Natl Acad Sci U S A. 1996 Aug 6;93(16):8253-7. [PubMed: 8710856]

Nakatsu Y, Asahina H, Citterio E, Rademakers S, Vermeulen W, Kamiuchi S, Yeo JP, Khaw MC, Saijo M, Kodo N, Matsuda T, Hoeijmakers JH, Tanaka K. *XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription*. J Biol Chem. 2000 Nov 10;275(45):34931-7. [PubMed: 10944529]

Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A, Mann M. *Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex*. Nat Genet. 1998; 20(1):46-50. [Pubmed: 9731529]

Nakayasu H, Ueda K. *Small nuclear RNA-protein complex anchors on the actin filaments in bovine lymphocyte nuclear matrix*. Cell Struct Funct. 1984 Dec;9(4):317-25. [PubMed: 6241503]

Nilsen TW, Graveley BR. *Expansion of the eukaryotic proteome by alternative splicing.* Nature. 2010 Jan 28;463(7280):457-63 [Pubmed: 20110989]

Novoyatleva T, Tang Y, Rafalska I, Stamm S. *Pre-mRNA missplicing as a cause of human disease.* Prog Mol Subcell Biol. 2006;44:27-46. [PubMed: 17076263] (Abstract)

Patel AA, Steitz JA. *Splicing double: insights from the second spliceosome*. Nat Rev Mol Cell Biol. 2003 Dec;4(12):960-70. [Pubmed: 14685174]

Patel N, Aldahmesh MA, Alkuraya H, Anazi S, Alsharif H, Khan AO, Sunker A, Al-Mohsen S, Abboud EB, Nowilaty SR, Alowain M, Al-Zaidan H, Al-Saud B, Alasmari A, Abdel-Salam GM, Abouelhoda M, Abdulwahab FM, Ibrahim N, Naim E, Al-Younes B, E AlMostafa A, AlIssa A, Hashem M, Buzovetsky O, Xiong Y, Monies D, Altassan N, Shaheen R, Al-Hazzaa SA, Alkuraya FS. *Expanding the clinical, allelic, and locus heterogeneity of retinal dystrophies.* Genet Med. 2015 Sep 10. doi: 10.1038/gim.2015.12 [Pubmed: 26355662]

Pedersen RA, Wu K, Bałakier H. *Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection*. Dev Biol. 1986 Oct;117(2):581-95. [Pubmed: 2428686]

Plessel G, Fischer U, Lührmann R. *m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain*. Mol Cell Biol. 1994 Jun;14(6):4160-72. [Pubmed: 8196654]

Qiu XB, Shao YM, Miao S, Wang L. *The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones.* Cell Mol Life Sci. 2006 Nov;63(22):2560-70.

Ralston A, Rossant J. *Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo*. Dev Biol. 2008 Jan 15;313(2):614-29. [Pubmed: 18067887]

Reed R. *Protein composition of mammalian spliceosomes assembled in vitro.* Proc Natl Acad Sci U S A. 1990 Oct;87(20):8031-5. [Pubmed: 2146679]

Sacco-Bubulya P, Spector DL. *Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing*. J Cell Biol. 2002 Feb 4;156(3):425-36. [PubMed: 11827980]

Santisteban P, Acebron A, Polycarpou-Schwarz M, Di Lauro R. *Insulin and insulin-like growth factor I regulate a thyroidspecific nuclear protein that binds to the thyroglobulin promoter*. Mol Endocrinol, 1992; 6:1310–1317. [Pubmed: 1406708]

Sharma A, Takata H, Shibahara K, Bubulya A, Bubulya PA. *Son is essential for nuclear speckle organization and cell cycle progression*. Mol Biol Cell. 2010 Feb 15;21(4):650-63 [PubMed: 20053686]

Schul W, van Driel R, de Jong L. *A subset of poly(A) polymerase is concentrated at sites of RNA synthesis and is associated with domains enriched in splicing factors and poly(A) RNA*. Exp Cell Res. 1998 Jan 10;238(1):1-12. [PubMed: 9457051]

Shopland LS, Johnson CV, Lawrence JB. *Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains*. J Struct Biol. 2002; 140: 131–139. [PubMed: 12490161]

Smith CW1, Valcárcel J. *Alternative pre-mRNA splicing: the logic of combinatorial control*. Trends Biochem Sci. 2000 Aug;25(8):381-8. [Pubmed: 10916158]

Smith KP, Moen PT, Wydner KL, Coleman JR, Lawrence JB. *Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific*. J Cell Biol. 1999; 144: 617–629. [PubMed: 10037785]

snRnp200 lethality: J:204739 International Knockout Mouse Consortium, MGI download of modified allele data from IKMC and creation of new knockout alleles. Database Download. 2014.

Spector DL. *Nuclear domains*. 2001. J Cell Sci; 114(Pt 16):2891-3. [Pubmed: 11686292]

Spector DL, Fu XD, Maniatis T. *Associations between distinct pre-mRNA splicing components and the cell nucleus*. EMBO J. 1991; 10(11):3467-81. [Pubmed: 1833187]

Staley JP, Woolford JL Jr. *Assembly of ribosomes and spliceosomes: Complex ribonucleoprotein machines*. Curr Opin Cell Biol. 2009 Feb;21(1):109-18. [Pubmed: 19167202]

Sutherland AE, Speed TP, Calarco PG. *Inner cell allocation in the mouse morula: the role of oriented division during fourth cleavage*. Dev Biol. 1990 Jan;137(1):13-25. [Pubmed: 2295360]

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009 May 1;25(9):1105-11. [Pubmed: 19289445]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010 May;28(5):511-5. [Pubmed: 20436464]

Trinkle-Mulcahy L, Ajuh P, Prescott A, Claverie-Martin F, Cohen S, Lamond AI, Cohen P. *Nuclear organisation of NIPP1, a regulatory subunit of protein phosphatase 1 that associates with pre-mRNA splicing factors*. J Cell Sci. 1999 Jan;112 (Pt 2):157-68. [PubMed: 9858469]

Trinkle-Mulcahy L, Sleeman JE, Lamond AI. *Dynamic targeting of protein phosphatase 1 within the nuclei of living mammalian cells*. J Cell Sci. 2001 Dec;114(Pt 23):4219-28. [PubMed: 11739654]

Umen JG, Guthrie C. *The second catalytic step of pre-mRNA splicing*. RNA. 1995 Nov; 1(9):869-85. Review. [PubMed: 8548652]

Wang Z, Burge CB. *Splicing regulation: from a parts list of regulatory elements to an integrated splicing code*. RNA. 2008 May;14(5):802-13. [Pubmed: 18369186]

Wansink DG, Schul W, van der Kraan I, van Steensel B, van Driel R, de Jong L. *Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus*. J Cell Biol. 1993 Jul;122(2):283-93. [PubMed: 8320255]

Ward AJ, Cooper TA. *The pathobiology of splicing*. J Pathol. 2010 Jan;220(2):152-63. [PubMed: 19918805]

Wei X, Somanathan S, Samarabandu J, Berezney R. *Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles*. J Cell Biol. 1999; 146: 543–558. [PubMed: 10444064]

Will CL, Lührmann R. *Spliceosome structure and function*. Cold Spring Harb Perspect Biol. 2011 Jul 1;3(7). [Pubmed: 21441581]

Xing Y, Johnson CV, Dobner PR, Lawrence JB. *Higher level organization of individual gene transcription and RNA splicing*. Science. 1993; 259: 1326–1330. [PubMed: 8446901]

Xing Y, Johnson CV, Moen PT, McNeil JA, Lawrence JB. *Nonrandom gene organization: Structural arrangements of specific pre-mRNAtranscription and splicing with SC-35 domains*. J Cell Biol. 1995§; 131: 1635–1647. [PubMed: 8557734]

Yang X, Bani MR, Lu SJ, Rowan S, Ben-David Y, and Chabot B. *The A1 and A1B proteins of heterogeneous nuclear ribonucleoparticles modulate 5' splice site selection in vivo*. Proc Natl Acad Sci U S A. 1994 91(15): 6924–6928 [Pubmed: 8041722]

Yonemasu R, Minami M, Nakatsu Y, Takeuchi M, Kuraoka I, Matsuda Y, Higashi Y, Kondoh H, Tanaka K. *Disruption of mouse XAB2 gene involved in pre-mRNA splicing, transcription and transcription-coupled DNA repair results in preimplantation lethality*. DNA Repair (Amst). 2005 Apr 4;4(4):479-91. [PubMed: 15725628]

Zannini M, Avantaggiato V, Biffali E, Arnone MI, Sato K, Pischetola M, Taylor BA, Phillips SJ, Simeone A, Di Lauro R. *TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation.* EMBO J. 1997 Jun 2;16(11):3185-97. [Pubmed: 9214635]

Zeng C, Kim E, Warren SL, Berget SM. *Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity*. EMBO J. 1997 Mar 17;16(6):1401-12. [PubMed: 9135155]

Zhao C, Bellur DL, Lu S, Zhao F, Grassi MA, Bowne SJ, Sullivan LS, Daiger SP, Chen LJ, Pang CP, Zhao K, Staley JP, Larssonson C. *Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs*. Am J Hum Genet. 2009 Nov;85(5):617-27. [Pubmed: 19878916]