CIBIR
Centro de Investigación Biomédica de La Rioja

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

C. ELEGANS INTEGRATOR COMPLEX: IDENTIFICATION AND ANALYSIS OF A ROLE BEYOND SMALL NVCLEAR RNA PROCESSING

TESIS DOCTORAL

BEATRIZ SÁENZ NARCISO

SALAMANCA, 2017

Dña. María de la Paz Sacristán Martín, Profesora Titular del Departamento de Microbiología y Genética de la Universidad de Salamanca

CERTIFICA:

Que la memoria titulada **"*C. elegans* Integrator complex: Identification and analysis of a role beyond snRNA processing"** presentada por la Licenciada en Biología **BEATRIZ SÁENZ NARCISO**, ha sido realizada en el Centro de Investigación Biomédica de La Rioja, CIBIR, y que reúne a su juicio, originalidad y contenidos suficientes para que sea presentada ante el tribunal correspondiente y optar al grado de Doctor por la Universidad de Salamanca.

Y para que así conste a efectos legales, expiden el presente certificado en Salamanca a 30 de Junio de 2017.

Fdo. María de la Paz Sacristán Martín

Tutora de la tesis

Don Juan Cabello Pardos, Investigador principal del grupo Proliferación y Diferenciación del Centro de Investigación Biomédica de La Rioja (CIBIR)

CERTIFICA:

Que la memoria titulada **"*C. elegans* Integrator complex: Identification and analysis of a role beyond snRNA processing"** presentada por la Licenciada en Biología **BEATRIZ SÁENZ NARCISO**, ha sido realizada en el Centro de Investigación Biomédica de La Rioja, CIBIR, y que reúne a su juicio, originalidad y contenidos suficientes para que sea presentada ante el tribunal correspondiente y optar al grado de Doctor en Biología.

Y para que así conste a efectos legales, expiden el presente certificado en Logroño a 29 de Junio de 2017.

Fdo. Juan Cabello Pardos

Director de la tesis

Dña. Eva María Gómez Orte, Investigadora post-doctoral del grupo Proliferación y Diferenciación del Centro de Investigación Biomédica de La Rioja (CIBIR)

CERTIFICA:

Que la memoria titulada **"*C. elegans* Integrator complex: Identification and analysis of a role beyond snRNA processing"** presentada por la Licenciada en Biología **BEATRIZ SÁENZ NARCISO**, ha sido realizada en el Centro de Investigación Biomédica de La Rioja, CIBIR, y que reúne a su juicio, originalidad y contenidos suficientes para que sea presentada ante el tribunal correspondiente y optar al grado de Doctor en Biología.

Y para que así conste a efectos legales, expiden el presente certificado en Logroño a 29 de Junio de 2017.

Fdo. Eva María Gómez Orte

Codirectora de la tesis

A mis padres,

José Luis y Rosa Mari,
por vuestro apoyo constante y cariño,

Y a los que ya no están...

Pedro y Luci,
José Luis y Marina,
Gracias por los momentos tan bonitos
que habéis dejado grabados en mi,
siempre os llevaré conmigo.

Abstract

In this work we described the *C. elegans* Integrator complex and the involvement of its INTS-6 subunit in DNA damage response.

The Integrator complex, which is comprised of at least fourteen subunits in human cells, is responsible for snRNA 3'-end processing. In addition, some of its subunits are involved in other steps of the RNAP II transcription cycle or other biological processes such as development or DNA repair.

Here, we demonstrated that the *C. elegans* Integrator complex is comprised of, at least, eleven subunits (INTS-1, INTS-2, INTS-3, INTS-4, INTS-5, INTS-6, INTS-7, INTS-8, INTS-9, INTS-11 and INTS-13). RNAi knockdown of any subunit leads to 3'-end processing defects that result in the formation of chimeric RNAs composed of an snRNA and an mRNA, which we have called "sn-mRNAs"'.

We also detected these chimeric "sn-mRNAs" upon gamma radiation. Finally, involvement of INTS-6 in DNA repair via the HR pathway was demosnstrated and we suggested a link between INTS-6 in DSB DNA repair and the formation of chimeric "sn-mRNAs".

TABLE OF CONTENTS

List of Figures

27. A) Fertility of the N2 versus the *t1903* mutant at 15ºC and 25ºC. B) Embryonic viability (%) of N2 versus *t1903* mutants, growing at 15ºC and 25ºC.

28. N2 embryonic development observed under 4D microscopy.

29. Embryonic development of the *t1903* mutant embryo number one observed under 4D microscopy.

30. Embryonic development of the *t1903* mutant number two observed under 4D microscopy.

31. Embryonic development of the *t1903* mutant number three observed under 4D microscopy.

32. A) Scheme of the pJC51 plasmid used to generate the JCP341 transgenic line. B) DIC-1 expression in the JCP341 transgenic strain.

33. A) Scheme of the pJC56 plasmid used to generate the JCP378 transgenic line. B) DIC-1 expression in the JCP378 transgenic strain. C) DIC-1 expression in embryos.

34. DIC-1 expresssion in the 505 transgenic strain.

35. RNA deep sequencing of *t1903* mutant versus N2 at 15ºC and 25ºC. A) RiboMinus RNA seq. (rmRNA-seq). B) Polyadenylated RNA seq (mRNA-seq).

36. Human Ints6 localizes in the nucleus. A) Scheme of the pBS15 (eGFP-Ints6) and pBS16 (Ints6-eGFP) plasmids. B) Immunostaining of U2OS cells with α-INTS6.

37. Human Ints6 protein expression levels are constant throughout cell cycle. A) The WB shows Ints6 and distinct cell cycle markers along the cell cycle. B) FACS to corroborate cell cycle phase (G1, G1/S, G2 and Mitosis).

38. RNAP II Co-immunoprecipitates with Ints6.

39. Depletion of human Ints6 affects snRNA 3'-end processing. A) Ints6 protein levels after the Ints6 depletion. B) RPE depleted cells for Ints-6 show 3'-end snRNA processing defects in U1 and U2 snRNAs.

40. RPE cells depleted for Ints6 increased in number.

41. Predicted *C. elegans* Integrator complex orthologs vs *H. sapiens* Integrator complex subunits.

42. INTS-6 Immunoprecipitation process and silver staining.

43. Phenotype of *C. elegans* Integrator complex predicted subunits after RNAi knockdown.

44. Knockdown of *C. elegans* Integrator complex subunits results in 3'-end snRNAs processing defects. A) U1 type snRNAs. B) U2 type snRNAs. C) SL type snRNAs. D) Actins.

45. RiboMinus RNA deep sequencing (rmRNA-seq) shows snRNA 3'-end processing defects upon Integrator complex knockdown.

46. Scheme of the plasmids made and integrated by mosSCI system in Chromosome II to check the possible formation of proteins. A) Plasmid pJC57 B) Plasmid pJC60 C) Plasmid pJC55 D) Plasmid pJC50 E) Plasmid pJC58.

47. WBs to check if chimeric "sn-mRNAs" are translated into proteins.

48. Scheme of the plasmids made and integrated into chromosome II using mosSCI system to check the possible formation of peptides. A) Plasmid pJC63 B) Plasmid JC64.

49. WBs to check if the chimeric “sn-mRNAs” are translated into peptides.

50. Phenotype of N2 worms knocked down for *ints-6* under normal growing conditions *and* 24h after IR (90Gy).

51. *ints-6* knockdown impairs RAD-51 recruitment to DSBs following IR.

52. *ints-6* knockdown abrogates Tyr15 CDK-1 phosphorylation in response to DNA damage.

53. Quantification of the nuclei within gonadal mitotic region in *ints-6* RNAi and L4440 RNAi backgrounds in non-irradiated worms and 24 hours following gamma radiation (Mean ± SD).

54. A) pHH3 positive cells quantification in the mitotic region of gonads in *ints-6* RNAi and L4440 RNAi backgrounds in worms non-irradiated and 24 hours following irradiation. B) Representative pictures of the α-pHH3 immunostained gonads.

55. A) Scheme of the pJC51 (“*ints-6*::3xFLAG::eGFP”) and pJC54(“*ints-6*(S850A)3xFLAG::eGFP”) plasmids. B) Phos-tag™ SDS-PAGE and WB against the 3xFLAG tag from worm protein extracts of the JCP383 and JCP382 strains under IR and non-IR conditions.

56. Schematic drawing of the transgenic strains to study INTS-6 Serine 850 possible phosphorylation.

57. RAD-51 recruitment and INTS-6 localization following irradiation in transgenics that mimic our hypothesized INTS-6 phosphorylation of JCP472 (“*ints-6*(S850E)::3xFLAG”) versus others that mimic the non-phosphorylated form of JCP483 (“*ints-6*(S850A)::3xFLAG”) compared to wild-type JCP462 “*ints-6*::3xFLAG”.

58. RAD-51 IRIF quantification following irradiation (90Gy).

59. Quimerics “sn-mRNAs” are formed in response to DNA damage.

60. Working model of the Integrator complex upon DNA damage and formation of chimeric “sn-mRNAs”.

List of Tables

Abbreviations

6-4 PP: 6-4 photoproduct

9-1-1 complex: Rad9-Hus1-Rad1 complex

A: Alanine

aa: amino acid

Ab: Antibody

APS: Ammonium persulfate

ARM: Armadillo

ATM: Ataxia-telangiectasia mutated

ATR: ATM and Rad3-related

BCA: Bicinchoninic acid

BER: Base Excision Repair

BLAST: Basic Local Alignment Search Tool

BLM: Bloom Syndrome Protein

bp: base pairs

BRCA2: Breast Cancer type 2

BSA: Bovine Serum Albumin

C. elegans: *Caenorhabditis elegans*

CAK: CDK-activating kinase

CDK: Cyclin Dependent Kinase

cDNA: complementary DNA

Chr: Chromosome

cis-Pt: Cisplatin

COIL: Coiled coil domains

CPD: Cyclobutane Pyrimidine Dimer

CPE: Core Promoter Elements

CPSF100: Cleavage and Polyadenylation Specificity Factor subunit 100

CPSF73: Cleavage and Polyadenylation Specificity Factor subunit 73

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

crRNA: CRISPR RNA

CRTAP: Cartilage associated protein

CstF: Cleavage Stimulation Factor

Ct: Carboxi-terminal

CTD: Carboxy Terminal Domain

CtIP: CtBP-interacting protein

DDRNAs: DNA Damage Response RNAs

Df: Deficiencies

DIC: Differential Interference Contrast

diRNA: Damage induced RNA

DMEM: Dulbecco's Modified Eagle's Medium

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

dNTPs: deoxynucleotides

DPE: Downstream Promoter Element

DSBs: Double Strand Breaks

DSE: Distal Sequence Element

DSIF: DRB (5,6-dichloro-1β-D-ribofuranosylbenzimidazole) Sensitivity Inducing Factor

DSS1: Deleted in Split-Hand/Split-Foot 1

DTT: 1,4-Dithiothreitol

DUF: Domain of Unknown Function

E: Glutamic

E. coli: *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

EGF: Epidermal Growth Factor

eGFP: enhanced Green Fluorescent Protein

EMS: Ethyl methanesulfonate

eRNAs: enhancer RNAs

FACS: Fluorescence Activated Cell Sorting

FBS: Fetal Bovine Serum

Fig.: Figure/ Figura

G1: Gap1

G2: Gap2

GFP: Green Fluorescent Protein

GTF: General Transcription Factors

***H. sapiens*:** *Homo sapiens*

HDE: Histone Downstream Element

HEAT: Huntingtin, Elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1

HEK: Human Embryonic Kidney,

HR: Homologous Recombination

hSSB1: human Single Stranded Binding protein 1

HVS: Herpesvirus saimiri

IEG: Immediate Early Genes

Inr: Initiator

IP: Immunoprecipitate, immunoprecipitation

IRIF: Irradiation induced foci

ISDCC: INTS6/SAGE/DDX26B/CT45 C-terminus

Kb: Kilobase

kDa: kilodalton

Lab: Laboratory

LB: Luria-Bertani

LC-MS/MS: Liquid Chromatography-Mass Spectrometry/Mass Spectrometry

LG: Linkage Group

lncRNAs: Long non-coding RNAs

L1: Larval stage 1

L2: Larval stage 2

L3: Larval stage 3

L4: Larval stage 4

M: Mitosis

M: Molar

mA: miliampere

mg: miligrame

Min: minutes

miRNA: microRNA

ml: milliliter

MM: Materials and Methods

mM: milimolar

mm: millimeters

MMC: Mitomycin C

mosSCI: Mos1 mediated Single Copy transgene Insertion

MPC: Monomethylphosphate cap

MRN complex: Mre11-Rad50-Nbs1 complex

mRNA: Messenger Ribonucleic Acid

NABP1/2: Nucleic Acid Binding Protein 1/2

ncRNA: non-conding RNA

NELF: Negative Elongation Factor

NER: Nucleotide Excision Repair

ng: nanograme

NGM: Nematode Growth Medium

NHEJ: Non-Homologous End-Joining

nm: nanometers

Nt: Amino-terminal

nt: nucleotide

O/N: Overnight

OB: Oligosaccharide/oligonucleotide-binding

ORF: Open Reading Frame

P-TEFb: Positive Transcription Elongation Factor b

P3H1: Prolyl 3-hydroxylase 1

P3H4: Prolyl 3-hydroxylase family member 4

PAF1: Polymerase-associated factor 1

PAM: Protospacer Adjacent Motif

PAS: Polyadenylation sequence

PBP: PSE-binding protein

PBS: Phosphate Buffer Saline

PCR: Polymerase Chain Reaction

pH: log1/[H+]

PHD: Plant homeodomain

pHH3: phospho histone H3

PIC: Preinitiation complex

PIKK: Phosphatidyl-Inositol-3-OH Kinase-like Kinases

PMSF: Phenylmethane sulfonyl fluoride

pre-miRNA: precursor miRNA

Pro: Proline

PSE: Proximal Sequence Element

psi: pounds per square inch

PTF: PSE-binding Transcription Factor

RMMBL: Zn-dependent metallo-hydrolase RNA specificity

RNA: Ribonucleic acid

RNAi: RNA interference

RNAP: RNA polymerase

RNP: Ribonucleoprotein

RPA1: Replication Protein A1

RPAP2: RNAP II associated protein 2

RPE: Retinal Pigment Epithelium,

rRNA: ribosomic RNA

RT: Reverse Transcriptase

RT-PCR: Reverse Transcription Polymerase Chain Reaction

RT-qPCR: Real Time-quantitative Polymerase Chain Reaction

S: Serina

S: Synthesis phase

SAC: Spindle Assembly Checkpoint

SD: Standard Derivation

SDS-PAGE: Sodium-DodecylSulfate-Polyacrilamide gel electrophoresis

SDS: Sodium Dodecyl Sulfate

Sec: seconds

SEC: Super Elongation Complex

Ser: Serine

SETX: Senataxin

sgRNA: single guide RNA

shRNA: short hairpin RNA

siRNA: small interfering RNA

SL: Sequence Leader

SLBP: Stem-Loop-Binding Protein

SMN: Survival motoneuron

SNAPc: snRNA activating protein complex

sncRNA: small non-coding RNA

snoRNA: small nucleolar RNA

SNP: Single Nucleotide Polymorphism

snRNA: small nuclear RNA

snRNP: small nuclear ribonucleoparticle

SOSS1: Sensor Of Single-Stranded DNA complex 1

TAFs: TBP-associated factors

TBE: Tris/Borate/EDTA

TBP: TATA-binding protein

TBS-T: Tris-Buffered Saline-Tween

TBS: Tris-Buffered Saline

TF: Transcription Factor

TGS1: Trimethylguanosine synthase 1

Thr: Threonine

TMG: Trimethylguanosine

tracrRNA: trans-activating crRNA

tRNA: transfer RNA

TRP: Tetratricopeptide

ts: thermosensitive

TSS: Transcription Start Site

Tyr: Tyrosine

U snRNA: Uridine rich small nuclear RNA

UTR: Untranslated region

UV: Ultraviolet

V: Volts

VWA: von Willebrand factor type A

WB: western blot

WT: Wild-type

µg: microgram

µl: microliter

µm: micrometers

µM: micromolar

Introduction

1. *Caenorhabditis elegans*

Caenorhabditis elegans is a tiny, free-living nematode found worldwide, predominantly in humid temperate areas (Andersen et al., 2012; Frezal & Félix 2015). *C. elegans*, has been mischaracterized as a soil nematode, where it is mostly found in a non-feeding stage called a dauer (Barriére & Félix, 2014), but feeding and reproducing stages can most easily be isolated from rotting vegetable matter, such as fruits and thick herbaceous stems (Félix & Duveau, 2012). These rotting substrates in late stages of decomposition provide abundant bacterial food for the worm.

1.1. *C. elegans* as a model organism

This nematode was first described in 1900 by Emilie Maupas and initial experiments on its mode of reproduction, meiosis and development were carried out by Maupas and later by the laboratory of Victor Nigon. However, it was not until 1965 when Sydney Brenner established *C. elegans* as a model organism for understanding questions of developmental biology and neurobiology (Riddle et al., 1997). During the next decades, Brenner and collaborators raised *C. elegans* to the status of a premier model organism. Brenner treated *C. elegans* hermaphrodites with ethyl methanesulfonate (EMS), a mutagenic alkylating agent. Approximately 300 mutations affecting behavior and morphology were identified. Most of them were recessive and allowed the characterization of more than a hundred genes (Brenner, 1974). In 1983, John Sulston and co-workers described the embryonic cell lineage from zygote to newly hatched larva, which is highly invariant and plays an important role in determining cell fate (Sulston et al., 1983). In 1986, Brenner and White along with co-workers published the article “The mind of a worm”, where the *C. elegans* nervous system architecture was deduced from reconstructions of electron micrographs of serial sections (White et al., 1986). In 1998, the *C. elegans* genome was the first genome of a multicellular organism to be completely sequenced (The *C. elegans* Sequencing Consortium, 1998). A comparison of the *C. elegans* genome to the human genome along with the extensive knowledge available on the molecular, cellular, developmental and behavioral biology of this worm, have revealed that evolution has maintained thousands of conserved genes that play similar, or in some

cases nearly identical functions in nematodes and other animals including humans (Corsi et al., 2015). For example, the vertebrate apoptosis regulator Bcl-2 can functionally substitute for its *C. elegans* ortholog *ced-9* (Hengartner & Horvitz, 1994).

Right from the beginning, studies of *C. elegans* led to key discoveries. Sydney Brenner, Robert Horvitz, and John Sulston were awarded the 2002 Nobel Prize in Physiology or Medicine for their discoveries concerning genetic regulation of organ development and programmed cell death. Moreover, *C. elegans* researchers have also made technical discoveries with a broad biological impact, such as gene silencing by RNA interference (RNAi) (Fire et al., 1998) that led to Andrew Fire and Craig Mello winning the 2006 Nobel Prize in Physiology or Medicine and the development of the green fluorescent protein (GFP) as a biological marker (Chalfie et al., 1994) for which Martin Chalfie shared the 2008 Nobel Prize in Chemistry.

Currently, *C. elegans* is widely used to study a variety of biological processes: apoptosis, cell cycle, cell signaling, gene regulation, aging, behavior, sex determination, metabolism, etc. (Kaletta & Hengartner 2006; Corsi et al., 2015).

The *C. elegans* reference wild-type strain is N2 and it was obtained by Sydney Brenner from the one originally isolated in Bristol, England (Riddle et al., 1997). In the laboratory, culturing *C. elegans* is very simple and working with it does not require especially expensive equipment beyond a good dissecting microscope and a compound microscope. Adults are approximately 1 mm long and 80 µm in diameter. They are grown on agar plates seeded with *Escherichia coli* as a food source (Fig. 1). As a self-fertilizing hermaphrodite, a single animal can give rise to an entire population. The *C. elegans* life cycle is very quick: 3.5 days at 20ºC. Their half-life is about 2 or 3 weeks under favourable conditions. Additionally, these animals can be grown at temperatures ranging from 12° to 25°C, making it possible to control the rate of animal development and to use of temperature-sensitive mutants (Riddle et al., 1997; Corsi et al., 2015). Moreover, the fact that *C. elegans* is transparent makes it possible to visualized individual cells and subcellular details using Nomarski (Differential Interference Contrast, DIC) optics. Another advantage of this organism is that strains can be stored in liquid nitrogen for years. Thus, scientists can maintain enormous mutant collections or transgenic strains and revive a desired strain when needed. *C. elegans* mutants have been classically obtained by chemical

mutagenesis, exposure to ionizing radiation or transposon insertion. At the present time new genome-editing techniques such as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), along with the ability to reduce gene activity using RNAi by feeding, allow for a deeper functional study of any gene. (Frøkjær-Jensen, 2013; Xu, 2015). All these features have made this organism an excellent model system for biological research.

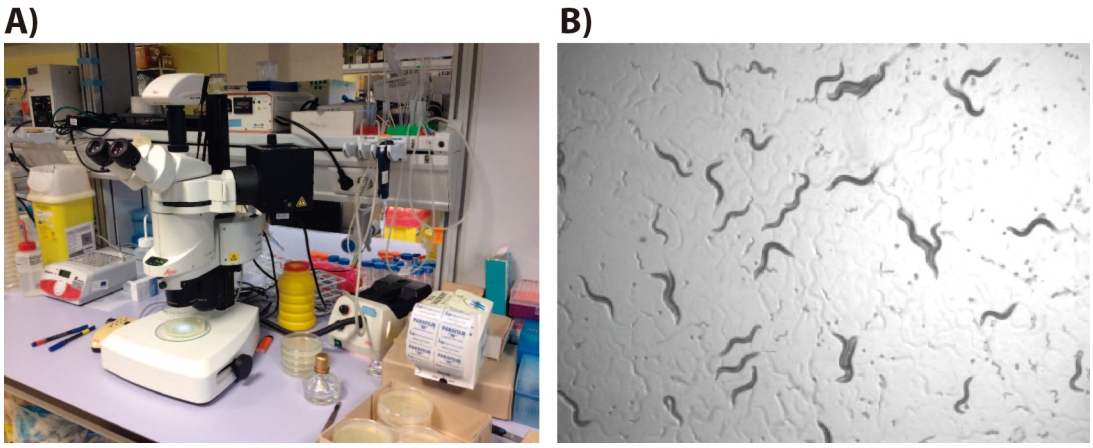


Figure 1. A) *C. elegans* visualization under the dissecting microscope and its culture on Petri dishes with agar. B) *C. elegans* visualized through the dissecting microscope. Worms are observed on Petri dishes while they move, eat, and lay eggs.

1.2. *C. elegans* anatomy

C. elegans has two sexual forms: self-fertilizing hermaphrodites and males (Fig. 2A and B). Basically, *C. elegans* body can be described as two concentric tubes separated by a space, the pseudocoelom. The outer tube consists of cuticle, epidermis (traditionally called the hypodermis), neurons and muscles surrounding a pseudocoelomic fluid-filled cavity that contains the intestine and the gonad (the inner tube) (Fig. 2C). A basement membrane separates epidermis from muscle. Additionally, the gonad, the intestine and the pharynx are also wrapped by basement membranes. In the pseudocoelomic cavity, *C. elegans* has six cells called coelomocytes, which act as scavengers in the body cavity similarly to the function of macrophages in vertebrates (Grant & Sato, 2006). The shape of the worm is maintained by internal hydrostatic pressure, controlled by an osmoregulatory system (Wood, 1988; Corsi et al., 2015).

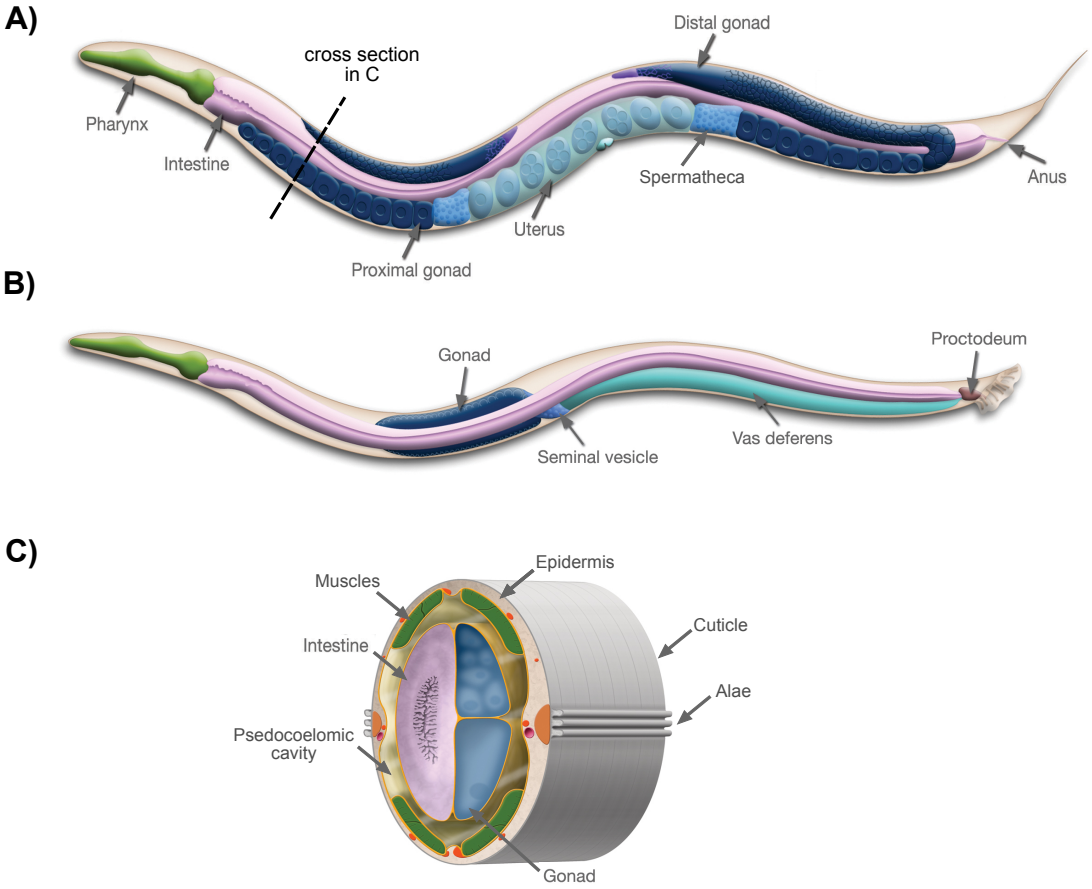


Figure 2. *C. elegans* schematic drawing of anatomical structures. A) *C. elegans* hermaphrodite. B) *C. elegans* male. C) Cross-section of the *C. elegans* hermaphrodite in the anterior region (location marked with a black line in A. Images are modified versions from those found at www.wormatlas.org.

- *C. elegans* epidermis and cuticle

The *C. elegans* epidermis is a model for the innate immune response, wound healing, cell-cell fusions and the establishment of epithelial layers in developing embryos (Chisholm & Hardin, 2005; Podbilewicz, 2006; Taffoni & Pujol, 2015). Epidermal cells secrete a cuticle, which is a syncytial tissue (made up of large multinucleate cells), composed of collagen, lipids and glycoproteins that form a protective layer (Chisholm & Hardin, 2005; Page & Johnstone, 2007). The cuticle serves as a model for extracellular matrix function and formation (Page & Johnstone, 2007). In adults, lateral, longitudinal cords of seam cells make special cuticular structures called alae on the cuticle surface (Fig. 2C). On solid media the worm crawls on one side with the alae contacting the substrate (Wood, 1988).

- *C. elegans* muscles

Interior to the epidermis and connected to it, are four strips of body-wall muscles that run along the length of the body, two dorsally and two ventrally (Fig. 2C). Regular contraction and relaxation of muscle cells lead to the “elegant” sinusoidal movement of the animal. These somatic muscles are obliquely striated (which is unusual) and mononucleate (muscle cells do not fuse as they do in vertebrates) with multiple sarcomeres per cell (Moerman & Fire, 1997). In addition, *C. elegans* has muscles that control eating (pharyngeal muscles), egg-laying (vulval and uterine muscles and the contractile gonad sheath), mating (male-specific tail muscles), and defecation (enteric muscles) (Wood, 1998).

- *C. elegans* digestive system

C. elegans feeds through a two-lobed pharynx which pumps food into the intestine, grinding it as it passes through the second lobe (Avery & You, 2012) (Fig. 2).

The animals’s pumping behavior depends on the availability and quality of the food. For example, animals pump more when they are hungry and less when they are full (Avery & Shtonda, 2003). The pharynx is wrapped by a thicker basement membrane than the others found in the animal, and it is composed of epidermal cells, muscle cells and neurons. These pharyngeal neurons form a nervous system that is practically autonomous. The study of pharyngeal development has been a model for organogenesis (Mango, 2007).

The pharynx is connected to the intestine through a valve that controls the amount of food that passes through it (Mango, 2007). The *C. elegans* intestine consists of 20 large polyploid epithelial cells arranged in pairs that surround a central lumen, which connects to the anus near the tail (Wood, 1988; Corsi et al., 2015). Interestingly, after the first larval stage, *C. elegans* intestinal cells undergo rounds of endoreduplication before each larval stage, which is thought to handle the increasing demands of the growing animal (Hedgecock & White, 1985). The *C. elegans* intestine has served as a model to study infection and response to infection by different pathogens that colonize the digestive system (Corsi et al., 2015).

- *C. elegans* excretory system

The excretory system has a rather simple structure in *C. elegans*. It consists of four distinctive cell types: one pore cell, one duct cell, one canal or excretory cell, and a fused pair of gland cells (Fig. 3). The soma of these cells is located in the head region. The canal or excretory cell, the largest in the nematode, is an H-shaped cell with four arms that run the length of the animal. All four arms, join at the cell body located ventrally to the pharynx that connects to the exterior through the pore cell or excretory pore. The role of the excretory cell is probably maintenance of osmotic balance and removal of metabolic waste, analogous to the renal system of higher animals. In addition, the excretory gland cells are connected to the same duct and pore, and they secrete materials from large membrane-bound vesicles of glycoproteins and hormones (Nelson et al., 1983; Altun & Hall, 2009).

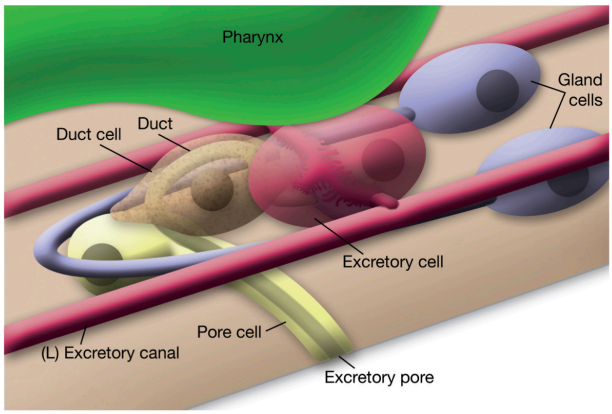


Figure 3. Schematic drawing of the *C. elegans* excretory system. The excretory system is composed of: the pore cell (yellow), the duct cell (brown), the canal (excretory) cell (red) and the fused pair of gland cells (purple). Image from www.wormatlas.org.

- *C. elegans* nervous system

The nervous system has been completely reconstructed using electron microscopy. It is essentially identical between individuals, even at the level of the positions of individual synapses (Ward et al., 1975; White et al., 1986; Hall & Russel, 1991). It can be said that *C. elegans* has two distinct and independent nervous systems: a large somatic nervous system and a small pharyngeal nervous system, connected through a single pair of interneurons (Fig. 4). The adult *C. elegans* hermaphrodite

has 302 neurons, whereas the adult male has 383. Cell bodies of most neurons are clustered in the a few ganglia in the head, in the ventral cord, and in the tail. The specialized male tail has the majority of the extra neurons (Ward et al., 1975; Sulston & Horvitz, 1977; Sulston et al., 1983; White et al., 1986).

In most of the *C. elegans* neurons, it is not possible to distinguish axons from dendrites because they both give and receive synapses, so they are generally called neurites or processes. Neurites, usually one or two neurites for each neuron, form an external ring (the nerve ring) around the pharynx or run along the length of the animal body (Sulston et al., 1983; White et al., 1986; Wood, 1988) (Fig. 4). In addition, *C. elegans* neurons do not send terminal branches with boutons to make synapses as in vertebrate systems. Most of the connections are made *en passant* (side by side, as neurites pass each other). The neurites form synapses to each other in four major areas: the nerve ring, the ventral nerve cord, the dorsal nerve cord, and the neuropil of the tail (Sulston et al., 1983; White et al., 1986; Wood, 1988) (Fig. 4).

Sensory neurons run anteriorly from the nerve ring to the sensory organs (sensilla) in the head. The nerve ring, which can be thought of as the animal's brain, receives input from the head region and sends its output primarily to the body-wall muscles. Nematodes are unusual in that motor neurons do not send processes that synapse onto muscle; instead, muscles send cellular projections to motor neurons to receive synapses (Wood, 1988; Corsi et al., 2015).

C. elegans neurons make more than 7000 chemical synapses (using many of the most common neurotransmitters: acetylcholine, glutamate, GABA, dopamine, serotonin, etc.) and gap junction connections (White et al., 1986; Hobert, 2013). In addition *C. elegans* neurons are modulated by numerous neuroendocrine signals (Li & Kim, 2010).

Although to a lesser extent than in vertebrates, *C. elegans* also has several glia-like support cells. Neurites of these glia-like cells have been found at neuronal junctions, suggesting that they could be involved in intercellular information transfer (Oikonomou & Shaham, 2010).

In last decades, *C. elegans* has become an important model for the study of

neurobiology in many aspects, from neuronal development to neuronal degeneration, passing through worm behaviour (Corsi et al., 2015).

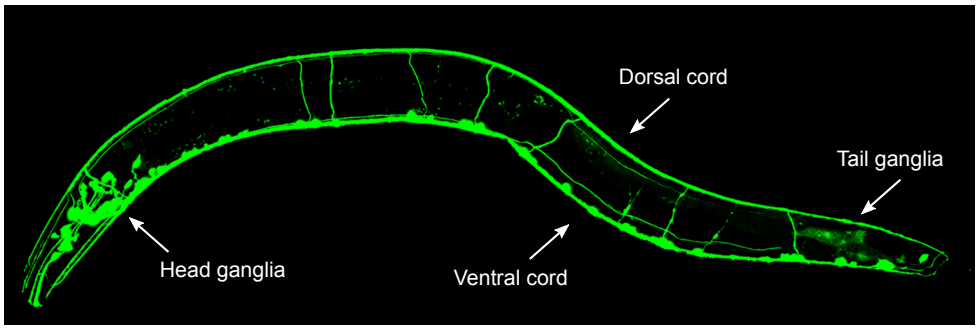


Figure 4. *C. elegans* nervous system. This picture is a fluorescent image of the *C. elegans* nervous system labelled with a GFP reporter (*sto-6::gfp*). Image modified from the one found in Corsi et al., 2015.

- *C. elegans* reproductive system

C. elegans has a very peculiar mode of reproduction called androdioecy. This nematode can reproduce either by having hermaphrodite self-fertilize or breed with males. Both sexes have six chromosomes: five autosomal chromosomes and one sex-determining chromosome. However, the hermaphrodite is diploid for the six chromosomes, five autosomal chromosomes (A) and one chromosome sex-determining (X), but males only have a single X chromosome (XO) (Nigon, 1949; Wood, 1988). Males are rare in both, in the nature and under laboratory conditions (Frezal & Félix, 2015).

Most offspring produced by self-fertilization are hermaphrodites. Each hermaphrodite can produce up to 300 offspring and only 0.1-0.2% of the progeny are male due to meiotic non-disjunction of the X chromosome. When a hermaphrodite is mated with a male, the reproduction capacity increases significantly, producing approximately 1000 offspring with an equivalent ratio between males and hermaphrodites (Riddle et al., 1997; Corsi et al., 2015).

The hermaphrodite reproductive system consists of a symmetrically bilobed U-shaped gonad, with one lobe extending anteriorly and the other posteriorly from the center of the animal. Each lobe contains an ovary and an oviduct terminating in a

spermatheca. Both lobes are joined to a common uterus and egg-laying apparatus in the midbody (Fig. 5A). The ovaries are syncytial, with germline nuclei, partially segregated by membranes, surrounding a central cytoplasmic core. As germ cells move proximally from the distal tip through the oviduct, the nuclei are first mitotic and then progress through the stages of meiosis. Oocytes are formed just before the bend of each gonadal lobe, when individual nuclei are almost completely enclosed by membranes. These oocytes are fertilized when they pass through the spermatheca, becoming embryos that will move into the uterus until they reach the vulva and exit (Fig. 5A).

The male gonad is a single J-shaped lobe, extending anteriorly from its distal end and then looping later and connecting with the cloaca near the tail. The J-shaped gonad consists of a testis, a seminal vesicle with a valve region, and the vas deferens duct that connects to the cloaca (Fig. 5B). At the gonad distal end, the germline nuclei are mitotic. Meiotic cells in progressively later stages of spermatogenesis are distributed sequentially along the gonad from the distal end to the seminal vesicle. Two meiotic divisions occur to produce the mature spermatids, which are stored in the seminal vesicle and released during copulation through the vas deferens to the cloaca (Wood, 1988; Emmons & Sternberg, 1997).

The male tail has specialized neurons, muscles, and specific epidermal structures for mating. The mail tail is composed of an elongated bursa, a cuticularized fan, and a proctodeum. Nine pairs of sensory rays are embedded in the fan. At the base of the tail are two spicules, which are inserted into the hermaphrodite vulva during copulation to aid in sperm transfer.

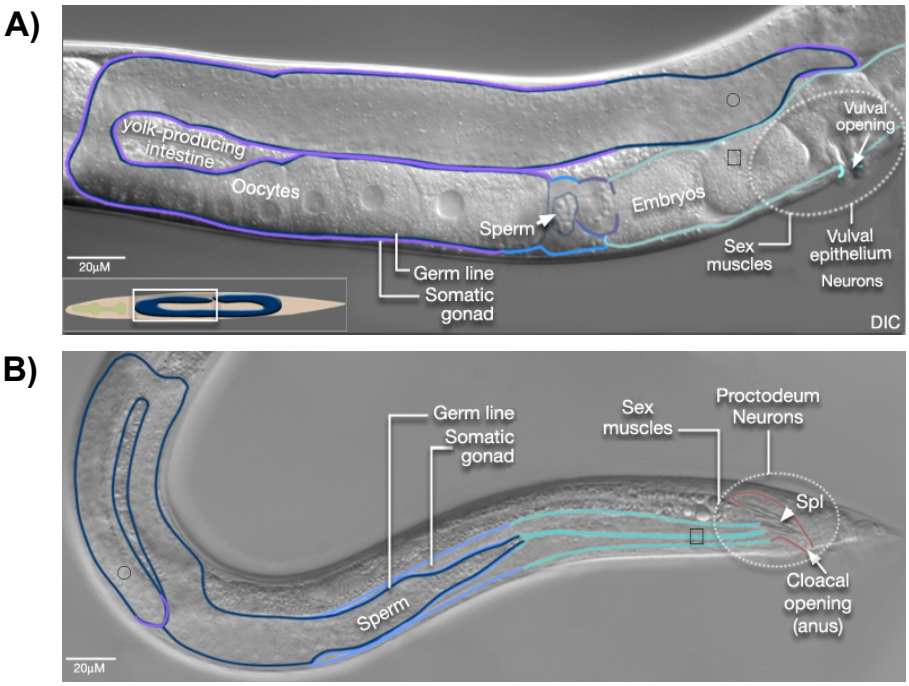


Figure 5. Nomarski microscopy images of hermaphrodite and male gonads. A) *C. elegans* hermaphrodite gonad. The syncytial germline that becomes oocytes before the bend of the gonad lobe can be observed. Oocytes are fertilized when they pass through the spermatheca to form embryos that will exit through the vulva. **B) *C. elegans* male gonad.** The picture shows the J-shaped male gonad. The seminal vesicle and the vas deferens that connects with the cloaca are marked. Images modified from those found at www.wormatlas.org.

1.3. *C. elegans* embryogenesis and life cycle

C. elegans embryogenesis takes approximately 14 hours at 22ºC. It starts with the fertilization of an oocyte. At this time, the anterior-posterior polarity of the embryo is set, as the entry site of the sperm pronucleus determines the posterior pole (Albertson, 1984). After fertilization, the maternal pronucleus that was arrested in prophase stage of the first meiosis ends meiosis (Albertson, 1984). Once the egg has been fertilized, an impermeable eggshell composed of three layers is formed: an inner layer of vitellogenin, another of chitin and a final one of lipids and proteins. This eggshell is required to generate an osmotic barrier between the embryo and the surrounding environment so that the embryo can develop independently from the mother. Additionally, this eggshell is required for correct polarization of the embryo and to minimize DNA segregation errors during meiosis (Johnston et al., 2006).

Embryogenesis can be divided into two phases of equal duration: (1) cell proliferation and organogenesis, and (2) morphogenesis. The first period takes approximately 7 hours and consists of rapid cellular proliferation with a precise temporal and spatial pattern in cell division, cell movements and cell death, invariant from one embryo to another, which gives rise to a fixed number of cells with rigidly determined fates. Five asymmetric cell divisions will give rise to the six precursor cells: AB, MS, E, C, D and P_4 (Sulston et al., 1983), each one with a specified fate (Fig. 6). During the next 7 hours, cell proliferation nearly ceases and morphogenesis begins. The body elongates, forming several folds. The embryo passes through two-fold stage to three-fold stage. This elongation is done by cytoskeletal rearrangements that form rings around the embryo, creating hydrostatic pressure (Ciarletta et al., 2009). At the same time, neural processes grow out and interconnect and, finally, the cuticle is secreted (Wood, 1988).

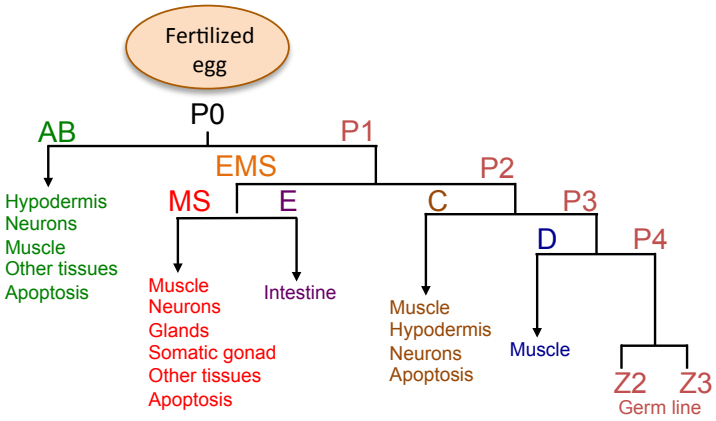


Figure 6. *C. elegans* embryo cell lineage. This figure represents the principal cell divisions that take place during embryogenesis. Five asymmetric cell division lead to the six cell precursor cells: AB, MS, E, C, D, P4. Each precursor cell forms specific tissues as is shown in the figure.

The embryo hatches and becomes a first stage larva (L1). This larva develops through four larval stages (L1-L4). At the end of each larval stage, a new cuticle is synthesized in a sleep-like period of inactivity called lethargus that ends with the molting of the old cuticle (Raizen et al., 2008). The L1 stage is ~12 hours long; the other stages are ~ 8 hours long. Approximately 8 hours after the L4 molt, adult hermaphrodites produce progeny for 2-3 days. Later, they can live several more weeks before dying of senescence (Corsi et al., 2015). Under some circumstances, when bacteria are depleted and the animals are crowded, L2 larva activate an alternative life cycle and enter a pre-dauer stage (L2d) followed by an alternative non-feeding L3 larval stage called the “dauer” larva (Golden & Riddle, 1984; Hu,

2007). This dauer larva is the *C. elegans* dispersal form most commonly encountered in the wild. They display active locomotion and a specific behaviour called nictation, where they stand on their tail and wave their body in the air. Remarkably, dauers may also congregate to form a column and nictate as a group (Félix & Duveau, 2012). The dauer larva can survive for many months. Its cuticle, which has enhanced resistance to chemicals, completely surrounds the animal and plugs the mouth, preventing the animal from eating and thereby arresting development. Upon return to more favorable conditions, they shed their mouth plugs, molt, and continue their development as slightly different L4 larvae (Corsi et al., 2015).

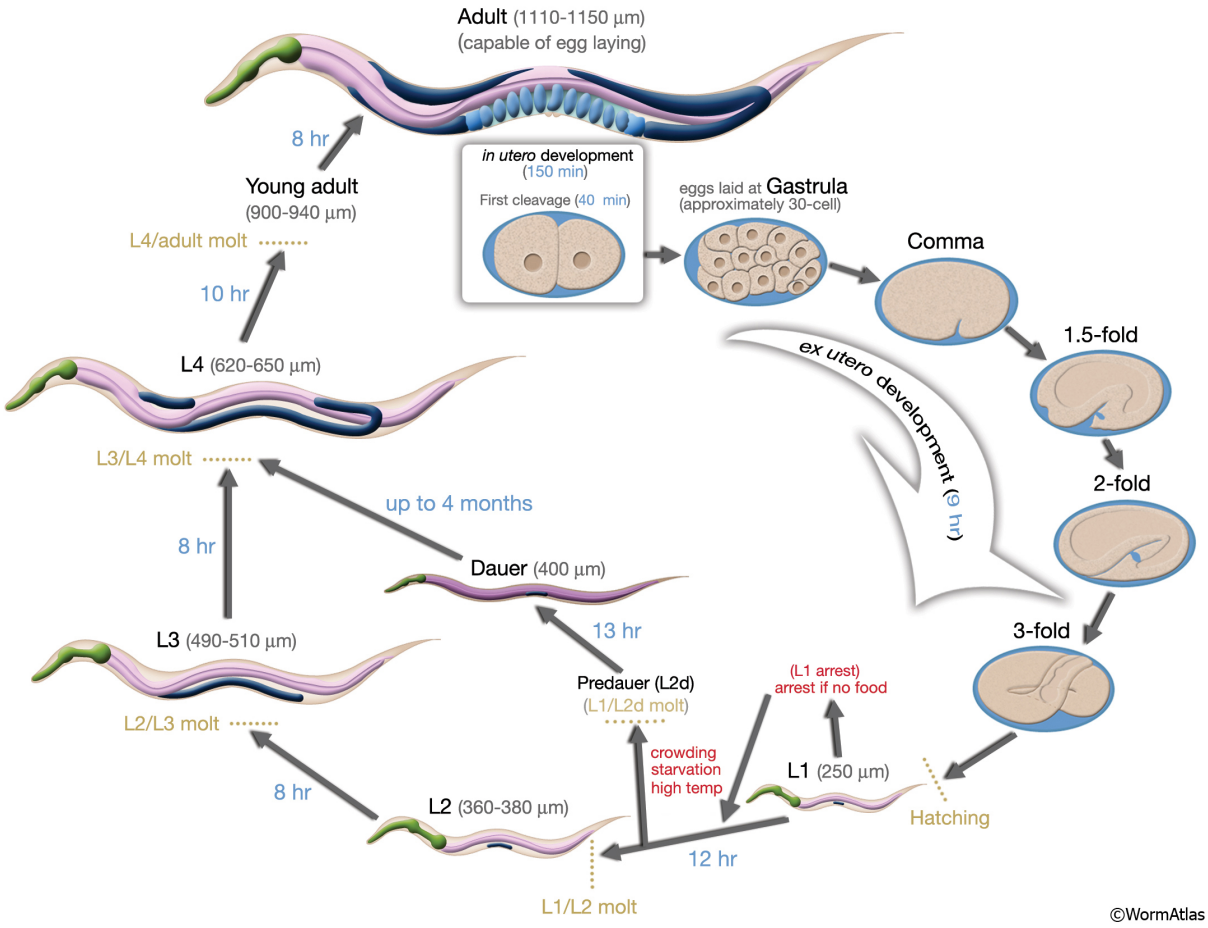


Figure 7. *C. elegans* life cycle at 22ºC. The fertilized egg begins embryogenesis inside the hermaphrodite uterus. The first asymmetric division takes place at 40 min post-fertilization. The eggs are laid at approximately the gastrula stage (150 min post-fertilization). Once the embryo hatches, the larva will develop through four larval stages (L1-L4). Under unfavourable conditions, L1 larva enter an alternative life cycle giving rise to dauer larva, a resistant non-feeding *C. elegans* stage which can live up to four months. Numbers in blue show the length of time the worms spend in a particular stage. The length of the animal at each stage is marked next to the stage name in micrometers (µm). Image taken from www.wormatlas.org.

1.4. *C. elegans* genome

The *C. elegans* genome was the first multicellular eukaryotic organism to be sequenced (*C. elegans* Sequencing Consortium, 1998). Its genome contains over 20,000 protein-coding genes and about 1,300 genes that are known to produce functional non-coding RNA (ncRNA) transcripts which include transfer RNA (tRNA) genes, ribosomal RNA (rRNA) genes, trans-spliced leader RNA (SL RNA) genes, microRNA (miRNA) genes, spliceosomal RNA genes, and small nucleolar (snoRNA) genes (Stricklin et al., 2005). The majority of *C. elegans* genes have been conserved throughout evolution. It is estimated that between 60-80% of this nematode´s genes have their corresponding orthologs in humans (Shaye & Greenwald, 2011).

Protein-coding genes are found similarly on either strand of DNA, and they are uniformly distributed throughout the six chromosomes (Spieth & Lawson, 2006). Both *C. elegans* sexes contain five autosomal chromosomes, also called linkage groups (LG) I, II, III, IV, and V and the X chromosome. Protein-coding genes are slightly denser on autosomes than on chromosome X and, in general, in the central regions of the autosomes (although the left arm of chromosome II is an exception) (Spieth & Lawson, 2006). *C. elegans* chromosomes are holocentric or polycentric, which means that during mitosis the microtubule spindle attaches to more than one position along the chromosome (Corsi et al., 2015).

C. elegans protein-coding genes are relatively small (the average gene size is 3 kb). Similarly to most eukaryotic protein-coding genes, *C. elegans* gene structure is defined by: a 5' untranslated (UTR) region, an open reading frame (ORF) and a 3'UTR region (Spieth & Lawson, 2006).

The *C. elegans* genome has two unusual aspects: most protein-coding mRNAs are *trans*-spliced (70%) and some genes are organized in operons (a cluster of genes under the control of a single promoter) (Zorio et al., 1994; Blumenthal, 2005). *Trans*-splicing is the addition of a 22-nucleotide leader sequences (SL1 or SL2) at the 5' end of mRNA (Blumenthal, 2005; Lasda & Blumenthal, 2011).

2. The Integrator complex: discovery, description and functions

In 2005, the Shiekhattar laboratory serendipitously discovered the Integrator Complex while searching for proteins that interacted with the Deleted in Split hand/Split foot 1 protein (DSS1). The initial affinity purification of the complex in human HeLa cells revealed that it was comprised of 12 subunits and associated with the carboxy-terminal domain (CTD) of the largest subunit (Rpb1) of the RNA polymerase II (RNAP II) (Baillat et al., 2005). This complex was found to function as the processing machinery for RNAP II-transcribed small nuclear RNAs (snRNAs) and was termed "Integrator complex" because it integrates the CTD of RNAP II largest subunit with the 3'-end processing of small nuclear RNAs U1 and U2. The 12 different subunits identified at that moment were named (Integrator 1-12) according to their predicted molecular weight. Later, a genome wide RNAi screen performed in *Drosophila* S2 cells found two additional subunits that were renamed Integrator 13 (also known as Asunder) and Integrator 14 (Chen et al., 2012).

Interestingly, most Integrator subunits do not have identifiable paralogs within the human genome. However, Integrator subunit 11 (Ints11) and Integrator subunit 9 (Ints9) are clearly homologous to the subunits of the cleavage and polyadenylation specificity factor, CPSF73 and CPSF100 respectively (Dominski et al., 2005), which are involved in the cleavage of pre-messenger RNAs and histone mRNAs (Xiang et al., 2014). Importantly, both belong to a large group of zinc-dependent nucleases called the β-CASP family (Callebaut et al., 2002). Ints11 and CPSF73 contain a β-CASP (metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2) β-lactamase domain that is essentially a β-lactamase domain that has been modified to allow the endonucleolytic cleavage of nucleic acids. Ints9 and CPSF100 also contain a β-CASP β-lactamase domain, however, it has been modified so as to render it inactive (Dominski et al., 2005). This relationship was very important to connect the Integrator complex with its function.

The involvement of the Integrator complex in the 3'-end processing of snRNAs was demonstrated in distinct studies. First, using RNA interference (RNAi), Baillat and colleagues depleted Ints11 (because it was the predicted catalytic subunit) or Ints1

(because it was the largest subunit) and processing of U1 or U2 snRNAs was assessed using the T1 ribonuclease protection assay. In both cases, an accumulation of primary snRNAs transcripts, observed by northern blot, was consistent with a defect in the processing of the 3'-end of snRNAs. Moreover, Ints11 was determined to be the catalytic subunit of the Integrator complex. An Ints11 mutant that lacked catalytic activity in the β-CASP domain was overexpressed, followed by depletion of endogenous Ints11, and an accumulation of primary snRNAs was observed (Baillat et al., 2005).

Later, an independent study developed a reporter system in *Drosophila* S2 cells to check the correct processing of snRNAs. In this study, the *Drosophila* U7 snRNA was chosen as the model snRNA gene because it is the smallest snRNA and it is predicted to have the least amount of secondary structure. The green fluorescent protein (GFP) gene and a polyadenylation signal (PAS) were placed downstream of the 3' box (the DNA sequence processing signal) of the U7 snRNA so that when transfected into cells, any U7 snRNAs transcribed from the mini reporter plasmid would be cleaved upstream of the 3' box and no GFP would be expressed. However, if the U7 was not properly cleaved, RNAP II would read-through to the PAS and the GFP would be expressed in these cells, indicating that snRNAs were being misprocessed (Ezzeddine et al., 2011). Using this system it was shown that depletion of Integrator 1, -4, -9 led to the highest levels of GFP expression and therefore misprocessing, followed by Integrator 11 and -7. Depletion of Integrator 2, -6, and -8 showed a weak response, while the depletion of Integrator 3 and -10 did not result in significant misprocessing of the reporter and depletion of Integrator 12 resulted in only modest expression of GFP. In addition it was shown that Integrator subunits knockdown affected endogenous *Drosophila* spliceosomal snRNAs by real RT-qPCR (Real Time quantitative Polymerase Chain Reaction) and northern blot.

The Integrator Complex appears to be broadly conserved throughout evolution. *In silico* analysis show that orthologs in other metazoans are really detectable. Some of the complex subunits are also present in many unicellular eukaryotes including amoebas and chromalveolates, suggesting an early evolutionary origin (Peart et al., 2013). However, the complex is absent in yeast such as *Saccharomyces cerevisiae* where the Nrd1/Nab3/Sen1 complex mediates snRNA 3'-end formation (Steinmetz et

al., 1996; Steinmetz et al., 2001; Peart et al., 2013). In certain fungal phyla such as Zygomycetes, some Integrator subunits can be identified, suggesting that the Integrator complex might have been lost through evolution in yeasts (Peart et al., 2013). A phylogenetic distribution of Integrator subunits is shown below (Table 1).

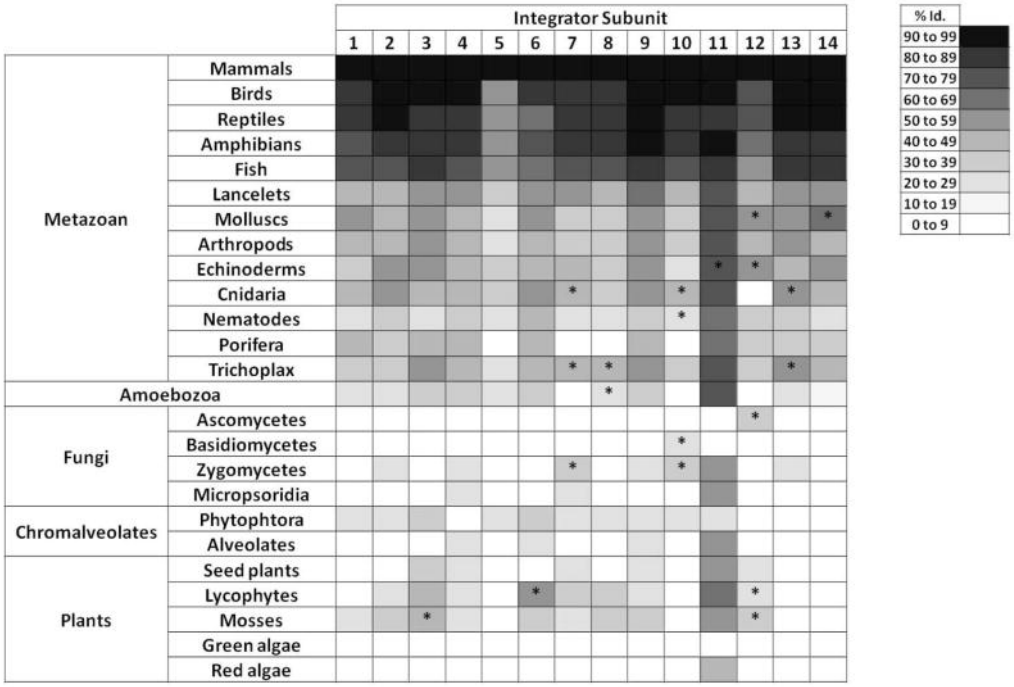


Table 1. Phylogenetic distribution of the Integrator complex subunits. Each column represents a subunit of the Integrator complex where the shading of each cell represents the level of identity between the human sequence and the considered organism(s). Asterisks indicate that a significant homology was detected only on a portion of the protein sequence. Image taken from Peart et al., 2013.

Following discovery of the Integrator complex, multiple studies implicating some members of the complex, either in snRNA processing or other biological functions, have emerged. The findings of these studies are summarized below.

Ints1 disruption in mouse embryos resulted in embryonic lethality suggesting that Ints1 is essential for survival. Further examination of the embryos showed an increase in misprocessed U2 snRNAs (Hata & Nakayama, 2007). Additionally, a recent study in murine embryonic stem cells differentiation suggested that Ints1 might regulate early hematopoietic development (Piazzi et al., 2015).

Interestingly, distinct studies demonstrated that Ints3 and Ints6 along with NABP1/2 and C9orf80, form a complex responsible for detecting DNA double-strand breaks

(DSBs) (Huang et al., 2009; Li et al., 2009; Skaar et al., 2009). The function of the Integrator complex related to DNA damage is further discussed in section 4.

Mutations in *Drosophila* Ints4 caused early developmental defects and embryos accumulated significant levels of misprocessed snRNAs (Ezzeddine et al., 2011).

In cultured human cells, at least Ints4 and Ints11 are required for the homeostasis of Cajal bodies, which are small nuclear bodies enriched with coilin protein and involved in RNA-related metabolic processes such as snRNPs biogenesis, maturation and recycling (Takata et al., 2012).

Zebrafish embryos treated with antisense morpholinos to Ints5 resulted in severe developmental defects. These embryos failed to develop hematopoietic cells. Aberrant splicing in the mRNAs of proteins necessary for hematopoiesis (Smad1/5 RNAs), probably caused because U1 and U2 snRNAs were not processed properly, was observed (Tao et al., 2009). In addition, Ints11 knockdown also led to improper Smad5 splicing, U1/U2 snRNA processing and arrested hematopoiesis (Tao et al., 2009).

Ints6 is the only member described before the Integrator complex was discovered. It was named *DICE1* (Deleted In Cancer 1). This gene was isolated as a candidate tumor suppressor in non-small cell lung carcinomas because it was located at a critical region of frequent loss of heterozygosity (human chromosome 13q14) in lung tumors and other tumors. Moreover, *DICE1* mRNA expression was reduced or undetectable in most non-small cell lung carcinoma cell lines tested (Wieland et al., 1999). Later, reduced *DICE1* expression was reported to be associated with hypermethylation of the CpG sites in the *DICE1* promoter in non-small-cell lung carcinoma cell lines and in prostate cancer cell lines (Wieland et al., 2001; Röpke et al., 2005). Additionally, *DICE1* exogenous re-expression in cancer cell lines led to inhibition of their capacity to form colonies *in vitro* (Wieland et al., 2004; Filleur et al., 2009).

In *C. elegans*, the Ints6/*DICE1* ortholog was called *dic-1* and its protein was reported to localize in the mitochondrial inner membrane where it promoted cristae tubules formation and was therefore related to mitochondrial activity (Han et al., 2006; Lee et al., 2009).

A recent *in vitro* study showed that Ints6 and Ints11 subunits were necessary for adipose differentiation. Expression levels of both subunits were increased in 3T3-L1 cells when the cells differentiated into adipocytes and their expression levels were reduced after the completion of cell differentiation (Otani et al., 2013).

In *Zebrafis*h a role for Ints6 in embryonic patterning was established. A recessive mutation was isolated and embryos from mutant mothers displayed delayed cell movements during gastrulation and severe dorsalization due to widespread de-repression of dorsal organizer genes (Kapp et al., 2013).

The *Drosophila* deflated/Ints7 gene is essential for normal development. Mutant individuals displayed early embryonic defects affecting many stages of development (Rutkowski & Warren, 2009; Ezzeddine et al., 2011). Previous reports from large-scale mutagenesis experiments and RNAi knockdown experiments demonstrated that the corresponding ortholog of Ints7 is essential for normal development in *Zebrafish* and *C. elegans* (Golling et al., 2002; Kamath et al., 2003).

Ints13, previously identified and named as Asunder (Anderson et al., 2009) is a critical regulator of dynein-mediated processes, either in *Drosophila* spermatogenesis/oogenesis (Anderson et al., 2009; Sitaram et al., 2014) or in cultured human cells at the onset of cell division (Jodoin et al., 2012). In addition to Ints13, other Integrator subunits play a role in dynein-mediated processes such as recruitment of cytoplasmic dynein to the nuclear envelope (Ints1, -6, -9, -11, or -12) (Jodoin, Sitaram et al., 2013) as well as in the dynein related process of ciliogenesis (Ints3, -4, -9, -11, -12) (Jodoin, Shboul et al., 2013).

Interestingly, a study revealed that the Integrator complex is necessary for a primate herpesvirus to generate microRNAs. The Herpesvirus saimiri (HVS) is a γ-herpesvirus that expresses Sm-class U RNAs in latently-infected marmoset T cells (Fickenscher & Fleckenstein, 2001). HVS produces microRNAs (miRNAs) by cotranscription of precursor miRNA (pre-miRNA) hairpins immediately downstream from viral small nuclear RNAs (snRNA). The Integrator complex of the host cell is required to recognize the 3'box and cut the snRNAs generating HVS pre-miRNAs (Cazalla et al., 2011). A later study identified a novel 3′ box-like sequence (miRNA 3′ box) downstream from HVS pre-miRNAs and demonstrated that 3′-end processing of

HVS pre-miRNAs also depends on Integrator activity (Xie et al., 2015).

Recent findings have extended Integrator functions into a broader spectrum of the RNAP II transcription cycle in addition to 3'-end processing, including transcription initiation, promoter-proximal pausing, elongation, and termination (Gardini et al., 2014; Stadelmayer et al., 2014; Lai et al., 2015; Skaar et al., 2015). These RNAP II-related functions are described in detail in the next section (3).

Strikingly, Lai and colleagues very recently demonstrated that Integrator complex is recruited to enhancers and super-enhancers (active enhancers densely clustered in a ~10–30 kb region) in a stimulus-dependent manner where it promotes the transcription of enhancer RNAs (eRNAs) and the enhancer–promoter chromatin looping (Lai et al., 2015). eRNAs are relatively short non-coding RNA molecules (50-2000 bp) transcribed from the DNA sequence of distal regulatory elements called enhancers (Li et al., 2016). Enhancers and their transcripts (eRNAs) play a key role in gene expression, allowing tissue- and temporal-specific regulation (Li et al., 2016).

In this study, serum-starved HeLa cells were treated with EGF (epidermal growth factor) to induce immediate early genes (IEGs). EGF-induced enhancers displayed bi-directional eRNAs that were predominantly not polyadenylated (Lai et al., 2015). These EGF-induced enhancers were occupied by a detectable amount of Integrator complex before EGF induction that resulted in further Integrator complex recruitment after EGF treatment. In addition, functional depletion with short hairpin RNAs (shRNAs) against Ints1 or Ints11 resulted in diminished eRNA levels induced by EGF stimulation, and an abrogated stimulus-induced enhancer–promoter chromatin looping, inhibiting enhancer and promoter communication (Lai et al., 2015).

Importantly, in the absence of Integrator, eRNAs remain bound to RNAP II and their primary transcripts accumulate, which suggests an Integrator complex-dependent 3'-end cleavage and transcription termination of eRNAs (Lai et al., 2015).

To learn more about Integrator complex functions and mechanisms, the sequences of the 14 Integrator subunits were closely examined for secondary structural features (Fig. 8) (Chen & Wagner, 2010; Baillat & Wagner, 2015). In addition to the β-CASP β-lactamase domains found in Ints9 and -11, a detailed analysis demonstrated that Ints4 had HEAT (Huntingtin, Elongation factor 3, protein phosphatase 2A, and the

yeast kinase TOR1) repeats, Ints4 and -7 had ARM (Armadillo) repeats, Ints6 and -14 had a VWA (von Willebrand factor type A) domain and a ISDCC (INTS6/SAGE1/DDX26B/CT45 C-terminus) domain, Ints8 had TPR (tetratricopeptide repeats) domains, Ints12 had a PHD (plant homeodomain finger) domain and Ints13 had a COIL domain (coiled coil domain).

Armadillo, HEAT, and TPR repeats as well as von Willebrand factor A domains are all involved in protein-protein interactions as well as intracellular transport. The PHD finger is a chromatin-binding domain. In addition, a few Integrator subunits (Ints1, Ints3) were shown to have domains of unknown function (DUF), areas identified as functional domains but with no correlation to any other family of existing known domains.

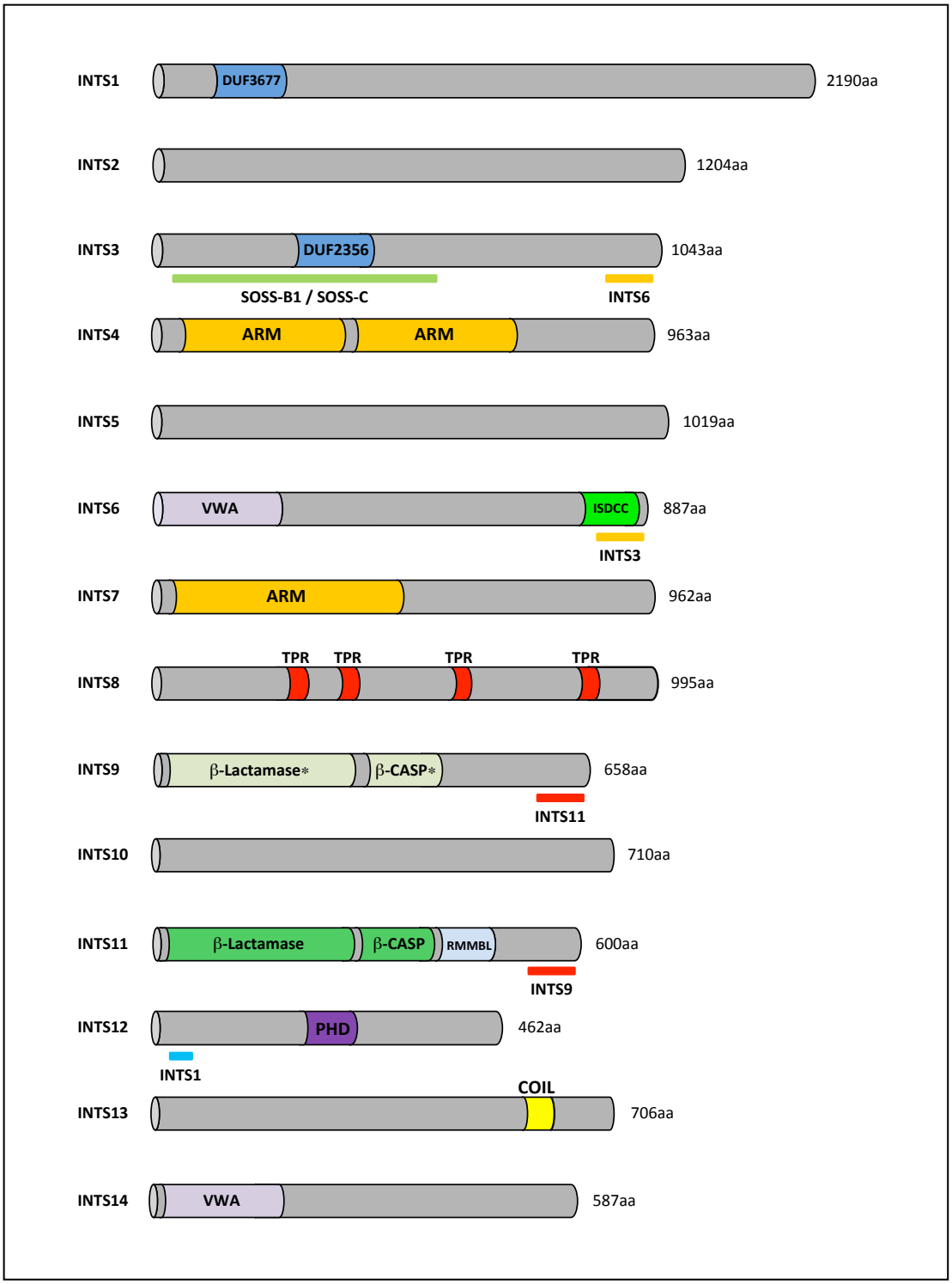


Figure 8. Scheme of the Integrator complex subunits and their predicted domains. Abbreviations: **ARM**, armadillo-like repeats; **COIL**, coiled coil domain; **DUF**, domain of unknown function; **ISDCC**, INTS6/SAGE1/DDX26B/CT45 C- terminus; **PHD**, plant homeodomain finger; **TPR**, tetratricopeptide repeats; **β-lactamase/β-CASP** (* indicates the presence of an inactive β-lactamase/β-CASP domain); VWA, von Willebrand type A-like domain. Identified interacting domains with other proteins are underlined. Image taken from Baillat & Wagner, 2015.

3. Transcription & the Integrator complex

Transcription is the process by which the information in a strand of DNA is copied into a new molecule of RNA. In eukaryotic cells, several chromatin-associated proteins are responsible for packaging the genome into the nucleus. Basically, chromosomes are wrapped around histone proteins to form nucleosomes, the major subunits of chromatin. To make a gene accessible for transcription, structural changes in the chromatin must occur. Chromatin does not only have a structural role but it is also a critical component of transcriptional regulation. For example, chromatin can repress gene expression by blocking the access of transcription factors to the DNA (Venkatesh & Workman, 2015).

In eukaryotes, three distinct DNA-dependent RNA polymerases: **RNA polymerase I** (RNAP I), **RNA polymerase II** (RNAP II) and **RNA polymerase III** (RNAP III) execute transcription. None of these binds DNA directly, but rather they are recruited to the DNA by other proteins. These three RNA polymerases transcribe different classes of genes. RNAP I is used for the synthesis of large ribosomal RNAs (rRNA). RNAP II transcribes messenger RNAs (mRNAs), the Sm-class of small nuclear RNAs (snRNAs) and other non-coding RNAs (ncRNAs). RNAP III transcribes all transfer RNAs (tRNAs), one rRNA (5S rRNA) and several small RNAs (including the Lsm-class snRNAs) (Cramer et al., 2008).

3.1. RNAP II mediated transcription

The eukaryotic RNAP II holoenzyme is a complex of 12 subunits. RNAP II transcribes mRNAs, the Sm-class of snRNAs (described in 3.1.2) and other ncRNAs. The transcription cycle in mRNAs has been the most studied and is the best known. Although the process of transcription is similar, there are some differences depending on the type of gene to be transcribed. Here, we will review the transcription cycle of mRNAs and the Sm-class of snRNAs.

Rpb1, the large subunit of RNAP II, contains a repeated motif on its C-terminal domain (CTD), which consists of tandem repeats of a heptapeptide sequence (Tyr_1–Ser_2–Pro_3–Thr_4–Ser_5–Pro_6–Ser_7) varying in number from 26 in yeast to 52 in vertebrates.

Transcription by RNAP II is a tightly regulated process. In metazoans, it can be regulated at several distinct steps along the transcription cycle. The RNAP II CTD domain plays an essential role in all the transcription regulation steps and also couples transcription and processing of the nascent RNA. All residues within the CTD can be modified either by phosphorylation (Tyr, Thr, Ser) or isomerization (Pro) (Hsin & Manley, 2012; Zaborowska et al., 2016).

The addition of a 5' cap (generally an N7-methylated guanosine linked to the first nucleotide of the RNA via a reverse 5′ to 5′ triphosphate linkage) and processing of the nascent transcript occurs co-transcriptionally. Once a gene is transcribed, it can undergo multiple rounds of transcription based on the efficiency of reinitiation (Shandilya & Roberts, 2012; Sainsbury et al., 2015; Jonkers & Lis, 2015; Porrua & Libri, 2015).

3.1.1. RNAP II transcription of mRNAs

The RNP II transcription cycle is divided into three distinct phases (Shandilya & Roberts, 2012; Sainsbury et al., 2015; Jonkers & Lis, 2015; Porrua & Libri, 2015):

1) **Initiation**: assembly of the preinitiation complex (PIC) (composed of general transcription factors: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, the Mediator complex and the RNAP II) and production of the first few phosphodiester bonds in the RNA transcript.

2) **Elongation**: progression of RNAP II through a locus as it transcribes the gene.

3) **Termination**: release of RNAP II when it reaches the end of the gene being transcribed.

- **Initiation:**

An essential step for initiating transcription is formation of the PIC at gene promoters. Each gene can be classified depending on its *core promoter elements* (CPE), which are DNA sequences located upstream or downstream the transcription start site

(TSS) or even within the coding region. The most studied core promoter element for RNAP II is the *TATA box*, a sequence rich in A and T nucleotides, located between 25 and 30 bases upstream of the TSS. The presence or absence of a TATA box is used broadly to classify promoters as TATA- containing or TATA-less (Mathis & Chambon, 1981; Shandilya & Roberts, 2012).

A critical step for PIC formation (Fig. 9) is the recruitment of TBP (TATA-binding protein) containing TFIID, a general transcription factor that recognizes the TATA sequence and binds to it specifically, along with several TBP-associated factors (TAFs). This process is tightly regulated, both positively and negatively. Upon binding to the promoter, TBP extensively distorts the TATA sequence and facilitates the ordered assembly of other general transcription factors (GTFs) (Sainsbury et al., 2015). However, most eukaryotic genes have TATA-less promoters. In these cases, the TFIID and other similar complexes can recognize other sequences such as Initiator (Inr) and downstream promoter elements (DPE) to bind to the DNA and form the PIC (Sainsbury et al., 2015).

The transcriptional coactivator complex Mediator (composed of 26 subunits in humans) was shown to stabilize and facilitate PIC formation (Kelleher et al., 1990; Kim et al., 1994; Baek et al., 2002; Esnault et al., 2008). The Mediator complex enables hypo-phosphorylated RNAP II recruitment via interaction with its CTD domain. The large size of the Mediator likely promotes PIC stability by interacting with multiple PIC factors (Allen & Taatjes, 2015).

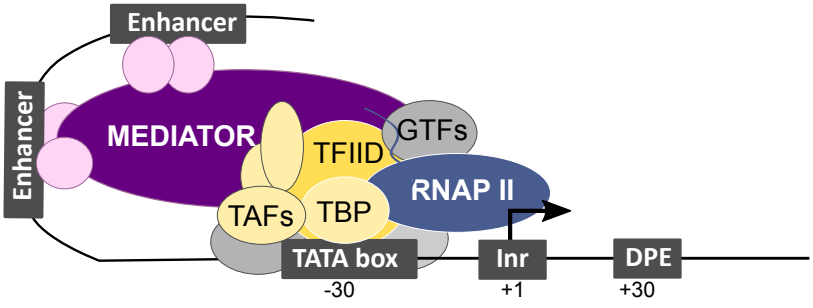


Figure 9. Schematic drawing of the PIC. The PIC is comprised of GTFs, the Mediator complex and the RNAP II. The TFIID containing the TBP is bound to the TATA box, along with several TAFs, located in the promoter. In genes that do not have TATA box, the TFIID could be bound to the Inr or DPE sequences. The transcriptional coactivator Mediator enables the recruitment of RNAP II through interaction with its CTD domain.

Once the PIC is formed, the ATPase and helicase activities within TFIIH create a negative superhelical tension in the DNA, resulting in the transition from a closed to an open PIC, allowing RNAP II to incorporate the first nucleotides into the nascent RNA. Also, TFIIH phosphorylates RNAP II CTD on Ser5 via its CDK7 kinase subunit, an important marker in transcription initiation (Shandilya & Roberts, 2012; Sainsbury et al., 2015).

When ~25 nucleotides of nascent RNA are synthesized, the 7-methylguanosine cap (m^7G) structure is attached to its 5′ end (Ramanathan et al., 2016). To proceed with elongation RNAP II must dissociate from PIC to escape the promoter. Although the full mechanism remains to be discovered, it is known that phosphorylation of the RNAP II CTD plays a key role in disrupting the Mediator–RNAP II interaction. The ability of TFIIH to phosphorylate the RNAP II CTD is dependent on the Mediator. Therefore, the Mediator triggers its own release from the RNAP II (Shandilya & Roberts, 2012; Sainsbury et al., 2015).

Importantly, the Mediator complex and TFs remain associated with the promoter as part of the scaffold complex to facilitate subsequent rounds of RNAP II recruitment and reinitiation (Yudkovsky et al., 2000; Sainsbury et al., 2015).

- **Elongation**

Early elongation: RNAP II transcriptional pause/release

Following transcription initiation in higher eukaryotes, the RNAP II can be paused at gene promoters (40-60 nt downstream of the TSS), predominantly on inducible and developmentally regulated genes. This RNAP II pausing serves to further regulate these genes depending on the frequency and probability of pause release. It is a rate-limiting step on more than 70% of metazoan genes (Adelman & Lis, 2012).

RNAP II pausing is regulated by **DSIF** (*DRB (5,6-dichloro-1β-D-ribofuranosylbenzimidazole) Sensitivity Inducing Factor*) and **NELF** (*Negative Elongation Factor*) proteins, which bind to and inhibit RNAP II function. Phosphorylation of the largest subunits of DSIF and NELF (Spt5 and NELF-E respectively) by a cyclin dependent kinase (CDK9) associated with the **p-TEFb**

(*Positive Transcription Elongation Factor b*) reverses their transcription blocking. NELF is evicted from RNAP II upon phosphorylation and DSIF becomes a positive elongation factor. In some cases, nucleosomes may also contribute to pausing (Jonkers & Lis, 2015).

In a recent study, Gardini and colleagues showed that the Integrator complex plays a critical role in transcription initiation and the release of paused RNAP II at immediate early genes (IEGs) following stimulation with epidermal growth factor (EGF) in HeLa cells (Gardini et al., 2014). Thus, a model emerged in which recruitment of the Integrator complex to pause transcription sites was necessary for the subsequent recruitment of p-TEFb (Fig.10).

IEGs are genes activated rapidly in response to a wide variety of sitimuli. Most IEGs, such as c-Fos are regulated through the RNAP II pause/release mechanism (Plet et al., 1995). The authors observed an increased accumulation of Integrator (Ints1, Ints9, Ints11) at TSS and throughout the body of IEGs after EGF stimulation, which was abrogated when Integrator subunits were knocked down (Ints1 and Ints11). As a result of the absence of the Integrator complex at the TSS, RNAP II failed to escape pausing and progress into productive elongation (Gardini et al., 2014). This study also showed that the Integrator was necessary for stimulus-dependent recruitment of the large super elongation complex (SEC) to promoter-proximal paused genes (Fig. 10). The SEC is required for rapid transcription in response to external stimuli such as heat shock, retinoic acid or serum treatment. SEC can vary in composition depending on the cellular context (Luo et al., 2012). It contains the most active form of p-TEFb, which leads to the release of paused RNAP II and resumption of productive elongation (Fig.10) (Gardini et al., 2014). SEC can also interact with a subset of co-activators such as the Mediator complex or a polymerase-associated factor 1 (PAF1) (Luo et al., 2012).

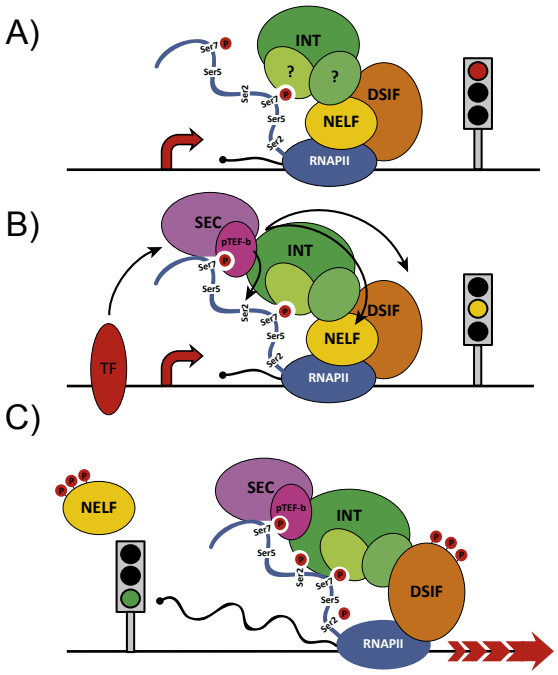


Figure 10. The role of the Integrator complex in RNAP II promoter-proximal pause-release. A) Under non-stimulated conditions, RNAP II initiates transcription and pauses 40–60nt downstream of the TSS. The DSIF, NELF and likely, the Integrator complex, inhibit RNAP II elongation. The Integrator complex is associated with the RNAP II CTD through Ser7P recognition. **B)** After a stimulus, the Integrator complex is further enriched at the pause site and recruits p-TEFb and SEC, which phosphorylates the DSIF, NELF, and Ser2 of the RNAP II CTD. **C)** Once phosphorylated, NELF is displaced, DSIF becomes a positive regulator of elongation, and the polymerase proceeds to elongate. Image taken from Baillat et al., 2015.

Another study conducted by Stadelmayer et al. revealed the interaction between the Integrator complex and Spt5 and NELF, the elongation machinery components (Stadelmayer et al., 2014). However, in this work a defect in promoter-proximal pausing following knockdown of Ints11 or Ints9 (but not Ints3) was observed in the genes studied, which were previously chosen because a correlation between RNAP II pausing at the TSS and Integrator and NELF binding closely was found. This same study also showed that Ints11 subunit was required for RNAP II processivity (Stadelmayer et al., 2014).

Regardless of the differences, both studies revealed a role for the Integrator complex in the transcriptional regulation of protein-coding genes.

Productive elongation:

Once RNAP II escapes the promoter, it goes to the coding region of the gene to elongate the transcript. During elongation, splicing factors are recruited to perform co-transcriptional splicing of the nascent pre-mRNA transcript (Saldi et al., 2016).

The CTD was phosphorylated on Ser5 during initiation. As RNAP II elongates, Ser2 is increasingly phosphorylated, first by the CDK9 subunit of p-TEFb and then by CKD12, while Ser5 phosphorylation is gradually removed by phosphatases. Although the process is considerably more complex, the phosphorylation pattern of Ser5P around the TSS, and Ser2P toward the ends of the transcribed genes functions to recruit the transcription-associated proteins required at different phases of the transcription cycle (Zaborowska et al., 2016).

- **Termination**

Transcription termination occurs when the RNAP II dissociates from the DNA template. In yeast and higher eukaryotes, the molecular mechanisms for dismantling the elongation complexes are not fully understood.

It is widely accepted that the 3′-end formation of mRNA precursors plays a central role in RNAP II transcription termination. In addition, it is important for mRNA export and stability. Addition of the poly(A) tail determines its translational efficiency. There are distinct 3'-end processing and termination mechanisms depending on the nascent transcript. Although the precise mechanism is poorly understood, three different complexes have been described for distinct nascent RNAs: poly(A) mRNAs, replication-dependent histone mRNAs and snRNAs. As mentioned previously, the snRNAs are processed by the Integrator complex. Their characteristics, transcription and maturation are described in detailed in the next section (3.1.2). Figure 11 represents the three distinct 3'-end processing machineries that have been described.

In poly(A) mRNAs, members of the Cleavage and Polyadenylation Specificity (CPSF) and Cleavage Stimulation Factor (CstF) complexes are vital to 3′-end processing and transcription termination (Mandel et al., 2008). Both CPSF and CstF complexes are recruited to the predominantly Ser2-phosphorylated CTD repeats of RNAP II (Ahn et al., 2004). Cleavage of the nascent transcript is produced 18-30 bp

downstream of the polyadenylation signal (PAS: 5'-AAUAAA-3') by CPSF73 (component of CPSF complex) (Proudfoot & Brownlee, 1976; Mandel et al., 2008).

Distinct models have emerged to explain the transcription termination coupled to polyadenylation of mRNAs. In the *Torpedo model*, termination occurs following cleavage and exonucleolytic degradation of the nascent RNA by the 5' to 3' exonuclease XRN2 (homolog of the yeast Rat1) (West et al., 2004). The helicase senataxin (SETX, homolog of the yeast Sen1) has been suggested to participate in termination by resolving R-loops (RNA-DNA hybrids) to allow the entry of XRN2 (Porrua & Libri, 2015). In the *Allosteric model*, when the elongating RNAP II encounters a polyA signal, a physical change in the complex is triggered by loss of associated proteins on its CTD that provokes termination (Logan et al., 1987, Zhang & Gilmour, 2006).

Although it remains a matter of debate, transcription termination in metazoans is thought to be associated with RNAP II pausing. Once RNAP II transcribes the polyadenylation signal, there is a marked reduction in its processivity, which leads to pausing further downstream. This might facilitate exonucleolytic degradation by XRN2 or conformational changes that lead to termination (Porrua & Libri, 2015, Loya & Reines, 2016).

In the case of metazoan replication-dependent histone mRNAs, whose transcripts lack of introns and which are the only known non-polyadenylated mRNAs in eukaryotes, their 3'-end processing leads to stable and translatable transcripts characterized by their terminal secondary structure. Although there are many similarities between the endonucleolytic cleavage of polyadenylated mRNAs and histone mRNAs, the mechanism is different.

The sequence of histone pre-mRNAs is characterized by a 3' end that contains a conserved stem-loop (SL) sequence, an AC-rich sequence after the stem-loop and a histone downstream element (HDE), which is a purine-rich sequence located about 15 nucleotides after the cleavage site. The SL sequence is recognized by the stem-loop-binding protein (SLBP) and the HDE is recognized by the U7 snRNA, which is a component of the U7snRNP. Afterwards, a cleavage factor is recruited that probably consists of CPSF73/CPSF100 and Symplekin (Marzluff et al., 2008). Notably, a

recent study revealed that apart from snRNA 3'-end processing defects, depletion of Integrator subunits also resulted in disruption of replication-dependent histone mRNAs 3'-end processing and as result of the read-through, snRNAs and replication-dependent histone mRNAs acquired cryptic poly(A) signals (Skaar et al., 2015).

Additionally, RNAP II transcribes several types of long non-conding RNAs of diverse functions whose 3' end processing mechanisms are still unknown.

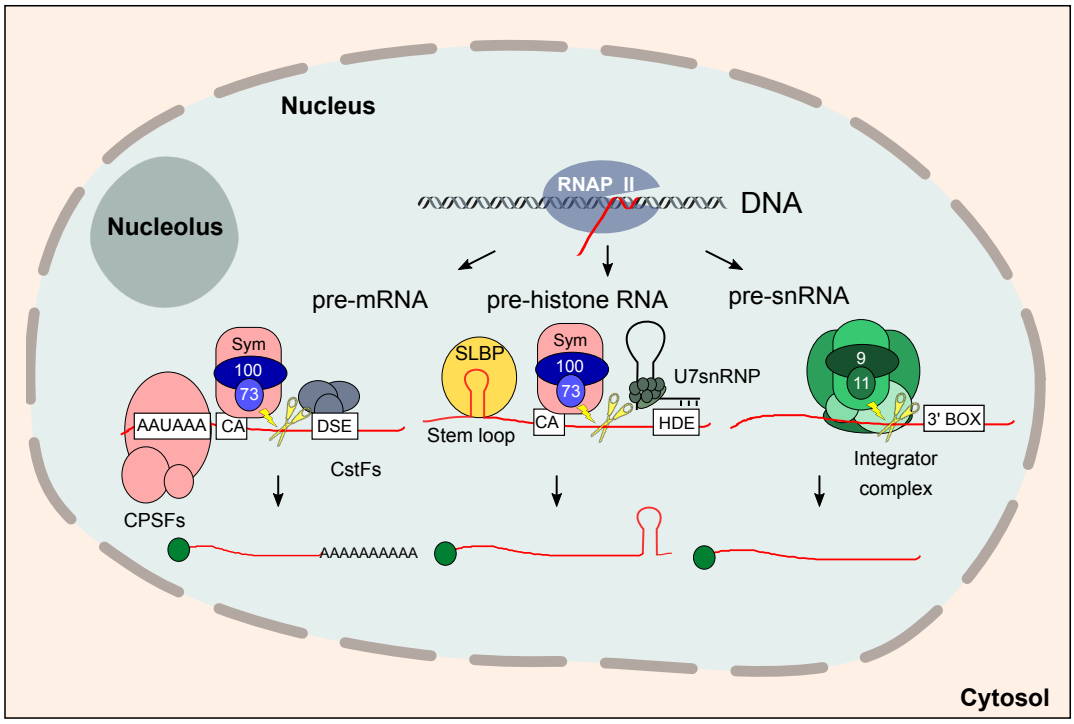


Figure 11. Scheme of the different complexes involved in 3'-end processing of the RNAP II transcripts: poly(A) mRNAs, histone RNAs and snRNAs. Formation of the 3'end of poly(A) pre-mRNAs is dependent on protein complexes binding to signals just upstream and downstream from the future cleavage site. The upstream site is the PAS signal (AAUAAA), the downstream sequence element (DSE) is a GU/U-rich sequence. A component of the CPSF complex binds to the PAS, and a component of the CstF complex binds to the DSE. Once these two sequences are recognized, the subunits CPSF73, CPSF100, and a large scaffold protein called Symplekin are recruited. In histone mRNAs, the SLBP recognizes the stem loop sequence and the U7 snRNP recognizes the HDE sequence. A cleavage complex containing CPSF73 (the endonuclease that performs the cleavage), CPSF100, Symplekin, and some unknown factors are recruited to cleave the pre-mRNA. Both cleavage reactions, pre-poly(A)mRNAs and pre-histone mRNAs, occur downstream of the CA nucleotide. In snRNAs, the Integrator complex, along with other factors, is responsible for the cleavage. Its Ints9 and Ints11 subunits are paralogs of the CPSF100 and CPSF73, respectively (the processing by the Integrator complex is described in next section 3.1.2).

❖ **Box.1 *C. elegans* RNAP II transcription**

RNAP II transcriptional regulation in *C. elegans* can be described as typical for eukaryotes. The large subunit of RNAP II in *C. elegans* has a CTD of 38 tandem repeats based upon the YSPTSPS sequence. The phosphorylation patterns of the CTD are similar to those of other metazoans, the Ser2P levels increase while the Ser5P levels decrease along transcriptional elongation.

The only major difference between the transcriptional machinery of *C. elegans* and other metazoans is the **absence of NELF**. However, evidence suggests that transcriptional regulation by RNAP II pausing also exits in *C. elegans*. While the release of RNAP II from initiation into elongation may not be a ubiquitous rate-limiting step during normal growth in *C. elegans,* recent genome-wide GRO-seq (Genomic Run-On sequencing) experiments in starved larvae revealed a significant number of genes showing a 5' accumulation of RNAP II (Reinke et al., 2013).

Interestingly, recent studies have identified **tissue-specific regulation** of Ser2 phosphorylation in *C. elegans*. Overall, in germline tissues, Ser2P does not require the activity of CDK-9 but rather CDK12, a pattern of Ser2P regulation surprisingly similar to that seen in yeast. In contrast, regulation of Ser2P in soma is similar to that seen in *Drosophila* and mammalian systems and requires the activity of CDK-9 (Bowman & Kelly, 2014).

3.1.2. RNAP II transcription of snRNA genes

➢ **Small nuclear RNAs (snRNAs): structure and function**

Small nuclear RNAs are commonly referred to as "uridine-rich small nuclear RNAs" (U snRNAs). The U snRNAs are small non-coding RNAs (60–200 nucleotides) that are ubiquitous, intronless, non-polyadenylated and generally highly expressed. They are predominantly located in the nucleus and that is also where they function

(Matera et al., 2007; Matera & Wang 2014). Basically, the snRNAs can be divided into two classes based on common sequence features and protein cofactors. The Sm-class (comprised of U1, U2, U4, U4atac, U5, U7, U11 and U12) is transcribed by RNAP II. The Lsm-class (comprised of U6 and U6atac) is transcribed by RNAP III (Matera et al., 2007). Additionally, in lower eukaryotes such as *C. elegans* or *Trypanosoma*, there is another class of snRNAs called SL snRNAs.

Structurally, Sm-class RNAs are characterized by a 5′-trimethylguanosine cap (TMG), a 3′ stem–loop and a binding site for a group of seven Sm proteins (the Sm site) that form a heteroheptameric ring structure (Fig. 12A). Lsm-class RNAs contain a 5'-monomethylphosphate cap (MPC) and a 3′ stem–loop, terminating in a stretch of uridines that form the binding site for a distinct heteroheptameric ring of Lsm proteins (Fig. 12A) (Matera et al., 2007).

These U snRNAs assemble with proteins to form small nuclear ribonucleoproteins (snRNPs). In metazoans, the Sm site and the 3′-stem-loop are required for recognition by the survival motor neuron (SMN) complex for assembly into stable core ribonucleoproteins (RNPs), whereas the TMG cap and the assembled Sm core are required for recognition by the nuclear import machinery (Matera et al., 2007).

Except for the U7 snRNP that is involved in the 3′-end processing of the replication-dependent histone mRNAs (Marzluff, 2005), the other U snRNPs are part of the major (comprised of U1, U2, U4, U5 and U6; secondary structures are shown in Fig. 12B) and minor (comprised of U11, U12, U4atac, U5 and U6atac) spliceosomes that mediate pre-mRNA splicing (Matera et al., 2007). The minor spliceosome is present in only a subset of eukaryotes and is less abundant. It splices the rare introns with different splice site sequences. Unlike the major spliceosome, it is found outside the nucleus. Spliceosome assembly occurs by the ordered interaction of the spliceosomal snRNPs and numerous other splicing factors. Both the conformation and composition of the spliceosome are highly dynamic, affording the splicing machinery accuracy and flexibility at the same time (Will & Luhrmann, 2011).

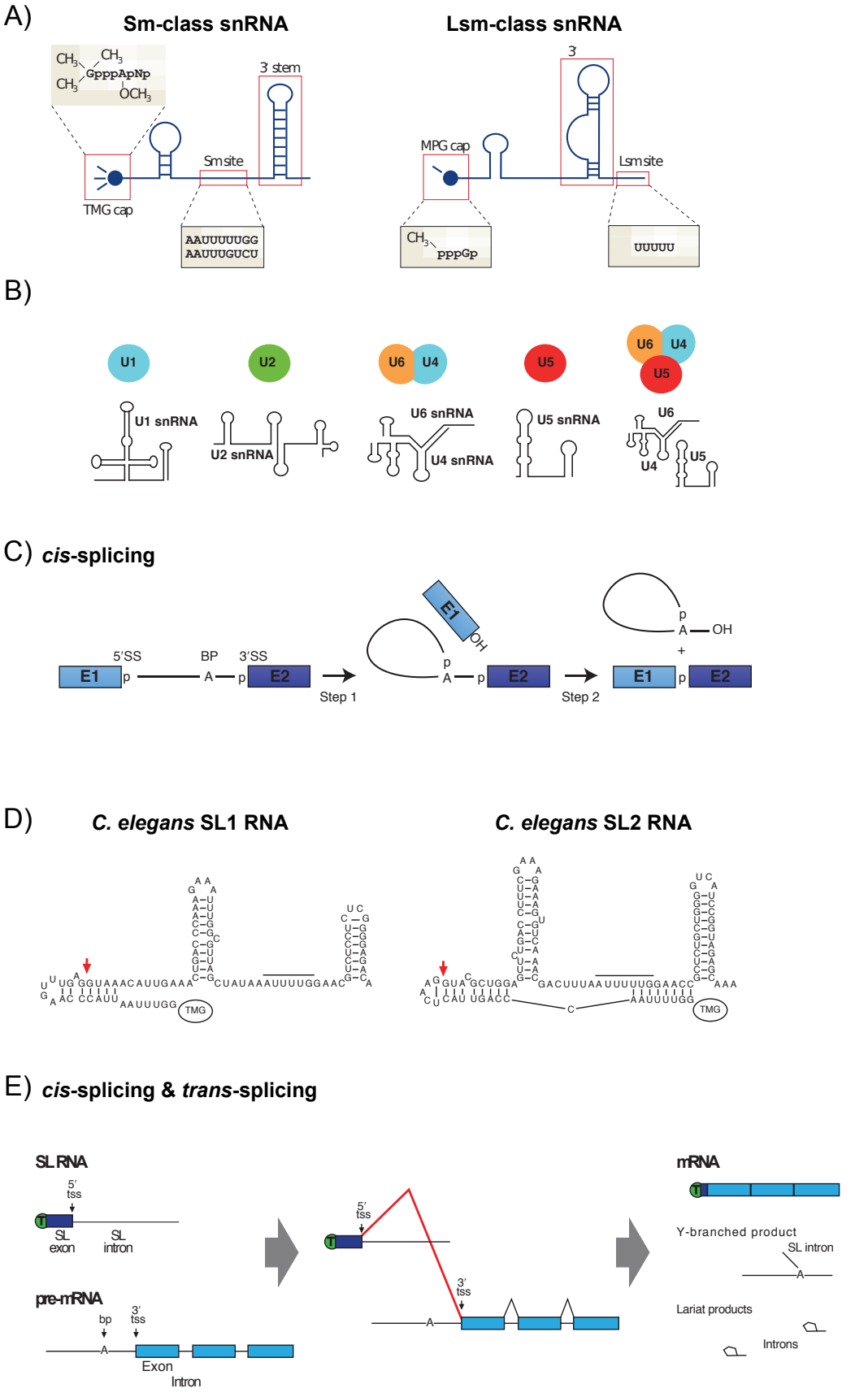


Figure 12. A) Features of Sm- and Lsm-class snRNAs. Sm-class snRNAs are transcribed by RNAP II and contain three important recognition elements (boxed): a TMG cap, an Sm-protein-binding site (Sm site) and a 3′ stem–loop structure. The consensus sequence of the Sm site and the

sequence of the specialized U7 Sm site are shown. The consensus Sm site directs assembly of the canonical heptameric protein ring (Sm core). Lsm-class snRNAs are transcribed by RNAP III and contain an MPG cap, a 3′ stem and terminate in a stretch of uridine residues (the Lsm site) that is bound by the Lsm core. U6 and U6atac are the only known Lsm-class snRNAs. Image taken from Matera et al., 2007. **B) snRNA secondary structures of the major human spliceosomal snRNPs.** Image taken from Will & Luhrmann, 2011. **C) Schematic representation of pre-mRNA *cis*-splicing.** The two consecutive trans-esterification reactions that lead to exon joining and intron removal are represented. Exons are indicated by blue boxes and introns are shown as lines. The branch site adenosine is indicated by the letter A and the phosphate groups at the 5′ and 3′ splice sites are shown by the letter p. Image taken from Will & Luhrmann, 2011. **D) *C. elegans* SL1 and SL2 sequences and predicted secondary structures.** Red arrows mark the 5′ trans-splice sites. The 5′ TMG cap is shown by a black oval. A horizontal line indicates predicted Sm protein binding sites. Image taken from Lasda & Blumenthal, 2011.
E) Schematic representation of SL *trans*-splicing and *cis*-splicing. (Left) SL RNA and pre-mRNA precursor molecules. (Middle) SL *trans*- and *cis*-splicing. SL *trans*-splicing (red lines) joins the 5′ tss located on the SL RNA with the 3′ tss located on the pre-mRNA. *Cis*-splicing joins exons of the same transcript. (Right) Resulting products of the splicing reactions. The mature mRNA has the SL exon and the TMG cap at its 5′ end, and can be exported to the cytoplasm and translated.
The Y-branched product that is the SL intron attached to the pre-mRNA at the branchpoint and the lariat introns that result from *cis*-splicing are rapidly degraded. Exons are represented by boxes and introns as lines. (tss) trans-splice site; (bp) branchpoint. Green circle with T represents the TMG cap on the SL exon. Image adapted from Lasda & Blumenthal, 2011.

Splicing or *cis*-splicing consists of the removal of introns (non-coding sequences) and ligation of exons (coding sequences) within an mRNA molecule (Fig. 12C) (Will & Luhrmann, 2011). Also, exons can be alternatively selected from the same RNA molecule to create variable forms of the final mRNA, which is called “alternative splicing”. This splicing method is prevalent in higher eukaryotes and increases the number of unique proteins from a single mRNA molecule (Nilsen & Graveley, 2010).

Introns are defined by short specific sequences within a pre-mRNA: a 5’ splice site (ss), a 3’ ss and a branch site (BS). The BS is typically located 18-40 nucleotides upstream from the 3’ ss and in higher eukaryotes, is followed by a polypyrimidine tract. Introns are removed by two consecutive *trans-esterification* reactions. First, the 2′ hydroxyl (OH) group of the conserved adenosine within the branching site 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 and release of the intron (Will & Luhrmann, 2011) (Fig. 12C).

In addition to the common splicing *in cis,* there is a process of *trans*-splicing or splicing *in trans* in which two exons that are not within the same mRNA molecule are

joined together. Whereas in mammals trans-splicing events are not frequent and are associated with cancers and translocations, a specialized form of *trans*-splicing called spliced leader (SL) *trans*-splicing is found in almost all genes in *Trypanosoma* and *Caenorhabditis* (Lasda & Blumenthal, 2011).

In SL *trans*-splicing, a short exon is donated from the 5′ end of an SL RNA and connected at or near the 5′ end of an mRNA, thus becoming the first exon of that transcript (Fig. 12E). The SL RNA molecules are snRNAs similar to U1, U2, U4, and U5 (Fig. 12D). SL *trans*-splicing uses the basic machinery of the major spliceosome and the same RNA sequence signals to indicate the splice site (Lasda & Blumenthal, 2011).

The sequences of snRNAs are highly conserved in all eukaryotes (Busch et al. 1982). However, the number and organization of snRNAs, typically found in clusters, (Card et al. 1982; Matera et al. 1990) is different among divergent species. For example, multiple copies of human U2 genes are located within 6.1 kb tandem repeats on chromosome 17 (Van Arsdell & Weiner, 1984; Lindgren et al., 1985). In contrast, a single U2 gene is present in yeasts genomes of *Saccharomyces cerevisiae* (Ares, 1986) and *Schizosaccharomyces pombe* (Brennwald et al., 1988), in which mRNA introns are rare.

❖ **Box.2 *C. elegans* snRNAs**

C. elegans has been found to have all snRNAs of the major spliceosome (Thomas et al., 1988). In general, they are spread throughout the genome. A small multigene family (6-12 members) encodes each spliceosomal snRNA.

Additionally, *C. elegans* has two distinct classes of snRNAs (SL1 and SL2) that are used in ***SL trans*-splicing** (Fig. 12E). Interestingly, the majority of *C. elegans* transcripts, ~70%, are SL trans-spliced, which means that the initial 5' UTR of the transcript is replaced by a 21-22 nt leader sequence (SL1 or SL2). SL trans-splicing uses the basic machinery of the major spliceosome involved in normal- or *cis*-splicing.

SL1 and SL2 RNAs form snRNPs that are bound by Sm proteins. They have a hypermethylated m2, 2,7GpppN trimethylguanosine (TMG) 5′ cap, a 22-nt SL exon, and a predicted three stem-loop secondary structure (Fig. 12D). In contrast to the normal spliceosomal snRNAs, SL RNAs contain a 5′ splice site that marks the borders of the SL exon and they are "consumed" during the splicing reaction (Fig. 12D).

One function for SL *trans*-splicing is the processing of polycistronic RNA transcripts such as those from operons (clusters of tandemly arranged genes transcribed from a single upstream promoter) into mature, monocistronic mRNAs. In *C. elegans* it can be distinguished the first gene of an operon (not SL or SL1 trans-spliced) versus downstream operon genes (SL2 trans-spliced). However, in *C. elegans* only ~9% of genes are found in operons. Another postulated function for SL *trans*-splicing is to enhance translational efficiency, which is triggered by the combination of the 5′-end TMG cap and the short 5′-end RNA stem-loop (Riddle et al., 1997; Lasda & Blumenthal, 2011).

➢ Current model of RNAP II-mediated snRNA transcription and 3'-end formation by the Integrator complex

The gene structure of the RNAP II-transcribed snRNAs (Sm class) is simple. They have short (~250 bp) TATA-less promoters, which contain a distal sequence element (DSE) that serves as an enhancer of transcription, and an essential proximal sequence element (PSE) located in the core promoter and common to all human snRNA genes (Fig. 13A). Transcripts of snRNA genes do not have introns, ORFs and their 3′-ends are not polyadenylated (Hernandez, 2001; Egloff et al., 2008; Jawdekar & Henry, 2008).

The RNAP II-dependent transcription cycles of snRNAs and mRNAs share several similarities including common GTFs needed to activate transcription (e.g., TBP, TFIIA, TFIIB, TFIIE, and TFIIF), the relative positioning of elements that control transcription and RNA processing. Similar to what happens with other RNAP II transcripts, capping of the 5'-end and cleavage of the 3′-end are thought to occur co-transcriptionally (Egloff et al., 2008, Jawdekar & Henry, 2008).

In addition to GTFs, the Oct-1 and Staf TFs bind to sites on the DSE. Oct-1 enhances the recruitment of $SNAP_C$ (snRNA activating protein complex), a multisubunit snRNA gene-specific TF, to the PSE sequence (Fig. 13B) (Egloff et al., 2008; Jawdekar & Henry 2008). $SNAP_C$ is also known as PTF (PSE-binding transcription factor) (Yoon et al., 1995) or PBP (PSE-binding protein) (Wanandi et al., 1993). The SNAPc is important for snRNA PIC assembly and is also a target for multiple transcription activators and repressors (Egloff et al., 2008; Jawdekar & Henry, 2008).

From 9 to 19 bp downstream of the snRNA-encoding region there is a specific sequence called the "3' box" (GTTTN0– 3AAARNNAGA), which is required for accurate snRNA 3'-end processing (Hernandez, 1985; Yuo et al., 1985; Egloff et al., 2008; Peart et al., 2013).

Since the 1980s, snRNA promoters were known to be somehow necessary for an accurate 3'-end processing of the snRNAs (Hernandez & Weiner, 1986). However, it was not until recently that the Integrator complex, which is recruited at snRNA gene promoters and is responsible for producing the cleavage of the pre-snRNAs, was

discovered.

Other recent experiments revealed that modifications of the CTD of RNAP II play a central role in Integrator complex recruitment and 3'-end processing. Ser2 phosphorylation by the CDK9 subunit of p-TEFb is necessary for recognition of the 3′-box *in vivo* (Medlin et al., 2003; Medlin et al., 2005) and Ser7 phosphorylation is required to recruit the Integrator complex to U1 and U2 snRNA genes (Egloff et al., 2007). In addition, the fact that mutation of Ser7 to an alanine residue produces a defect in transcription of snRNAs, but does not affect RNAP II recruitment, suggests that the Integrator complex recruitment to promoters is necessary for transcription initiation of snRNAs (Egloff et al., 2007).

Interestingly, a recent study demonstrated that the DSIF elongation factor, specifically its largest subunit, Spt5, and NELF, participate in the RNP II-mediated transcription of snRNAs through their association with the Integrator (Fig. 13B) (Yamamoto et al., 2014). Knockdown of Spt5 results in a substantial reduction of RNAP II at snRNAs loci suggesting that it functions early in the transcription cycle. In contrast, NELF most likely functions in snRNA 3'-end processing and transcription termination because its knockdown results in misprocessing and aberrant polyadenylation of snRNAs (Yamamoto et al., 2014). Participation of NELF in 3'-end processing is not unique for snRNAs. A previous report showed a NELF requirement in the 3'-end processing of replication-dependent histone mRNAs (Narita et al., 2007).

Although the exact molecular mechanism by which pre-snRNAs are transcribed and processed in their 3'-end remains unknown, a model to explain the function of the Integrator complex function in snRNAs RNAP II transcription and 3'-end processing has emerged.

The CTD of RNAP II is initially phosphorylated on Ser5 and Ser7 by the CDK7 subunit of TFIIH at the snRNA promoter (Akhtar et al., 2009). This Ser5 phosphorylation mark is important for capping (addition of m^7G) of the 5′-end of nascent transcripts, which occurs co-transcriptionally (Egloff & Murphy, 2008). Then, the Ser7P mark would be recognized by the RNAP II associated protein 2 (RPAP2),

homolog of the yeast phosphatase Rtr1 (Mosley et al., 2009; Egloff et al., 2007). A recent alternative for recruitment of RPAP2 suggests that the binding of a dimer of proteins (RPRD1A/B) to two Ser7P residues is responsible for recruiting RPAP2 (Ni et al., 2014). Once RPAP2 is bound to the CTD, it removes the Ser5P mark. This interaction leads, in turn, to the recruitment of at least a subset of Integrator subunits (Fig. 13B) (Egloff et al., 2012). Later, during the transcription cycle, the Ser2P by the CDK9 subunit of p-TEFb would enable the recruitment of the remaining Integrator subunits that would be loaded onto the RNAP II CTD through recognition of the Ser7P/Ser2P dyad. Although the identity of Integrator subunit(s) that recognize the phosphorylation pattern of the CTD is not known, a strong preference for the Ser7P/Ser2P dyad has been clearly established (Egloff et al., 2010).

In the end, when the terminal stem-loop and the 3' box emerge from the nascent transcript, these RNA elements are recognized (through an unknown mechanism) and the endonuclease activity of the Integrator subunit 11 cleaves the pre-snRNA (Peart et al., 2013; Baillat & Wagner, 2015). In a reporter assay, the Cdk8-CyclinC heterodimer was shown to have snRNA 3′ processing activity and it physically interacted with the Integrator complex, but if it phosphorylates the RNAP II or any Integrator subunit remains unknown (Chen et al., 2012). The process of transcription termination coupled to 3'-end processing is not clear. The Integrator complex and NELF participate, but the details of their functions are not known. Recently, it was proposed that the cap binding complex (CBC) and its associated factor, ARS2, specifically function in the recognition of 3′ signals early in transcription, perhaps mediating a link between the cap and the elongating polymerase. This complex would be able to sense the distance from the TSS, the phosphorylation pattern of the CTD and promote termination (Porrua & Libri, 2015).

The CBC is also important for exporting the 3' extended pre-snRNA generated to the cytoplasm, where the pre-snRNA is further trimmed on its 3'-end and the m^7G cap is hypermethylated by trimethylguanosine synthase-1 (TGS1) to form a 2,2,7-trimethylguanosine (TMG) cap structure to obtain the mature snRNA (Matera et al., 2007).

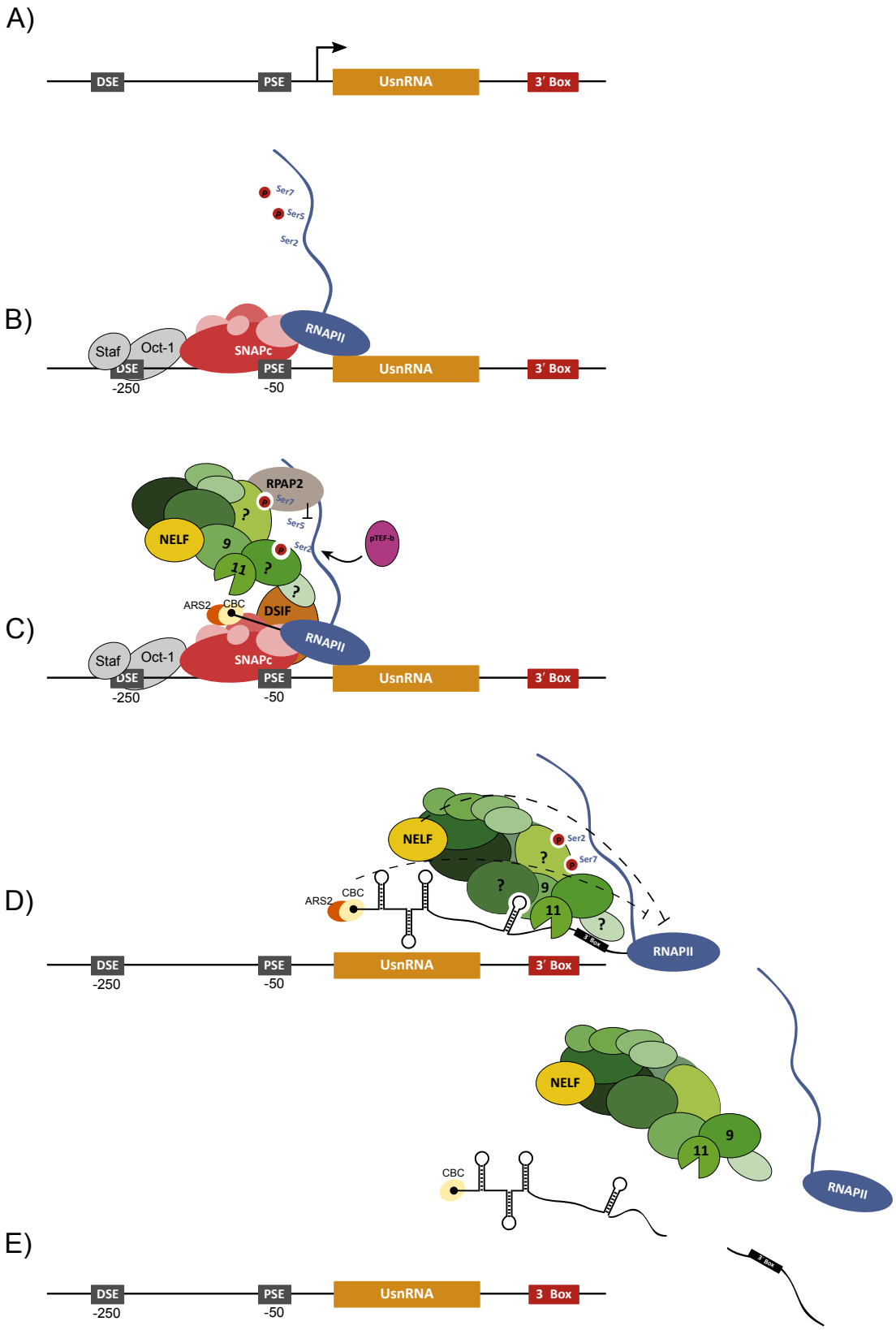


Figure 13. Current model of snRNA transcription and 3'-end formation by the Integrator complex. A) Gene structure of an snRNA transcribed by RNAP II. The DSE and PSE are located

approximately 250 and 50 bp upstream of the start TSS. The 3′ box is located 9-19 bp downstream of the mature end of the U snRNA.

B) Transcription initiation. In addition to the GTFs, snRNA transcription requires the Oct-1 and Staf TFs that bind to the DSE. Oct-1 enhances the recruitment of SNAPc that binds to the PSE sequence. The CTD of RNAP II is phosphorylated on Ser5 and Ser7. Ser5 phosphorylation is linked to capping of the 5′-end nascent transcripts.

C) Integrator complex recruitment and transcription elongation. After transcription initiation, DSIF is recruited to RNP II and p-TEFb phosphorylates the RNAP II CTD on Ser2. The Ser7 phosphorylation mark is recognized by RPAP2 that removes the Ser5P mark. The pattern of Ser7P/ Ser2P is recognized by the Integrator complex, and it is then recruited through its interactions with the phosphorylated-CTD, and possibly DSIF. NELF is also recruited, possibly as a result of interaction with the Integrator complex.

D) snRNA 3'-end processing and transcription termination. The terminal stem loop and the 3'box emerging from the nascent transcript are recognized, and the Integrator complex (catalytic subunit: Ints11) produces the cleavage. Coupled with or subsequently to the 3′-end cleavage, transcription termination occurs by an unknown mechanism that requires the Integrator complex, NELF, the cap-binding complex (CBC) and its associated factor ARS2.

E) RNAP II and the Integrator disassembly. The RNAP II and the Integrator complex can be used for subsequent rounds of snRNA transcription. The pre-snRNA generated will be exported to the cytoplasm where it will be further processed (3' trimming and TMG cap) to acquire its mature form.

4. DNA damage & the Integrator complex

Our genome is constantly challenged by different types of DNA damage such as endogenous cellular metabolites or exogenous environmental hazards. Moreover, our genome is further defied by mutagenic processes such as replication errors. Although mutations can be beneficial on an evolutionary scale, accurate repair of all types of DNA lesions is necessary to ensure genomic stability. As a result, multiple DNA repair pathways have evolved to handle this inevitable and constant threat.

Cells can detect when their DNA is damaged. Among other signals, their DNA double helix conformation is altered. Once the damage is localized, DNA repair molecules are recruited near or at the site of damage to repair the DNA. If it is not possible to repair or tolerate the alterations to the DNA, cells may undergo apoptosis, necrosis, autophagy and senescence or even cell transformation (Sperka et al., 2012; Marechal & Zou, 2013; Ta & Gioeli, 2014).

4.1. Type of lesions and their repairs

A scheme of the most relevant DNA damage repair pathways and the DNA damaging agents is shown in Figure 14. Some lesions arise during DNA synthesis when base errors are not corrected by proofreading activity. These errors in the newly synthesized DNA strand are quickly recognized by *mismatch repair systems* that remove, resynthesize and link the DNA (Larrea et al., 2010). Other types of DNA lesions, for example UV light exposure, cause DNA pyrimidine dimers and 6,4-photoproducts, which are mainly repaired by the *nucleotide excision repair pathway* (NER) where the damaged nucleotide is removed, together with a little segment upstream and downstream, and is later replaced by the sequence complementary to the undamaged DNA strand (Scharer, 2013). Non-helix distorting base lesions, such as those caused by misincorporation of Uracil into DNA, are corrected by the *base excision repair pathway* (BER) (Dengg et al., 2006), where DNA glycosylases recognize and remove specific damaged or inappropriate bases leaving empty pyrimidine or purine sites that posteriorly will be cleaved by endonucleases. The new strand is synthesized using the complementary strand as a template (Carter &

Parsons, 2016).

The most common lesions are breaks in a DNA strand or single-strand breaks (SSBs), which can be generated by direct attack of intracellular metabolites. SSBs are also repair by the BER pathway.

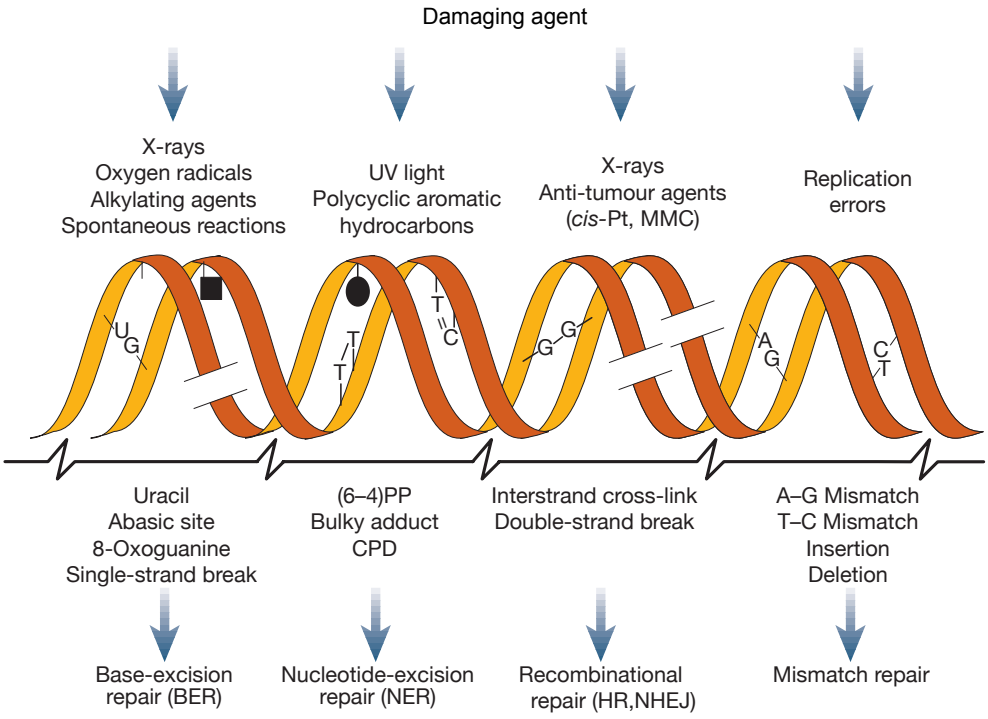


Figure 14. DNA damage and repair mechanisms. The most common DNA damaging agents are represented followed by examples of DNA lesions induced by these agents and the most relevant reparation pathways. Abbreviations: cis-Pt, cisplatin; MMC, mitomycin C; (6–4)PP, 6–4 photoproduct; CPD, cyclobutane pyrimidine dimer; HR, homologous recombination; NHEJ, non-homologous end joining. Image adapted from Hoeijmakers, 2001.

Double-strand breaks (DSBs) are the most toxic forms of DNA damage because even a single DSB has the potential to activate cell cycle arrest and can be lethal to the cell, altering its growth and metabolism (Bennett et al., 1993). Improper repair of DSBs can lead to chromosomal loss, fusion and translocation. However, some cells deliberately create DSBs to induce genetic variation, as seen in lymphocytes during V(D)J recombination or in sexually reproducing organisms during meiosis (Mladenov et al., 2016; Schatz & Swanson, 2011).

To repair DNA DSBs and safeguard genome integrity, two main repair mechanisms are used in eukaryotes: *homologous recombi*nation (HR) and *non-homologous end-joining* (NHEJ). The election of one DSB repair pathway over the other depends on

the repair context, stage of the cell cycle, and the state of the broken DNA (Li & Heyer, 2008; Jasin & Rothstein, 2013). The HR pathway (Fig. 15) is a high-fidelity repair route that uses an undamaged homologous DNA template from a sister chromatid or a homologous chromosome to provide the sequence information lost at the break site. It operates exclusively during the S and G2 phases of the cell cycle, when a homologous sister chromatid is available as a template for repair. In contrast, NHEJ (Fig.15) is an error-prone mechanism that directly joins broken DNA ends together and may result in the addition or removal of nucleotides at the repair site (Li & Heyer, 2008; Jasin & Rothstein, 2013; Schwertman et al., 2016).

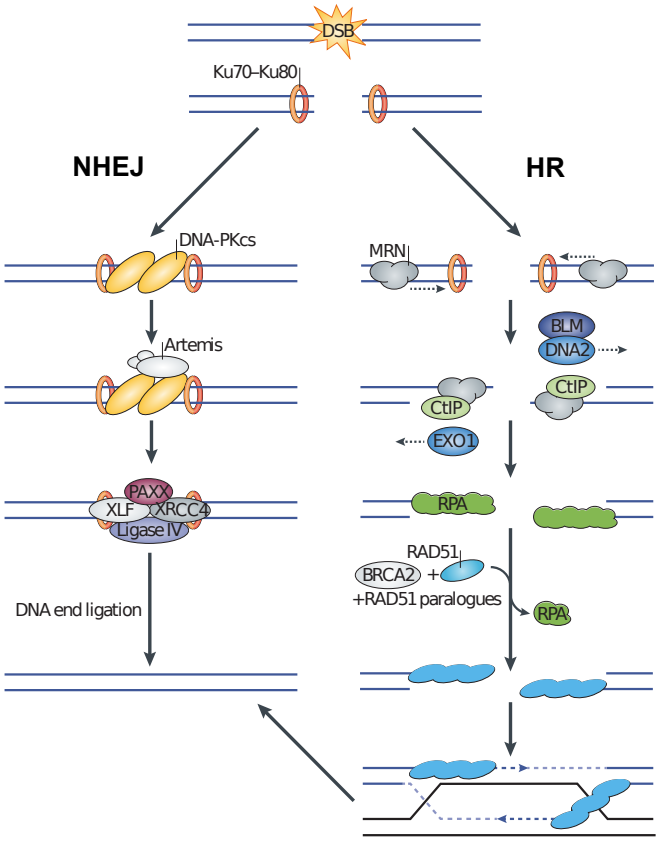


Figure 15. Illustration of the main DSB repair pathways in eukaryotic human cells: NHEJ and HR. In **NHEJ** (left) broken DNA ends are rapidly bound and protected by the Ku70-Ku80 heterodimer, providing a recruitment platform for the DNA-dependent protein kinase (DNA-PKcs). Phosphorylation by DNA-PKcs triggers recruitment of end-processing factors such as the Artemis nuclease, that trims the DNA ends, making them compatible for ligation by a ligase complex. In **HR** (right) the MRE11–RAD50–NBS1 (MRN) complex promotes end resection (via the 3′–5′ endonuclease activity of MRE11) in close proximity to the break site, paving the way for limited 5′–3′ resection mediated by the CtBP-interacting protein (CtIP). Extensive resection of the broken ends by the EXO1 and BLM (Bloom syndrome protein)–DNA2 exonucleases then produces longer stretches of single-strand DNA (ssDNA) that are rapidly coated by replication protein A (RPA). With the help of mediator proteins, including breast cancer type 2 susceptibility protein (BRCA2) and the RAD51 paralogues, RPA is subsequently exchanged with RAD51. The resulting RAD51 filament performs homology search and strand invasion, allowing DNA synthesis at the resected strand and subsequent repair. The resulting

joint molecule generated is processed by resolvases to terminate the repair process. Image taken from Schwertman et al., 2016.

4.2. Cell cycle regulation, a grand master of DNA damage control

Maintenance of genome stability acquires higher relevance during cell division because any change in proliferating cells will be propagated throughout their progeny. Cell cycle regulation is essential for normal development, cellular homeostasis and tumor suppression. Indeed, it is widely known that accumulation of genomic mutations and misregulation of cell cycle are among the main characteristics of tumor cells and a hallmark of cancer disease (Kastan & Bartek, 2004).

The mitotic cell cycle is defined as the series of events that take place in a cell leading to cell growth and division into two daughter cells. The cell cycle is tightly regulated, in eukaryotes is divided into three periods (Lodish et al., 2013):

- Interphase, where the cell grows, accumulating nutrients and duplicating its DNA. This phase is the longest (usually 90% of the total time required for the cell cycle). Interphase is subdivided in distinct phases (Fig.16):
    - G1 phase (Gap 1): the cell increases its supply of proteins, increases the number of organelles (such as mitochondria and ribosomes), and grows in size. In this phase, cells can also enter quiescence (or G0), a state of replicative dormancy.
    - S phase (Synthesis): the nuclear genome is duplicated. During this phase, DNA replication is completed as quickly as possible to minimize the exposure of base pairs to harmful external factors such as mutagens.
    - G2 phase (Gap 2): is a period of protein synthesis and rapid cell growth to prepare the cell for mitosis.
- Mitotic phase (M), where chromosomes are condensed, sorted and then equally distributed to daughter cells.
- Cytokinesis, where the physical division of the cytoplasm yields two new daughter cells.

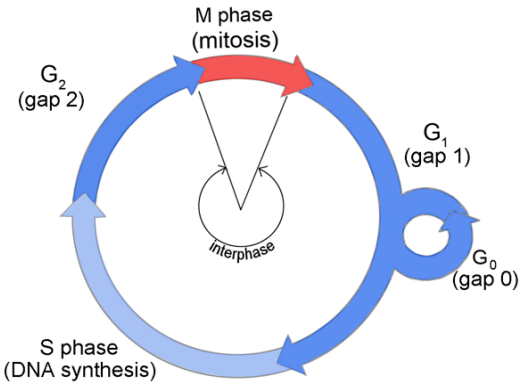


Figure 16. Mitotic cell cycle progression in higher eukaryotes. The drawing illustrates the Interphase (in blue) and the Mitotic phase (in red). Interphase is subdivided into G1 phase, the S phase and the G2 phase. In the G1 phase, the cell grows in size, increasing the number of organelles. At this point the cell can enter a state of dormancy called G0. In the S phase, the cell duplicates its DNA content. In the G2 phase, the cell synthesizes proteins and grows rapidly to prepare for the Mitotic phase where chromosomes are equally distributed to daughter cells. Image taken from http://cyberbridge.mcb.harvard.edu.

During *C. elegans* development, variations of this typical somatic cell division cycle are used to fulfill specific requirements. These include rapid embryonic cell cycles that lack the G1 and G2 phases, meiotic cell cycles that allow formation of haploid gametes, and endoreduplication cycles in the intestine and hypodermis during development where S phases are not followed by mitosis, thus doubling the DNA ploidy with each additional cycle (van den Heuvel, 2005).

Cell cycle progression through the diverse phases is controlled by the activation and inactivation of cyclin-dependent kinases (CDKs). CDKs are small serine/threonine protein kinases that require association with a cyclin subunit for their activation. Cyclins are short half-life proteins expressed in specific phases of the cell cycle (Morgan, 1995). Once cyclin-CDKs holoenzymes are activated, they phosphorylate effector proteins allowing cell cycle progression. Their activity is generally low in G1 and rises progressively until it reaches a maximum upon mitotic entry. Later, the destruction of specific cyclins (by the Anaphase Promoting Complex/Cyclosome, an E3 ubiquitin ligase) triggers mitotic exit, thereby re-establishing the G1 state (Malumbres & Barbacid, 2009).

CDKs are very well conserved from yeast to mammals, however the number of kinases is different among organisms. In yeast, a single CDK (known as Cdc28 in

Saccharomyces cerevisiae and Cdc2 in *Schizosaccharomyces pombe)* acts with different cyclins to promote cell cycle progression (Malumbres & Barbacid, 2009). In *C. elegans*, studies have revealed that at least two CDKs, CDK-1 and CDK-4, are essential for cell-cycle progression (Boxem et al., 1999; Boxem & van den Heuvel, 2001; Park & Krause, 1999; van den Heuvel, 2005).

To ensure tight control over cell-cycle progression, many levels of regulation act on the CDKs. Such regulation involves the availability of cyclins (which rely on their controlled synthesis and ubiquitin-dependent proteolysis), the controlled expression and destruction of inhibitory proteins that associate with CDKs, or CDK/cyclin complexes, and post-translational modifications that either activate or inactivate CDKs (Malumbres & Barbacid, 2009). For example, CDKs activation requires phosphorylation by CAK (CDK-activating kinase) (Kaldis, 1999), whereas inactivation of CDK1 and CDK2 is caused by phosphorylation of inhibitory residues (Thr14 and Tyr15) (Parker et al., 1992; Parker & Piwnica-Worms, 1992; Mueller et al., 1995).

Additionally, genomic integrity and elements required for cell division are verified at different points during the cell cycle to ensure correct cell division. These control mechanisms are called checkpoints.

The progress of a cell through the mitotic cycle is monitored at different moments by cell cycle checkpoints. Checkpoints guarantee that each stage of the cell cycle is initiated once the previous one is completed, as well as monitor the integrity of chromosomes and their accurate segregation during mitosis (Lodish et al., 2013). Cells usually cells by-pass these surveillance mechanisms. However, under some circumstances such as when there is DNA damage, the checkpoints are activated and the cell cycle is arrested to repair any defect, thus preventing its transmission to cellular progeny. If cells are not able to repair the damage (e.g., genetic defects in either the checkpoint or the DNA repair machinery), they may enter senescence or undergo apoptosis. Alternatively, accumulation of DNA alterations may result in premature aging or even cell transformation and cancer (Sperka et al., 2012).

Checkpoints were originally defined as surveillance mechanisms within the cell cycle. However, it is more accurate to define them as signaling pathways because they also have a key role on the coordination of various cellular processes such as the activation of DNA repair pathways, activation of transcriptional programs and apoptosis upon DNA damage (Branzei & Foiani, 2008; Bertoli et al., 2013).

There are four main checkpoints (Lodish et al., 2013) (Fig.17):

- **G1 checkpoint**: restriction point or start point: at the end of the G1 phase, the cell checks the quality of the DNA, factors required for DNA replication and cell size.
- **Intra S-phase checkpoint:** where continual control of DNA synthesis and fidelity allows the repair of possible errors that occur during replication.
- **G2/M checkpoint:** at the end of the G2 phase, the cell checks DNA quality, correct completion of DNA replication and if there are conditions for mitosis (cell size, DNA integrity and nutrients).
- **Spindle assembly checkpoint (SAC):** the cell checks proper chromosome segregation.

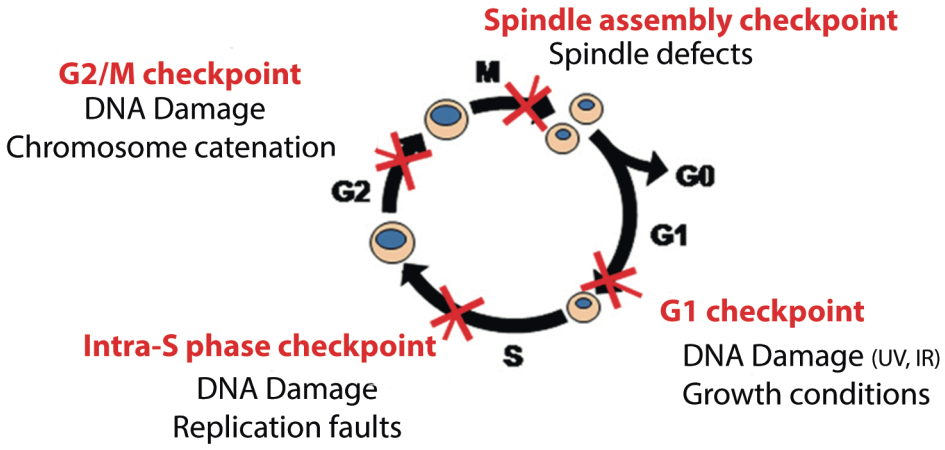


Figure 17. Cell cycle checkpoints. Represented in red are the four main checkpoints (G1 checkpoint, intra-S phase checkpoint, G2/M checkpoint and spindle assembly checkpoint). Below are the factors that activate them. DNA damage activates the checkpoints during interphase at G1, S and G2. However, during mitosis the SAC is not activated by DNA damage unless it affects the centromere region. Image adapted from Gabrielli et al., 2012.

Following DNA damage, cell cycle checkpoints are activated (Fig.17). The first evidences that cell-cycle transitions were under genetic control in response to DNA damage were observed in the SOS DNA damage response pathway in *E. coli*

(George et al., 1975). In mammals, cells from patients with ataxia telangiectasia (an inherited syndrome associated with elevated cancer incidence), which were defective for the ataxia telangiectasia mutated (ATM) gene, failed to arrest cell-cycle after X-irradiation, establishing a connection between cell cycle progression and cancer (Painter et al., 1980). These control mechanisms for delaying the cell cycle where initially characterised in yeast. The first checkpoint genes were identified while studying fission yeast mutants isolated in screens for UV-sensitivity (Hannan & Nasim, 1976) and the term checkpoint was coined later by Weinert and Hartwell in 1988 while studying a rad9 mutant in budding yeast with a similar defect (Weinert et al., 1988).

The DNA damage checkpoint can be defined as a network of interacting pathways operating in concert to recognize alterations in the DNA and eliciting a response that ultimately leads to CDK inhibition and cell cycle arrest, allowing the cell time to repair the damage before mitosis (Elledge, 1996; Iliakis et al., 2003). DNA damage checkpoints are activated during interphase at G1, S and G2 (Beishline & Azizkhan-Clifford, 2014). However, during mitosis the SAC is not activated by DNA damage unless it affects the centromere region (Rieder et al., 1994; Rieder et al., 1995; Beishline & Azizkhan-Clifford, 2014).

Once the repair is completed, the DNA damage checkpoint response is inactivated and the cells re-enters the cell cycle in a process known as recovery (Clémenson & Marsolier-Kergoat, 2009).

The DNA damage checkpoint pathway is highly conserved between lower eukaryotes and higher multicellular eukaryotes.

The ATM (ataxia-telangiectasia mutated) and ATR (ataxia-telangiectasia and Rad3-related) kinases are members of the phosphatidyl-inositol-3-OH kinase-like kinases (PIKK), are the two main central regulators that function in the DNA damage response (Marechal & Zou, 2013; Awasthi et al., 2015). Their activation leads to cell cycle arrest while the cell activates repair pathways and facilitates an open chromatin structure needed for repair. If the damage is excessive or sustained, the kinases may also promote apoptosis. Which kinase functions during damage recognition depends on the type of lesion and the stage of the cell cycle (Fig. 18)

(Marechal & Zou, 2013; Awasthi et al., 2015).

- Damage-Induced G1/S Checkpoint

The G1/S checkpoint prevents the replication of damaged DNA by blocking entry into the S-phase. Basically, the cell has two mechanisms, transcriptional and non-transcriptional, to halt the cell cycle in the G1 phase when there has been DNA damage.

The non-transcriptional control mechanism is fast and reversible. It relies on the fact that cells cannot progress through the G1 phase into the S phase in the absence of proper Cdk activity. Following DNA damage, the ATM kinase phosphorylates and activates the cell cycle checkpoint protein Chk2 (Fig. 18). The Chk2 kinase phosphorylates Cdc25A, inducing its degradation. Ccd25A is a phosphatase that keeps the Cdk2–Cyclin E complex in its active form by dephosphorylation, which causes progression of the cell cycle from G1 to S. Due to the degradation of the Cdc25A phosphatase, the Cdk2–Cyclin E complex remains in its hyperphosphorylated inactive form, culminating in G1/S arrest (Fig.18) (Beishline & Azizkhan-Clifford, 2014).

In addition, when there is DNA damage, ATM has been shown to target Cyclin D to modulate its ubiquitin-mediated degradation (Hitomi et al., 2008).

The transcriptional control mechanism is slower but essential for complete inhibition of entry into the S-phase. The most relevant mechanism is the p53-dependent pathway. ATM phosphorylates the p53 transcription factor and its negative regulator Mdm2 (Lakin & Jackson, 1999; Zhang & Xiong, 2001; Maya et al., 2001). In addition, once ATM phosphorylates and activates Chk2, this kinase can further phosphorylates p53 (Fig. 18) (Hirao et al. 2000). As a result of these post-translational modifications, p53 is active and stable, which results in the upregulation of factors such as p21, which inhibits the activity of CDKs by binding to cyclins (Fig. 18) (Harper et al., 1995).

- Damaged-induced Intra-S phase checkpoint

Activation of the intra S phase checkpoint is triggered by replication stress generated from DNA lesions (DNA adducts, breaks or certain DNA modifications produced by alkylating agents) or when there is a deficiency in substrates needed for DNA replication, such as dNTPs (Grallert & Boye, 2008; Labib & Piccoli, 2011).

DNA replication is initiated at the replication origins, which are specific sites defined by a number of proteins that ensure that the origins fire (start replicating) once per cell cycle. When the polymerase and its associated proteins encounter a blockade to progression, they remain stably associated with the replicating chromatid so that replication can resume once the blockade is removed (Barnum & O'Connell, 2014).

The intra S phase checkpoint acts in three ways: first, decreasing the firing of late origins; second, slowing the replication fork progression; and third, stabilizing the replication fork machinery (Beishline & Azizkhan-Clifford, 2014).

The prevalent pathway functioning during the intra-S phase checkpoint is the ATR pathway. When DNA replication is blocked, binding of RPA to the increased ssDNA recruits ATR via its accessory factor, ATRIP (Zou & Elledge, 2003). Subsequently, the 9-1-1 complex (composed of Rad9-Hus1-Rad1) is recruited to the ssDNA-RPA sites around the replication fork and supports ATR activation (Fig. 18) (Yan & Michael, 2009). Once ATR is activated, this kinase transduces a signaling cascade activating other substrates. Among them Chk1 activation plays the dominant role (Zhao et al., 2002) (Fig. 18). Chk1 suppresses Cyclin–Cdk-dependent activities through inhibition of Cdc25A phosphatase (which normally activates CDK2 to allow DNA replication progression) and other replication promoting functions such as inhibition of Cdc45 association with chromatin, which is required for origin firing (Liu et al., 2006; Liu et al., 2010; Beishline & Azizkhan-Clifford, 2014).

Sometimes, collapse of the stacked replication forks can generate DSBs. In this case, the ATM kinase acts in parallel to ATR but Chk1 is also required for amplification of the initial signaling carried by the ATM-Chk2-Cdc25A pathway. Indeed, ATM targets involved in HR repair such as Mre11, Brca2, Rad51 and Nbs-1 have been shown to be important for proper S phase checkpoint functions (Zhao et al., 2000; Falck et al., 2002; Hashimoto et al., 2010; Schlacher et al., 2011).

- Damaged-induced G2/M phase checkpoint

The G2/M checkpoint is essential for preventing cells from going into mitosis with unresolved DNA lesions, thus ensuring transmission of DNA integrity to cellular progeny. Activation and maintenance of the G2 cell cycle arrest includes a variety of actions: targeting phosphatases which promote mitosis, kinases that block Cdk1 function and regulating of cyclin B levels and their subcellular localization, as well as other regulatory processes involved in the normal progression of the cell from G2 into M phase (Beishline & Azizkhan-Clifford, 2014).

In response to DSBs in G2, ATM is activated and recruited to DSBs by the Mre11–Rad50–Nbs1 (MRN) complex. Later, DSBs resection generates the formation of RPA coated ssDNA that triggers ATR recruitment. This recruitment of ATM and ATR to the damage sites provides downstream activation of the Chk1 and Chk2 kinases, respectively (Fig. 18) (Jazayeri et al., 2006; Shiotani & Zou, 2009). Although the ATM-Chk2 pathway is activated immediately, it seems that downstream activation of the ATR–Chk1 pathway is necessary to sustain cell cycle arrest during G2 to allow the cell time for the slower DNA repair by HR, therefore ensuring DNA repair fidelity (Beishline & Azizkhan-Clifford, 2014).

As part of the DDR, the Chk2 and Chk1 kinases are activated at the G2/M checkpoint by the same factors as during the G1/S and intra-S phase checkpoints, although their contribution could be different depending on the damage that triggers the response (e.g., DSBs or ssDNA accumulation) (Lobrich & Jeggo, 2007).

Maintenance of the Cdk1–Cyclin B1 complex in its inactive state blocks entry into mitosis. Therefore, Cdk1 is tightly controlled at the G2/M phase. Cdk1 activation involves removal of inhibitory phosphorylations on Tyr15 and Thr14, added earlier in the cycle by Wee1 and Myt1, respectively (Parker & Piwnica-Worms, 1992; Booher et al., 1997; Liu et al., 1997). Because these residues are within the ATP-binding domain, they inactivate Cdk1 kinase activity if phosphorylated. Counteracting these inhibitory phosphorylations, Cdk1 is dephosphorylated by Cdc25c phosphatase (Draetta & Eckstein, 1997) as well as Cdc25A phosphatase during the G2/M checkpoint (Zhao et al., 2002).

Therefore, Chk1 and Chk2 converge in signaling cascades that increase the

phosphorylation of Cdk1 by Wee1/Myt1 or prevent of its dephosphorylation by Cdc25 phosphatases inducing the cell cycle arrest.

In addition, to obtain a sustained Cdk1/cyclin B inhibition, the transcriptional induction of Cdk1 inhibitors is required via p53-dependent (e.g., p21, Gadd45, and 14-3-3σ) or -independent (e.g., via BRAC1) mechanisms. The p53 protein also supports G2 delay by repressing Cdk1 and Cyclin B transcription (Powell et al., 1995; Hermeking et al., 1997; Chan et al., 1999; DeSimone et al., 2003; Iliakis et al., 2003; Dai & Grant, 2010).

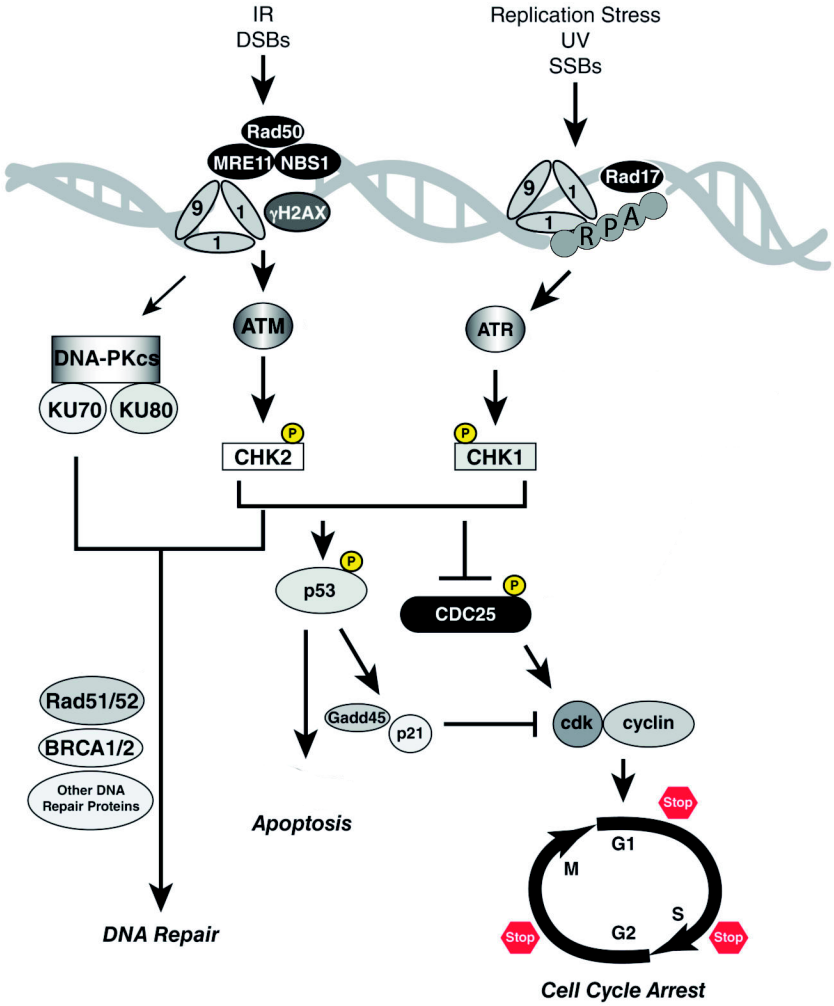


Figure 18. Simplified scheme of DNA damage checkpoint regulatory pathways in mammals. Exogenous or endogenous cellular insults activate the DNA damage response pathways. The presence of DSBs, which can be caused by IR, activate the ATM/CHK2 pathway. However, SSBs, which can be caused by UV light or replication stress, activate the ATR/CHK1 pathway. Chk1 and Chk2 converge in signaling pathways that inhibit CDC25 phosphatases to block the dephosphorylation of Cdk-cyclin complexes and also activate p53, which promotes the transcriptional induction Cdk inhibitors (Gadd45, p21). Therefore, these signaling pathways induce cell cycle arrest at the G1, S and G2/M to provide the cell time for DNA repair. DSBs activate DNA-PKcs to repair the

damage by the NHEJ pathway. Otherwise, proteins such as RAD51 and BRAC1 are rapidly recruited to DSB sites where they promote strand invasion and mediate the repair by HR. However, if the repair is unsuccessful, the cell will probably undergo apoptosis (via p53-dependent or independent induction). Image adapted from Ta & Huy, 2014.

4.3. Integrator complex in DNA repair

One interesting aspect of the Integrator Complex is the recently described involvement of at least, two of its subunits, Ints3 and Ints6, in DNA damage response (Skaar et al., 2009; Zhang et al., 2009; Zhang et al., 2013). Indeed, the involvement of members of the transcriptional machinery in DNA repair is not a striking issue and has a long history (Lainé & Egly, 2006; Derheimer et al., 2007).

Initially, distinct studies showed that Ints3 interacts with hSSB1 (human Single Stranded Binding protein 1) and the uncharacterized protein C9orf80, forming a stable complex that was called the *INTS3–MISE–hSSB1 complex* or *SOSS1 complex* (Sensor of single-stranded DNA complex 1), depending on the author. Both studies defined a core complex composed of hSSB1, Ints3 and C9orf80 apart from other Integrator subunits binding in a substoichiometric manner (e.g. Ints6, -5, -8, and -1) (Huang et al., 2009; Skaar et al., 2009).

Later, Zhang and colleagues revealed that Ints6 was indeed a major subunit of this core hSSB1 complex and therefore it was renamed the *hSSB1/INTS complex* (Zhang et al., 2013). Recently, Skaar and colleagues postulated that the hSSB1/INTS complex is part of the Integrator complex (Skaar et al., 2015). It is still a matter of debate whether it is an independent “subcomplex” or if this “subcomplex” is a part of the Integrator complex (Skaar et al., 2015).

Previously, studies on hSSB1 (a 211-amino acid polypeptide with an N-terminal oligosaccharide/oligonucleotide-binding (OB) domain) demonstrated that after DNA damage, hSSB1 is phosphorylated by ATM kinase and localizes to DNA DSBs. hSSB1 was observed to play an essential role for efficient repair of DSBs via the HR pathway. Although the exact mechanism remains unclear, hSSB1 is involved in recruitment of DNA repair proteins such as Rad51 and BRCA1 at DSBs. Moreover, distinct experiments revealed that hSSB1 activation and stabilization generates a positive feedback loop to amplify the ATM-dependent signaling cascade. Therefore, cells deficient in hSSB1 display defective checkpoint activation, diminished capacity

for DNA repair and enhanced genomic instability (Richard et al., 2008; Richard et al., 2011; Yang et al., 2013).

Interestingly, it was observed that Ints3 regulates the normal transcription of hSSB1 and also helps to stabilize its protein levels. Thus, Ints3-depleted cells showed a similar phenotype to hSSB1-depleted cells exhibiting increased radiosensitivity, chromosomal instability and reduced activation of the ATM signaling pathway (Skaar et al., 2009; Huang, 2009).

Whether there is damage or not, the hSSB1/INTS complex is stable. The N-terminal (Nt) region of Ints3 interacts with the OB fold domain of hSSB1 (which is essential for directing hSSB1 to damaged DNA sites) and also with C9orf80, whereas Ints3 C-terminal (Ct) region binds to the Ct region of Ints6 (Fig. 19). The hSSB1/INTS complex is localized in the nucleoplasm of cells but following ionizing radiation (IR), all its members (hSSB1, Ints3, Ints6 and C9orf80) are re-localized to DSB sites forming IR-induced foci (IRIF). Generally, the proteins that form IRIF are physically recruited to the damaged DNA sites and become chromatin bound.

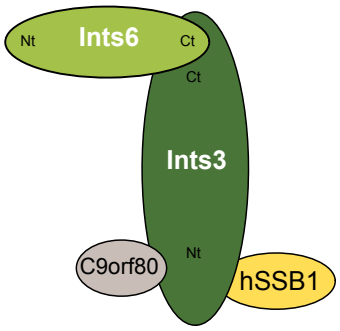


Figure 19. Scheme of the hSSB1/INTS complex. The hSSB1-INTS complex is localized in the nucleoplasm of cells, but following IR it re-localizes to DSB sites. The Ct region of Ints6 is bound to the Ct region of Ints3. The Nt region of Ints3 is bound to the OB fold domain of hSSB1 and C9orf80. Image adapted from Zhang et al., 2013.

Interestingly, hSSB1 and Ints6 genes have paralogs in the human genome, hSSB2 and DDX26b, respectively. Although the transcription of these paralogs is much lower than that of hSSB1 and Ints6, they form analogous complexes with Ints3 and C9orf80, and are implicated in DNA repair by HR. The protein levels of hSSB1 and hSSB2 seem to have an inverse relationship indicating that they might play

complementary roles (Li et al., 2009; Zhang et. al, 2013).

The interaction between hSSB1 and Ints3 is important to stabilize each other at DNA damage sites. It was observed that Ints3-depleted cells reduced hSSB1 and hSSB2 IRIF. Meanwhile, the combination of hSSB1 and hSSB2 knockdowns also impaired the Ints3 IRIF. However, Ints6 and DDX26b depletion did not affect the Ints3 or hSSB1 IRIF, but the Ints6 IRIF were abrogated in Ints3- or hSSB1-depleted cells. Therefore, the ability of hSSB1/INTS complex members to form IRIF is influenced by their interactions with each other (Skaar et al., 2009; Zhang et al., 2013).

Importantly, Ints3 was reported to interact with Nbs1, a member of the MRN complex. The MRN complex, comprised of Rad50, Mre11 and Nbs1, is recruited to DSBs where promotes DNA end resection and the generation of ssDNA, which is critical for the recruitment of RPA and HR repair. Depletion of the MRN complex resulted in significantly reduced IRIF of Ints3 and hSSB1/2, suggesting that the hSSB1/INTS complex acts downstream of the MRN complex in the HR pathway (Huang et al., 2009).

In addition, the hSSB1/INTS complex might also be involved in the ATR signaling pathway, which responds to ssDNA lesions. Initially, it was observed that either hSSB1 or Ints3 knockdown cells showed defective phosphorylation of Chk1 after DNA damage (Richard et al., 2008; Zhang et al., 2013). Moreover, the capacity of TopBP1 (a member of ATR pathway), to form IRIF was significantly reduced in Ints3-depleted cells (Zhang et al., 2013). Later, in the absence of RPA, hSSB1 and Ints3 were observed interacting with the ATR-ATRIP complex and recruiting it to DNA Damage the sites (Kar et al., 2015).

Discovery of the hSSB1/INTS complex has led to the establishment of new network in DNA damage response pathways. It is probable that pathogenic mutations in hSSB1/INTS complex members will be found given the high frequency of DNA repair proteins affected in cancer and radiosensitivity syndromes.

4.4. *C. elegans* as a model system for studying the DNA damage response

The *Caenorhabditis elegans* germline (where the only proliferative cells of the adult worm are located) is well established as a model system for studying DNA damage response processes in a multicellular organism. The hermaphrodite gonad is a syncytium where germ cells are organized into a spatial/temporal gradient along the distal-proximal axis. Mitotic cells are located at the distal end of the gonad and they progress through meiosis as they are moved along the germline (Fig. 20) (Craig et al., 2012).

Interestingly, competence to perform cell cycle arrest and apoptosis induction is spatially separated in the worm germline. Cell cycle arrest only occurs in proliferating mitotic germ cells at the distal end of the gonad, while the ability to undergo apoptosis is restricted to late-stage meiotic pachytene cells near the gonad bend (Fig. 20) (Craig et al., 2012).

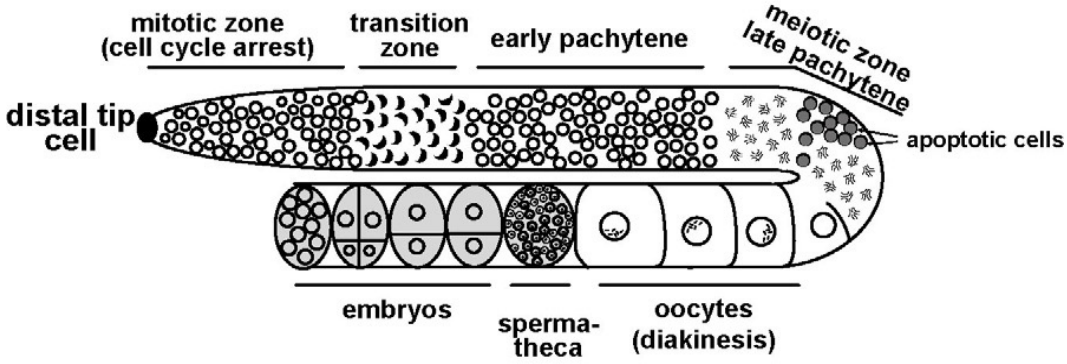


Figure 20. Illustration *C. elegans* hermaphrodite gonad. Germ cell proliferation occurs in the mitotic zone and cells progress through meiosis as they are moved along the germline. The distal tip cell is a single somatic cell located at the tip of the gonad where it regulates germline stem cell proliferation and entry into meiosis. In the transition zone (nuclei in leptotene/zygotene) the homologous chromosome pairing results in the polarized redistribution of chromosomes that gives rise to the characteristically crescent-shaped DNA. The nuclei move from the transition zone into the pachytene region where paired and aligned homologous chromosomes are distributed around the periphery of each nucleus. In the oocytes, chromosome condensation continues during diakinesis and the 6 bivalents can be easily visualized within the nucleus. The oocytes are fecunded while they pass through the spermatheca and early stage embryos are formed within the gonad.
In response to DNA damage cell cycle arrest occurs only in the mitotic zone, while only late pachytene cells can to undergo apoptosis. Image taken from Craig et al., 2012.

Both cell cycle arrest and apoptosis responses depend on conserved DNA damage checkpoint proteins. Apoptotic cell corpses induced by DNA damage are morphologically indistinguishable from all other forms of apoptosis induced by other

stimuli such as environmental stress or pathogens (Aballay & Ausubel, 2001; Salinas et al., 2006). Similarly, apoptosis occurs only in pachytene-stage germ cells, and requires the core apoptotic genes *ced-9*, *ced-4*, and *ced-3*. However, unlike physiological apoptosis, DNA damage-induced germ cell death is largely dependent on *cep-1(C. elegans p53-like)* and its transcriptional targets, *egl-1* and *ced-13* (Gartner et al., 2000; Schumacher et al., 2001; Craig et al., 2012).

The worm homologs of the mammalian ATM and ATR kinases (ATM-1 and ATL-1) have also been implicated in DNA damage checkpoint signaling (Garcia-Muse & Boulton, 2005; Stergiou et al., 2007). However, CHK-1 and CHK-2, downstream checkpoint kinases, have distinct functions in the *C. elegans* germ line. CHK-1 has been involved in cell cycle arrest and apoptosis induced by IR (Kalogeropoulos et al., 2004) while CHK-2 is needed for chromosome pairing during early meiosis and for apoptosis induced by UV (MacQueen & Villeneuve, 2001).

Following DNA damage, HR is the main form of DNA DSB repair in the worm germline (presumably to ensure the correct transmission of genomic information from one generation to the next) and in proliferating somatic cells (Clejan et al., 2006). In contrast, NHEJ is the major repair pathway used in postmitotic somatic cells (Clejan et al., 2006). NHEJ is based on DNA end protection: the Ku70/Ku80 heterodimer stabilizes the dsDNA ends and prepares the DSB for direct ligation by DNA ligase IV. In contrast, HR is based on DNA 5'-3' end resection. The *C. elegans* MRN complex (comprised of MRE-11 RAD-50 and COM-1), along with other nucleases, degrades the dsDNA to generate 3' ssDNA tails, which are coated by RPA-1 and then replaced by RAD-51 recombinase, forming a nucleoprotein filament that invades an intact donor DNA. A D-loop structure is formed and new DNA is synthesized using the intact donor strand as a template. BRC-2, the *C. elegans* ortholog of the human BRCA2, is crucial for proper recruitment of RAD-51, nucleoprotein filament stability and stimulation of RAD-51-mediated D-loop formation (Fig. 21) (Lemmens & Tijsterman, 2011).

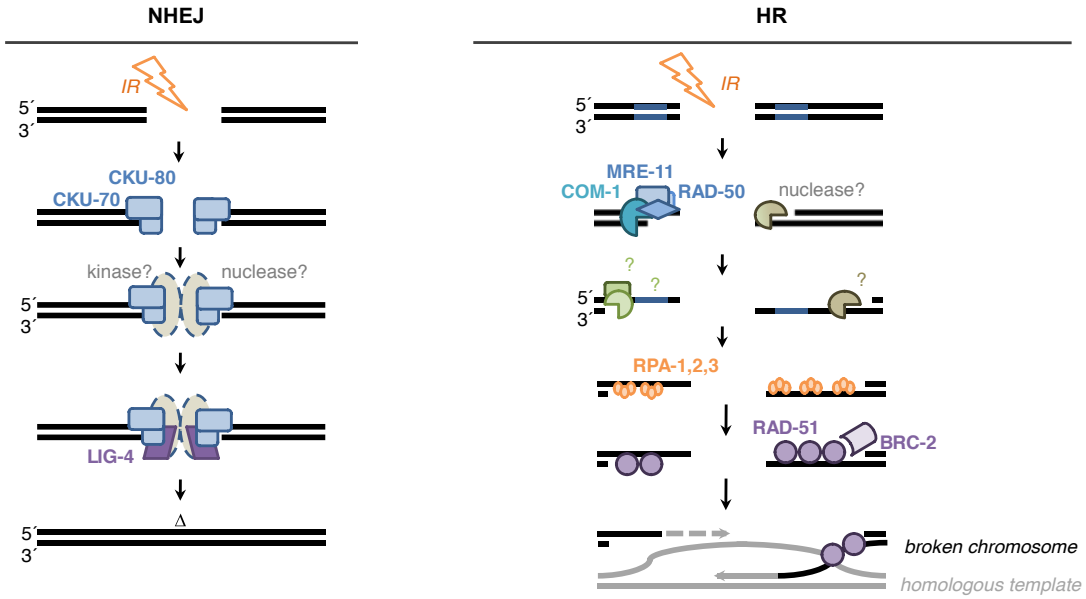


Figure 21. Simplified scheme of the main DSB repair pathways in *C. elegans* after IR: NHEJ and HR. HR is the main form of DNA DSB repair in the worm germline whereas NHEJ is the major pathway used for repair in post-mitotic somatic cells. In NHEJ the Ku70/Ku80 heterodimer stabilizes the dsDNA ends that are ligated by DNA ligase IV. In HR the MRN complex and other nucleases degrade the dsDNA to generate 3' ssDNA tails, which are coated by RPA-1 and then replaced by the recombinase RAD-51 forming a nucleoprotein filament that invades an intact donor DNA. A D-loop structure is formed and new DNA is synthesized using the intact donor strand as a template. Image adapted from Lemmens & Tijsterman, 2011.

In worms, defects in DSBs repair often result in developmental abnormalities, altered chromosome morphology, and/or DNA damage sensitivity, which are phenotypes that can be easily detected. Furthermore, elevated chromosomal instability in the germline often manifests an increased X chromosome non-disjunction, which in *C. elegans* results in high male progeny (XO) or him phenotype (high-incidence-of-males) (Kelly et al., 2000; Tang et al., 2010; Lemmens & Tijsterman, 2011).

Objectives

1. Identification of the *C. elegans* Integrator complex.

2. Analysis of the transcriptional effect downstream of the snRNA loci upon knockdown of any subunit of the *C. elegans* Integrator complex.

3. Analysis of the involvement of the Integrator subunit 6 in DNA damage caused by γ-radiation.

4. Description of the transcriptional effect downstream of the snRNA loci in response to γ-radiation.

Results

1. Screening of mutants with embryonic lethality

The *t1903* mutant was discovered in a screening performed to find embryonically lethal mutants with a maternal effect. Therefore, mutants had defects in pathways involved in the process of embryogenesis. This screening was conducted in the laboratory of Dr. Ralf Schnabel (Institute of Genetics, TU Braunschweig, Germany), and consisted of mutating the following *C. elegans* strain with ethyl methanesulfonate (EMS):

him-9 (e1487) II; unc-24 (e138) IV; dpy-11 (e224) V/ nT1 (m345) (let) (IV;V)

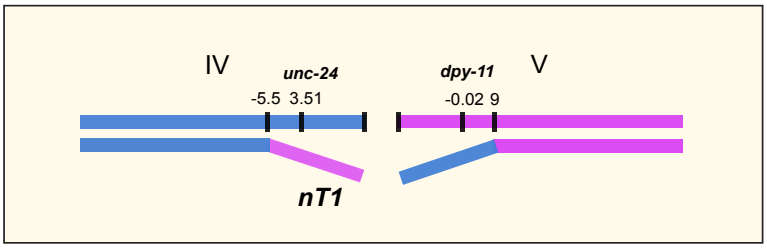


Figure 22. Strain used in the screening to find and select mutants with embryonic lethality. This strain was treated with the EMS mutagen. The strain had two genetic markers in the balanced region by the *nT1* balancer, one in chromosome IV and the other in chromosome V: *unc-24 (e-138)* and *dpy-11 (e224)* respectively.

EMS is the most commonly used mutagen for *C. elegans* studies. EMS usually induces point mutations, changing one nucleotide for another (GC → AT transitions are favored) (Kohalmi et al., 1993). These changes can generate different genetic alterations, for example: the change of one amino acid for another can result in modification of a protein structure, or generate a nonsense codon or a premature stop codon that will interrupt mRNA translation.

After EMS mutagenesis, a screening of the mutants generated was performed in order to select maternally rescued lethal mutants. The strain used in the screening had the *nT1 (m345) (let)* balancer in which one region from chromosome IV was swapped with another from chromosome V (Fig. 22).

Genetic balancers are genetic constructs or chromosomal rearrangements that allow lethal mutations to be stably maintained as heterozygotes. There are different types of genetic balancers: translocations, duplications,

transgenes, inversions etc. Balanced strains are commonly used in *C. elegans* to allow a mutated lethal strain in homozygosity be viable and maintained in heterozygosity because the wild-type copy of the gene is provided by the balancer. In *C. elegans*, phenotypes associated with lethal mutations range from zygotic or larval lethality to adult sterility and maternal effect lethality, and can include conditional effects such as temperature sensitivity. Thus, lethal mutations constitute a rich source of information about basic biological processes in this nematode.

The screening was done in a balanced strain to select both conditional and non-conditional mutants affecting the balanced region. These mutants were embryonically lethal in homozygosity but viable in heterozygosity.

In addition, the strain chosen for screening had two genetic markers: *unc-24 (e138)* on chromosome IV and *dpy-11 (e224)* on chromosome V (Fig. 22). These genetic markers produce a characteristic phenotype that can be easily visualized under a dissecting microscope. In the case of the genetic marker *unc-24 (e138),* worm movements are uncoordinated while the genetic marker *dyp-11 (e224)* produces animals thicker and shorter than the wild type. Genetic markers are used to easily identify the maternally rescued homozygotic worms. Currently, these genetic markers have been replaced by the use of balancers labeled with the GFP.

Mutants with high embryonic lethality in homozygosis were selected. The development of each mutant selected was observed under a 4D-microscope to characterize its phenotype. To verify if mutants had different alleles of the same gene, genetic complementation experiments among all the mutants were performed. In that way, the likelihood that mutants with a similar phenotype, and which were not able to complement among themselves, had a mutation in the same gene was increased.

One of the selected mutants was *t1903* (strain GE3632). This mutant was thermosensitive (ts): the embryos died at 25ºC, showing abnormalities in cell fate specification.

2. Mapping and cloning of the gene affected in the *t1903* mutant

Mapping and cloning of the gene affected in the *t1903* mutant was done in the laboratory of Dr. Ralf Schnabel (Institute of Genetics, TU Braunschweig, Germany).

2.1. Complementation with deletions

The *t1903* mutation was linked to the genetic marker *unc-24(e138),* present in the balanced region of chromosome IV. To clone the gene containing the *t1903* mutation, the first step was to perform complementation tests on the mutant with some deletions (also called “deficiencies”, Df) located on the balanced region of chromosome IV (Fig. 23).

The following deletions were used:

- nDf41 (IV: 1.68 to 3.42)
- stDf7 (IV: 2.42 to 3.52)
- eDf19 (IV: 3.65 to 4.57)
- sDf2 (IV: 4.6 to 6.24)
- sDf21 (IV: 5.7 to 10.05)

All these deletions could to complement the *t1903* mutant and rescue its embryonic lethality, providing the heterozygous *t1903*/Df worms with the wild-type copy of the gene, except in the case of the eDf19 deletion. Therefore, this deletion, which is in the CB3824 strain, had to include the gene affected in the *t1903* mutant because it was not able to rescue the embryonic lethality of this mutant. In conclusion, the *t1903* mutation had to be in the genetic interval IV: (3.65, 4.57) of chromosome IV. This genetic interval is of approximately 2900 kb.

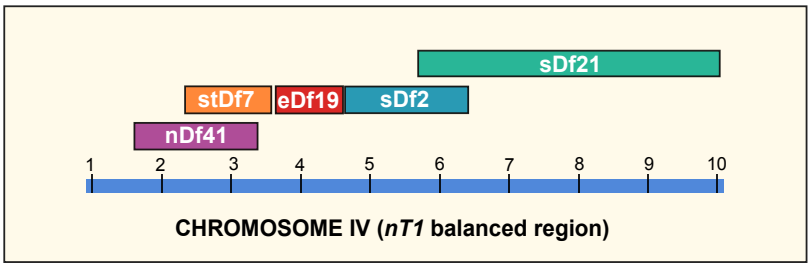


Figure 23. Deletions used mapped to the balanced region of *nT1*. The only deletion that was not able to rescue the embryonic lethality of the *t1903* thermosensitive mutant was eDf19. Thus, the mutated gene had to belong to this genetic region.

2.2. snip-SNP mapping of the *t1903* mutation

The *C. elegans* wild type strain commonly used in the laboratory is N2. It was isolated in Bristol and differs in certain polymorphisms that affect single nucleotides (single nucleotide polymorphisms, SNPs) of a closely related strain, the Hawaiian strain, CB4856 (isolated in Hawaii). Some of these SNPs affect the target of a restriction enzyme so that digestion with the restriction enzyme in the same genetic region will result in different digestion products according to the strain from which it comes. Polymorphisms that affect a restriction site are called snip-SNPs (Wicks et al., 2001). In *C. elegans* the mapping technique using snip-SNPs has been widely used to easily clone genes with a mutation because it permits the researcher to limit the genetic interval in which the desired mutated gene is located. SNPs were first used in *C. elegans* in 2001 (Wicks et al., 2001). Today, the process of linking a mutant phenotype to a gene can be done much more rapidly due to new techniques based on whole-genome sequencing (Doitsidou et al., 2010; Zuryn & Jarriault, 2013; Smith et al., 2016).

The snip-SNIP technique consists of two phases: in the first one, some snip-SNPs are used to perform a general mapping in order to verify in which chromosome the mutated gene resides; in the second one, mapping is more detailed to limit the genetic interval in the selected chromosome (Davis et. al, 2005).

In the case of the *t1903* mutant, the first phase of genetic mapping was not

necessary. It had been previously verified that the mutation was on chromosome IV because its segregation was linked to the genetic marker *unc-24 (e138)*. The second phase of the genetic mapping was done to ascertain the small genetic interval where the mutated gene was located.

First, a three-factor genetic strain was created in an N2 genetic background (Fig. 24). To generate this three-factor strain, the GE3632 strain (*unc-24 (e182); dic-1 t1903*) was crossed with the CB2223 strain (*unc-5(e53) dpy-20(e1282)*). Only the animals carrying the mutation flanked by two genetic markers, *unc-24(e138)* and *dpy-20(e1282)*, were selected. These genetic markers provide the worms with visible phenotypes. Whereas *unc-24(e138)* animals showed uncoordinated movements, *dpy-20(e1282)* worms were thicker and smaller than their wild-type counterparts.

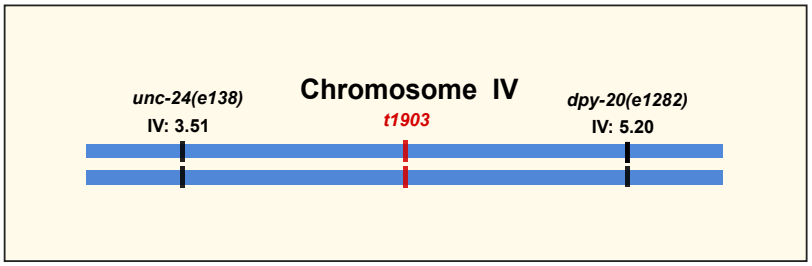


Figure 24. Three-factor strain used for genetic mapping of the mutant gene. A three-factor genetic strain was created in which the *t1903* mutation was flanked by two different genetic markers: *unc-24(e138)* and *dpy-20(e1282)*, both of which provide the animals with visible phenotypes so that it was possible to distinguish their presence in the progeny.

Then, the three-factor strain was crossed with the Hawaiian strain CB4856 in order to have heterozygotes in both genetic backgrounds, as well as the *t1903* mutation (Fig. 25).

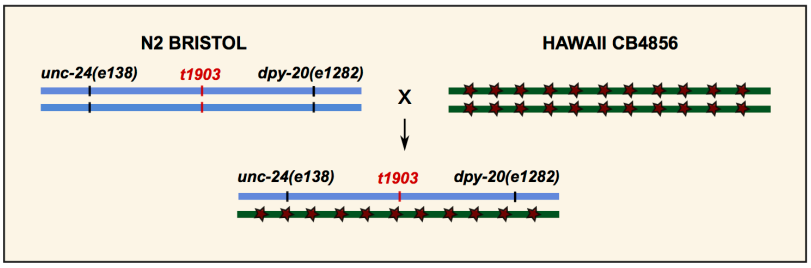


Figure 25. Crossing of the three-factor strain and the Hawaiian strain CB4856. The Bristol N2 and Hawaiian CB4856 were crossed in order to have heterozygotes of both genetic backgrounds, as well as the two genetic markers (*unc-24(e138)* and *dpy-20(e1282)*) and the *t1903* mutation.

The progeny of this heterozygotic worm gave rise to: heterozygotic animals as hermaphrodite mothers (phenotype: wild type), homozygotic animals in the Hawaiian strain genetic background (phenotype: wild type) and homozygotic animals in the three-factor strain background (phenotype: *unc-24(e138)* and *dpy-20(e1282),* thermosensitive). Additionally, some of the progeny had genetic recombination so they were homozygotic for one genetic marker but heterozygotic for the other (Fig. 26).

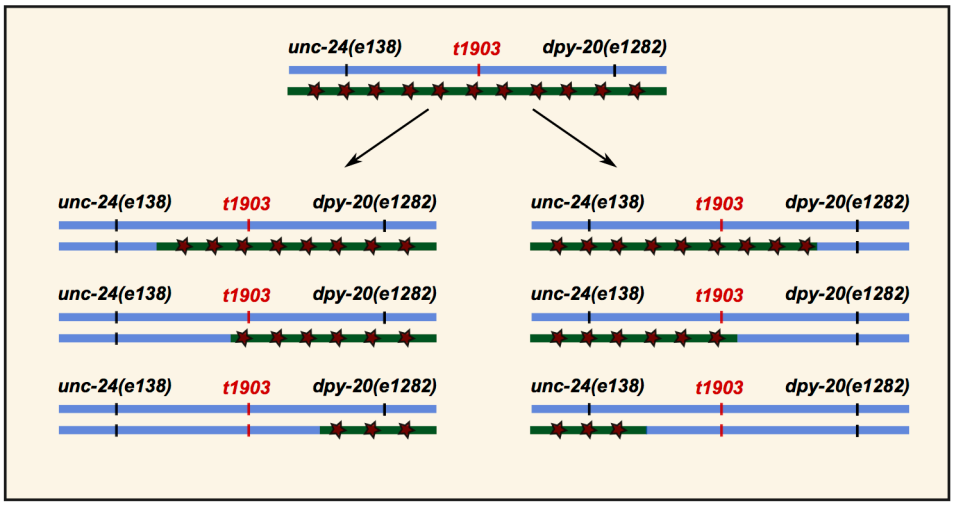


Figure 26. Possible recombinants in the progeny of the heterozygotic worms obtained by crossing the three-factor strain with the Hawaiian strain. The progeny that underwent genetic recombination were homozygotic for one genetic marker but heterozygotic for the other. In addition, they could be homozygotic for the mutation *t1903* (the progeny of these recombinants was not going to be viable) or heterozygotic. All the recombinants were used to mark the genetic interval where the mutation was located by snip-SNP mapping.

These genetically recombinant worms were used to mark the genetic interval where the mutation was located. Single worms that had either the *unc-24(e138)* phenotype or the *dpy-20(e1282)* phenotype were selected for analysis using snip-SNP. Basically, DNA from each recombinant worm was extracted and then amplified by PCR using specific primers (targeting different snip-SNIPs belonging to the desired genetic interval: 3.65-4.57). Afterwards, it was easy to distinguish if these PCR products came from a Bristol N2 or a Hawaiian genetic background because each snip-SNIP had a modification of a restriction enzyme site so when the PCR products were digested with the corresponding restriction enzyme, different bands were obtained, depending on the genetic background.

Snip-SNIPs analysis revealed that the *t1903* mutation was located between the regions 3.79 cM and 4.17 cM. The *t1903* mutant strain was transformed with cosmids containing genomic DNA fragments that overlapped with this genetic interval (these cosmids came from the library used to sequence the *C. elegans* genome). Only the F08B4 and H23L24 cosmids could rescue the *t1903* mutant phenotype, suggesting that the mutated gene was in these two cosmids and furthermore, that it was in the overlapping region between these two cosmids.

2.3. Mutated gene cloning

As the mutated gene was present in the overlapping region between cosmids F08B4 and H23L24, individual wild-type genes belonging to this region were cloned in plasmids and complementation experiments were performed. Only the *dic-1* gene was able to rescue the phenotypes of the *t1903* mutant. Moreover, when the GE3632 strain (*unc-24 (e182); dic-1 t1903*) was crossed with the NB327 strain (a *dic-1* deletion strain), the heterozygote double mutant *t1903/tm1615* was not viable, reaffirming that the *t1903* mutation was in the *dic-1* gene.

The NB327 strain was obtained from the GCG, its genotype was: *dic-1(tm1615)* IV/*nT1 [qIs51]* (IV;V). This *tm1615* strain was only viable in a heterozygous state as *dic-1(tm1615)* homozygotes arrested at L3 larval stage.

Later, *dic-1* was cloned from the mutant strain and sequenced. *dic-1* was confirmed to have a mutation, a swap from a C to T that resulted in a change from the amino acid serine 850 to phenylalanine.

3. Phenotypic characterization of the *t1903* mutant

Our lab continued working with the *t1903* mutant, which had a point mutation in the *dic-1* gene.

The *C. elegans dic-1* gene is the ortholog of the human gene *DICE1* (*deleted in cancer 1*), which was first described by Wieland in 1999. *DICE1* was isolated as a putative tumor suppressor gene in non-small cell lung carcinomas because *DICE1* is localized at a critical region in the chromosome where there is frequent loss of heterozigosity in lung carcinomas. Moreover, *DICE1* mRNA expression was reduced or undetectable in most non-small cell lung carcinoma cell lines tested (Wieland et al., 1999). Later, reduced *DICE1* expression was observed to be associated with hypermethylation of the CpG sites in the *DICE1* promoter region in non-small-cell lung carcinoma cell lines and in prostate cancer cell lines (Wieland et al., 2001; Röpke et al., 2005). Additionally, *DICE1* exogenous re-expression in cancer cell lines led to inhibition of their capacity to form colonies *in vitro* (Wieland et al., 2004; Filleur et al., 2009).

DICE1 was renamed in 2005 as Ints6 when Baillat and colleagues discovered the Integrator complex. Ints6 is a member of the Integrator complex, which is comprised of at least 14 subunits called Ints (Ints1 to Ints14) and it is associated with the CTD of Rpb1, the largest subunit of RNAP II. It was called Integrator because it integrates the CTD of RNAP II largest subunit of RNAP II with the 3'-end processing of U1 and U2 snRNAs (Baillat et al., 2005; Chen et al., 2012; Baillat & Wagner, 2015).

Prior to working with the *t1903* mutant (strain GE3632), we crossed it with N2 males to eliminate the *unc-24* genetic marker. Once we obtained the *t1903* mutant strain without genetic markers, it was backcrossed with N2 males three more times to achieve a "clean" genetic background generating the JCP294 strain.

To phenotypically characterize the thermosensitive (ts) *t1903* mutant, we analyzed its fertility (the capacity to produce offspring), embryonic viability (the percentage of hatched larvae) and the embryonic development.

To check the *t1903* fertility at 15ºC and 25ºC, we put single L2/L3 stage worms (n=3 worms per condition) on individual plates at the desired temperature and transferred these worms daily to a fresh plate until no further eggs were produced. Fig. 27A shows that the *t1903* mutant had a statistically significantly reduced fertility at 15ºC and 25ºC compared to the N2 wild type at 15ºC and 25ºC respectively. It is widely known that even when optimal conditions are present, N2 worms growing at 25ºC have reduced fertility compared to those growing at 15ºC (Hirsh et al., 1976). Fig. 27A shows the same results.

To analyse the viability of eggs laid by the ts *t1903* mutant at 15ºC and at 25ºC, N2 worms and *t1903* worms were grown at the desired temperature since L2/L3 stage. To obtain the percentage viability under each condition, the number of hatched larvae was divided by the number of dead embryos plus the number of hatched larvae. Fig. 27B (n= number of eggs laid) illustrates that the ts *t1903* mutant shows 100% embryonic lethality at 25ºC.

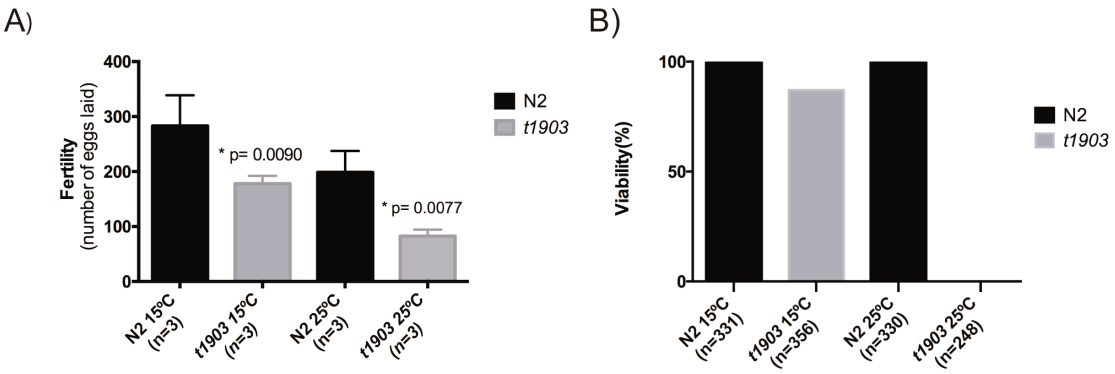


Figure 27. A) Fertility of the N2 versus the *t1903* mutant at 15ºC and 25ºC. Three single L2/L3 stage worms (n=3) of each genetic background were placed on individual plates at 15ºC and 25ºC and they were transferred daily to fresh plates until the worms stopped laying eggs. The total number of eggs laid was counted for each worm. The graph shows the number of eggs laid by N2 worms (in black; n=3) compared to those laid by *t1903* worms (in gray; n=3) growing at 15ºC and 25ºC (Mean ± standard derivation (SD)). The differences between N2 and *t1903* are statistically significant. P-values correspond to the Student's t-test. **B) Embryonic viability (%) of N2 versus *t1903* mutants, growing at 15ºC and 25ºC.** N2 and *t1903* worms were grown at the desired temperature starting at L2/L3 stage. The percentages of viability were obtained by dividing the total number of hatched larvae by the total number of eggs counted (hatched and not hatched) under each condition (n= number of

eggs). The graph shows the percentage of hatched larvae in N2 worms (in black) compared to the percentage of hatched larvae in *t1903* worms (in gray) growing at 15ºC and 25ºC.

Development of *t1903* mutant embryos was observed in detail under Nomarski 4D microscopy and compared to that of N2 embryos (Figs. 28, 29, 30 & 31; videos attached). Because the *t1903* mutant is ts, worms of both N2 and JCP294 strains, were grown at 15ºC. Then the worms were transferred to 25ºC and allowed to grow overnight (O/N) at the higher temperature prior to selecting the embryos.

As described in Material & Methos (MM 8.1.1), embryos from two to four-cell-stage were selected and prepared. Development of these embryos was observed under Nomarski/DIC optics at 25ºC. Images on 30 focal planes (1micron/section) were taken every 30 seconds for 16 h.

The *t1903* embryos showed morphogenetic defects as well as necrosis (Figs. 29, 30 & 31).

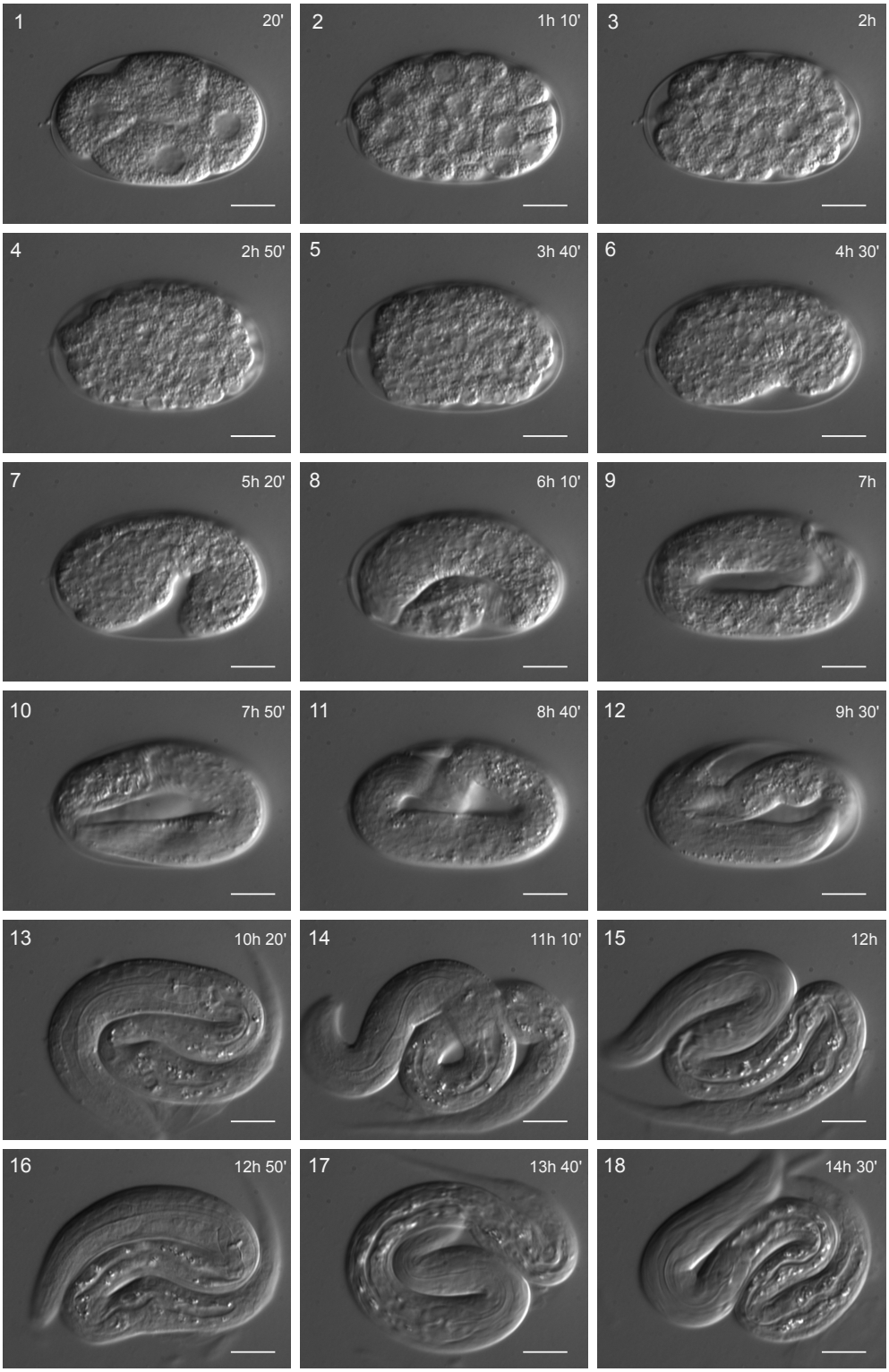


Figure 28. N2 embryonic development observed under 4D microscopy. *C. elegans* wild-type embryonic development from four-cell-stage embryo to hatching. The embryonic development (form image 1 to 18) takes 12 hours at 25ºC. DIC images taken every 50 min of development are shown. Developmental time is indicated on each image. Morphogenesis starts in image 6. Scale bar: 10 µm. (Video attached).

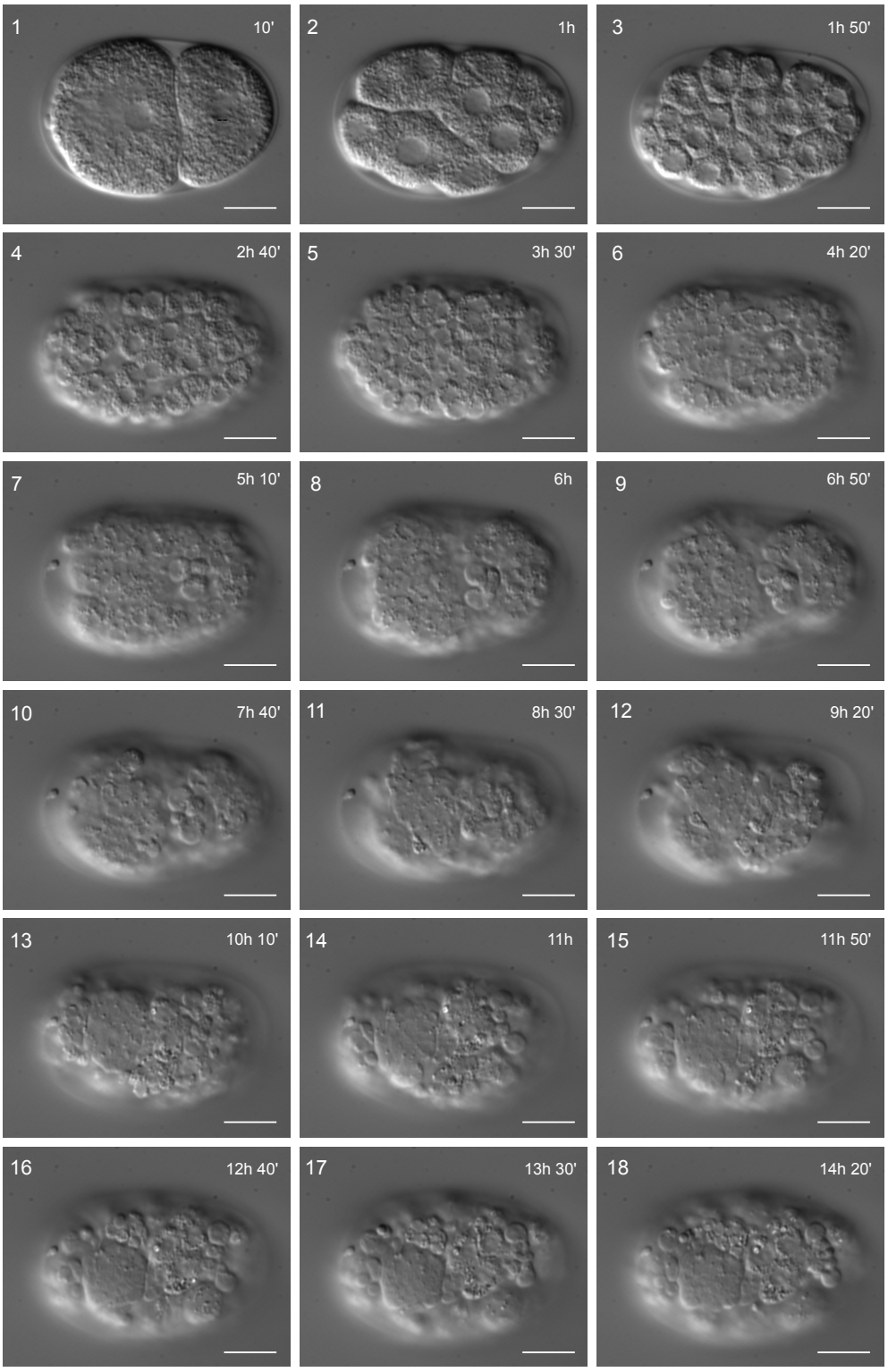


Figure 29. Embryonic development of the *t1903* mutant embryo number one observed under 4D microscopy. *t1903* embryonic development from two-cell stage embryo. DIC images taken every 50 min of development are shown. Developmental time is indicated on each image. This embryo shows morphogenesis defects: Hypodermal ventral enclosure defects can be observed in image 8. These defects result in a GEX (gut on the exterior) phenotype (images 9,10 and 11). Scale bar: 10 µm. (Video attached).

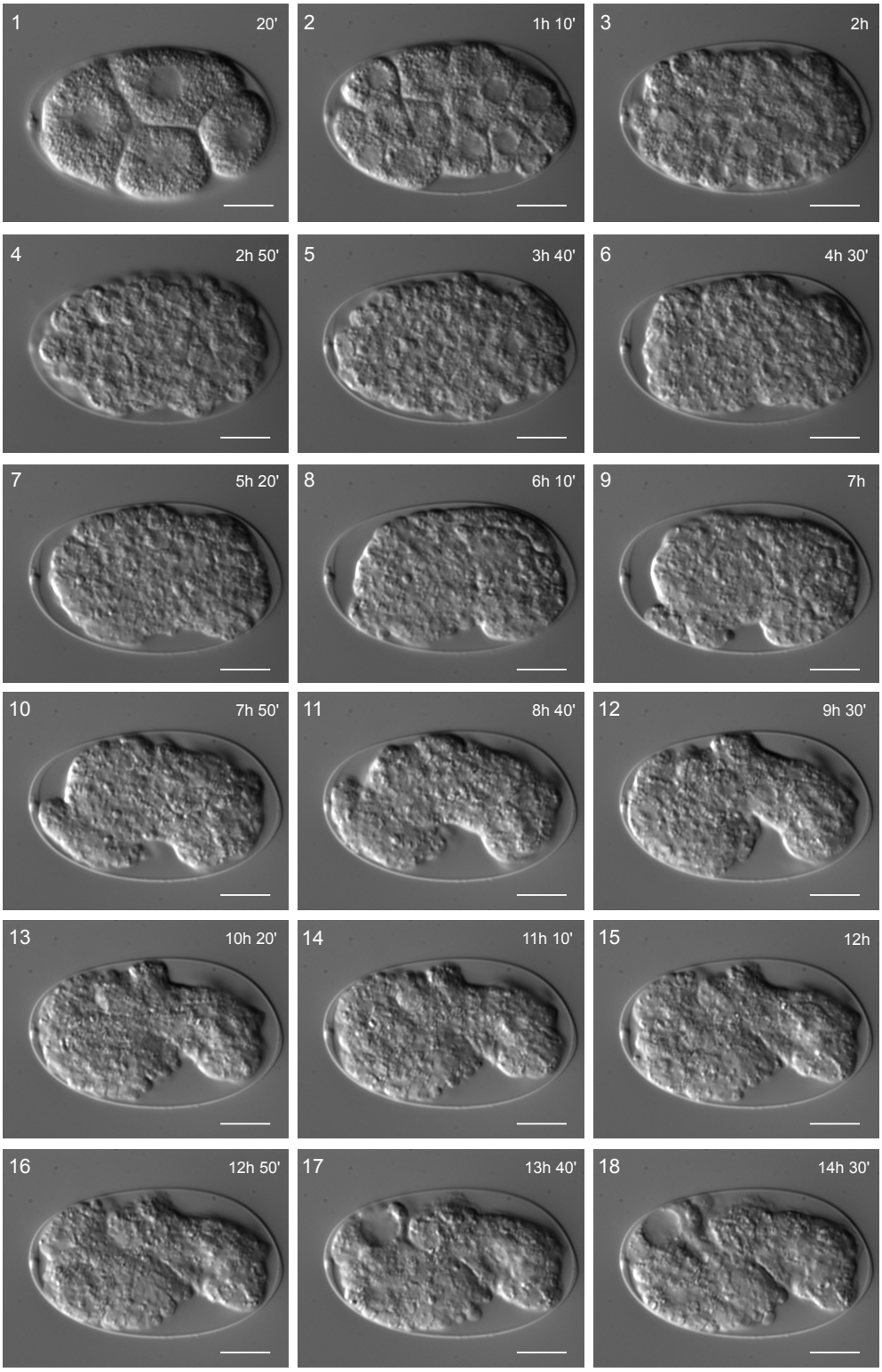


Figure 30. Embryonic development of the *t1903* mutant embryo number two observed under 4D microscopy. *t1903* embryonic development from four-cell stage embryo. DIC images were taken every 50 min of development are shown. Developmental time is indicated on each image. This embryo shows morphogenesis defects. In addition to this phenotype, necrosis became obvious in the late stages (images 17 and 18). Scale bar: 10 µm. (Video attached).

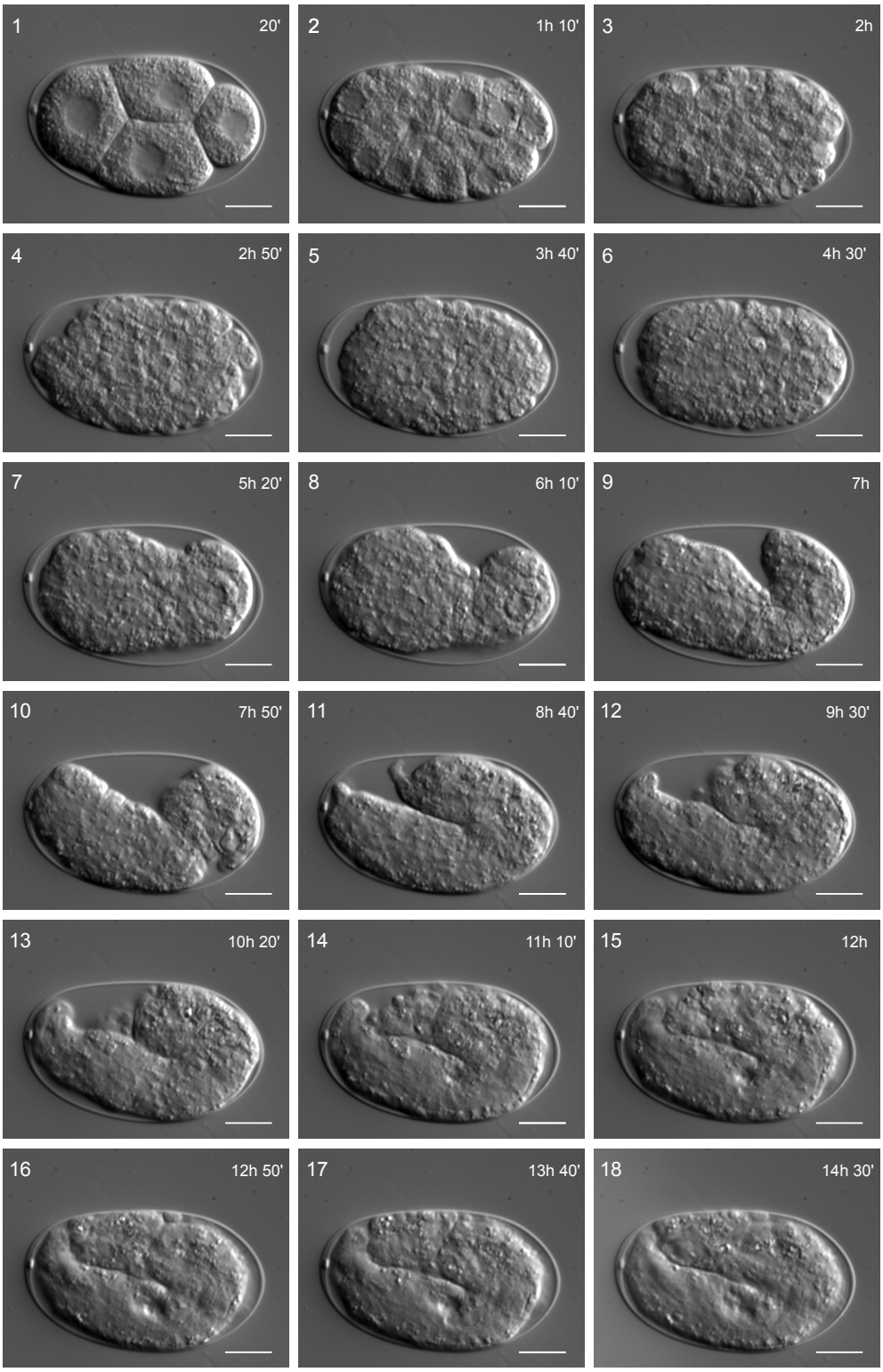


Figure 31. Embryonic development of the *t1903* mutant embryo number three observed under 4D microscopy. *t1903* embryonic development from four-cell stage embryo. DIC images taken every 50 min of development are shown. Developmental time is indicated on each image. This embryo shows morphogenesis defects. In addition to this phenotype, necrosis became obvious in late stages (images 14 and 18). Scale bar: 10 µm. (Video attached).

4. Localization of DIC-1

The human Ints6/*DICE1* gene is present in the nucleus as its function in U1 and U2 snRNA processing suggests (Wieland et al., 2004; Baillat et al., 2005). However, the *C. elegans* ortholog of Ints6/*DICE1,* called DIC-1*,* was reported to localize in the mitochondrial inner membrane where promotes cristae tubules formation and therefore is related to mitochondrial activity (Han et al., 2006; Lee et al., 2009). According to these studies, the *C. elegans* DIC-1 clearly presents differences in its subcellular location and function from its human homolog.

Because we wanted to study *dic-1*, we decided to corroborate the localization of the *C. elegans* DIC-1 as well as that of the human Ints6 gene (Images shown in section 6; Fig. 36).

To check DIC-1 localization, we created different *C. elegans* transgenic lines. First, *t1903* mutants (strain JCP294) were rescued by bombardment with the plasmid pJC12 (*dic-1*p::*dic-1*::GFP). The extrachromosomal arrays obtained could rescue the embryonic lethality phenotype, once again confirming that the *dic-1* mutation was responsible for the *t1903* phenotype. The transgene showed a nuclear localization for DIC-1::GFP, although the expression level was very low and varied from one experiment to another.

Because the extrachromosomal arrays are not stable, we decided to construct integrated transgenic lines under the control of the endogenous promoter and the strong promoter *eft-3*. The *eft-3* (homolog to the translation elongation factor 1-alpha) promoter has been reported to drive expression to the pharynx, intestine and body wall muscles (www.wormbase.org).

pJC51 (*dic-1*p::*ints-6*::3xFLAG::eGFP::*dic-1*UTR, *unc-119*(+)) and pJC56 (*eft-3*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR, *unc-119*(+)) plasmids (Figs. 32 & 33; full plasmids in MM 4.10.2.2) were used for microinjection into the gonad syncytium to integrate them into worms using mosSCI (mos1-mediated Single Copy Insertion) system to place them in chromosome II (Frokjaer-Jensen et al., 2008). This procedure (MM 6.3.1) generated the JCP341 & JCP378 strains respectively.

In addition, pJC56 plasmid was used to transform the *C. elegans* HT1593

(*unc-119(ed3)* III) strain by bombardment (MM 6.3.2). We chose the *unc-119* mutant genetic background to ease selection of the positive transformants because only the transformed worms had a wild-type phenotype. We obtained extrachromosomal arrays with higher expression levels and the GFP could be detected even under the dissecting microscope.

DIC-1 localization was nuclear in all transgenic lines. Nomarski and fluorescence microscopy images showing *dic-1* expression in the JCP341, JCP378 and JCP505 transgenic lines are shown in Figs. 32, 33 and 34 respectively. Care should be taken not to confuse cellular nuclei with gut granules that emit an intense blue fluorescence under UV light (Klass, 1977; Gerstbrein et al., 2005).

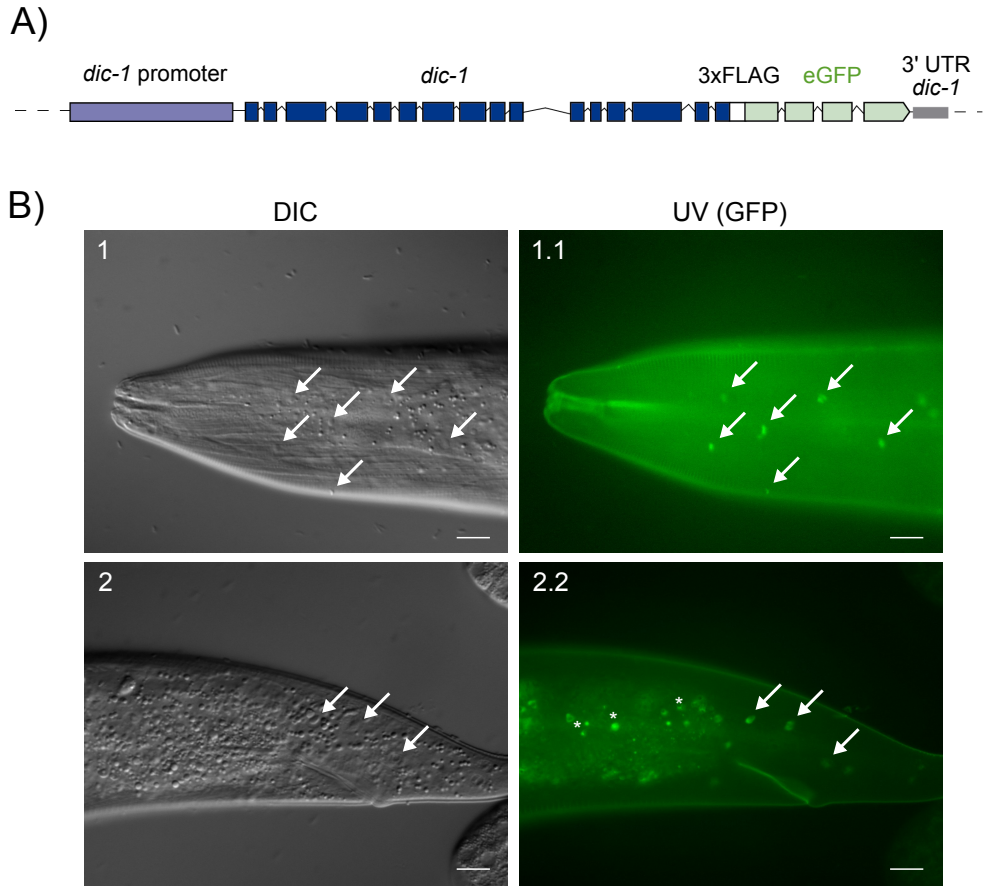


Figure 32. A) Scheme of the pJC51 plasmid used to generate the JCP341 transgenic line. The plasmid contains the endogenous *dic-1* promoter, the *dic-1* genomic sequence fused to 3xFLAG and eGFP followed by the *dic-1* 3'UTR. The full plasmid is shown in MM 4.10.2.2. **B) DIC-1 expression in the JCP341 transgenic strain**. Pictures show details of the head and tail. Images 1 and 1.1 show a head under DIC optics and ultraviolet (UV) light respectively. Arrows indicate cellular nuclei. Images B2 and B2.2 show a tail under DIC optics and UV light respectively. Arrows indicate cellular nuclei. Asterisks indicate gut granules that emit an intense blue fluorescence under UV light. Scale bar: 20 μm.

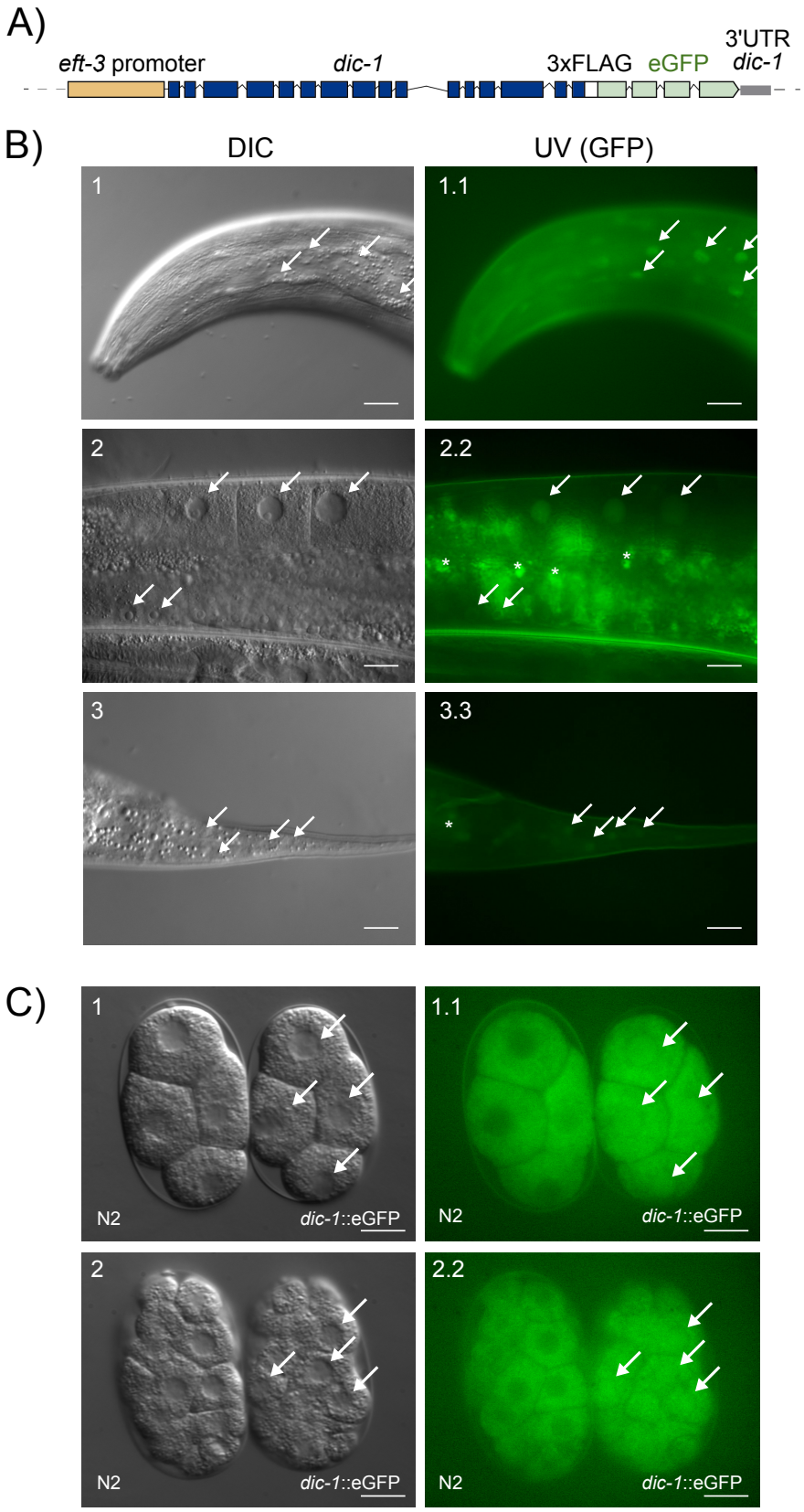


Figure 33. A) Scheme of the pJC56 plasmid used to generate the JCP378 transgenic line. The plasmid contains the *eft-3* promoter, the *dic-1* genomic sequence fused to 3xFLAG and eGFP followed by the *dic-1* 3'UTR. The full plasmid is shown in MM 4.10.2.2. **B) DIC-1 expression in the JCP378 transgenic strain**. Pictures show details of the head, gonad and tail. Images 1 and 1.1 show a head under DIC optics and UV light respectively. Arrows indicate cellular nuclei. Images 2 and 2.2 show a gonad under DIC optics and UV light

respectively. Arrows indicate cellular nuclei that can be observed in the syncytium of the gonad and the oocytes. Images 3 and 3.3 show a tail under DIC optics and UV light respectively. Arrows indicate cellular nuclei. Scale bar: 20μm **C) DIC-1 expression in embryos.** The long UV exposure time needed to acquire the images resulted in cellular autofluorescence, hence in each image there is an N2 embryo (left) compared to a JCP378 embryo (right). Arrows indicate cellular nuclei. Asterisks indicate gut granules that emit an intense blue fluorescence under UV light. Scale bar: 10 μm.

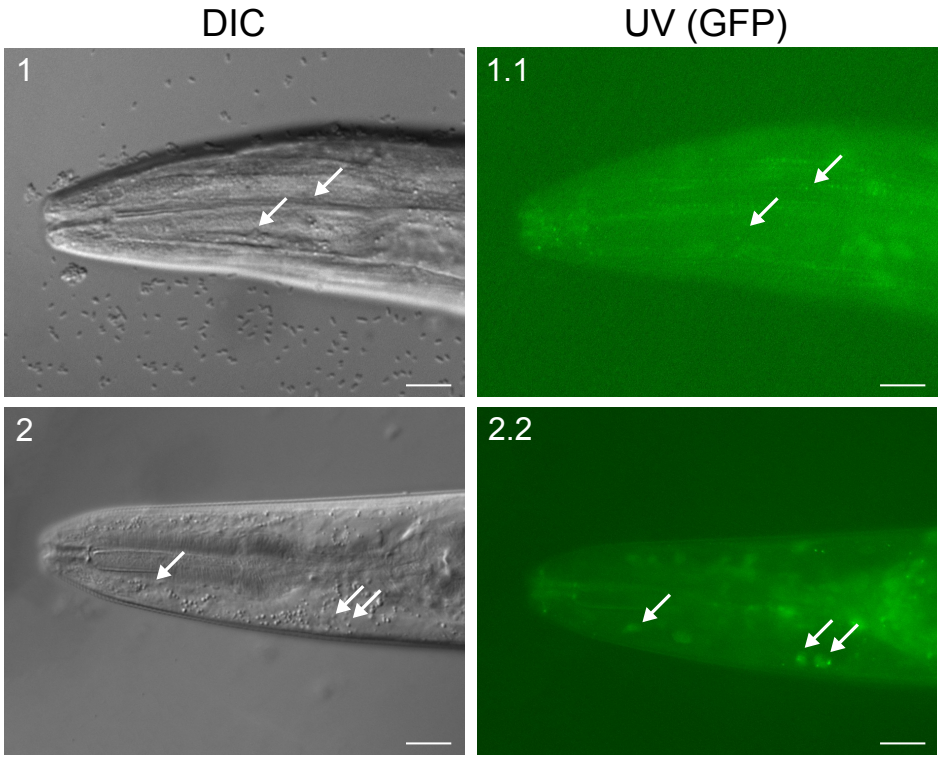


Figure 34. DIC-1 expression in the 505 transgenic strain. Pictures show details of the head. Transgenic strain JCP505 was created by bombardment of the HT1593 strain with the pJC56 plasmid. Images 1 and 2 show worm heads under DIC optics and 1.1 and 2.1 show the same heads under UV light. Arrows, on image 2 and 2.2, indicate cellular nuclei. Additionally, on image 1 and 1.1, arrows indicate the eGFP that might also localize to mitochondria in the body wall muscles. Scale bar: 20μm.

DIC-1 showed clear nuclear localization in all transgenic lines obtained. However, GFP expression could not be detected in embryos or germlines of transgenic worms carrying the endogenous promoter, probably because of low *dic-1* expression levels.

5. Transcriptomic analysis of the *t1903* mutant

We hypothesized that *dic-1* could function in snRNA 3'-end processing as its human ortholog because of its homology and nuclear localization. Therefore, we thought that the *t1903* mutant could have defects in snRNA 3'-end processing and as a result, splicing defects.

To analyze the possible involvement of *dic-1* in snRNA maturation and splicing we decided to analyze *t1903* ts mutant RNA by deep sequencing and compared it to N2 wild type RNA. Two different RNA-sequencing techniques were chosen for this purpose: RiboMinus RNA-sequencing (rmRNA-seq) and poly(A)-selected RNA-sequencing (mRNA-seq). The RiboMinus sequencing allows both, polyadenylated and non-polyadenylated transcripts, such as the snRNAs, to be studied. By contrast, poly(A)-selected RNA-sequencing only looks at the polyadenylated transcripts which were previously selected.

Because the *t1903* mutant is ts, worms were collected under two different conditions: one group was grown at 15ºC and another group was grown at 15ºC, but the night before collecting them, they were shifted to 25ºC. N2 wild type worms were grown and collected under the same conditions.

Next, total RNA was extracted using the *mirVana* ^{TM} *miRNA isolation kit* that enables the recovery of small RNAs such as snRNAs (MM 4.1). RNA deep sequencing was carried out by the Genomic Platform of the CIBIR (Center for Biomedical Research of La Rioja). Two separate sequencing libraries were prepared to run RiboMinus and poly(A)-selected RNA sequencing (MM 4.2). RiboMinus library was built to enrich the whole spectrum of RNA transcripts by selectively depleting ribosomal RNA molecules. The vast majority of rRNA was eliminated to allow the analysis of less abundant transcripts. The poly (A) library contained only polyadenylated transcripts that were isolated using poly-T oligo-attached magnetic beads. The resulting libraries were sequenced on the Illumina Genome Analyzer IIx platform to generate 150 bp single-end reads. We visualized the results of deep-sequencing using the IGV (Integrative Genomics Viewer) software. The results obtained from the rm-RNAseq showed that snRNAs 3'-end processing was affected in the U1, U2, U4, U5, U6 and SL families. This phenotype was already detected in *t1903*

worms growing at 15ºC and it was higher in the worms shifted to 25ºC the previous night. Figure 35A shows an example of rm-RNAseq reads visualized using IGV of a U1 and a U2.

Interestingly, major splicing defects were not found because the reads mapped the exons either in the rm-RNA seq (Fig. 35A) or the mRNAseq (Fig. 35B).

Correct splicing might be explained because, as observed in Fig. 35A, a fraction of the total number of snRNAs is processed correctly and this is enough for splicing. Additionally, snRNPs have been reported to have a long half-life and they are also recycled (Stutz et al., 1993; Fury & Zieve, 1996; Matera et al., 2007), which supports the idea that a small number of mature snRNAs is enough properly splice the mRNA of the cell.

The results of snRNA processing defects indicate that the *C. elegans dic-1* is involved in 3'-end processing of snRNAs and suggest the existence of a *C. elegans* Integrator complex responsible for the 3'-end processing of snRNAs. Therefore, the *C. elegans dic-1* will be called the Integrator subunit 6 or *ints-6* from this point on in this dissertation.

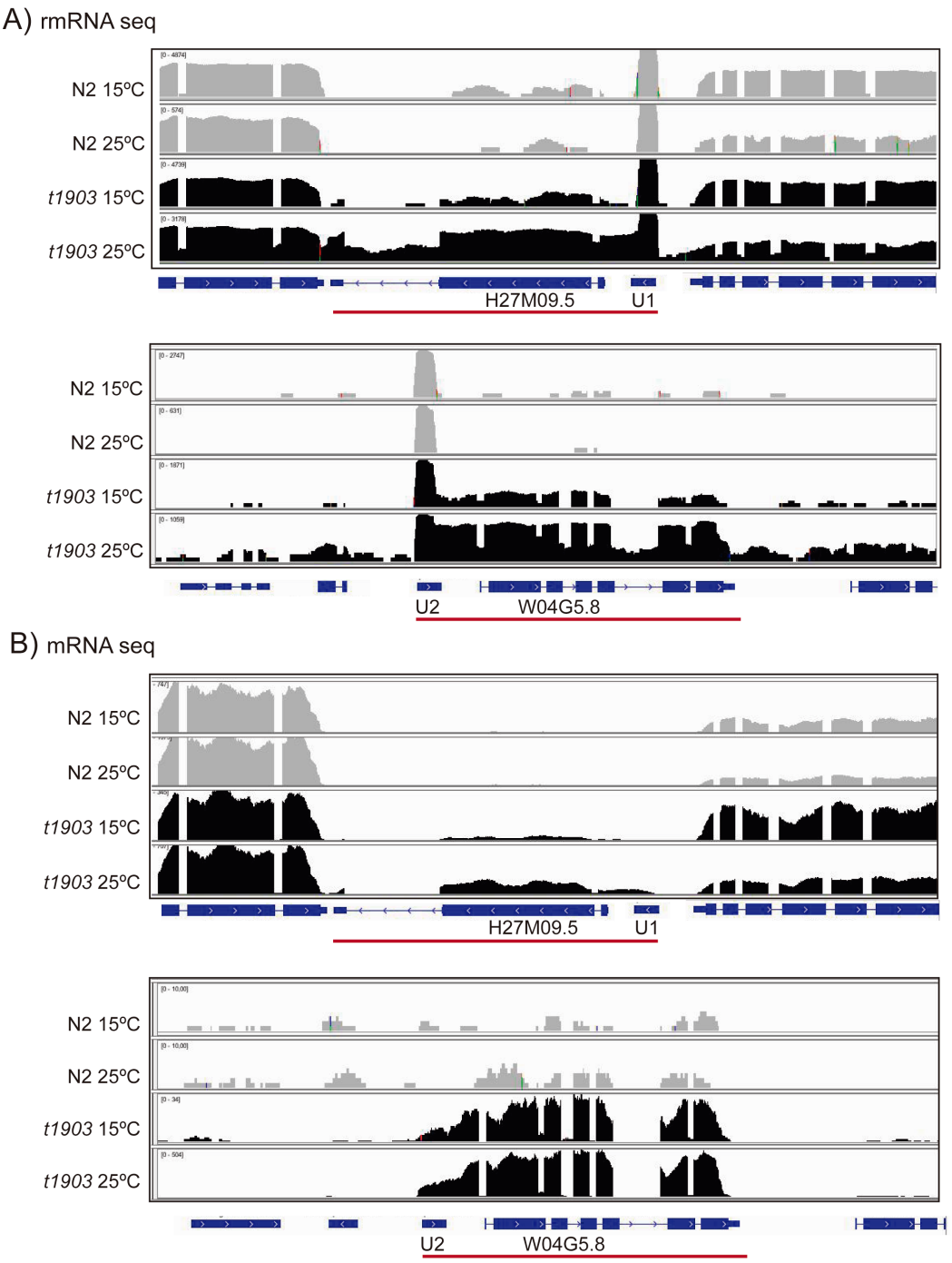


Figure 35. RNA deep sequencing of the *t1903* mutant versus N2 at 15ºC and 25ºC. A) RiboMinus RNA seq. (rmRNA-seq). B) Polyadenylated RNA seq (mRNA-seq). The figure shows the RNA deep sequencing reads aligned with the *C. elegans* genome in the region of the U1: H27M09.6 and the U2: W04G5.11 visualized in IGV software for each deep sequence technique, rmRNA-seq in A or mRNA-seq in B. N2 reads are shown in gray whereas *t1903* mutant reads are black. Underneath each graph, the *C. elegans* genome is represented in blue. The exons are shown as blue boxes and the introns as lines. Splicing in the *t1903* mutant is mainly correct because the reads mapped to the exons as in N2. As illustrated in section A, part of the reads belonging to U1 or U2 in the *t1903* mutant continue on to the downstream gene located *in sense*, which indicates that U1 and U2 are not cleaved at their 3-end. In addition, there is an overexpression of the gene downstream of U1 (H27M09.5) and U2 (W04G5.8) in the *t1903* mutant. This difference between N2 and the *t1903* mutant is underlined in red in sections A and B.

Strikingly, in the *t1903* mutant we observed that the reads downstream of the snRNA 3'-end continued and they reached the closest gene, suggesting that transcription termination is not occurring and the RNAP II continues transcribing (Fig. 35A). There are two possible scenarios: that the gene located downstream of the snRNAs is orientated *in sense* or that it is orientated in the *antisense* direction. In the former case, the lack of snRNA 3'-end processing and transcription termination would lead to the generation of chimeric "sn-mRNAs" that contain the snRNA at the 5' end of the sequence, continue on the intergenic region and have the mRNA sequence of the downstream gene at the 3'-end, properly spliced and polyadenylated (Fig. 35A & B). In the latter case, if the gene was in the opposite orientation, this would lead to the transcription of chimeric RNAs that possess the snRNA sequence at the 5'-end and an antisense RNA of the downstream gene, at the 3'-end.

We analyzed this effect on all of the snRNAs, one by one, using the deep sequencing results of *t1903* compared to N2. Tables 2 and 3 show the data obtained in the analysis of U1 and U2, respectively. The gene name, the chromosome, the position in the genome, the strand (sense or antisense: 1, -1) and the 3'-end processing (affected or not) are all marked. One can also see details about the next gene: the strand (sense or antisense), whether its expression was affected or not, the expression of the downstream gene in N2 and its function or ortholog if known.

In total, we found around 30 genes that were over-expressed downstream of the snRNAs. The function of most of these genes was unknown, but among them we found serpentine receptors or enzymes involved in nucleic acid biosynthesis pathways. Some of the genes whose function was unknown were specific for the *Caenorhabditis* phylum and did not have the corresponding ortholog in humans.

Gene name	Chr.	Genomic position	U Str	Proc.	Gene downstream	G.Str	Effect	N2	Function
H27M09.6	I	6847138 - 6847302	-1	NO	H27M09.5	-1	↑	NO	Ortho: SREK1
F58G1.10	II	12944734 - 12944898*	-1	NO	F58G1.7*	-1	↑(exon)		
C15F1.9	II	6968516 - 6968352*	-1	NO	C15F1.5*	-1	↑(exon)		
T27E4.11	V	9086999 - 9086836	-1	NO	---	---	---	---	---
T27E4.12	V	9086999 - 9091432	-1	NO	T27E4.1	1	↑	NO	Ortho: SYT13/7
T27E24.22	V	9091432 - 90914601	1	NO	T27E4.20	-1			
F40G12.13	V	14267692 -14267528	1	NO	F40G12.4	-1			
F40G12.14	V	14277723 -14277559	1	NO	F40G12.8	-1			
F08H9.10	V	14463337 -14463173	-1	NO	F08H9.3	-1	↑	NO	Hsp
F08H9.11	V	14468718 -14468554	-1	NO	F08H9.12, F08H9.4	-1;-1	↑;↑	NO	*srz-97*;Hsp
T08G5.11	V	14028620 -14028456	-1	NO	T08G5.3	-1	↑	NO	unknown

Table 2. U1 snRNA analysis of *t1903*. This table contains the following information for each U1 (left to right): **Gene name**, the name of the U1; **Chr.**, the chromosome; **Genomic position**, the position in the genome; **U Str.**, the strand (sense: 1 or antisense: -1); **Proc**., whether the 3'-end is processed or not; **Gene downstream**, the gene/s downstream of U1 (asterisks indicate that the U1 overlaps with an intron of an mRNA); **G. Str**., the strand of the downstream gene (sense: 1 or antisense: -1); **Effect**, the effect of the RNAP II read-through on the downstream gene, ↑ indicates that it is overexpressed; **N2**, whether the gene downstream is expressed in the N2; **Function**, the function of the overexpressed genes or the human ortholog if known. Dashed lines (---) indicate the absence of a gene. Abbreviations: Ortho, ortholog; Hsp, heat shock protein. Gene names: SREK1, Splicing Regulatory Glutamic Acid and Lysine Rich Protein 1; SYT13, Synaptotagmin-13; SYT7, Synaptotagmin-7; *srz-97,* serpentine receptor class Z.

Gene name	Chr.	Genomic position	U Str	Proc.	Gene downstream	G.Str	Effect	N2	Function
Y47H9A.2	I	11628159-11628345	-1	NO	Y47H9A.1	1			
F15H9.5	I	12144974 - 12145160	1	NO	F15H9.3 *	1	↑(exon)	NO	
F15H9.9	I	12149570 - 12149756	-1	NO	F15H9.4	-1	↑	NO	*sri-16*
R05D7.6	I	12173716 - 12173902	-1	NO	R05D7.3	-1	↑	NO	Ortho: TMEM200B
F56H6.16	I	12285737 - 12285923	1	NO	---				
F56H6.14	I	12286579 - 12286765	1	NO	F56H6.2	1	↑	NO	unknown
F56H6.15	I	12288212 - 12288398	-1	NO	F56H6.2; F56H6.4	1			
C47F8.9	I	12324749 - 12324935	-1	NO	C47F8.5	-1	↑	NO	Ortho: B3GALNT1
T14B4.10	II	6716133 - 6716318	-1	--	---				
Y54G9A.8	II	13717497 - 13717683	-1	NO	Y54G9A.4	-1	↑	NO	*zipt-2.3*
F08G2.10	II	13829236 - 13829422	-1	NO	F08G2.3	-1	↑	YES	*his-42*
F08G2.9	II	13835865 - 13836049	-1	NO	F08G2.6; F08G2.5	-1;-1	↑	NO	*ins-37;* in.im.re.
F08G2.11	II	13852715 - 13852901	1	NO	F08G2.8 *	1	↑(exon)	NO	
W07G1.9	II	13948898 - 13949084	-1	N. cl.	W07G1.2				
W07G1.8	II	13948898 - 13949084	1	NO	W07G1.2*	1	↑(exon)	NO	
C05G6.4	IV	624083 - 624268	1	N. ex.	---				
Y57G11C.35	IV	14764816 - 14765002	-1	NO	Y57G11C.5 *	-1	↑(exon)	NO	
F11A5.14	V	16225681 - 16225867	-1	NO	---	1			

Table 3. U2 snRNA analysis in *t1903*. This table contains the following information for each U2 (left to right): **Gene name**, the name of the U2; **Chr.**, the chromosome; **Genomic position**, the position in the genome; **U Str.**, the strand (sense: 1 or antisense: -1); **Proc**., whether the 3'-end is processed or not; **Gene downstream**, the gene/s downstream of the U2 (asterisks indicate that the U2 overlaps with an intron of an mRNA); **G. Str**., the strand of the downstream gene (sense: 1 or antisense: -1); **Effect**, the effect of the RNAP II read-through on the downstream gene, ↑ indicates that it is overexpressed; **N2**, whether the gene downstream is expressed in the N2; **Function**, the function of the genes overexpressed or the human ortholog if known. Dashed lines (---) indicate the absence of a gene. Abbreviations: Ortho, ortholog; N. cl., not clear; N. ex. not expressed; in.im.re., innate immune response. Gene names: *sri-16*, serpentine receptor class I; TMEM200B, (transmembrane protein 200B); B3GALNT1, beta-1,3-N-acetylgalactosaminyltransferase 1; *zipt-2.3* (Zrt (ZRT), Irt- (IRT-) like Protein Transporter; *his-42*, histone 42.

6. Briefly characterization of human Ints6

We decided to briefly characterize human Ints6 to confirm its location and function in U1 and U2 3'-end processing as described in the scientific literature (Baillat et al., 2005).

6.1. Localization of human Ints6

We verified the Ints6 localization in two distinct cell lines, HEK293T (human embryonic kidney) and U2OS (human bone osteosarcoma) cells. HEK293T cells are widely used in cell biology, and they were chosen because they are easily to transfect. Instead, U2OS cells were chosen because they are bigger and therefore good candidates in which to observe the localization of a protein.

Human Ints6 was cloned into the backbone of commercial vectors pEGFP-C1 and pEGFP-N1, generating the pBS15 plasmid, which has the eGFP tagged to the Ints6 Nt domain and pBS16 plasmid, which has the eGFP tagged to the Ints6 Ct domain, respectively (Fig. 36A; MM 4.10.1). HEK293T cells were transfected (MM 6.2) with pBS15 or pBS16 and eGFP fluorescence was checked *in vivo* the following day. Ints6 localization was clearly nuclear whether the eGFP was tagged in its Nt or Ct domain (Figure 36A.1).
To confirm the localization in another cell line, U2OS cells were transfected separately with both plasmids (MM 6.2). Forty-eight hours later, these U2OS cells were fixed and stained with DAPI. As shown in Figure 36 A.2, the localization of Ints6 was nuclear. In addition, U2OS cells were fixed, immunostained using Ints6 antibody and counterstained with DAPI (MM 5.1.1), and again, the localization was observed in the nucleus (Fig. 36B).

Using different approaches, transfection and immunostaing, we concluded that Ints6 localization is nuclear in both cell lines, HEK293T and U2OS.

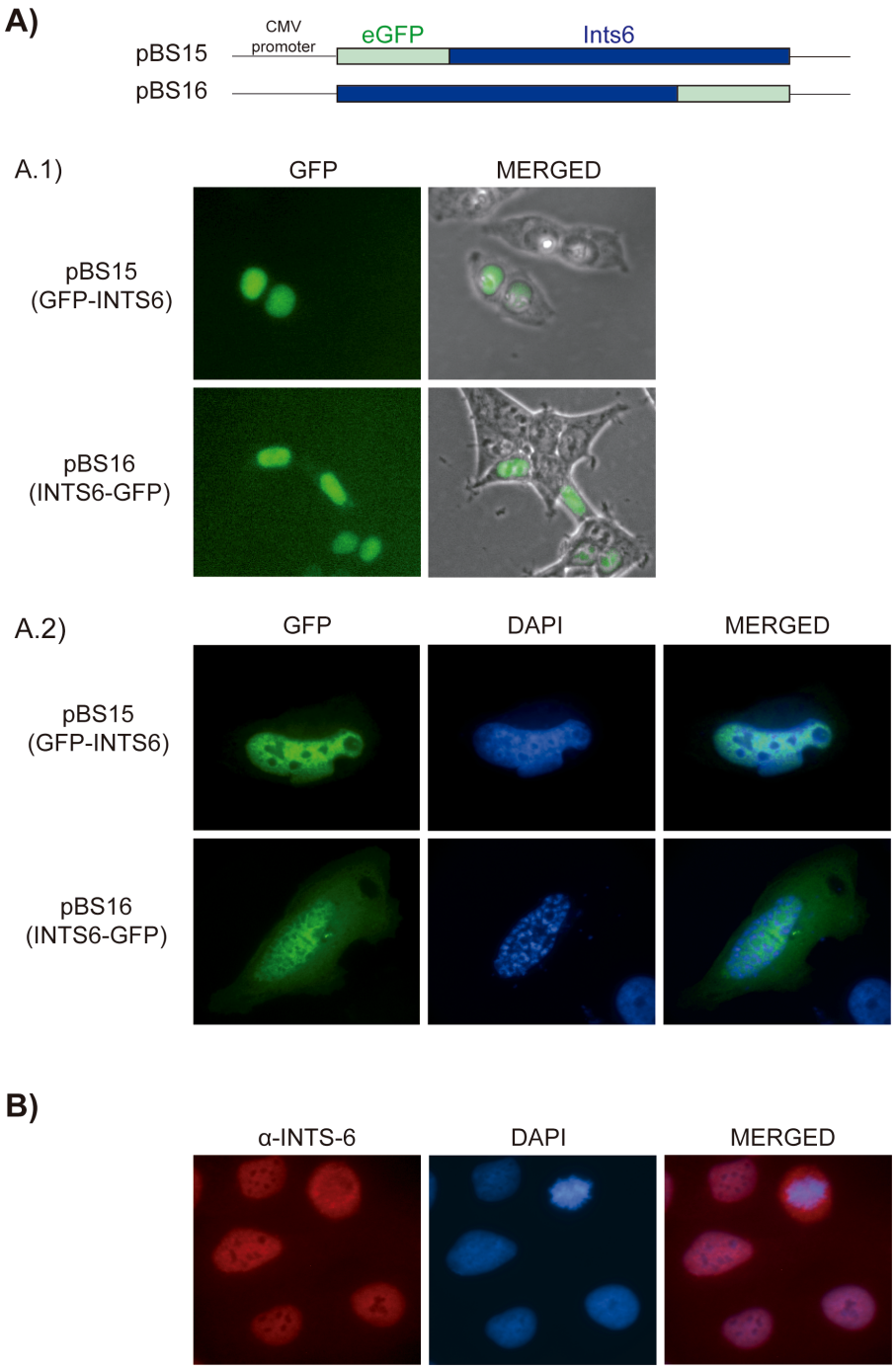


Figure 36. Human Ints6 localizes in the nucleus. A) Schematic drawing of the pBS15 plasmid (eGFP-Ints6) and pBS16 plasmid (Ints6-eGFP). These plasmids were used to transfect the HEK293T and U2OS cell lines. **A.1) *In vivo* HEK293T cells transfected with pBS15 or pBS16.** Images show eGFP fluorescence, checked 24h post-transfection, under UV light and the merging of UV light plus bright field. Ints6 localization was clearly nuclear in both transfections. **A.2) Fixed U2OS cells transfected with pBS15 or pBS16.** 48h post-transfection, U2OS cells were fixed and stained with DAPI. Images show GFP fluorescence, DAPI and merged. Ints6 localizes to the nucleus in both transfections. The images of the transfection with pBS16 (bottom) show that the DNA is starting to condense and, on this occasion, a minor amount of Ints6 is also observed in the cytoplasm. **B) Immunostaining of U2OS cells with α-INTS6.** U2OS cells were fixed, immunostained with α-INTS6 (Bethyl Laboratories, Inc.), and counterstained with DAPI. Ints6 localization was clearly nuclear. In the dividing cell, the DNA is clearly condensed, there is no nucleus and Ints6 is in the cytoplasm.

6.2. Human Ints6 protein levels are constant throughout the cell cycle

We decided to check the dynamics of Ints6 protein expression levels throughout the cell cycle because previous studies reported a link between Ints6 function and fundamental pathways involved in cell cycle regulation (Filleur et al., 2009).

We synchronized U2OS cells by performing a double thymidine block procedure, followed by a release in the presence of nocodazole (MM 10.1). U2OS cells were collected in distinct phases of the cell cycle (G1, G1/S, G2, and M) and also while the cells were growing asynchronously. Once collected, cells were analysed by FACS (fluorescence activated cell sorting) and WB. FACS DNA content analysis (MM 10.2) corroborated that the cell cycle phase collected in each case was correct (Fig. 37B). The WB was performed to check Ints6 protein expression levels throughout the distinct phases of the cell cycle and to corroborate the cell cycle phase collected in each case. As it can be observed in Figure 37A, there are no differences in Ints6 protein expression levels throughout the distinct phases of the cell cycle.

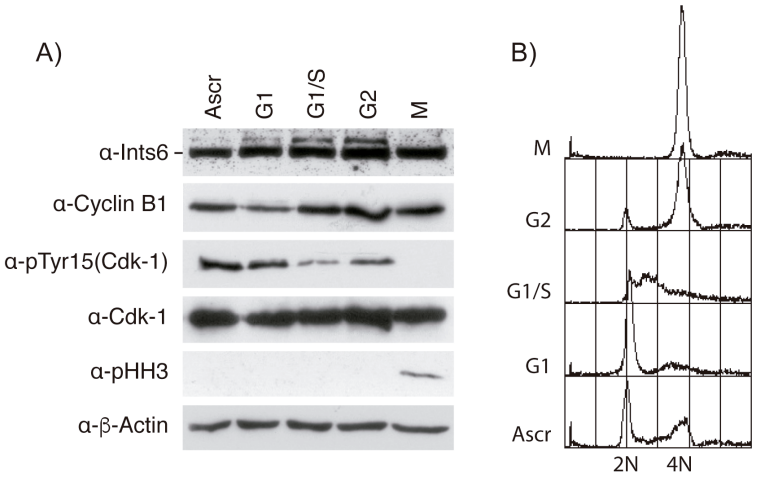


Figure 37. Human Ints6 protein expression levels are constant throughout the cell cycle. U2OS cells were synchronized by a double thymidine block treatment and later released in the presence of nocodazole. **A) The WB shows Ints6 and distinct cell cycle markers throughout the cell cycle.** The cell cycle markers Cyclin B1, pTyr15-Cdk1, Cdk1 and pHH3 show that the cell cycle phase collected in each case is correct. Only mitotic cells are phosphorylated on histone 3 (Ser 10) and not phosphorylated on Tyr-15 Cdk1. Ints6 protein expression levels are constant throughout the cell cycle. **B) FACS to corroborate cell cycle phase (G1, G1/S, G2 and Mitosis).** The DNA content in each case shows that the cells were collected correctly in each phase, G1, G1/S, G2 and mitosis.

Interestingly, an upper band on the WB of Ints6 seems to be cell-cycled regulated, which we do not know if it is specific.

In addition, the distinct cell cycle markers used: Cyclin B1, pTyr15-Cdk1, Cdk1 and pHH3, reassured us that the cell cycle phase collected in each case was correct. Only mitotic cells are phosphorylated on histone 3 (Ser 10) and not phosphorylated on Tyr-15 Cdk1.

6.3. RNAP II Co-Immunoprecipitates with human Ints6

During snRNA transcription, the Integrator complex is bound to RNAP II CTD (Egloff & Murphy, 2008). Therefore, Ints6 should interact (directly or indirectly) with RNAP II. To verify the interaction, we immunoprecipitated Ints6 and checked for the presence of RNAP II. First, we transfected HEK293T cells with pBS15 (GFP-Ints6) to overexpress Ints6 (MM 6.2). Next, protein was extracted and Ints6 was immunoprecipitated using the GFP antibody as described in MM 5.7.1. As observed in Figure 38, we could detect a subtle band that corresponds to RNAP II in the Ints6 immunoprecipitate whereas in the input control, although it is little bit dirty, RNAP II is not visible.

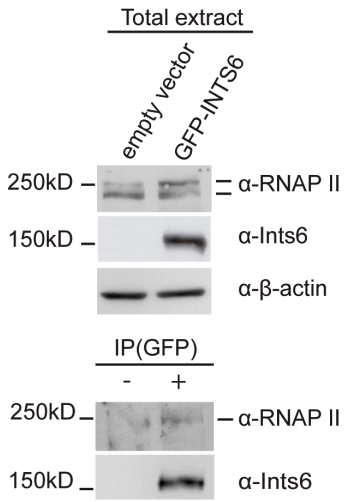


Figure 38. RNAP II Co-immunoprecipitates with Ints6. HEK293T cells were separately transfected with pBS15 (eGFP-Ints6) and the empty vector as a negative control. Cells were collected 48 hours later and lysed. Cell extracts were used to immunoprecipitate Ints6 with α-GFP (monoclonal, Living Colours® Clontech). Immunoprecipitates were analysed by WB (α-RNAP II (H-224) polyclonal sc; α-Ints6: Bethyl Laboratories, Inc.). The upper WB shows RNAP II, Ints6 and β-actin in total extracts. The lower WB shows RNAP II (subtle band) and Ints6 in the immunoprecipitate from cell extracts transfected with pBS15, whereas in the cells transfected with the empty vector no bands were detected.

6.4. Depletion of human Ints6 affects snRNA 3'-end processing

We wanted to corroborate the function of human Ints6 in U1 and U2 3'-end processing as previously reported in the scientific literature (Baillat et al., 2005). To study the function of Ints6, we chose RPE (retinal pigment epithelial) cells because they are non-tumoral. Ints6 was depleted using distinct siRNAs targeting Ints6 (siRNA Ints6 number 5: 5'-GAAGAGCACUCGCAGAUUU-3' and siRNA Ints6 number 8: 5'-GAGCCGAUCACAUGGUUUA-3'). RPE cells were transfected with each Ints6 siRNA separately. An siRNA against the Luciferase (5'-NNACGUACGCGGAAUACUUCGAA-3') was used as the negative control (MM 7.1). RPE cells were collected 48h post-transfection. First, we corroborated by WB that after transfection of the RPE cells with the two distinct siRNAs, Ints6 protein expression levels were lower, as shown in Figure 39A.

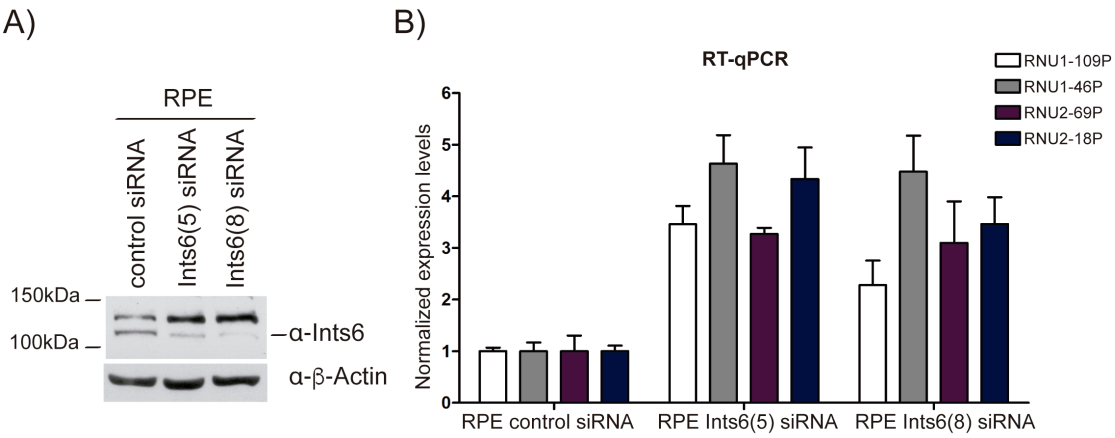


Figure 39. Depletion of human Ints6 affects snRNA 3'-end processing. A) Ints6 protein levels after Ints6 depletion. RPE cells were transfected with two siRNAs targeting Ints6 (Ints6 siRNA number 5 and Ints6 siRNA number 8). An siRNA against the Luciferase was used as the control. The WB shows that Ints6 protein levels are lower after transfection of RPE cells with the siRNAs targeting Ints6 compared to the RPE cells transfected with the control. **B) RPE depleted cells for Ints6 show 3'-end snRNA processing defects in U1 and U2 snRNAs.** Total RNA was extracted from RPE cells depleted of Ints6 and RT-qPCR using primers designed on the 3' region of two U1 (RNU1-109P & RNU1-46P) and two U2 (RNU2-69P & RNU2-18P) was performed. Primers targeting the 18S were used to normalize the data. The figure shows that there is an increase in the amplification of the 3' region in both siRNAs targeting Ints6 compared to the siRNA control.

Afterwards, total RNA extraction of the depleted cells was done using *mirVana* ^{TM} *miRNA isolation kit* that enables recovery of small molecules such

as snRNAs. Real time quantitative PCR (RT-qPCR) using primers designed on the 3' region (downstream of the 3'box) of two U1 (RNU1-109P & RNU1-46P) and two U2 (RNU2-69P & RNU2-18P) was performed (Fig. 39B). Data obtained from amplification of the U1 and U2 3' regions were normalized against the data from the amplification of primers targeting the 18S. As shown in Figure 39B, there is an increase in the amplification of the 3' region in both siRNAs targeting Ints6 compared to the control siRNA. Therefore, Ints6 depletion caused snRNA 3'-end processing defects.

6.5. Depletion of human Ints6 results in over-proliferation

Because we observed that RPE cells depleted of Ints6 grew faster than control siRNA treated RPE cells, we decided to quantify that difference in a proliferation assay. Separately, we transfected RPE cells with the two siRNAs previously used that target Ints6 (number 5 and number 8). Again, the siRNA targeting the Luciferase was used as the control. Later, the number of cells was quantified at different times: 24h, 48h and 72h. At 24h, the number of cells was slightly higher in the cells depleted of Ints6. That difference increased at 48h post-transfection, and at 72h, the difference was very high (Fig. 40).

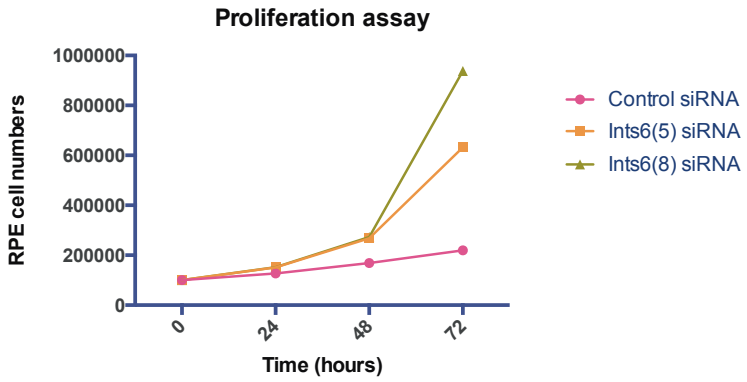


Figure 40. RPE cells depleted of Ints6 increased in number. RPE cells were transfected with two distinct siRNAs targeting Ints6 (Ints6 siRNA (5) and Ints6 siRNA Ints6 (8)). Control RPE cells were transfected with an siRNA against the Luciferase. RPE cells were collected and counted at different times post-transfection: 24h, 48h and 72h. The graph shows the over-proliferation phenotype of the RPE cells depleted of Ints6 compared to RPE cells treated with control siRNA.

7. Identification of the *C. elegans* Integrator complex

7.1. The Integrator complex is evolutionarily conserved in metazoans.

The Integrator complex was discovered in 2005 in human cell lines. Twelve subunits were identified at that time, and they belonged to a complex involved in U1 and U2 snRNA processing. Integrator subunits were named Ints1 to Ints12 according to their molecular weight from the highest to the lowest.
Later, another study in *Drosophila* revealed the existence of two new subunits and they were named Ints13 and Ints14 (Baillat et al., 2005; Chen et al., 2012).
Because our previous results suggested the existence of a *C. elegans* Integrator complex, we bioinformatically searched the Integrator subunit orthologs. The protein sequences of *H. sapiens* Integrator subunits were taken from the Uniprot database (http://www.uniprot.org/) and their corresponding orthologs were searched using BLAST (Basic Local Alignment Search Tool) with default parameters for the *C. elegans* genome (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The ortholog results found were corroborated with those proposed in www.orthodb.org. The same analysis was performed in the following species: *M. musculus*, *G. gallus*, *D. rerio* and *D. melanogaster*. The table below shows the percentages of identity (in bold) and similarity (in brackets) found for each Integrator complex subunit in each of the species compared to *H. sapiens* Integrator subunits.
The Ints9 and -11 subunits, are the most conserved in all species, and these two subunits are homologous to CPSF100 and CPSF73 respectively, which are involved in cleavage of pre-messenger RNAs (Xiang et al., 2014).
In *C. elegans,* after INTS-9 and -11 subunits, the INTS-1, -6, -7, -13, -3 and -2 are the best conserved with a similarity of over 40%. However, in INTS-4, -5, -8, -10 and -12 subunits, the homology was found only a portion that covered less than 50% of the protein.
Surprisingly, INTS-14 is very well conserved in all species, but no orthologs were found by homology in the *C. elegans* genome.

	H. sapiens	*M. musculus*	*G. gallus*	*D. rerio*	*D. melanogaster*	*C. elegans*
Ints1	NP_001073922	NP_081024 **90%** (94%)	XP_414768 **84%** (92%)	NP_003198200 **74%** (85%)	NP_611875 **37%** (56%)	NP_491739 **29%** (49%)
Ints2	NP_065799	NP_081697 **96%** (97%)	NP_001025893 **91%** (94%)	NP_001018378 **78%** (87%)	NP_573232 **38%** (58%)	NP_001256569 **19 %** (41%)
Ints3	NP_075391	NP_663515 **99%** (99%)	XP_001235108 **95%**(98%)	NP_001038397 86% (93%)	NP_001036437 **39%** (59%)	NP_490863 **23%** (45%)
Ints4	NP_291025	NP_081532 **97%** (97%)	XP_417220 **90%** (94%)	XP_001923719 **77%** (86%)	NP_572488 **36%** (54%)	NP_493396 * **32%** (49%)
Ints5	NP_085131	NP_789813 **95%** (97%)	XP_00495052 **58%** (67%)	XP_002664505 **51%** (65%)	AAR96207 **30%** (42%)	NP_507625 **22%** (40%) *
Ints6	NP_036273	NP_032741 **96%** (97%)	XP_417071 **86%** (91%)	NP_1119900 **67%** (76%)	NP_ 572253 **46%** (62%)	NP_501492 **28%** (45%)
Ints7	NP_056249	NP_848747 **94%** (97%)	NP_001006399 **92%** (91%)	NP_775374 **77%** (88%)	NP_648352 **35%** (53%)	NP_496477 **28%** (46%)
Ints8	NP_060334	NP_001153067 **96%** (98%)	NP_001035318 **89%** (95%)	NP_001018592 **70%** (82%)	NP_611162 **26%** (44%)	NP_001252168 **21%** (40%) *
Ints9	NP_060720	NP_700463 **97%** (99%)	NP_001026271 **93%** (97%)	NP_001070738 **80%** (89%)	NP_648838 **52%** (70%)	NP_504953 **35%** (57%)
Ints10	NP_060612	NP_81866 **95%** (98%)	NP_81866 **96%** (98%)	NP_997804 **67%** (82%)	NP_647830 * **27%** (47%)	NP_500453 * **23%** (39%)
Ints11	NP_060341	NP_82296 **97%** (98%)	NP_001012854 **92%** (95%)	NP_001018457 **84%** (92%)	NP_651721 **69%** (81%)	NP_495706 **60%** (75)
Ints12	NP_065128	NP_082203 **91%** (95%)	NP_0420500 **75%** (82%)	NP_001002161 **59%** (70%)	NP_651507 * **29%** (45%)	NP_505182 * **39%**(60%)
Ints13	NP_060634	NP_620096 **92%** (94%)	NP_004938067 **94%** (97%)	NP_956177 **80%** (89%)	NP_524379 **41%** (61%)	NP_499881 **26%** (44%)
Ints14	NP_001129515	NP_780362 **97%** (98%)	XP_015147580 **94%** (97%)	NP_956128 **79%** (91%)	NP_608579 **33%** (48%)	

% Id.
100
91-99
81-90
71-80
61-70
51-60
41-50
31-40
21-30
11-20
0 -10

Table 4. Integrator subunits in *H. sapiens* and their orthologs in the following species: *M. musculus*, *G. gallus*, *D. rerio*, *D. melanogaster* and *C. elegans*. The Integrator subunit protein sequences were searched for in the Uniprot database (http://www.uniprot.org/) and their corresponding orthologs were searched for using the BLASTp tool with default parameters against the corresponding genome (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Results were corroborated at www.orthodb.org. Percentages of identity (in bold) and similarity (in brackets) are shown. Each cell in the table is colored depending on the percentage of identity between the human sequence and the organism being considered. Black means 100% homology and white means no significant homology. Asterisks indicate that significant homology was found on a portion covering less than 50% of the protein sequence.

In addition, we performed a deeper analysis to compare *H. sapiens* vs *C. elegans* Integrator complex members found by homology. Motifs for each subunit were searched using the Pfam database (Fig. 41).

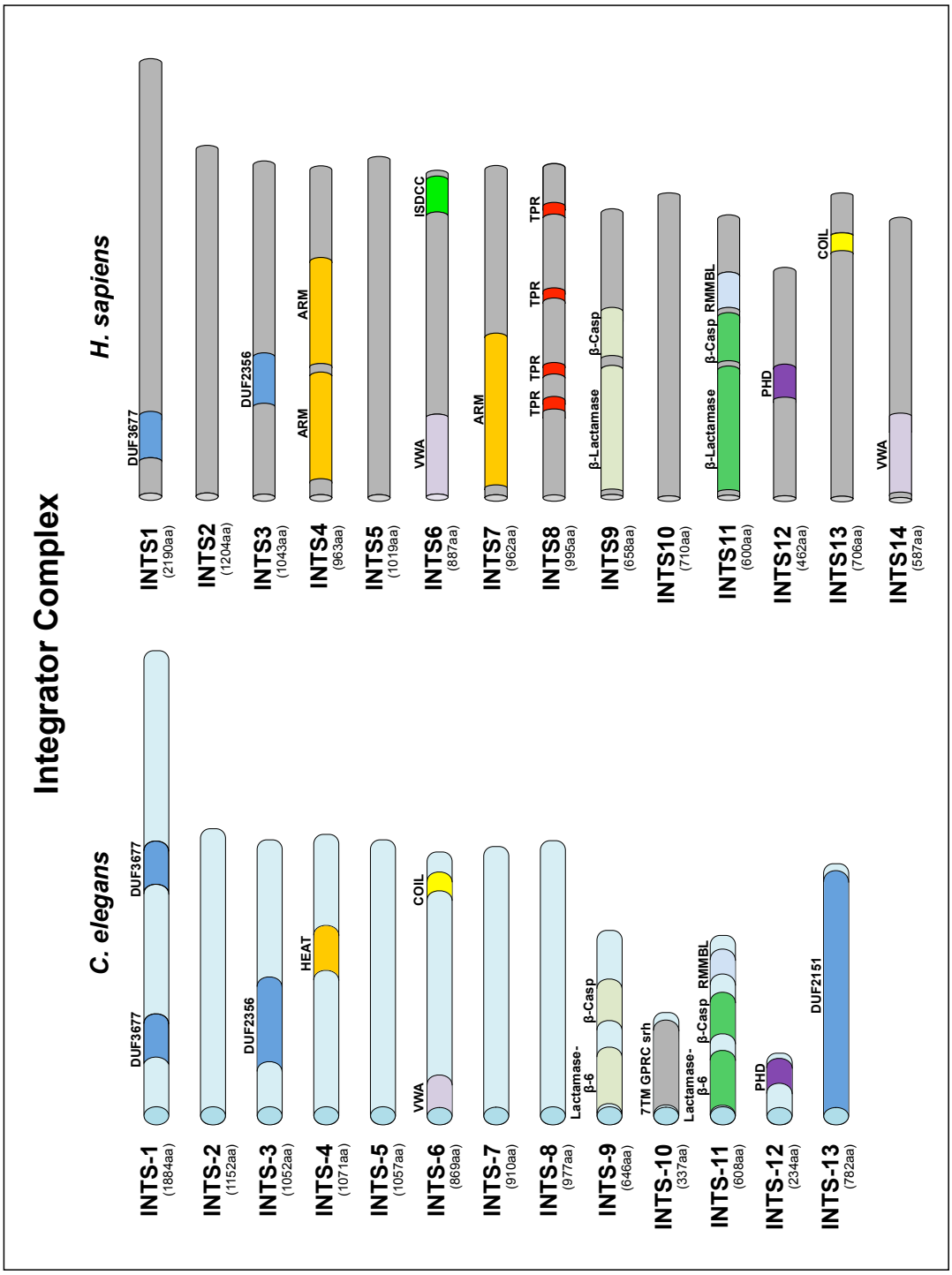


Figure 41. Predicted *C. elegans* Integrator complex orthologs vs *H. sapiens* Integrator complex subunits. Protein domains were searched using the Pfam database. The length of each subunit is indicated in amino acids (aa). Abbreviations: **DUF**, domain of unknown function; **HEAT**: Huntingtin, Elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1; **ARM**: armadillo-like repeats; **COIL**, coiled coil domain; **VWA**: von Willebrand type A domain. **β-lactamase/β-CASP** (*indicates the presence of an inactive β-lactamase/β-CASP domain); **RMMBL**: Zn-dependent metallo-hydrolase, RNA specificity domain; 7TM GPCR srh: seven-transmembrane G-protein-coupled receptor, serpentine receptor class h; PHD, plant homeodomain finger; **ISDCC**, INTS6/SAGE1/DDX26B/CT45 C- terminus; **TPR**, tetratricopeptide repeats.

The majority of predicted *C. elegans* Integrator subunits share the protein domains of their respective ortholog. The most common of these are α-helical repeats such as HEAT, ARM or VWA, which are usually involved in protein-protein interactions. Importantly, *C. elegans* INTS-11 and INTS-9 have domains of the β-Casp family, a large group of zinc-dependent nucleases.

The human Ints6 has a VWA domain in its Nt, and in its Ct it has a domain called ISDCC because it is found in Ints6 and also in sarcoma antigen 1 (SAGE1), protein DDX26B and members of the cancer/testis antigen family 45.
In *C. elegans* INTS-6, the VWA domain is conserved in its Nt but in its Ct, there is a COIL domain (765-785aa). A coiled coil is a structural motif in which 2–7 alpha-helices are coiled together like strands of a rope. Experimental studies have confirmed that COIL domains play a fundamental role in subcellular infrastructure, in the tethering of transport vesicles and in enzymes, where they function as molecular rulers, positioning catalytic activities at fixed distances (Truebestein & Leonard, 2016).

7.2. Immunoprecipitation of *C. elegans* INTS-6 and detection of its Co-IP interactors

To identify *C. elegans* Integrator complex subunits, we immunoprecipitated INTS-6 and detected its Co-immunoprecipitated protein partners by LC-MS/MS (Liquid Chromatography-Mass Spectrometry/Mass Spectrometry). The JCP378 strain (Si19[pJC56(*eft-3*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))]II;*unc-119(ed3)*III) was used to immunoprecipitate INTS-6 along with its protein partners. The N2 strain was used as a negative control. Worm extracts from both strains were immunoprecipitated using *anti-FLAG M2®Magnetics beads* (MM 5.7.2). Importantly, no detergents were used during the washing steps to avoid losing weak interactions. IPs from JCP378 and N2 worm extracts were eluted from the beads either by boiling the samples in SDS-PAGE sample buffer or by competitive elution with the 3xFLAG peptide.

Afterwards, both IPs were resolved by SDS-PAGE, stained with Coomassie Blue and analyzed by LC-MS/MS. This experiment was repeated four separate times. In addition, a small portion of the eluted IP (immunoprecipitation) from the JCP378 worm extracts was silver stained (Fig. 42).

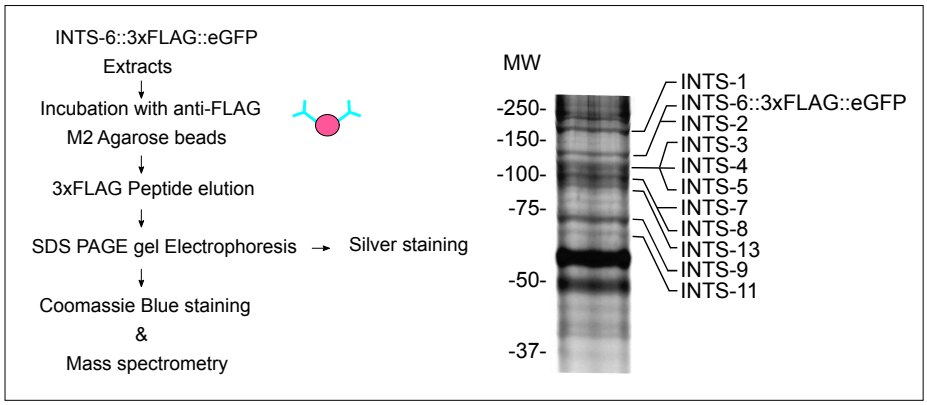


Figure 42. INTS-6 Immunoprecipitation process and silver staining. Protein extracts were obtained from the JCP378 strain. INTS-6 was immunoprecipitated using *α- FLAG M2® Magnetic beads*. On this occasion, the elution from the antibody was performed by competition with the 3xFLAG peptide. The eluted immunoprecipitate was run on SDS-PAGE electrophoresis that was stained with Coomassie Blue and analyzed by LC-MS/MS. In addition, a small amount of the immunoprecipitate was run on SDS-PAGE electrophoresis and silver stained. The Integrator complex members detected by mass spectrometry are indicated on the silver stained gel according to their expected molecular weight. Peptides were detected for all *C. elegans* Integrator subunit orthologs except for INTS-10 and INTS-12.

To examine the data obtained from the mass spectrometry analysis, all the proteins belonging to peptides detected in N2 IPs were deleted from the list of proteins detected in JCP378 IPs, thus eliminating all possible false positives. Proteins with at least two unique peptides and found in at least two IPs were considered a positive match.

All *C. elegans* Integrator subunit orthologs except INTS-10 and INTS-12 were found along with other proteins (Figs. 42 & 43).

Protein	ΣCoverage (%)	Σ# Unique Peptides	Σ# Peptides	Detected in X out of 4 IPs
INTS-1 (C06A5.1)	25.9	42	42	4
INTS-2 (ZC376.6)	21.9	24	24	4
INTS-3 (Y92H12A.4)	56.27	16	54	4
INTS-4 (W04A4.5)	15.87	13	13	4
INTS-5 (Y51A2D.7)	27.72	22	22	4
INTS-6/DIC-1	56.42	42	42	4
INTS-7 (D1043.1)	16.59	12	12	4
INTS.8 (Y48G10A.4)	29.89	24	24	4
INTS-9 (F19F10.12)	19.20	11	11	3
INTS-11 (F10B5.8)	2.80	2	2	4
INTS-13 (R02D3.4)	11.38	10	10	2
LAF-1	40.54	21	24	3
Y73F8A.26	51.68	19	19	2
R07E4.5	18.85	12	12	3
F37C4.5	11.51	6	6	3
C06G3.8	59.85	5	5	3
ASP-4	13.06	4	4	2
rpl-7A	13.06	3	3	3
Y56A3A.31	8.72	3	3	2
C02B10.4	15.17	2	2	3
RARS-1	3.90	2	2	2
ZK856.16	39.53	2	2	2

Table 5. Mass spectrometry analysis (LC-MS/MS) of four separate INTS-6 IPs. INTS-6 was immunoprecipitated from protein extracts of the JCP378 strain four separate times. Protein extracts from the N2 strain were immunoprecipitated equally and used as a negative control. IPs were eluted either by competition with the 3XFLAG peptide or by boiling them in SDS-PAGE sample buffer. The eluates were run on SDS-PAGE electrophoresis, stained with Coomassie Blue and analyzed by LC-MS/MS. Positive proteins were considered to be those with at least two unique peptides per immunoprecipitate and found in at least two IPs. For each protein detected, the following information is indicated: Σ coverage (%), Σ# unique peptides, Σ# peptides and the number of IPs in which they were detected. Eleven orthologs of *C. elegans* Integrator complex subunits were found, along with other proteins.

In preliminary experiments with INTS-6 IP we observed that there were only minor differences between the elutions methods used (boiling the samples in SDS-PAGE sample buffer, competitive elution with 3xFLAG peptide or under acidic conditions with 0.1 M glycine HCl, pH 3.0). However, if the IP washes were performed under astringent conditions (increasing the salts and using detergents) the only Integrator complex ortholog detected, apart from INTS-6, was INTS-3, suggesting that the interaction between INTS-3 and INTS-6 is strong.

7.3. Knockdown of *C. elegans* Integrator subunits

To study our proposed *C. elegans* Integrator complex, snRNA 3'-end processing was verified upon knockdown of the *C. elegans* Integrator complex members found by orthology and/or Co-IP.

Integrator complex orthologs were knocked down by RNAi feeding in worms starting with the L1 stage (MM 7.2). The phenotypes observed in the adult stage are shown in Figure 43.

Distinct phenotypes were observed among the different subunits and these fall into the following categories: a) worms depleted of INTS-2 and INTS-4, which arrested at L2-L3 stage, manifested the most aggressive phenotype, b) worms depleted of INTS-5, INTS-9 and INTS-11 that arrested at L3, L4 stage, c) worms depleted of INTS-1, INTS-6, INTS-7, INTS-8 and INTS-10, reached adult stage and laid eggs although all presented a percentage of embryonic lethality, and d) worms depleted of INTS-3, INTS-12 and INTS-13, no phenotype was observed. In addition, we could observe some malformation defects, such as protruding vulvas in worms depleted of subunits INTS-2, INTS-4, INTS-5, INTS-6, INTS-7 and INTS-8.

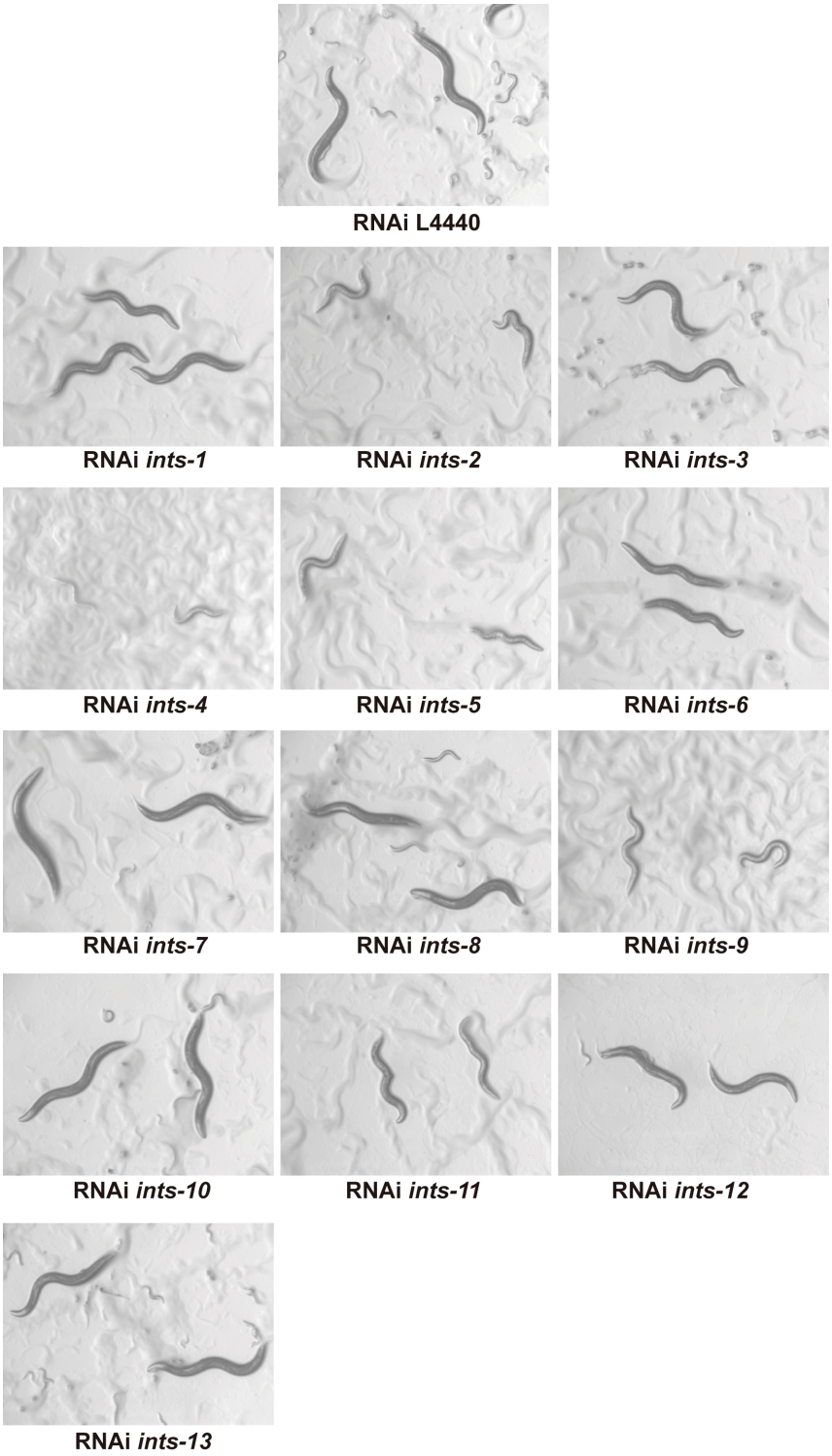


Figure 43. Phenotype of *C. elegans* Integrator complex predicted subunits after RNAi knockdown. N2 worms in L1 stage were fed bacterial RNAi clones of each predicted member of the Integrator complex by homology and/or Co-IP. Worms were grown at 15ºC and images were taken on the sixth day (adult stage). The figure shows that worms fed the bacterial RNAi clones of *ints-2* and *-4* arrested at the L2-L3 stage, whereas worms fed the bacterial RNAi clones of *ints-5*, *-9* and *-11* arrested at the L3, L4 stage. In contrast, worms fed the bacterial RNAi clones of the other subunits reached the adult stage.

Worms were collected on the sixth day. Total RNA was extracted using the *mirVana* ™ *miRNA isolation*. Later, RT-PCRs (Reverse Transcription Polymerase Chain Reaction) in the 3' region of snRNAs were performed. We chose five snRNA coding genes for this purpose based on the results obtained from the *t1903* mutant transcriptomic analysis: two U1 snRNAs (H27M09.8, F08H9.14), two U2 snRNAs (W04G5.11, F08G2.9) and one *sls-2* (*sls-2.8*).

All these genes showed strong snRNA misprocessing as well as RNAP II read-through that continued, resulting in transcription of the closest gene located *in sense*. Primer pairs were designed to amplify from the 3'-end of the snRNA until the closest gene located *in sense* (the forward primer on the snRNA and the reverse primer on the next gene). The RT-PCRs (Fig. 44) demonstrated that the knockdown of any *C. elegans* Integrator complex subunit resulted in snRNA 3'-end processing defects (though to different degrees) and the transcription reached the closest gene, suggesting the formation of chimeric "sn-mRNAs" as in the *t1903* mutant. The INTS-1 subunit showed weak defects. The INTS-3, INTS-10 and INTS-12 subunits only showed subtle defects (Fig. 44).

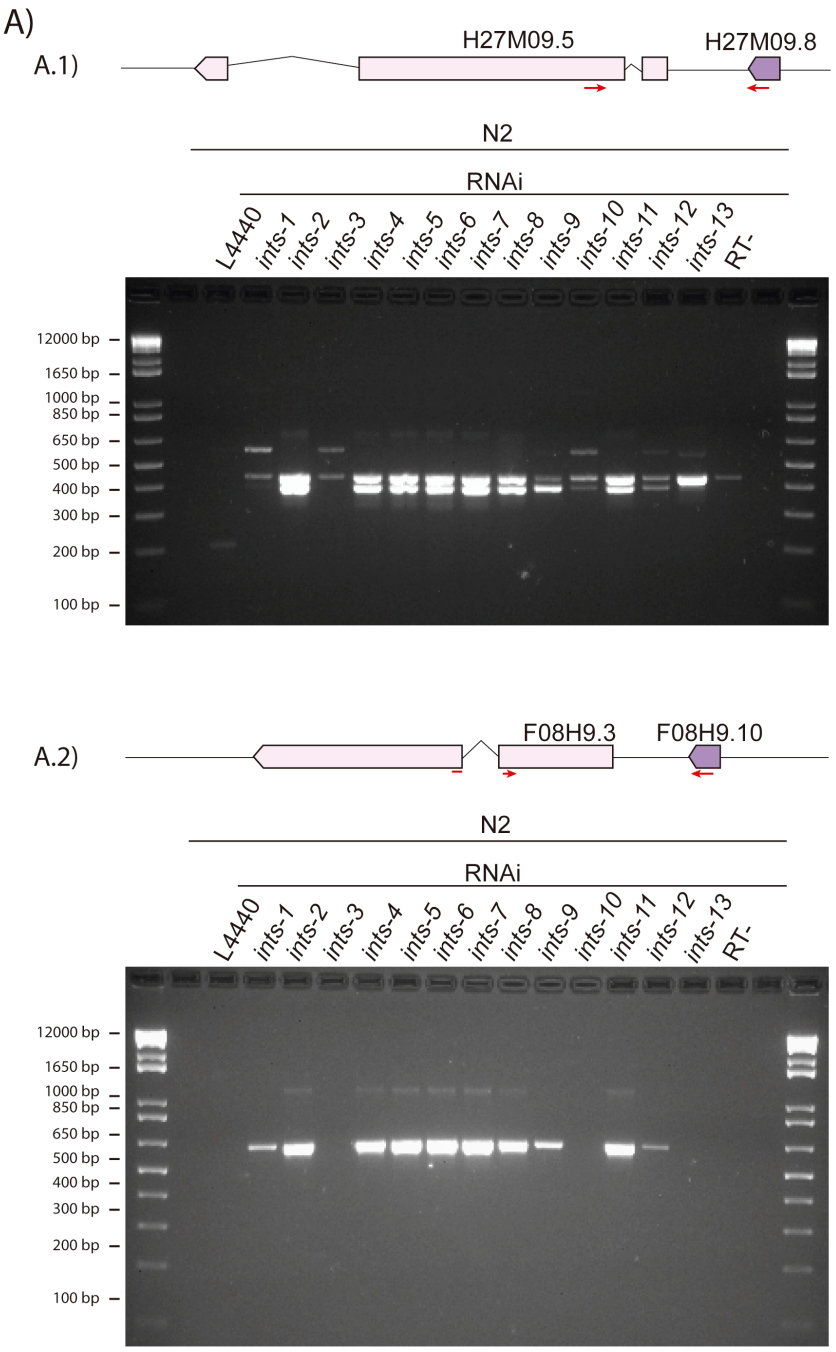


Figure 44A. Knockdown of *C. elegans* Integrator complex subunits results in 3'-end snRNAs processing defects. A) U1 type snRNAs. A1) Upper panel: schematic representation of the snRNA gene H27M09.8 gene and the downstream H27M09.5 gene. Red arrows represent the primer pairs used. Lower panel: Gel electrophoresis showing the PCRs that were run to amplify the region downstream of the snRNA H27M09.8. RNAi feeding with the empty control vector L4440 does not affect snRNA processing, whereas RNAi feeding with *ints-2*, *-4*, *-5*, *-6*, *-7*, *-8, -9* and *-11* abrogates 3'-end snRNA processing. RNAi feeding with *ints-10* and *-12* also abrogates 3'-end snRNA processing, though to a lesser degree. PCR expected length: 396 bp (genomic: 439 bp)**. A2)** Upper panel: schematic representation of the snRNA F08H9.10 gene and the downstream F08H9.3 gene. Red arrows represent the primer pairs used. Lower panel: Gel electrophoresis showing the PCRs performed to amplify the region downstream of the snRNA F08H9.10. RNAi feeding with the empty control vector L4440 does not affect snRNA processing whereas RNAi feeding with *ints-1*, *-2*, *-4*, *-5*, *-6*, *-7*, *-8, -9, -11* and *-12* abrogates 3'-end snRNA processing. PCR expected length: 637 bp.

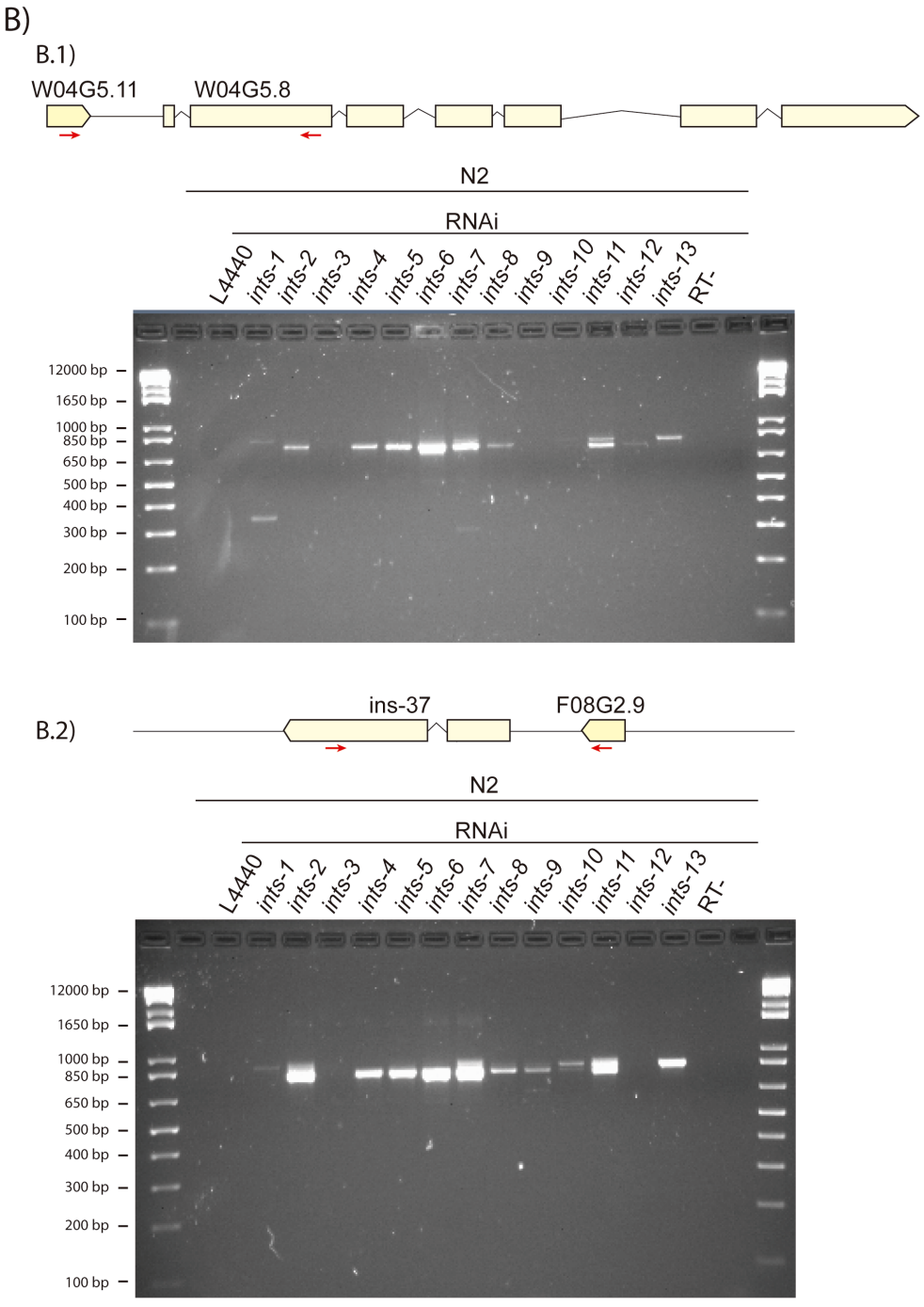


Figure 44B. Knockdown of *C. elegans* Integrator complex subunits resulted in 3'-end snRNAs processing defects B) U2 type snRNAs. B1) Upper panel: schematic representation of the W04G5.11 snRNA gene and the downstream W04G5.8 gene. Red arrows represent the primer pairs used. Lower panel: Gel electrophoresis showing the PCRs performed to amplify the region downstream of the snRNA W04G5.11. RNAi feeding with the empty control vector L4440 does not affect snRNA processing whereas RNAi feeding with *ints-2*, *-4*, *-5*, *-6*, *-7*, *-8*, *-11* and *-12* bacterial RNAi clones abrogates 3'-end snRNA processing. PCR expected length: 782 bp (genomic: 840 bp). **B2)** Upper panel: schematic representation of the snRNA gene F08G2.9 and the downstream gene *ins-37*. Red arrows represent the primer pairs used. Lower panel: Gel electrophoresis showing the PCRs that were run to amplify the region downstream of the snRNA F08G2.9. RNAi feeding with the empty control vector L4440 does not affect snRNA processing whereas RNAi feeding with *ints-2*, *-4*, *-5*, *-6*, *-7*, *-8, -9, -10, -11* and *-12* abrogates 3'-end snRNA processing. PCR expected length: 804 bp (genomic: 856 bp).

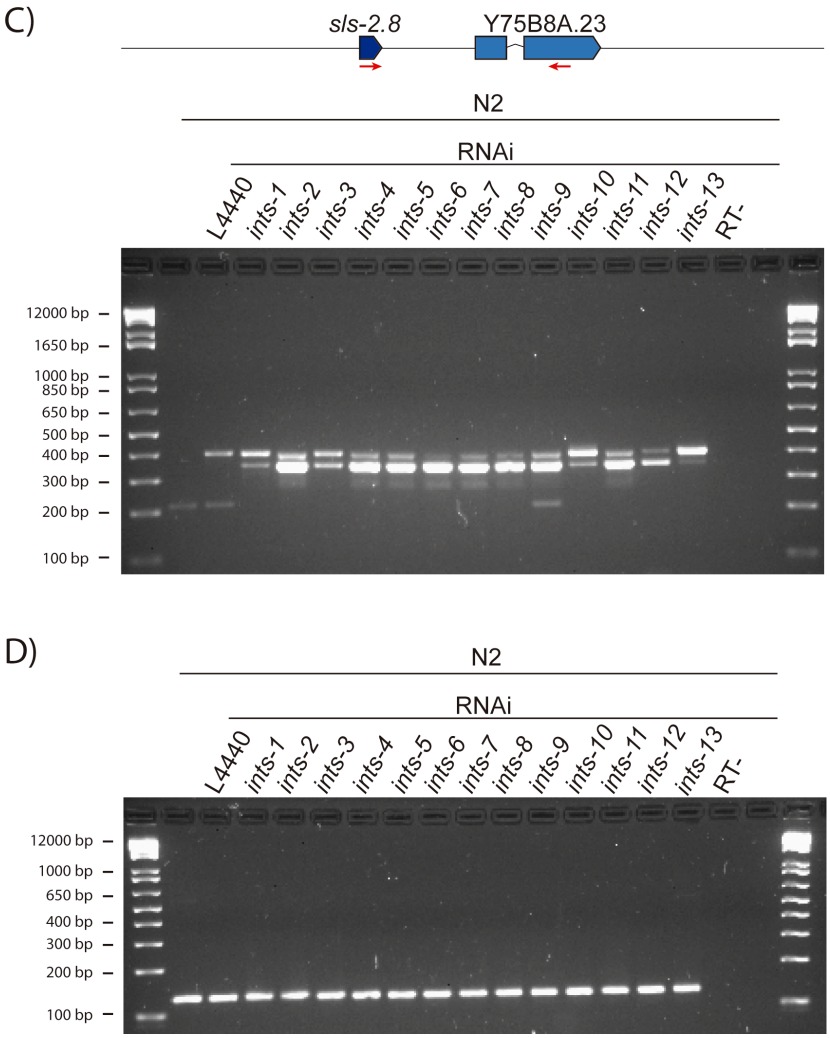


Figure 44C and D. Knockdown of *C. elegans* Integrator complex subunits results in 3'-end snRNAs processing defects C) SL type snRNAs. Upper panel: schematic representation of the *sls-2.8* gene snRNA and the downstream gene Y75B8A.23. Red arrows represent the primer pairs used. Lower panel: Gel electrophoresis showing the PCRs that were run to amplify the region downstream of the snRNA *sls-2.8*. RNAi feeding with the empty control vector L4440 does not affect snRNA processing whereas RNAi feeding with any member of the Integrator complex abrogates 3'-end snRNA processing, though to different degrees. RNAi feeding with *ints-1*, *-3*, *-10* and *-13* show low 3'-end processing defects. PCR expected length: 358 bp (genomic 410 bp) **D) Actins.** Actins were used as a control to corroborate that the amount of cDNA among the samples was equal.

In addition, we decided to validate the results obtained from the RT-PCRs by performing RiboMinus RNA deep sequencing. This rmRNA-seq analysis would allow us confirm the results of Integrator subunits' function in 3'-end processing of the snRNAs, the formation of chimeric "sn-mRNAs" and very importantly, it would give us information about possible gene regulatory roles of the Integrator subunits in *C. elegans* or a hint about other biological roles, which would open new lines of research in our lab. An example of the rmRNA-

seq reads visualized on IGV software is shown on Fig. 45 on the region of the U1: H27M09.6 and the U2: W04G5.11. As shown below, there are 3'-end processing defects on the knockdown of Integrator genes: *ints-1*, *-2*, *-4*, *-5*, *-6*, *-7*, *-8*, *-9*, and *ints-11.* Upon the depletion of these subunits, transcription continues and the closest gene located *in sense* is properly spliced and overexpressed.

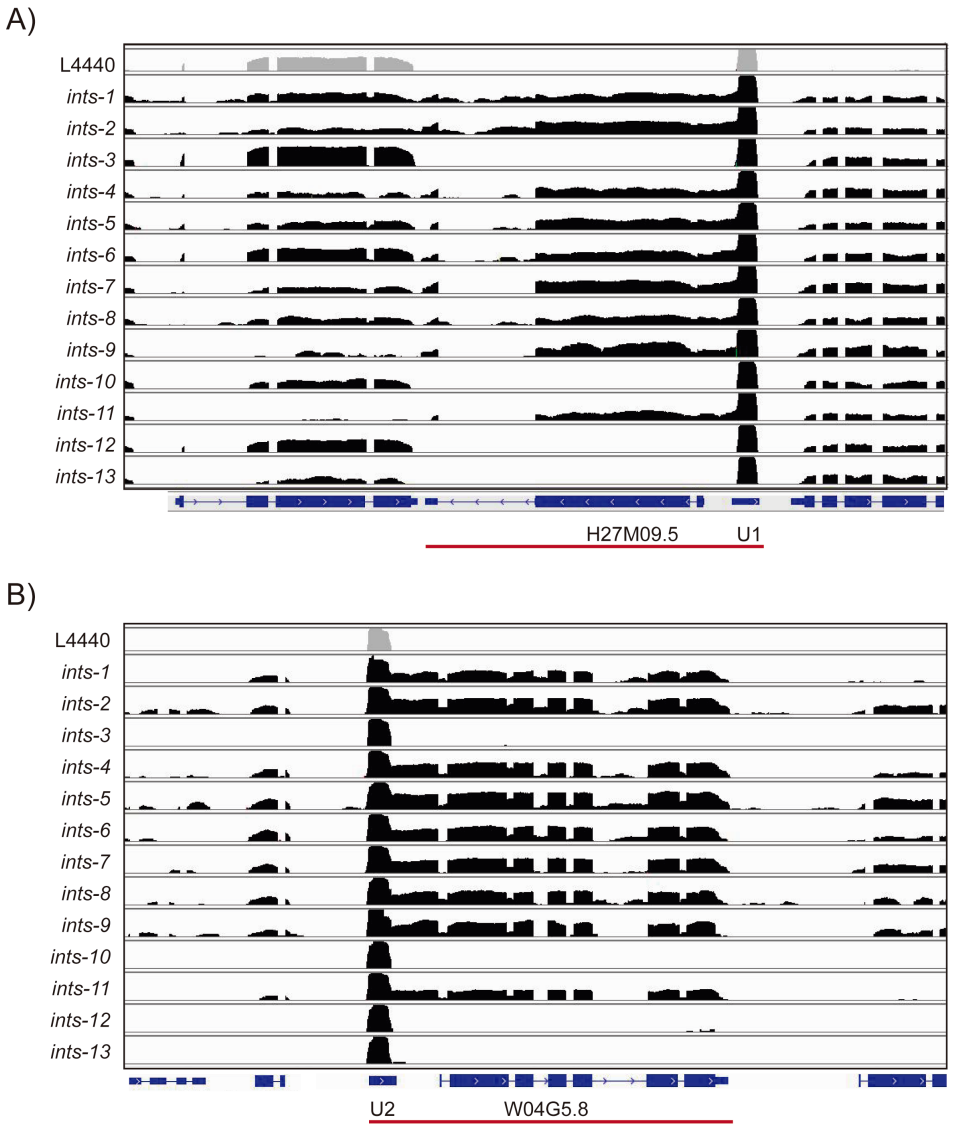


Figure 45. RiboMinus RNA deep sequencing (rmRNA-seq) shows snRNA 3'-end processing defects upon Integrator complex knockdown. The figure shows two examples of RiboMinus RNA. The deep sequencing reads aligned to the *C. elegans* genome in the U1: H27M09.6 (in A) and the U2: W04G5.11 (in B) regions are visualized on IGV software. Reads from the RNA of worms fed the bacterial RNAi clones of the empty L4440 vector are represented in gray. Reads from RNA of worms fed the bacterial RNAi clones of each Integrator subunit are shown in black. Underneath each panel, the *C. elegans* genome is shown in blue, the exons are represented as blue boxes and the introns as lines. As shown in the figure, RNAi knockdown of the Integrator genes *ints-1, -2*, *-4*, *-5*, *-6*, *-7*, *-8*, *-9* and *-11* leads to snRNA misprocessing, but no 3'-end processing defects were observed upon RNAi

knockdown of the Integrator genes ints-*3*, *-10*, *-12* and *-13*. In all the cases where we observed defects related to 3'-end cleavage, the transcription continues and the closest gene located *in sense* is overexpressed (indicated by a red underline). In addition, it is clear that the splicing is mostly correct in all cases because the reads mapped to the exons as in the L4440 control.

8. Are the chimeric RNAs translated into proteins or peptides?

We hypothesized that the multiple chimeric RNAs formed upon knockdown of any *C. elegans* Integrator complex subunit could be translated. We observed that the genes downstream of the snRNAs within the chimeric RNAs were properly spliced, polyadenylated and very importantly, that they have the typical structure of a translatable mRNA.

In *C. elegans*, translation initiation occurs at an AUG codon, but as in other metazoans, there is a consensus in the preceding nucleotides where *C. elegans* prefers A residues at each of the four positions preceding the AUG ("AAAAAUG") (Riddle et al., 1997). We observed that the mRNAs within the chimeric RNAs, which are not expressed under normal *C. elegans* growing conditions, had "A" residues before the AUG codon.

We assumed that these chimeric "sn-mRNAs" had a 5'cap, either m^7G or TMG, because it is a continuation of the transcription from the snRNA that already has a cap. To decipher if the chimeric RNAs could be translated, different transgenic lines were created to add the eGFP and/or the 3xFLAG tag to the Ct of the genes overexpressed downstream of the snRNAs. Thus, we could detect the formation of proteins by WB technique.

First, the eGFP and/or the 3xFLAG tag were fused in frame to the Ct region of the mRNA genes downstream of the previously chosen snRNAs: two were downstream of U1 (H27M09.5, F08H9.3), another two were downstream of U2 (W04G5.8, *ins-37*) and one was downstream of the *sls-2.8* (Y75B8A.23). A schematic drawing of the plasmids generated is represented in Figure 46 (full plasmids in M.M. 4.10.2.2). These plasmids were integrated into chromosome II by the mosSCI system (Frokjaer-Jensen et al., 2008), generating the strains JCP387, JCP405, JCP343, JCP301 and JCP394 respectively.

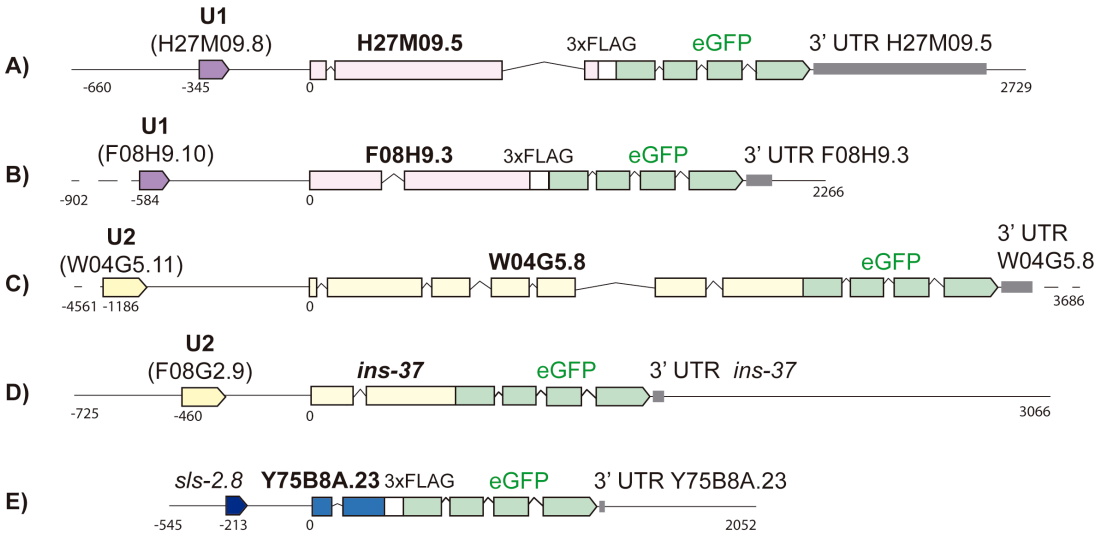


Figure 46. Scheme of the plasmids made and integrated by the mosSCI system in chromosome II to check for the possible formation of proteins.
A) Plasmid pJC57 shows a genomic region amplified upstream of the U1 H27M08.9 until the downstream region of the H27M09.5 gene. The 3xFLAG::eGFP tag was inserted in-frame downstream of the last exon of the H270M9.5 gene. This plasmid was used to generate the JCP387 strain.
B) Plasmid pJC60 shows a genomic region amplified upstream of the U1 F08H9.10 until the downstream region of the F08H9.3 gene. The 3xFLAG::eGFP tag was inserted in-frame downstream of the last exon of the F08H9.3 gene. This plasmid was used to generate the JCP405 strain.
C) Plasmid pJC55 shows a genomic region amplified upstream of the U2 W04G5.11 until the downstream region of the W04G5.8 gene. The eGFP tag was inserted in-frame downstream of the last exon of the W04G5.8 gene**.** This plasmid was used to generate the JCP343 strain.
D) Plasmid pJC50 shows a genomic region amplified upstream of the U2 F08G2.9 until the downstream region of the *ins-37* gene. The eGFP tag was inserted in-frame downstream of the last exon of the *ins-37* gene. This plasmid was used to generate the JCP301 strain.
E) Plasmid pJC58 shows a genomic region amplified upstream of the *sls-2.8* until the downstream region of the gene Y75B8A.23. The 3xFLAG::eGFP tag was inserted in frame downstream the last exon of the Y75B8A.23 gene. This plasmid was used to generate the JCP394 strain.

We performed RNAi feeding of each *C. elegans* Integrator complex subunit in each of the transgenic strains generated to check whether the chimeric RNAs formed, were translated into proteins. Worms were synchronized and fed the corresponding bacterial RNAi clone beginning at the L2 stage. Adult stage worms were collected and proteins were extracted. Next, WBs against the 3xFLAG tag or the GFP tag were run to check for the possible protein

formation (Fig. 47). The OP217 strain (ddIs172 [*aly-2*::TY1::eGFP::3xFLAG + *unc-119* (+)]) was used as a positive control of the WB to detect the 3xFLAG tag or the eGFP tag.

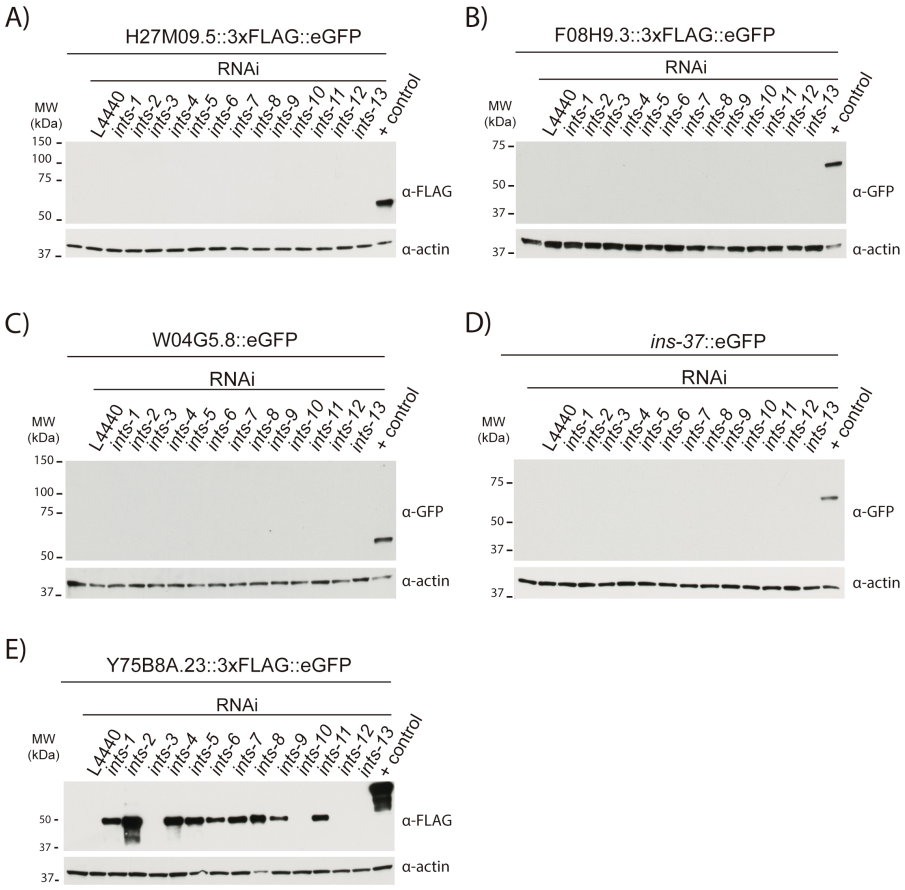


Figure 47. WBs to check if chimeric "sn-mRNAs" are translated into proteins. The figure shows five WBs, labeled A to E, of protein extracts from worms of the transgenic strains JCP387, JCP405, JCP343, JCP301 and JCP394, respectively. Worms from each transgenic strain background were synchronized and starting at the L2 stage, they were fed bacterial RNAi clones of the *C. elegans* Integrator complex subunits and the empty L4440 vector. They were grown at 20ºC and collected at the adult stage. Next, proteins were extracted followed by WBs against the GFP or the 3xFLAG tag to verify whether the mRNAs within the chimeric RNAs were translated into proteins. Only the Y75B8A.23 gene (WB E, JCP394 strain), which is preceded by the *sls-2.8* snRNA, was translated into a protein when the Integrator subunits INTS-1, -2, -4, -5, -6, -7, -8, -9, -11 were depleted. Estimated molecular weights: A) H27M09.5::3xFLAG::eGFP, 68.8kDa; B) F08H9.3::3xFLAG::eGFP, 46.6 kDa; C) W04G5.8::eGFP, 72.7kDa; D) *ins-37*::eGFP, 45.6kDa; E) Y75B8A.23::3xFLAG::eGFP, 39.5kDa; positive control: 54.3 kDa.

Lack of 3'-end processing of U1 and U2 snRNAs in the assayed transgenes caused by depletion of the Integrator complex led to the transcription of long chimeric RNAs containing the snRNA, the intergenic region and the

downstream tagged-genes properly spliced and polyadenylated (data not shown). However, these U1 and U2 derived chimeric RNAs were not translated into proteins as determined by WB (Fig. 47).
Only the Y75B8A.23 gene, which is preceded by an SL snRNA, was translated into a protein when the Integrator subunits INTS-1, -2, -4, -5, -6, -7, -8, -9, -11 were depleted (Fig. 47 E).
The Y75B8A.23 gene is uncharacterized. It is only found in the *Caenorhabditis* phylum and RNA sequencing studies indicate that this gene is enriched in the germline, DTC of the gonad male and the amphid sheath cell, a cell of the nervous system located in the pharynx (www.wormbase.org).

We observed that, there were not ATG codons in either the *sls-2.8* or the intergenic region until the Y75B8A.23 gene. Under these circumstances, when the Integrator complex is depleted, the first start codon that could be used to initiate translation of the long chimeric RNA is the initial ATG of the Y75B8A.23 gene. In contrast, the U1 and U2 snRNA genes contained several ATGs in their sequences and also in the intergenic region.
Therefore, we hypothesized that prior ATGs within the snRNA or the intergenic region could be acting as the start codon to translate peptides although the most probable ATG to initiate translation was the first one of each mRNA.

To determine whether other ATGs in these long chimeric “sn-mRNAs” could serve as the start codon for translating peptides, two distinct transgenes were generated which contained a cassette with the HA, MYC (+1b) and TY (+2b) tags in each of the 3 open reading frames (ORFs). Thus, the HA tag and the MYC tag were fused in-frame respectively with the first and the second ATG of the U1 snRNA gene (F08H9.3) and the U2 snRNA (W04G5.8). A schematic drawing of this and the constructs generated is shown in Figure 48. Full plasmids are shown in MM. 4.10.2.2
The pJC63 and pJC64 plasmids were integrated using the mosSCI system into chromosome II, generating the transgenic strains JCP479 and JCP504, respectively.

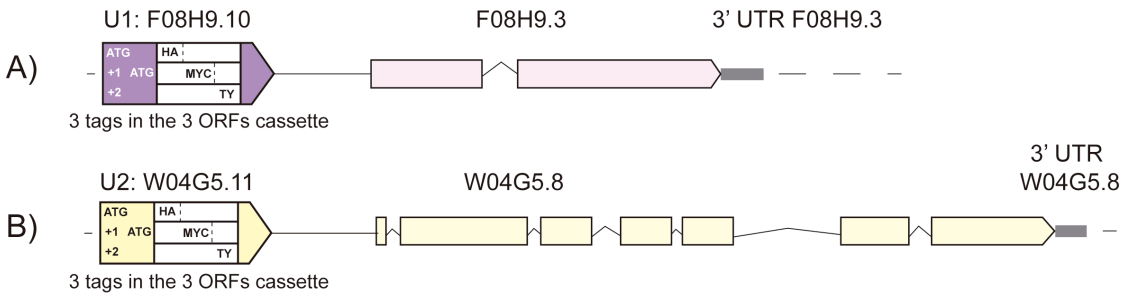


Figure 48. Scheme of the plasmids made and integrated into chromosome II, using the mosSCI system, to check for the possible formation of peptides. A) Plasmid pJC63 shows a genomic region amplified upstream of U1 F08H9.10 until the downstream region of the F08H9.3 gene. A cassette with the HA, MYC (+1b) and TY (+2b) tags in each of the 3 ORFs was inserted into U1 F08H9.3 so that the HA tag and the MYC tag were in-frame respectively with the first and the second ATGs of the U1 F08H9.3. This plasmid was used to generate the JCP479 strain. **B) Plasmid pJC64** shows a genomic region amplified upstream of U2 W04G5.11 until the downstream region of the W04G5.8 gene. The same cassette with the HA, MYC (+1b) and TY (+2b) tags in each of the 3 ORFs was inserted into the U2 W04G5.11. This plasmid was used to generate the JCP504 strain.

RNAi feeding of the Integrator complex members *ints-2*, *ints-6* and *ints-11* was performed in both transgenic strains to check whether the chimeric "sn-mRNAs" formed could be translated into peptides. The INTS-2, INTS-6 and INTS-11 subunits were chosen because their depletion produced a strong misprocessing. Therefore, they were good candidates to check the formation of peptides, though we expected the same result with other Integrator members that also produced defects in snRNA processing upon their knockdown.

Worms were synchronized and fed the corresponding RNAi starting at L2 stage. They were collected when they reached the adult stage and proteins extracts were performed. Next, WBs against the HA tag, the MYC tag and the TY tag were run to check for the possible formation of peptides (Fig. 30). The OH1117 (otEx4963 [lsy-6p(fosmid-delta 150 bp downstream)::GFP + *ttx-3*:mCherry]. otIs306 [*hsp-16.2*::*che-1*::3xHA + rol-6]), SJ4199 (zcIs40[*dve-1*p::*dve-1*::3xMYC-HIS tag + *myo-3*p::GFP]) and OP217 strains were used as positive controls of in the WB for the HA, MYC and TY tags respectively.

No peptide formation was detected.

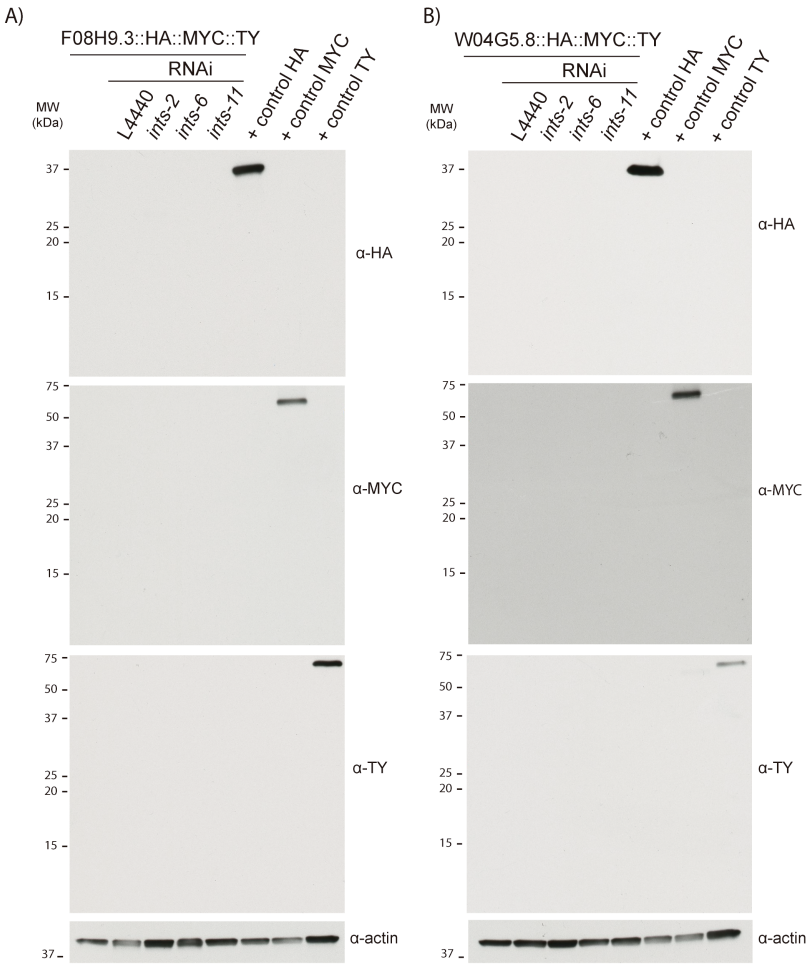


Figure 49. WBs to check if the chimeric "sn-mRNAs" are translated into peptides. Worms in each transgenic strain background (JCP479 and JCP504) were synchronized and starting at the L2 stage, they were fed bacterial RNAi clones of the *C. elegans* Integrator complex subunits *ints-2*, *-9* and *-11* and the empty L4440 vector. They were grown at 20ºC and collected at the adult stage. Next, proteins were extracted followed by WBs against the HA, MYC or TY tags to check whether the mRNAs within the chimeric RNAs are translated into proteins or not. No peptide formation was observed. Expected molecular weights: JCP479/JCP504 for HA (1^{ST} ORF): 7 kDa; MYC (2^{nd} ORF): 8.7 kDa; TY (3^{rd} ORF): 5.5 kDa. Positive control HA: 28.4 kDa; positive control MYC: 61.5 kDa; positive control TY: 54.3 kDa.

9. Does *ints-6* function in DNA repair in *C. elegans*?

Our data from INTS-6 immunoprecipitation experiments (Table 5) suggested that a possible *NABP1-INTS complex,* composed of at least INTS-6 (F08B4.1), INTS-3 (Y92H12A.4) and NABP1/2 (C06G3.8), is also present in *C. elegans*. The *NABP1-INTS complex* plays a key role in DNA repair by the HR pathway in humans (Skaar et al., 2009; Zhang et al., 2013).

To verify the involvement of *C. elegans ints-6* in DNA repair by the HR pathway we checked for RAD-51 IRIF formation (irradiation induced foci) in response to gamma radiation in the gonadal mitotic region upon *ints-6* knockdown. The RAD-51 protein plays a central role in the repair of DSBs by the HR pathway, RAD-51 is a recombinase that catalyzes homology recognition and mediates the strand invasion between homologous DNA molecules to allow re-synthesis of the damaged region (Lemmens & Tijsterman, 2011) (Figs. 15 and 21 Introduction section).

L1 stage worms were fed the bacterial RNAi clone of *ints-6* and the bacterial RNAi clone of the empty vector L4440. At the L4 stage, worms were exposed to gamma radiation (90 Gy). Twenty-four hours after irradiation, worms depleted of INTS-6 showed dramatically reduced progeny compared to the control, which suggests their inability to repair DNA damage (Fig. 50).

Gonads were dissected and immunostained with the RAD-51 antibody to check RAD-51 IRIF formation (MM 5.1.2) (Fig. 51). RAD-51 could be detected following IR (irradiation) in the gonadal mitotic region of worms fed with bacterial RNAi clone of the empty vector. However, no RAD-51 IRIF were detected in the mitotic region of the gonads depleted of INTS-6 (Fig. 51), which supports the hypothesis that INTS-6 is a key component of the *C. elegans* DNA damage response pathway and that it plays an essential role in RAD-51 recruitment to DSBs and DNA repair by the HR pathway.

To further investigate the function of *ints-6* in the DNA damage response, immunostaining with pTyr15 Cdk-1 antibody was performed (Fig. 52).

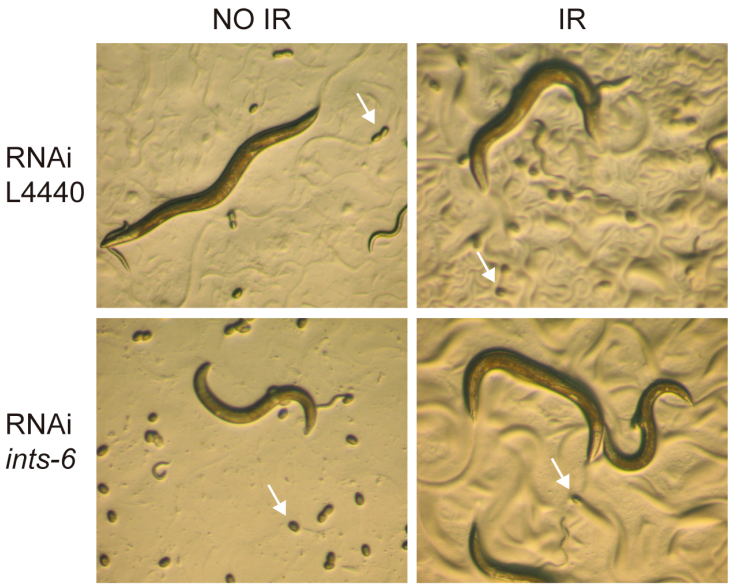


Figure 50. Phenotype of N2 worms knocked down for *ints-6* under normal growing conditions and 24h after IR (90Gy). N2 worms were synchronized and starting at the L1 stage, they were fed the *ints-6* bacterial RNAi clone and the empty L4440 vector bacterial RNAi clone that was used as the control. Worms were IR (90 Gy) at the L4 stage. Images were taken 24 hours later. Following irradiation, it is clear that the worms fed the RNAi *ints-6* bacterial clone hardly laid any eggs. Arrows mark the eggs.

In response to DNA damage, cells activate cell cycle checkpoints to maintain cells arrested in the G1, S or G2/M phase in order to allow the cell time to repair its DNA damage. Cdk-1 is tightly controlled at the G2/M phase where its activation is necessary for entry into mitosis. Cdk-1 activation involves the removal of inhibitory phosphorylations on Tyr15 and Thr14, added earlier in the cycle. Because these residues are within the ATP-binding domain, they inactivate the kinase activity of Cdk1 if phosphorylated. Therefore, after DNA damage, cells increase pTyr15 CDK-1 levels to induce cell cycle arrest at G2/M (Parker et al., 1992; Parker & Piwnica-Worms, 1992; Mueller et al., 1995, Malumbres & Barbacid, 2009).

N2 worms were synchronized and starting at the L1 stage, they were fed the *ints-6* bacterial RNAi clone and the empty L4440 vector used as the control. At the L4 stage, worms were irradiated (90Gy) and twenty-four hours later, gonads of irradiated and non-irradiated worms were dissected, fixed, immunostained using pTyr15 Cdk1 antibody and counterstained with DAPI (MM 5.1.2).

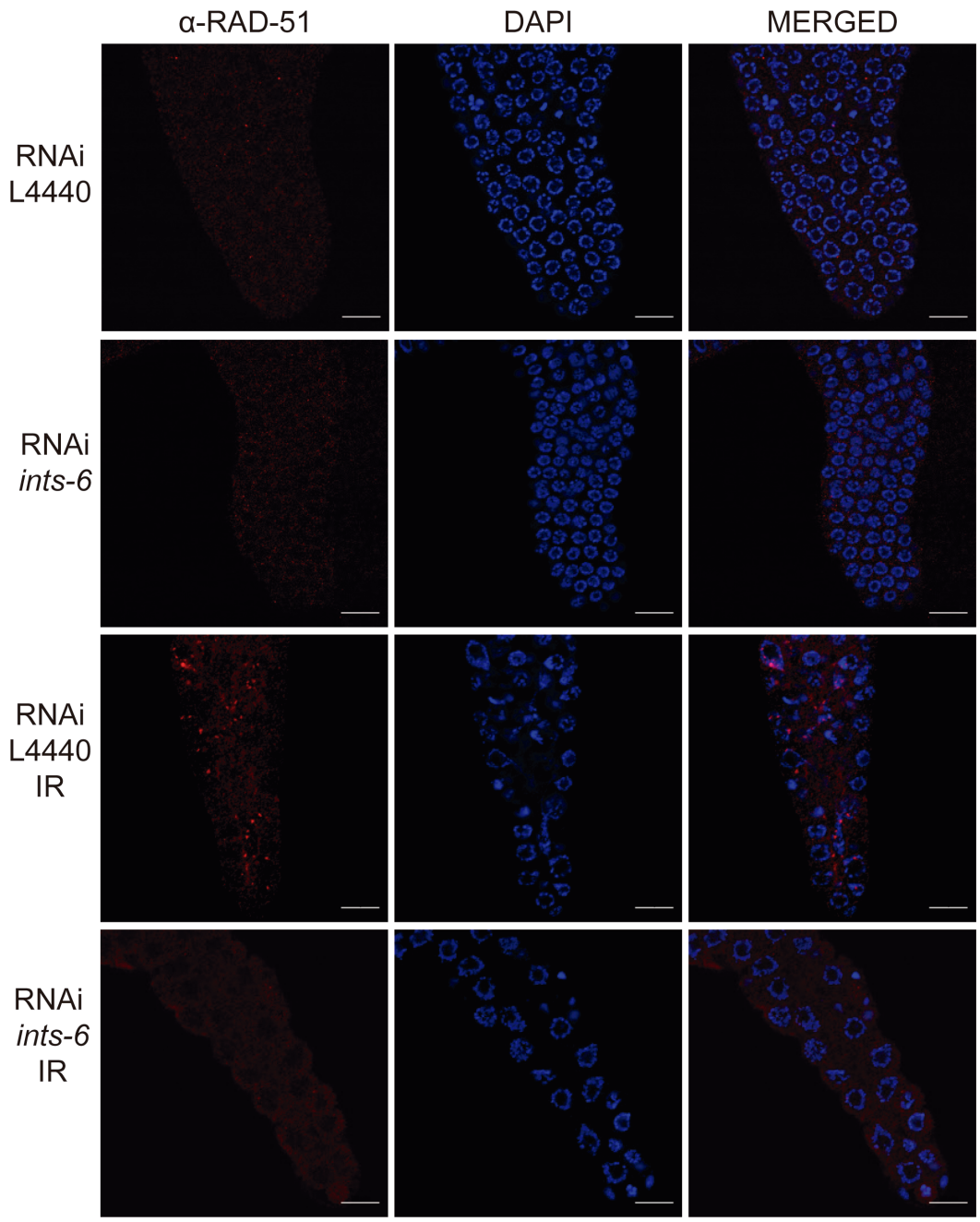


Figure 51. *ints-6* knockdown impairs RAD-51 recruitment to DSBs following IR. N2 worms were synchronized and starting at the L1 stage, they were fed the *ints-6* bacterial RNAi clone and the empty L4440 vector bacterial RNAi clone used as the control. At the L4 stage, worms were irradiated (90Gy). After 24 hours, gonads of irradiated and non-irradiated worms were dissected, fixed, immunostained with α-RAD-51 and counterstained with DAPI. As shown in the pictures, there are no RAD-51 foci before IR, either in worms fed the L4440 bacterial RNAi clone or fed the *ints-6* bacterial RNAi clone. Following irradiation, RAD-51 foci can be observed in the gonads of worms treated with the L4440 bacterial RNAi clone but no RAD-51 foci are observed in the mitotic region of gonads belonging to worms treated with the bacterial RNAi clone of *ints-6*. Scale bar: 10 μm.

Images of the gonadal mitotic region, where the only proliferative cells are located, were taken. As shown in Fig. 52, Tyr15 CDK-1 phosphorylation was not detected before irradiation either in the control or in the gonads knocked down for *ints-6*. Following IR, phosphorylation of Tyr15 CDK-1 was clearly detected in the nuclei of control gonads, as expected. However, worms fed the bacterial RNAi clone of *ints-6* did not induce phosphorylation on Tyr15 CDK-1.

Experiments demonstrating RAD-51 defective recruitment and defective CKD-1 Tyr-15 phosphorylation following DNA damage in RNAi *ints-6* knockdown worms clearly indicate that *ints-6* plays a key role in the DNA damage response.

Moreover, the defect to induce CDK-1 Tyr-15 phosphorylation in response to DNA damage in worms depleted of INTS-6 suggests that the DNA damage induced G2/M checkpoint could be affected in the absence of INTS-6. Therefore, the cells might not be totally arrested when *ints-6* is knocked down.

In *C. elegans,* cell cycle arrest within the gonadal mitotic region is evidenced by fewer nuclei that are larger in size (Gartner et al., 2004). To check cell cycle progression in worms at L4 stage with an *ints-6* knockdown background (under normal conditions and after DNA damage), the number of nuclei within the mitotic region of the gonads (Fig. 53) as well as the number of dividing nuclei within the gonad (Fig. 54), were quantified. N2 worms were synchronized and at the starting L1 stage they were fed the bacterial RNAi clone of *ints-6* or the empty L4440 vector bacterial RNAi clone. Next, worms were irradiated (90Gy) during the L4 stage and twenty-four hours later, gonads were dissected, fixed, immunostained with phosphor-histone H3 (Ser10) (pHH3) antibody and counterstained with DAPI. Ser 10 phosphorylation of histone 3 is used as a mitotic marker because it is only phosphorylated on Ser 10 during mitosis (Hendzel et al., 1997).

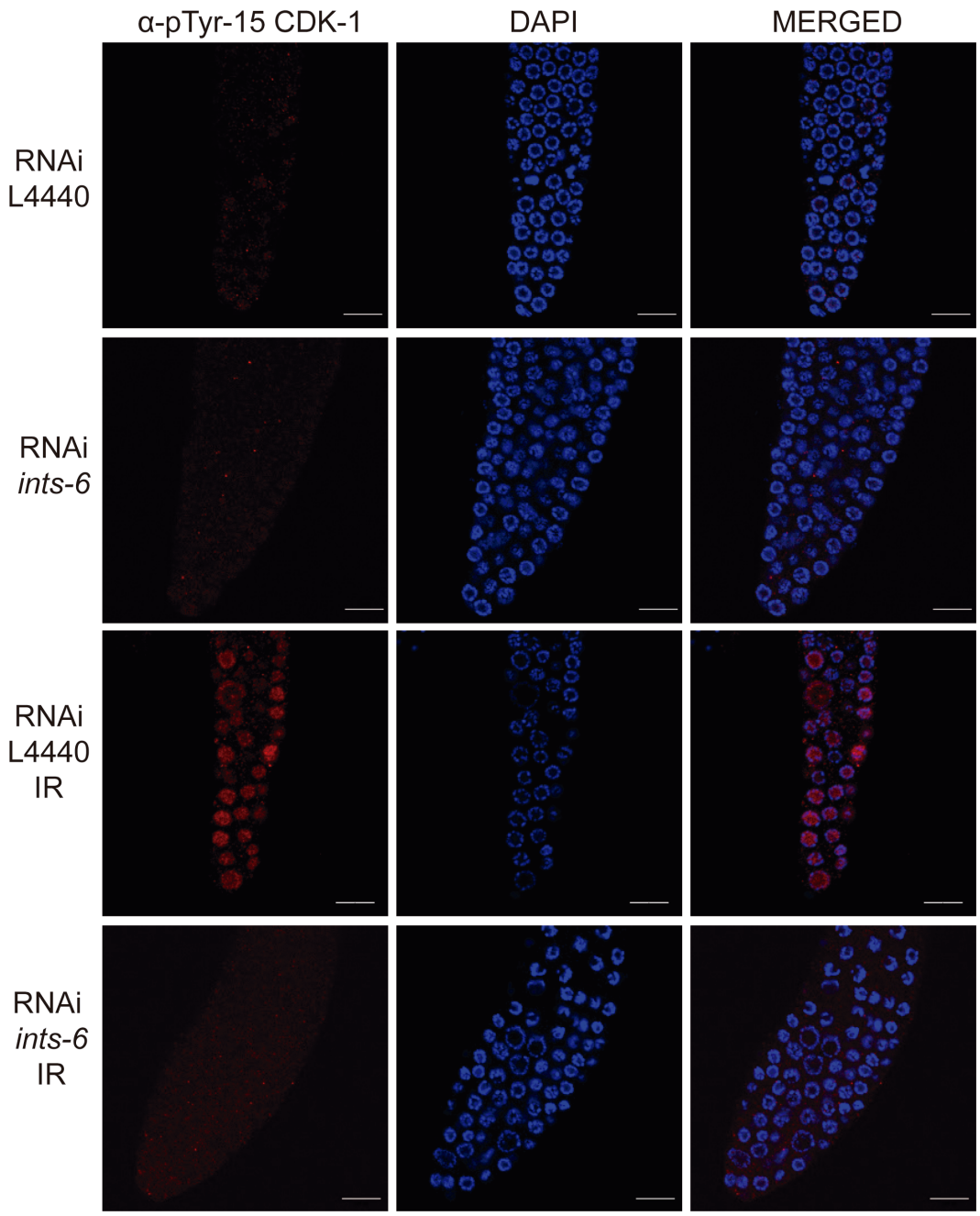


Figure 52. *ints-6* knockdown abrogates Tyr15 CDK-1 phosphorylation in response to DNA damage. N2 worms were synchronized and beginning at the L1 stage, they were fed the *ints-6* bacterial RNAi clone and the bacterial RNAi clone of the empty L4440 vector that was used as a control. At the L4 stage, worms were irradiated (90Gy). After 24 hours, gonads of irradiated and non-irradiated worms were dissected, fixed, immunostained with α-pTyr15 Cdk-1 and counterstained with DAPI. Images of the mitotic region were taken. As shown in the pictures, Tyr15 CDK-1 phosphorylation was detected in the nuclei of control gonads after irradiation. However, in the gonads of worms knocked down for *ints-6*, phosphorylation of Tyr15 CDK-1 following IR was not detected. Scale bar: 10 μm.

Images of the gonadal mitotic region were acquired in Z-stack and the number of nuclei, stained with DAPI, was manually counted. As observed in Fig. 53, the number of cells is significantly lower following IR both in worms knocked down for *ints-6* and in control worms. Therefore, proliferating cells within the gonadal mitotic region knocked down for *ints-6* are able to arrest in response to IR, just like cells in the control gonads.

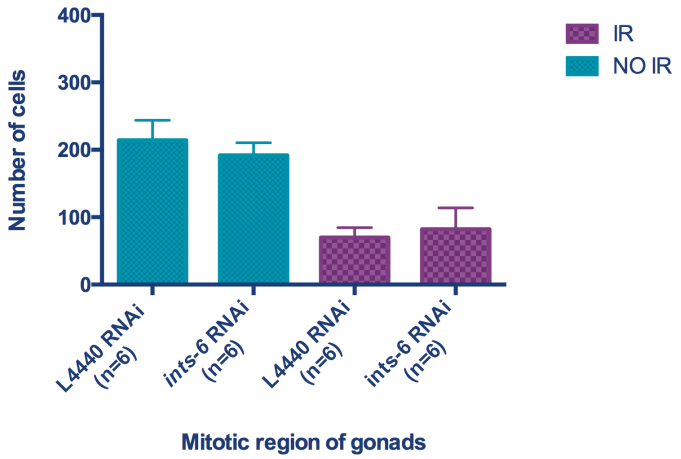
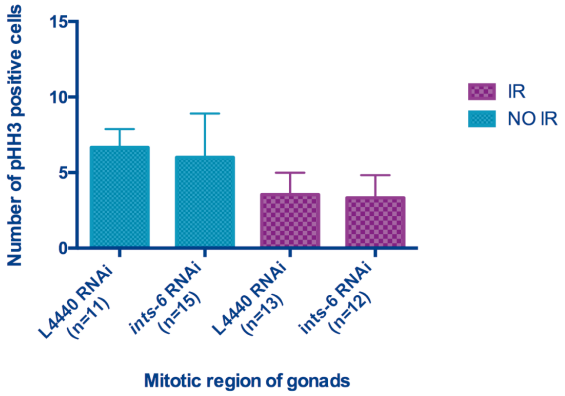


Figure 53. Quantification of nuclei within the gonadal mitotic region in *ints-6* RNAi and L4440 RNAi backgrounds in non-irradiated worms and 24 hours following gamma radiation (Mean ± SD). L1 stage N2 worms were fed the *ints-6* bacterial RNAi clone or the empty L4440 vector bacterial RNAi clone. At the L4 stage, worms were exposed to gamma radiation (90 Gy). Twenty-four hours later, gonads (n=6 per condition and genetic background) were dissected, fixed and stained with DAPI. Images were acquired in Z-Stack and afterwards the nuclei were manually counted. The number of nuclei is significantly lower following IR in both cases (in the control gonads and in the gonads depleted of INTS-6).

In addition, the number of mitotic cells was lower following IR in both cases (Fig.54), supporting the result observed in Fig. 53. It seems that germline proliferating cells in an *ints-6* knockdown genetic background are able to arrest, although they do not induce Cdk-1 phosphorylation on Tyr15.

A)



B)

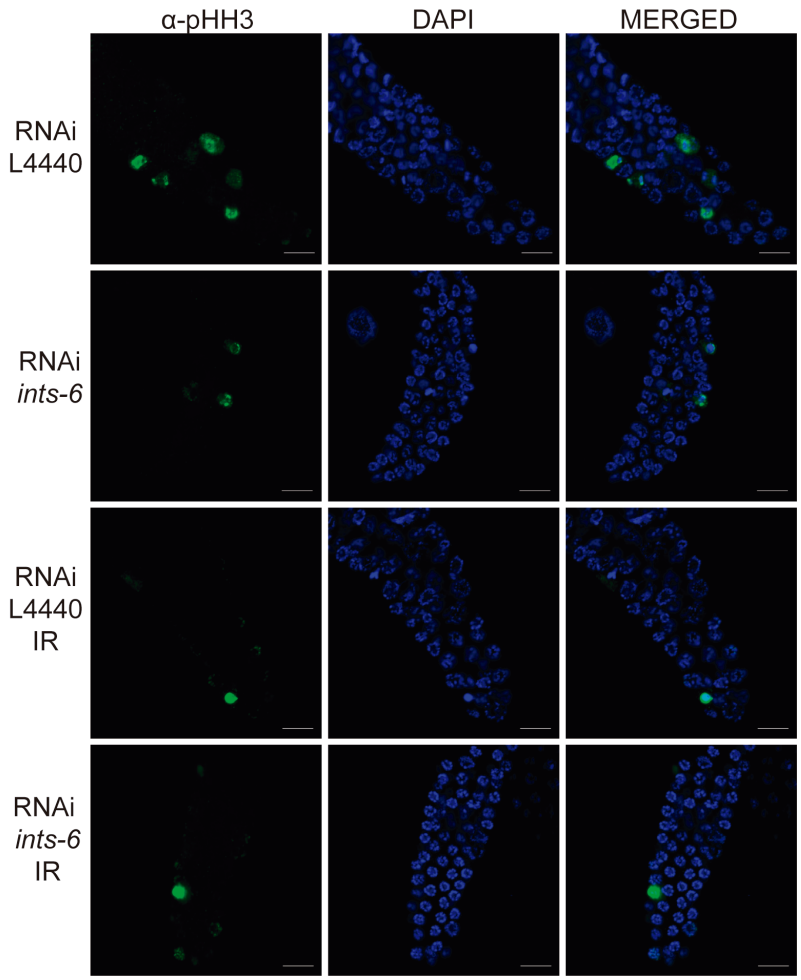


Figure 54. A) pHH3 positive cell quantification in the gonadal mitotic region in *ints-6* RNAi and L4440 RNAi backgrounds in non-irradiated worms and 24 hours following irradiation. B) Representative pictures of the α-pHH3 immunostained gonads. L1 stage N2 worms were fed the *ints-6* bacterial RNAi clone or the L4440 bacterial RNAi clone. At the L4 stage they were exposed to gamma radiation (90 Gy). Twenty-four hours later, gonads were dissected, fixed, immunostained with α-pHH3 and counterstained with DAPI. The number of positive pHH3 cells was manually counted in Z-Stack. The graph in A shows the quantification of pHH3 positive cells in all the optical stacks within the gonadal mitotic region. The number of mitotic cells is lower after IR in both genetic backgrounds. Scale bar: 10 μm.

10. Could INTS-6 Serine 850 be phosphorylated in response to DNA damage?

Our next objective was to delve deeper into understanding the mechanism of action of *ints-6* in response to irradiation. The single point *t1903* mutant has a phenylalanine instead of a serine at the 850 residue of the INTS-6 protein. *In silico* analysis showed that this 850 serine has a high probability of being phosphorylated (0.954 score, NetPhos 3.1 Server, www.cbs.dtu.dk/Services/NetPhos/). In addition, one of the kinases predicted to phosphorylate that serine residue is the ATM kinase (0.549 score), which is a key regulator of DSB DNA damage response, phosphorylating multiple targets to start cell cycle arrest (e.g. phosphorylation of p53 and CHK2) and coordinating DNA repair (e.g. phosphorylation of hSSB1, BRCA1, H2AX, and the MRN [Mre11-Rad50-Nbs1] complex) (Lavin, 2007).

To determine if the INTS-6 Serine 850 could be phosphorylated in response to DNA damage, different strategies were followed. First, it was decided to generate an antibody against the INTS-6 phosphorylated Ser 850. A polyclonal antibody against the peptide "N"_CLRES(P)QRFKLKQLTER_"C" was made by the IMMUNOSTEP enterprise. A specific polyclonal antibody against the phosphorylated epitope was obtained by affinity column purification. The flow-though was collected to use it as an antibody against the non-phosphorylated peptide. The anti-phospho Ser850 INTS-6 antibody specifically recognized the epitope of the phosphorylated peptide. The anti-INTS-6 antibody recognized both, the phosphorylated and the non-phosphorylated peptides. However, none of them were able to recognize INTS-6 from worm protein extracts.

To figure out if INTS-6 could be phosphorylated, we came up with an indirect strategy. Knowing that the glutamic (E) mimics the conformation of a phosphorylated serine (Maciejewski et al., 1995), we decided to generate a phosphomimetic strain carrying a glutamic, and also, the strain that mimics the non-phosphorylated form, carrying an alanine (A) residue instead. In addition, a 3xFLAG::eGFP tag was fused to the *ints-6* 5'-end to ease

biochemical experiments. The 3xFLAG::eGFP tag was also added to *ints-6* WT to work with all the strains in the same genetic background. The strains were generated by microinjection of the desired plasmid into the gonad syncytium. The plasmid was integrated into chromosome II using mosSCI system (Frokjaer-Jensen et al., 2008) (a schematic drawing of the pJC51 "*ints-6*::3xFLAG::eGFP" and pJC54 "*ints-6*(S850A)::3xFLAG::eGFP" plasmids is shown in Fig. 55A). Afterwards, each strain obtained was crossed with the *tm1615 dic-1/ints-6* deletion mutant (strain NB327) to end up with only one copy of the *ints-6* gene.

First, we decided to check the possible Ser850 phosphorylation in response to gamma radiation using a biochemistry technique called Phos-tag™ SDS-PAGE. This technique is based on a high molecular weight compound (Phos-tag™ ligand) that binds to divalent cations such as Zn^{2+} forming a complex that acts a selective phosphate-binding tag molecule. The Phos-tag™ ligand is added to normal acrylamide gels, along with Zn^{2+}. The phosphorylated and non-phosphorylated forms of a protein can be distinguished based on their mobility shift on a general SDS-PAGE. The protein phosphate groups will be bound to the phos-tag Zn^{2+} complex and their migration velocity will be slower.

Worms of the JCP383 ("*ints-6*::3xFLAG::eGFP") and JCP382 ("*ints-6*(S850A)::3xFLAG::eGFP") strains were synchronized. At L4 stage they were irradiated (120Gy) and collected at different time-points following IR (5 min, 10 min and 30 min). Next, worm protein was extracted followed by Phos-tag™ SDS-PAGE and WB against the 3xFLAG tag (Fig. 55B). Unfortunately, we did not detect any difference in the mobility of INTS-6 between the JCP383 ("*ints-6*::3xFLAG::eGFP") and JCP382 ("*ints-6*(S850A)::3xFLAG::eGFP") protein extracts.

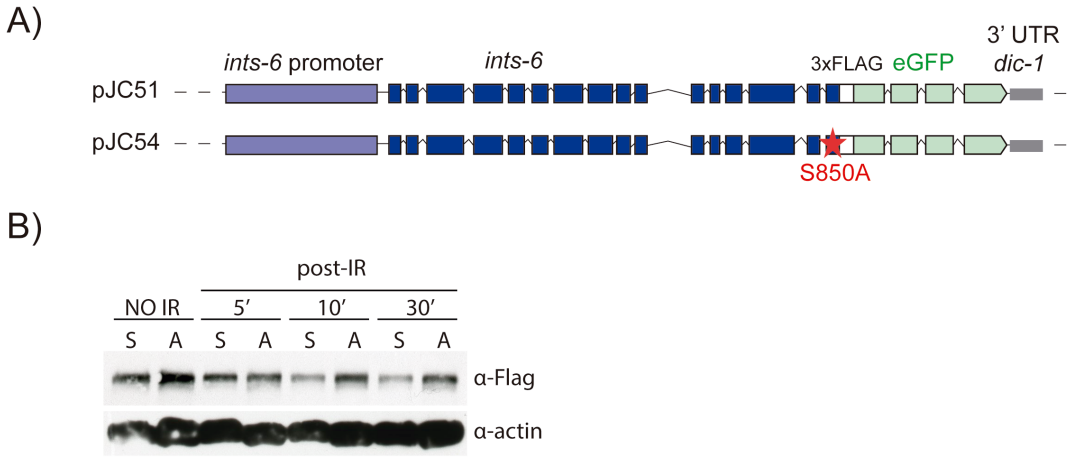


Figure 55. A) Scheme of the pJC51 ("*ints-6*::3xFLAG::eGFP") and pJC54 ("*ints-6*(S850A)3xFLAG::eGFP") plasmids. Plasmids pJC51 and pJC54 were used to generate the JCP341 and JCP342 transgenic strains. These strains were crossed with the *tm1615 dic-1/ints-6* deletion mutant strain (to attain only the desired copy of *ints-6* in the *C. elegans* genome) generating the JCP383 and JCP382 strains respectively. **B) Phos-tag™ SDS-PAGE and WB against the 3xFLAG tag from worm protein extracts of the JCP 383 and JCP 382 strains under IR and non-IR conditions.** Worms of the JCP383 ("*ints-6*::3xFLAG::eGFP"), represented by an "S" and JCP382 "*ints-6*(S850A)::3xFLAG::eGFP" represented by an "A", strains were synchronized and irradiated (120Gy) at the L4 stage. Next, they were collected at 5, 10 and 30 min post-IR and protein extracts of IR and non-IR worms were obtained. These protein extracts were run on Phos-tag™ SDS-PAGE electrophoresis, followed by WB against the 3xFLAG tag. As shown in the figure, no differences in the mobility shift of INTS-6 between the two strains were detected.

Even though we could not detect any difference in the Phos-tag™ SDS-PAGE between the JCP383 ("*ints-6*::3xFLAG::eGFP") and the JCP382 ("*ints-6*(S850A)::3xFLAG::eGFP") strains in the mobility shift of the INTS-6 protein, we hypothesized that if the INTS-6 Serine 850 is phosphorylated in response to DNA damage, there should be a difference between the transgenic strain that mimics the phosphorylated form of the protein (having a E in S850) and the one that cannot be phosphorylated (having an A in S850) either in RAD-51 recruitment to DSBs or in INTS-6 localization.

In order to approach our phosphorylation study in a manner that is most similar situation to the natural environment, we obtained the glutamic residue and the alanine residue of the serine 850 of INTS-6 only fused only to the 3xFLAG. We also used the WT INTS-6 that was fused to 3xFLAG to have all the transgenic strains in the same genetic background. These strains were constructed using the CRISPR technique (MM 6.3.1). A scheme of them is shown in Fig. 56.

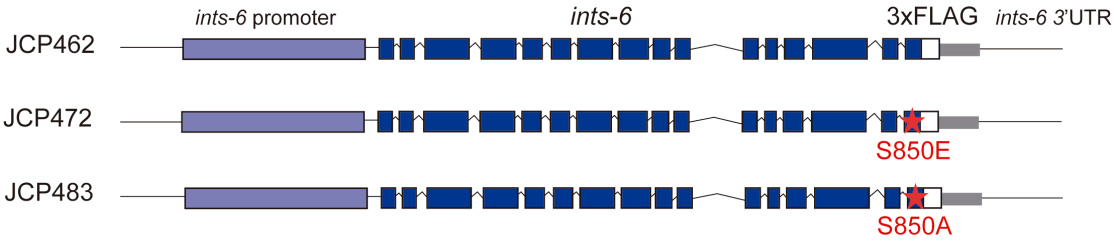


Figure 56. Schematic grawing of the transgenic strains generated to study the possible phosphorylation of INTS-6 Serine 850. This scheme represents the JCP462 "*ints-6*::3xFLAG", JCP472 "*ints-6*(S850E)::3xFLAG" and JCP483 "*ints-6*(S850A)::3xFLAG" strains.

Worms from transgenic strains JCP462 "*ints-6*::3xFLAG", JCP472 "*ints-6*(S850E)::3xFLAG" and JCP483 "*ints-6*(S850A)::3xFLAG" were synchronized and when they reached the L4 stage, they were exposed to gamma radiation (90 Gy). Twenty-four hours later, gonads from irradiated and non-irradiated worms were dissected, fixed, immunostained with α-RAD-51 and α-FLAG and counterstained with DAPI (Fig. 57).

We observed that after irradiation, INTS-6 formed foci in the "*ints-6*::3xFLAG" strain, supporting its role in DNA damage repair (Fig. 57). We also detected INTS-6 foci in the "*ints-6*(S850E)::3xFLAG" and "*ints-6*(S850A)::3xFLAG" strains. However, INTS-6 localization was generally more diffused.

Worms were able to form the RAD-51 IRIF in the mitotic region of all genetic backgrounds following IR (Fig. 57). There were no differences in the RAD-51 total foci number among the strains. However, we observed a difference in the "*ints-6*(S850A)::3xFLAG" strain. In contrast to "*ints-6*::3xFLAG" and "*ints-6*(S850E)::3xFLAG", no gonads were found in the "*ints-6*(S850A)::3xFLAG" with big nuclei having more than 16 RAD-51 foci, which suggests that the phosphorylation of this residue could be important for an efficient RAD-51 recruitment and stabilization at DSBs, and maybe cell cycle arrest as well (Fig. 58).

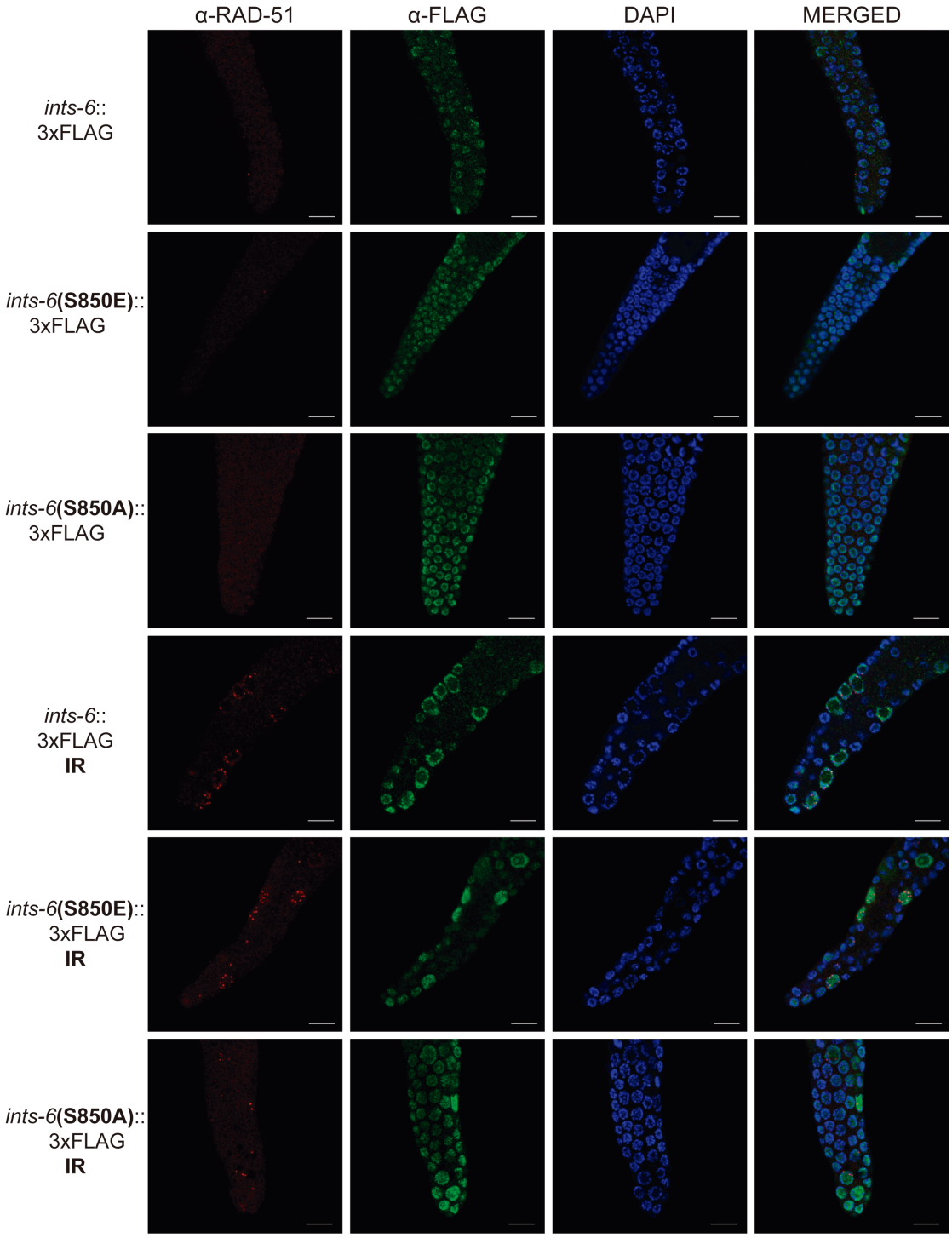


Figure 57. RAD-51 recruitment and INTS-6 localization following irradiation in transgenics that mimic our hypothesized INTS-6 phosphorylation of JCP472 ("*ints-6*(S850E)::3xFLAG") versus others that mimic the non-phosphorylated form of JCP483 ("*ints-6*(S850A)::3xFLAG") compared to WT JCP462 "*ints-6*::3xFLAG". L1 stage worms from each strain were fed the bacterial RNAi clones of *ints-6* or the empty L4440 vector. At the L4 stage, worms were exposed to gamma radiation (90 Gy). Twenty-four hours later, gonads were dissected, fixed, immunostained with α-RAD-51, α-FLAG and counterstained with DAPI. INTS-6 localization was observed in the nuclei in all transgenic backgrounds. Following IR, INTS-6 formed foci in all transgenic backgrounds but in the "*ints-6*(S850E)" and "*ints-6*(A850E)" backgrounds, the localization was generally more diffused within the nuclei. RAD-51 IRIF were observed following IR in all genetic backgrounds.

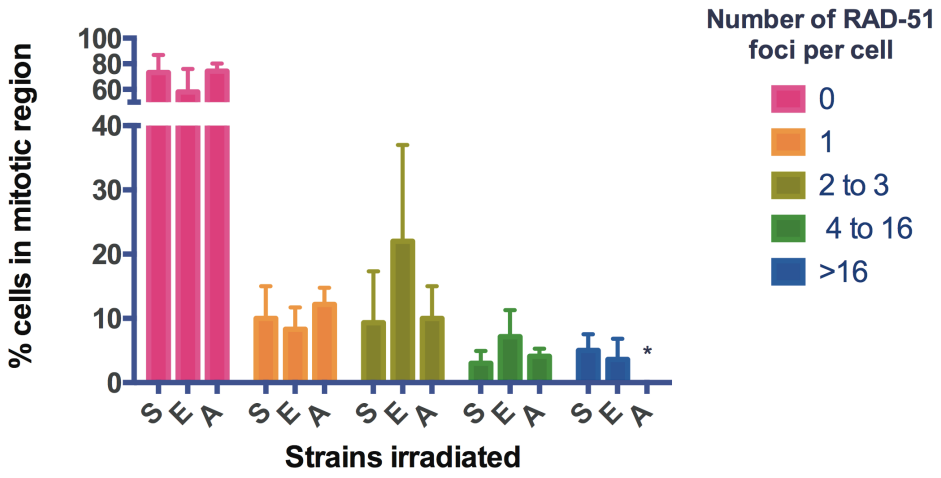


Figure 58. RAD-51 IRIF quantification following irradiation (90Gy). Images from gonadal mitotic regions from the irradiated transgenic strains: "*ints-6*::3xFLAG" represented by S, "*ints-6*(S850E)::3xFLAG" represented by E and "*ints-6*(S850A)::3xFLAG" represented by A, were taken in Z-stack and the number of RAD-51 foci per nuclei was manually counted. There were no differences in the percentage of nuclei with no foci (pink), one foci (orange), two foci (green) and four to sixteen foci (dark green). However, a statistically significant difference was found between strains A and S in nuclei with more than sixteen foci (blue) (P value: 0.0019).

11. Chimeric "sn-mRNAs" are detected in response to DNA damage

As we observed in Figure 57, *C. elegans* INTS-6 forms an IRIF, as described for human Ints6 (Zhang et al., 2013). We hypothesized that if INTS-6 forms IRIF, it could leave the Integrator complex in response to DNA damage caused by irradiation. This hypothesized absence of INTS-6 in the Integrator complex could lead to snRNA 3'-end processing defects, a read-through of the RNAP II and consequently the formation of chimeric "sn-mRNAs" as we observed in the *t1903* mutant background or in a genetic background knockdown for *ints-6*.

To verify whether these chimeric "sn-mRNAs" are formed in response to gamma radiation, which induces DSBs, N2 worms were synchronized and when they reached the L4 stage, they were irradiated (90Gy) and collected at different time points: 5 min, 15 min, 1 h, 3 h, 6 h, 24 h. Afterwards, total RNA was extracted using the *mirVana* ^{TM} *miRNA isolation kit*. Next, RT-PCRs were done to check the existence chimeric "sn-mRNAs" (Fig. 59). Primer pairs

used were the same as in the experiment of Fig. 44C. The forward primer on the snRNA F08H9.10 and the reverse primer on the next gene located *in sense* F08H9.3. As shown in the figure below, we detected the formation of chimeric "sn-mRNAs" following DNA damage. The strongest expression was observed at 24 hours post-irradiation.

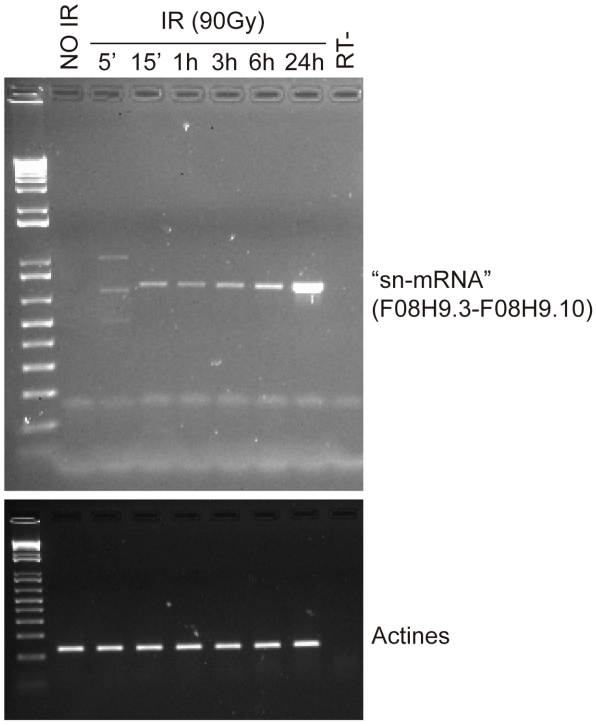


Figura 59. Chimeric "sn-mRNAs" are formed in response to DNA damage. N2 worms were synchronized and at the L4 stage they were irradiated (90Gy). Afterwards, worms were collected at different time points: 5min, 15min, 1h, 3h, 6h, 24h. The formation of a chimeric "sn-mRNA" is shown. The RT-PCR shows the amplification from the snRNA F08H9.10 until the F08H9.3 gene. The highest response was observed at 24 hours after DNA damage.

Discussion

In 2005, Baillat and coworkers described the Integrator complex in human cell lines (Baillat et al., 2005). Ints6 was the only subunit described prior to the discovery of the Integrator complex, which was named *DICE1* and described as a putative tumor suppressor, because of its location in the human genome, a region with frequent LOH in tumors, and also because it was downregulated in lung tumors (Wieland et al., 1999).

We started working with its *C. elegans* ortholog *dic-1*. A previous report in the scientific literature stated that the *C. elegans* DIC-1 localized in the mitochondria and functioned in the process of mitochondrial morphogenesis (Han et al., 2006; Lee et al., 2009). Thus, the authors suggested a divergent evolution between the *C. elegans* DIC-1 protein and the mammalian DICE1/ Ints6. Importantly, in this study we determined that *C. elegans* DIC-1 is indeed a member of the Integrator complex and functions in 3'-end processing of snRNAs as its human ortholog. Therefore, we have renamed this *C. elegans* gene as *ints-6*.

1. The *t1903* mutant shows defects in 3'-end processing of snRNAs. What happens with splicing?

Very importantly, the RNA deep sequencing analysis of the *t1903* mutant led us to discover the function of *dic-1* in 3'-end processing of the following snRNAs: U1, U2, U4, U5, U6 and SL. Thus, we could renamed *dic-1* as the Integrator subunit 6, *ints-6*. There were reports that in other species the Integrator complex was the machinery responsible for the 3'-end processing of U1 and U2, U4 and U5 (Baillat et al., 2005; Ezzeddine et al., 2011; Yamamoto et al., 2011). Although it was assumed that other snRNAs would be similarly processed, we described the processing mechanism for U6 and SL snRNAs, which had not been experimentally demonstrated until our study.

Surprinsingly, the splicing in the *t1903* mutant was mostly done correctly although a fraction of the snRNAs were not properly processed. We assumed that this could be explained because there were enough correctly processed snRNAs to splice properly or because of the reported long half-life of

spliceosomal snRNAs, which exceeds 60 hr (Fury & Zieve, 1996).

Later, RNAi knockdown experiments on the Integrator subunits followed by RT-PCRs led us to suspect that the splicing might not be totally correct. In those experiments, apart from the different degrees of misprocessing defects obtained depending on the Integrator subunit knocked down, the mature snRNAs could come from the initial developing embryos before they were affected by RNAi in the L1 stage. We designed one primer based on the snRNA and another on the second exon of the downstream gene so that we could distinguish between the possible amplification of any remaining genomic DNA from the initial RNA extraction and the cDNA that would be properly spliced. In the majority of PCRs, we obtained two bands. The size of the lower band corresponded to the cDNA (properly spliced) whereas the size of the upper one corresponded to genomic DNA. However, in the negative controls that were done without the reverse transcriptase (RT-) enzyme, either we did not observed any amplification or it was much lower (Fig. 44). This suggests that the upper bands correspond to the RNA not spliced and copied into cDNA. Therefore, we might say that there could be minor splicing defects upon knockdown of the Integrator complex. Actually, an improper splicing of the gene Smad5 gene upon Ints11 knockdown in *Zebrafish* was reported (Tao et al., 2009).

2. Localization of the Integrator subunit 6 in *C. elegans* and humans

A previous report in the scientific literature said that DIC-1/INTS-6 localized in the mitochondria and functioned in the process of mitochondrial morphogenesis (Han et al., 2006; Lee et al., 2009). Thus, the authors suggested a divergent evolution between the *C. elegans* DIC-1 protein and the mammalian DICE1/Ints6. In this study, we determined that the *C. elegans* DIC-1 is indeed a member of the Integrator complex and functions in the 3'-end processing of snRNAs, just as its human ortholog. Therefore, we expected a nuclear localization in accordance with its function in 3'-end processing of the snRNAs.

As anticipated, we observed that the localization of DIC-1 was nuclear. We used different techniques: *in vivo* localization of DIC-1 fused to eGFP in its Ct domain and immunostaining with α-FLAG. As we have already mentioned, in the study of Han and colleagues, the localization was observed in the mitochondria. This controversy might be explained because the approach used to observe DIC-1 localization is a little bit different. First, whereas we fused the eGFP to the Ct domain of the protein, they fused it to the Nt domain. The eGFP could be interfering with the molecular conformation of the protein, altering its localization. Second, the antibodies used to detect DIC-1 are different. We used a monoclonal 3xFLAG antibody to immunostain the JCP462 (*ints-6*::3xFLAG) strain and they used a polyclonal mouse antiserum (raised against the 143-570 aa of DIC-1) to immunostain N2 worms. Although the α-3xFLAG is widely used and has proved to be specific for the FLAG epitope in *C. elegans* tissues or worm extracts, we do not discard the possibility that DIC-1 also localizes to the mitochondria. Importantly, based on bioinformatic predictions, both nematode and mammalian proteins can localize to mitochondria as well as to nuclei. We must note that, similarly to what was observed in the other study, we saw morphogenetic defects upon RNAi knockdown of *dic-1*. Our *t1903* mutant also showed defects in embryonic morphogenesis as well as necrosis.

We also checked the localization of Ints6 in different humans cells (HEK293T and U2OS cell lines) and by different techniques, and in all the cases the localization was observed in the nucleus in accordance with previous reports in the scientific literature (Filleur et al., 2009).

3. Human Ints6 depletion results in over-proliferation and snRNAs misprocessing

We briefly characterized the Ints6 human Integrator subunit because we were interested in comparing the similarities and differences between *C. elegans* and *H. sapiens.*

Upon depletion of Ints6 in RPE cells, we observed an over-proliferation

phenotype which supports previous results in literature suggesting a link between Ints6 and cell cycle regulation (Filleur et al., 2009). Although we did not detect differences in Ints6 protein expression levels throughout the cell cycle, we observed on a WB (Fig. 37) that a higher band appeared in G1, increased in G1/S, reached a maximum in G2/M and disappeared in mitosis. Although this band could be non-specific, it is interesting to speculate that the cell-cycle regulated band could correspond either to either Ints6 with a post-translational modification that is cell-cycle dependent or the paralog of Ints6, DDX26B, that exists in the human genome and has been reported to have lower expression levels.

When we depleted Ints6 by RNAi in RPE cells to check U1 and U2 processing, we were able to knockdown *ints6* (protein expression levels were lower) and corroborate its function in 3'-end processing. Interestingly, we observed that on the WB in Fig. 39 that an upper band increased its expression upon depletion of Ints6. We corroborated that the siRNAs used were designed against the sequence of Ints6 but they did not target the paralog of Ints6, DDX26B. The antibody used against Ints6 targets the Ct region, which is mostly conserved in DDX26B. It would be exciting to discover if the upper band that we observed corresponds to DDX26B. We could specifically deplete the paralog of Ints6 and check if it plays a redundant role in 3'-end processing or if it is cell-cycle regulated.

4. *C. elegans* Integrator complex composition. Is it a modular complex?

We described the other members of the *C. elegans* Integrator complex by orthogy, Co-IP and function in 3'-end processing. Those experiments led us to conclude that the *C. elegans* snRNAs were processed by the Integrator complex similarly to humans but not to yeast, where it is involved the Nrd1/Nab3/Sen1 complex.

We proposed that the in *C. elegans* Integrator complex is comprised of the INTS-1 to INTS-13 subunits, although the INTS-10 and INTS-12 subunits were not detected the mass spectrometry analysis, probably because they are

the smallest subunits with MWs of 9 kDa and 8.5 kDa respectively. The short length of the protein sequence probably increased the difficulty in find peptides within the eluates. As observed in the homology Table 4, INTS-10 and INTS-12 only present homology to their respective human orthologs within a part of the protein sequence.

In the case of Ints12, the plant homeodomain (PHD) motif is present in *C. elegans* and *H. sapiens* proteins. Although there are not specific studies of the Ints12 function, several proteins with a PHD domain have been found in the nucleus and are involved in chromatin-mediated gene regulation (Sanchez & Zhou, 2011).

In the case of Ints10, the putative *C. elegans* ortholog found (NP_500453/F47C12.3) presents a serpentine receptor domain (7TM GPCR), which possesses seven transmembrane domains. The homology between *C. elegans* and *H. sapiens* for this protein was found in a region of two transmembrane domains. However, no protein domains were found in the analysis of human Ints10. It is surprising to find a serpentine receptor as a possible Integrator subunit. These receptors have been shown to detect molecules outside the cell such as pheromones or neurotransmitters and activate signal transduction pathways (Ritter & Hall, 2009). However, it is exciting to speculate that Integrator could be a modular complex regulating a transcriptional response under specific circumstances in *C. elegans*. There is evidence in the scientific literature that the Integrator complex regulates RNAP II transcription of IEGs upon stimulation with EGF (Gardini et al., 2014). Therefore, it would make sense to have Ints10, or at least a certain amount of Ints10, at the cellular membrane to receive signals from the outside that could be translocated to the nucleus in response to a stimulus where it could function in transcriptional regulation.

Homology of the putative Ints10 ortholog was only found to cover 20% of the protein sequence of *H. sapiens* with a 39% similarity and we only found subtle defects in 3'-end processing. Thus, we cannot ensure that a *C. elegans* ortholog exists for Ints10. Nevertheless, depletion of the *Drosophila* Ints10 also leads to subtle misprocessing of snRNAs (Ezzeddine et al., 2011) in

accordance with our results.

In regard to the molecular phenotypes in 3'-end processing that we observed upon RNAi knockdown of the Integrator complex subunits, there are differences in the degree of misprocessing depending on the subunit. In accordance with *Drosophila* studies, we observed subtle defects upon depletion of INTS-10 and -3. Slightly more misprocessing was perceived when *ints-12* was knocked down. In contrast to *Drosophila* where the depletion of the INTS-1, -4, -9 subunits led to the highest misprocessing levels, we observed weak snRNA processing defects in the *ints-1* RNAi and the highest defects were observed in the *ints-2*, *-4*, *-5*, *-6* and *-7* RNAi, followed by medium defects in the *ints-8*, *-9* and *-11* RNAi.

One of the advantages of studying the Integrator complex in *C. elegans* versus studying it in human or *Drosophila* S2 cells is that we could observe and compare the effect of knocked down Integrator subunits on the entire organism. The phenotypes of larval arrest that we perceived do not exactly correlate with the levels of misprocessing in each subunit. For example, the RNAi knockdown of *ints-4* and *ints-7* leads to high and similar levels of misprocessing, respectively. However, *ints-4* presents an aggressive larval arrest phenotype at L2-L3 stages, whereas INTS-7 depleted worms reach adulthood. This suggests that the Integrator complex in *C. elegans* plays additional roles in development from the processing snRNAs.

Even though there is no correlation between the levels of misprocessing and the phenotypes observed, it is also true that in worms where the subunit had subtle or weak misprocessing defects (INTS-1, -3, 10, -12, -13), these did not lead to arrest or manifestation of a strong phenotype.

We were surprised because there are no reports specifically about INTS-2, which is very well conserved throughout evolution. Keeping in mind that mutants (in *C. elegans* or other species) have not been found even though it manifests one of the most aggressive phenotypes of larval arrest upon *ints-2* RNAi in *C. elegans,* it could be that mutations in *ints-2* easily lead to lethality. It would be interesting to do an in-depth study of the function of *ints-2* in *C.*

elegans and other species.

5. Possible partner proteins of the *C. elegans* Integrator complex

Apart from the Integrator complex subunits and the ortholog of NABP1 (C06G3.8), other INTS-6 interactors, direct or indirect, were found by mass spectrometry analysis.

One of them is Y56A3A.31, which is an ortholog to the uncharacterized human C7orf26 protein. In mammals, reciprocal IPs against Integrator subunits were able to pull down C7orf26 (Malovannaya et al., 2010). These results, along with our analysis, suggest that C7orf26 could be another member of the core Integrator complex. Therefore, it would be interesting to analyze if depletion of the *C. elegans* Y56A3A.31 protein produces snRNA 3'-end processing defects.

Another INTS-6 interactor is LAF-1 (DEAD-box RNA helicase), which is required for embryonic development and sex determination. In embryos, it localizes predominantly to the cytoplasm with additional localization seen in P granules where LAF-1 is important for promoting P granule assembly. P granules are a class of perinuclear RNA granules specific to the germline. It would be interesting to study whether any Integrator subunit localizes to P granules and if they are important for maintaining the integrity of P granules.

Another protein found, C02B10.4, is the ortholog of the THO complex subunit 4, which is part of the TREX complex that is involved in coupling transcription to export mRNAs to the cytoplasm.

Interestingly, orthologs of ZK856.6 and R07E4.5 proteins belong exclusively to the nematode phylum and F37C4.5 belongs to nematodes and arthropods. Microarray and RNA sequencing studies indicate that R07E4.5 and F37C4.5 are enriched in the germline and in neurons. Related to localization in neurons, the ASP-4 protein, which encodes an aspartyl protease conserved in higher eukaryotes and is required for degenerative cell death in neurons of *C.*

elegans, was detected (Tavernarkis et al., 2002). In mammals, a recent study linked the Integrator complex with neuronal migration through its association with other proteins and TFs to directly regulate transcription of neuronal migration genes (van den Berg et al., 2017). It would be interesting to figure out if there could be a relationship between the Integrator complex and neurons in *C. elegans*.

Surprisingly, we found three cytoplasmic proteins. Two of them are involved in translation: RARS-1, which encodes an arginyl-tRNA synthetase and rpl7a, which encodes a large ribosomal subunit. The other one, Y73F8A.26, is an ortholog of human P3H4 (prolyl 3-hydroxylase family member 4 (non-enzymatic)), P3H1 (prolyl 3-hydroxylase 1) and CRTAP (cartilage associated protein). In *C. elegans*, microarray studies indicate that Y73F8A.26 is enriched in muscle cells.
Altogether, after filtering out nonspecific contaminants, our data from INTS-6 IP and mass spectrometry analysis provides a source of possible Integrator complex interactors. Future studies will help discover unknown aspects of the Integrator complex molecular functions in *C. elegans* and in humans.

6. Integrator complex knockdown and transcription termination defects

In our study we focused our efforts on deciphering the meaning of the chimeric “sn-mRNAs” formed upon Integrator complex knockdown.
The exact transcription termination mechanism coupled to 3’-end processing of snRNAs in metazoans is still to be determined. The Integrator complex and NELF participate, but the details of their functions are not known.
One interesting difference in snRNA transcription termination between *C. elegans* and mammals is that there is no ortholog for NELF in *C. elegans*. Therefore, snRNA transcription termination process should be a little bit different in *C. elegans*.

In yeast, where there is no Integrator complex or NELF. Transcription termination of snRNAs is done by the Nrd1-Nab3-Sen1 pathway. Nrd1 and Nab3 are RNA binding proteins and Sen1 is an RNA-DNA helicase. The

release of the polymerase occurs by a mechanism that strictly requires the action of the Sen1 helicase. Perhaps in *C. elegans*, the 3'-end processing coupled to transcription termination of snRNAs also requires an RNA-DNA helicase which is still to be determined.

In addition, another termination mechanism of snRNAs that was proposed to function as a fail-safe mechanism was described in yeast. The Rnt1 endonuclease, the ortholog of the bacterial RNase III, provides an entry point for the Rat1 exonuclease by cleaving nascent transcripts at defined sites, which triggers termination by the torpedo mechanism (Porrua & Libri, 2015). Strikingly, the disruption of *RNT1* results in the production of a 3' extended U2 that is polyadenylated (Abou Elela & Ares, 1998).

Upon inhibition of NELF in mammals, a similar phenotype that disrupts processing and induces the production of aberrant polyadenylated histone mRNAs and U1 or U2 snRNAs through the use of cryptic downstream poly(A) signals was found. The same phenotype of aberrant polyadenylation was observed upon depletion of Integrator (Narita et al., 2007; Yamamoto et al., 2014).

In the work of Yamamoto and colleagues, the polyadenylation sites were mapped in the NELF- and Integrator knockdown cell samples. There was no detectable PAS (AAUAAA or its variants) around these sites. Polyadenylation was found at different positions and therefore the transcripts differed in length. In addition, poly(A) tail lengths were different.

By contrast, in our "sn-mRNAs" phenotype, deep sequencing analysis showed that the reads finalized at the end of the downstream genes, both in the *t1903* mutant and upon RNAi knockdown of Integrator subunits, which suggests that the transcription termination is coupled to polyadenylation at the correct PAS site.

Interestingly, studies in yeast have uncovered a role for RNA polyadenylation in nuclear RNA degradation, which is in direct contrast to the well-known function of poly(A) tails in stimulating mRNA export to the cytoplasm where they promote mRNA stability and translation. Short poly(A) tails are added by

the TRAMP complex (Trf4/Air2/Mtr4p Polyadenylation complex), which is a multi-protein complex consisting of the RNA helicase (Mtr4/Dob1), a poly(A) polymerase (PAP) (either Trf4 or Trf5) and a zinc knuckle protein (either Air1 or Air2) that facilitates the degradation of substrates by the exosome (Schmidt & Butler, 2013; Kilchert et al., 2016).

The eukaryotic RNA exosome is an evolutionarily conserved ribonucleolytic complex that consists of 10 or 11 subunits and is a key component of the RNA-surveillance machinery. The exosome complex degrades different types of aberrant RNA transcripts, either in the nucleus or the cytoplasm. In addition, it is also responsible for the 3' trimming of nuclear precursors to several RNA species such as the rRNAs, snRNAs and snoRNAs (Houseley LaCava et al., 2006; Kilchert et al., 2016).

The addition of short poly(A) tails by the non-canonical polymerases of the TRAMP complex (Trf4 or Trf5) is thought to provide a 'grip' for the Mtr4 helicase, which then unwinds the RNA and feeds it into the exosome complex.

The difference between canonical polyadenylated transcripts and the polyadenylated transcripts for nuclear RNA degradation might be in the length of the poly(A) tails, which might lie in the processivity of the reactions.

The canonical cleavage and polyadenylation machinery is highly processive, (adding ~250 nt in humans and 60–90 nt in yeast). This might ensure that no free 3′-end is available until a very long tail has been synthesized and covered by poly(A)-binding proteins such as Pab1. By contrast, the polyadenylation by poly(A) polymerases of the TRAMP complex showed lower processivity, and thus the 3′-ends of the tails might frequently be free and available to the exosome and also might never be long enough for Pab1 binding. Orthologs of the yeast exosome and TRAMP complexes seem to be conserved throughout evolution even though they are not yet fully characterized (Houseley LaCava et al., 2006; Kilchert et al., 2016; Zinder & Lima, 2017).

Our results suggest that chimeric "sn-mRNAs" contain long poly(A) tails

that are added by the canonical machinery because transcription is coupled to termination of the downstream gene. The fact that these transcripts are easy to detect upon Integrator complex knockdown suggests that they are stable in the cell and are not a target for degradation by RNA-surveillance mechanisms such as the putative *C. elegans* TRAMP complex or any other RNA quality control mechanisms. The stability of these chimeric "sn-mRNAs" could be derived from the secondary structure of the snRNA and the poly(A) tail. Nevertheless, it would be interesting to check the half-life of the chimeric RNAs observed.

In addition, it has been reported that sites of non-conventional termination in yeast are enriched downstream of genes and are thought to prevent read-through transcription from one gene to an adjacent gene. Thus, redundant transcription termination pathways function as 'fail-safes' that protect neighboring genes from transcriptional interference (Gullerova & Proudfoot, 2010). The *C. elegans* genome is also very compact but these possible fail-safe mechanisms are not acting to stop the transcription downstream of the snRNAs when we knocked down the Integrator. Intriguingly, the transcribed transcripts that we observed in our study could have evolved a function in the *C. elegans* system (under specific conditions such as DNA damage) and as a consequence, the "fail-safes" mechanisms would have evolved to converge functionally and therefore be absent downstream of the snRNAs.

Importantly, the fact that the Y75B8A.23 gene is translated into a protein upon Integrator complex knockdown reveals that at least this "sn-mRNA" is transported to the cytoplasm. This observation seems to indicate that the cell's nuclear RNA quality control mechanisms of the cell do not recognize the transcript as aberrant and that the transcript is not unstable.

Other genes downstream of the SL2 type snRNA that are over-expressed would be also transported outside the nucleus and translated. These include: R12E2.11, a Uridine 5'-monophosphate synthase involved in the synthesis of nucleic acids, H41C03.3, an acetylglucosamine transferase involved in lysosome transportation and *hgo-1,* which encodes a putative homogentisate 1,2-dioxygenase, that is required for normal resistance to hypertonic stress in

humans.

Because the chimeric transcripts studied are comprised of U1 and U2 and their respective downstream genes were not translated into proteins, there should be a difference between them and the chimeric RNA formed by the SL snRNA *sls-2.8* and the Y75B8A.23 downstream gene, which was translated.

One difference between snRNAs and mRNAs is the 5' structure. Whereas mRNAs contain a 7-methylguanosine (m^7G) cap, both the Sm class snRNAs, (involved in splicing) and SL snRNAs (involved in *trans*-splicing) contain a trimethylguanosine (TMG) cap.

Recruitment and binding of the CBC to the m^7G cap is necessary to coordinate processes such as spliceosome assembly, 3′ processing or RNA export. Also, the majority of cellular mRNA translation is initiated by the cap-dependent mechanism. Upon exiting into the cytoplasm, CBC stays bound to the mRNA cap and recruits translation initiation factors to the 5'-end of the mRNA. These translation initiation factors interact with the poly(A) binding protein PABP1, which is bound to the poly(A) tail of mRNA, and create a pseudo-circular structure of translating mRNA (Ramanathan et al., 2016).

In *C. elegans*, 70% of genes are trans-spliced and possess a TMG cap provided by the SL exon. The function of *trans*-splicing remains unknown although it has been suggested that the SL exon improves the translational efficiency of the transcript. We supposed that all the chimeric RNAs, whether they have a TMG cap or an m^7G cap on at the 5'-end, had the potential to be translated.

One possibility is that chimeric RNAs formed by "U1-mRNA" and "U2-mRNA" would be not translated because they would remain in the nucleus. It could also be that once they are exported to the cytoplasm, maturation events of the Sm-class snRNAs take place such as Sm ring assembly to form snRNP. The Sm ring would physically impede degradation by the exosome in the case of an aberrant transcript. Also, the poly(A) tail would block the normal 3' trimming of the snRNAs by the exosome.

The Sm-core and TMG-cap have been reported to function as import signals for Sm RNPs (Matera, 2007). Formation of the TMG cap from a former m^7G cap in these transcripts occurs in the cytoplasm (re-capping) following assembly of the Sm core. We do not know if the chimeric "U1-mRNA" or "U2-mRNA" observed have a TMG cap or an m^7G cap, nor do we know how the cap would affect their possible function as ncRNAs.

Strikingly, in the study by Ezzedine and colleagues, the misprocessing was studied in a cell-based system where the GFP and a PAS sequence were tagged downstream of a U7 3'box so that they could quantify the processing defects by measuring the fluorescence levels. In contrast to our results of the chimeric "U1-mRNA" and "U2-mRNA", this artificial minigene could produce a protein upon Integrator knockdown. Ezzedine and coworkers chose the *Drosophila* U7 snRNA as the model snRNA gene because it is the smallest snRNA and is predicted to have the least amount of secondary structure. Perhaps the U1 and U2 secondary structure dampens translation in the chimeric RNAs we observed.

In *C. elegans* there is no U7, which in other metazoans is involved the 3'end processing of histones mRNAs. Thus, in *C. elegans* there should be a different mechanism that is still to be determined.

7. INTS-6 involvement in DNA damage response

It has been suggested that transcription is a major source of chromosome fragility and DNA damage. Regions with an open chromatin state have been found to be prone to genome instability such as UsnRNA genes or promoter-proximal pause sites (Yu et al., 1998; Yu et al., 2000; Li et al., 1998). The fact that cells must ensure the stability of genetic information during transcription is evident. Several studies have shown that members of the RNAP II-associated basal transcription machinery are involved in different DNA damage responses (Lainé & Egly, 2006; Derheimer et al., 2007). Additionally, large scale genetic and proteomic screenings for proteins involved in DNA damage response have revealed enrichment in RNA processing proteins.

This may indicate that RNA metabolism and DNA repair pathways functionally converge (Montecucco & Biamonti, 2013).

In dealing with the Integrator Complex, it has been speculated that the potential presence of hSSB1, through its interaction with Ints3 and Ints6 at transcriptional pause sites, might play a role in maintaining genome integrity (Baillat & Wagner, 2015). We wanted to go a step further and we hypothesized that the chimeric "sn-mRNAs" that we observed upon knockdown of Integrator subunits in *C. elegans* could be a response to DNA damage. We thought that if INTS-6, by itself or with other Integrator subunits, goes to DSBs as it does in humans, the Integrator complex would not process the snRNAs properly and therefore would produce chimeric "sn-mRNAs" as a result of RNAP II read-through. This hypothesized model is shown in Fig. 60.

We know that through bioinformatics analysis that the *C. elegans* INTS-6 Ser 850 has a high probability of being phosphorylated and one of the probable kinases for phosphorylating that residue is ATM. One of the first steps in DNA repair by the HR pathway is the activation of the ATM kinase, which rapidly phosphorylates various DNA repair factors and initiates a global DNA damage response in the cell. We tried to verify if Ser 850 is phosphorylated in response to DNA damage but we could not come to viable conclusions.

We checked if *C. elegans ints-6* was involved in DNA damage as it is in humans. We concluded that *ints-6* is a key component of DNA repair using the HR pathway because recruitment of RAD-51 to DSBs following IR was totally abrogated in an *ints-6* RNAi genetic background. In addition, we observed that depletion of INTS-6 interrupted phosphorylation of Tyr15 Cdk1 to induce cell cycle arrest in G2/M following IR. However, cells within the proliferation region of the gonads were able to arrest in response to IR although they did not increase Cdk1 Tyr15 phosphorylation. This suggests that redundant mechanisms are acting to induce cell cycle arrest.

The fact that RNAi *ints-6* knockdown in *C. elegans* abrogates recruitment of RAD-51 to DSBs following IR suggests that INTS-6 acts upstream of RAD-51 in the HR DNA repair pathway. Studies in human cells demonstrated that depletion of Ints6, along with its paralog DDX26B, totally disrupted recruitment

of Rad51 and reduced the accumulation of RPA IRIF (Zhang et al., 2013). It has also been postulated that the hSSB1/INTS complex should act downstream of the MRN complex in the HR pathway because the depletion of the MRN complex resulted in a significantly reduced IRIF of Ints3 and hSSB1/2 (Huang et al., 2009). Thus, INTS-6 should be recruited to DNA damage sites following recruitment of the MRN complex, which acts in DNA end resection of the DNA strands to generate 3'ssDNA tails and before recruitment of RAD-51, which replaces RPA and mediates the strand invasion to synthesize new DNA using the homologous strand as the template.

Recently, a study in fission yeast revealed the formation of RNA-DNA hybrids, which had previously been associated with genome instability, as a key step during DSB repair (Ohle et al., 2016). The appearance of short ssDNA segments might induce RNAP II transcription and RNA-DNA hybrid formation that might be involved in strand resection and RPA recruitment. RNase H activity is subsequently required to degrade these RNA-DNA hybrids allowing completion of the DSB repair process (Ohle et al., 2016).

It would be very exciting to study if INTS-6 or any other members of the Integrator complex, which is absent in yeast, has a role either in stimulating the possible production of RNA-DNA hybrids at DSBs or in removing their formation because both processes are necessary for Rad51 recruitment in yeast. It is also possible that exosome complex exonucleases, which are involved in degradation of unstable or aberrant transcripts as well as in the 3'-end processing of transcripts, could be required at this step to remove the RNA. It would be intriguing to analyze INTS-6 protein interaction partners after DNA damage because we would probably detect different protein interactors compared to those that we found in this study under normal growing conditions. The new protein partners could provide us a hint as to the molecular mechanism of INTS-6 in DNA repair.

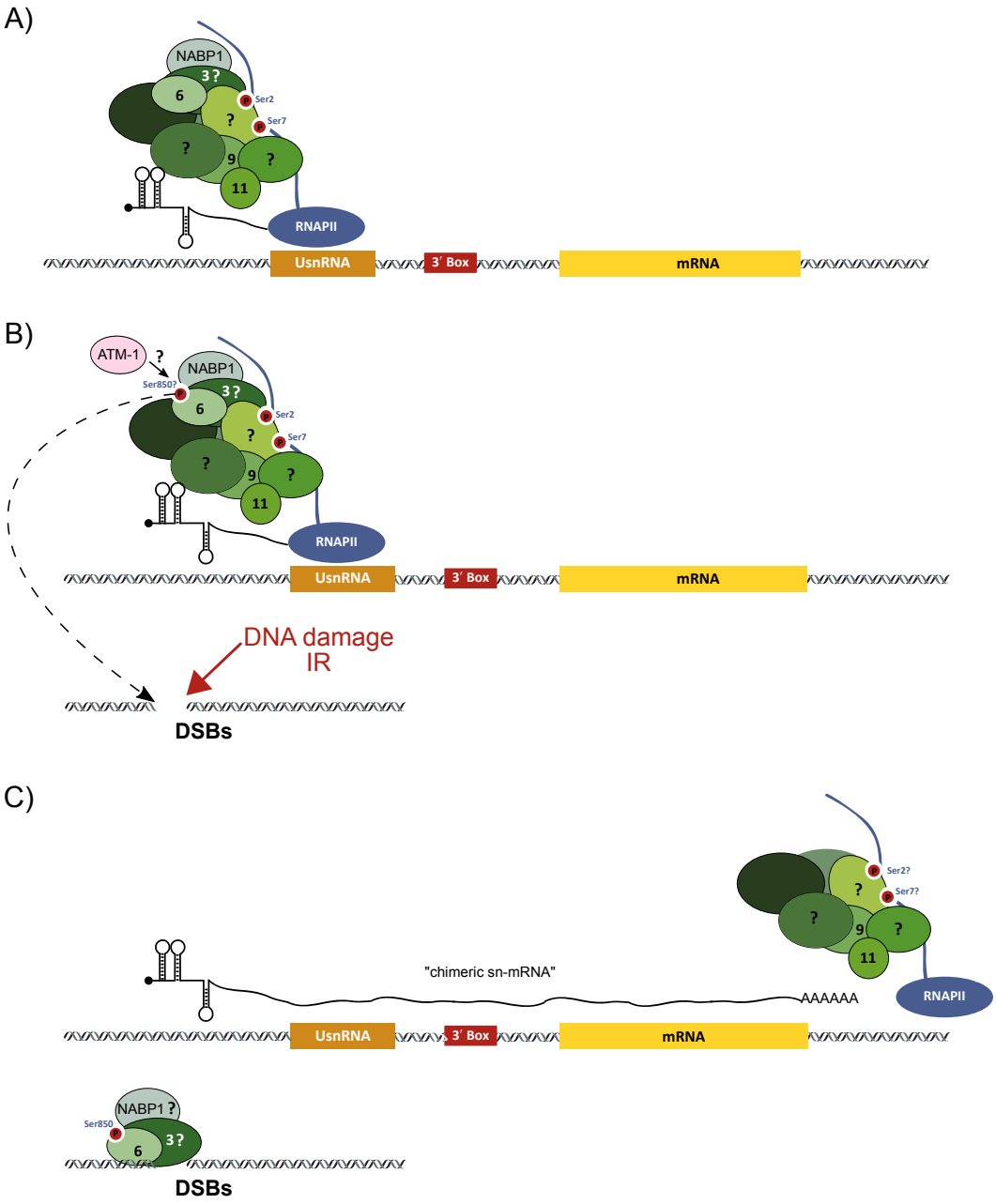


Figure 60. Working model of the Integrator complex upon DNA damage and the formation of chimeric "sn-mRNAs". A) The Integrator complex is transcribing an snRNA under normal conditions. B) Upon irradiation, DSBs are produced in the DNA. The ATM-1 kinase is activated and could phosphorylate INTS-6 in S850. INTS-6 would be recruited to DNA damage sites, probably with INTS-3 and NAPB1. C) Translocation of part of the Integrator complex to DNA damage sites would result in a read-through of the RNAP II and the formation of chimeric "sn-mRNAs".

Using RT-PCR we checked the existence of chimeric "sn-mRNAs" following IR. We detected their formation at different time points post-IR. Interestingly, at 15 min, they could already be detected and their expression increased at 24 h. Further work should be done to demonstrate whether these chimeric RNAs have a function in the DNA damage response or if they are merely a consequence.

Emerging evidence indicates the importance of RNA function on DNA damage response and repair, supporting the idea that these chimeric RNAs are not aberrant. For example, recent experiments revealed that while transcription in the proximity of DSBs is downregulated, at sites of DNA damage it is induced, producing small-noncoding RNAs (sncRNAs) with the DNA sequence flanking the DSB. These sncRNAs have been found in different species at very low copy numbers per cell, supporting the idea that *de-novo* transcription may be kept at very low levels and may be compatible with coordinated inhibition of promoter-driven transcription. They are named DNA-damage response RNAs (DDRNAs) and DSB-induced RNA (diRNA) (Francia et al., 2012; Wei et al., 2012; Francia et al., 2016; Capozzo et al., 2017). DDRNAs have been discovered in mammals. They are generated by cleavage of a presumably longer precursor RNA and participate in DDR signaling as they are required for full DNA damage response activation, which in turn results in transcriptional silencing of the damaged locus (Francia et al., 2012; Francia et al., 2016). diRNAs were discovered in plants and have been proposed to participate in DNA repair (Wei et al., 2012).

Another example of RNAs acting upon DNA damage is found in the DNA damage-induced long non-coding RNAs (lncRNAs) such as the lincRNAp21, which regulates DNA damage gene expression programs (Huarte et al., 2010).

The long non-coding RNAs are RNA molecules longer than 200 nt that do not appear to have protein coding potential, though they could have small ORFs that produce peptides. The lncRNAs have been shown to play functional roles in numerous biological processes. They form secondary structures, undergo

post-transcriptional processing (5'capping, splicing and polyadenylation), their sequence is poorly conserved between species and they show very low to moderate expression levels (Mercer et al., 2009; Engreitz et al., 2016).

In contrast to lncRNAs, the genes within the chimeric RNAs observed in this study possess well-conserved ORFs. The fact that the protein sequence has been preserved from mutations suggests that these genes found downstream of the snRNAs should encode for proteins at some points during the *C. elegans* life cycle. Perhaps, these proteins are regulated in a tissue and/or time specific manner or even in response to a stimulus, making their detection difficult. It could be that they have totally independent functions when translated to proteins and as chimeric "sn-mRNAs" following DNA damage. Maybe these chimeric RNAs act as decoys for microRNA regulation or scaffolds for DNA repair proteins or the chromatin remodeling machinery.
In the future, further effort should be made to determine whether these chimeric "sn-mRNAs" are just an artefact because of INTS-6 re-localization after DNA damage or if indeed there is regulated coordination and they form an active part of the DNA damage response and/or DNA repair.

Conclusions

1. The *C. elegans* ts *t1903* mutant has a point mutation in the *ints-6* gene (C →T) that results in a swap from Ser 850 to a Phe. The embryos of this mutant die at 25ºC due to embryonic morphogenetic defects.

2. The *C. elegans ints-6* gene and its human ortholog, Ints6, show a major nuclear localization.

3. There is an Integrator complex in *C. elegans* comprised of at least eleven subunits (INTS-1, INTS-2, INTS-3, INTS-4, INTS-5, INTS-6, INTS-7, INTS-8, INTS-9, INTS-11 and INTS-13), that functions in the 3'-end processing of snRNAs (U1, U2, U4, U5, U6 and SL).

4. Knockdown of human Ints6 in the RPE cell line results in 3'-end processing defects of at least U1 and U2 snRNAs and an over-proliferation phenotype.

5. Either the *t1903* mutation or the knockdown of any of the *C. elegans* Integrator complex subunit leads to 3'-end snRNA processing defects and a read-through of the RNAP II. As a result of this read-through, the genes downstream of snRNAs located *in sense* formed chimeric "sn-mRNAs", which are mostly correctly spliced and polyadenylated.

6. The chimeric "sn-mRNAs" composed of "U1-mRNAs" or "U2-mRNAs" are not translated into proteins or peptides. However, the "SL2-mRNA" composed of *sls-2.8* and the Y75B8A.23 gene is translated into a protein.

7. *C. elegans ints-6* is involved in DNA repair. INTS-6 formed foci upon DNA damage by γ-radiation. *C. elegans* INTS-6 knockdown abrogates RAD-51 recruitment to DSBs and phosphorylation on Tyr15 Cdk-1 upon DNA damage by γ-radiation. Additionaly, INTS-6 Ser850 might be phosphorylated in response to DNA damage.

8. In *C. elegans*, chimeric "sn-mRNAs" are formed upon γ-radiation. Formation of these RNAs could be a physiological process mediated by specific regulation of the Integrator complex in response to DNA damage.

Materials & Methods

1. Strains

1.1. *E. coli* strains

E. coli strains used in this work are shown in the table below which specifies their genotype, purpose and use.

STRAIN	GENOTYPE/DESCRIPTION	APPLICATION	SOURCE
OP50*	E. coli B, Uracil auxotroph, ampicillin resistant	*C. elegans* food	CGC*
DH5α	*F- endA1 glnv44 thi-1 recA1 relA1 gyrA96 deoR nupG 80dlacZ M15 (lacZYA- argF) U169, hsdR17(rK- mK+), λ -*	plasmid amplification; cloning	CIC
NovaBlue	*endA1 hsdR17*(r_{K12}^{-} m_{K12}^{+}) *supE44 thi-1 recA1 gyrA96 relA1 lac* F'[*proA*$^{+}$B^{+} *lac*I^{q}ZΔ*M15*::Tn*10*] (Tet^{R})	plasmid amplification; cloning	Novagen
HT115(DE3)	E. coli F-, mcrA, mcrB, IN(rrnD-rrnE)1, rnc14::Tn10(DE3 lysogen: lavUV5 promoter -T7 polymerase) (IPTG-inducible T7 polymerase) (RNAse III minus), tetracycline resistant	RNAi feeding	CIC

Table 6. *E. coli* strains used in this work.

1.2. Cell lines strains

Cell lines used in this work are shown in the table below which specifies their names, origin, source and the media used.

CELL LINE	ORIGIN	MEDIUM	SOURCE
HEK293T	Human Embryonic Kidney	DMEM + 10% FBS	Dr. Sacristán lab
U2OS	Human Bone Osteosarcoma	DMEM + 10% FBS	Dr. Sacristán lab
hTERT RPE-1	Retinal Pigment Epithelial	DMEM-F12 + 10% FBS	Dr. Sacristán lab

Table 7. Cell lines used in this work.

1.3. *C. elegans* strains

The *C. elegans* strains used in this work are shown in the table below which specifies their genotype and source.

STRAIN	GENOTYPE	SOURCE
N2	*C. elegans* wild isolate	CGC
GE3632	*unc-24*(*e138*) *ints-6*(*t1903*) IV	Schnabel´ s Lab
NB327	*dic-1(tm1615)* IV/*nT1[qIs51]V*	CGC
HT1593	*unc-119(ed3)* III	CGC
OH11117	otEx4963 [lsy-6p(fosmid delta 150bp downstream)::GFP + *ttx-3*:mCherry]. otIs306 [*hsp-16.2*::*che-1*::3xHA + rol-6]	CGC
OP217	*unc-119(ed3)* III; ddIs172 [*aly-2*::TY1::eGFP::3xFLAG(92C12) + *unc-119*(+)]	CGC
SJ4199	zcIs40[*dve-1*p::dve-1::3xMYC-HIS tag + *myo-3*p::GFP]	CGC
JCP294	*ints-6(t1903)* IV	Our lab
JCP301	jcpSi3[pJC50(*ins-37*p::*ins-37*::eGFP::*ins-37*UTR,*unc-119*(+))] II; *unc-119(ed3)* III	Our lab
JCP341	jcpSi10[pJC51(*ints-6*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))] II;*unc-119(ed3)* III	Our lab
JCP342	jcpSi11[pJC54(*ints-6p*::*ints-6*(S850A)::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))] II;*unc-119(ed3)* III	Our lab
JCP343	jcpSi12[pJC55(W04G5.8p::W04G5.8::eGFP::W04G5.8UTR, *unc-119*(+))] II;*unc-119(ed3)* III	Our lab
JCP378	jcpSi19[pJC56(*eft-3*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))]II;*unc-119(ed3)*III	Our lab
JCP387	jcpSi24[pJC57(H27M09.5p::H27M09.5::3xFLAG::eGFP::H27M09.5UTR,*unc-119*(+))]II;*unc-119(ed3)*III	Our lab
JCP382	jcpSi11[pJC54(*ints-6p*::*ints-6*(S850A)::3xFLAG::eGFP::*ints-6* UTR,*unc-119*(+))]II;*unc-119(ed3)*III; *dic-1(tm1615) IV*	Our lab
JCP383	jcpSi10[pJC51(*ints-6*p::*ints-6*(wt)::3xFLAG::eGFP::*ints-6*UTR,*unc-119(+)*)] II;*unc-119(ed3)* III; *dic-1(tm1615) IV*	Our lab
JCP394	jcpSi31 [pJC58 (Y75B8A.23p::Y75B8A.23::3xFLAG::eGFP::Y75B8A.23UTR, *unc-119*(+))]II; *unc-119(ed3)*III	Our lab
JCP405	jcpSi37 [pJC60(F08H9.3p::F08H9.3::3xFLAG::eGFP::F08H9.3UTR, *unc-119*(+)]II; *unc-119(ed3)*III	Our lab
JCP462	*ints-6*(jcp1)[*ints-6*::3xFLAG)]	Our lab
JCP472	*ints-6*(jcp4)[*ints-6*(S850E)::3xFLAG)]	Our lab
JCP479	jcpSi53 [pJC63 (3-tags-in-3-frames (HA:MYC::TY) in promoter of F08H9.3, *unc-119*(+))] II; *unc-119(ed3)* III	Our lab
JCP483	*ints-6*(jcp6) [*ints-6*(S850A)::3xFLAG)]	Our lab
JCP483	jcpSi55 [pJC64 (3-tags-in-3-frames (HA::MYC::TY) in promoter of W04G5.8, *unc-119*(+))] II;*unc-119(ed3)* III	Our lab
JCP505	jcpEx5[pJC56(*ints-6*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))] II;*unc-119(ed3)* III	Our lab

Table 8. *C. elegans* strains used in this work.

2. Growth medium and culture conditions

2.1. *E. coli* culture

E. coli cultures were grown at 37ºC O/N in LB (Luria-Bertani) liquid medium with constant shaking (220-240 rpm). Moreover, the required antibiotic was added to LB liquid medium, usually 100 mg/ml ampicilin.

2.2. Cultivating cell lines

Cell lines were cultivated on cell culture plates, growing in a monolayer at 37ºC and in an atmosphere of 5% CO_2 and 98% relative humidity. The culture mediums employed were supplemented with 10% Fetal Bovine Serum (FBS), 2mM L-glutamine and penicillin (100 units/ml) and streptomycin (100 µg/ml) antibiotics to prevent any microbial contamination.

Manipulation of the cultured cells was performed under sterile conditions in laminar flow hoods using sterile materials for direct contact with cell cultures.

After reaching 90% confluence, cells were split by washing once with PBS and treating them with 0.05% Trypsin-EDTA (GIBCO Invitrogen) for 5 to 10 min. Trypsin digestion was stopped by adding medium. Next, cells were plated in appropriate dilutions for further maintenance.

2.3. *C. elegans* culture on agar plates

C. elegans strains were maintained as described by Brenner (1974) on Nematode Growth Medium (NGM) agar plates seeded with a lawn of *E. coli* OP50*. NGM agar was prepared by mixing 0.3% (w/v) NaCl, 0.25% (w/v) peptone, 1.7% (w/v) agar and Milli-Q water. Then, the medium was autoclaved and subsequently the salts and the supplements necessary for proper growth of nematodes were added: 1 mM $CaCl_2$, 1 mM $MgSO_4$, 25 mM K-phosphate buffer, pH 6.0 and 5 mg/l cholesterol. In addition, ampicilin (100 mg/ml) and nystatin (0.004%) were added to *E. coli* OP-50* growth cultures to prevent any contamination of bacteria or fungi respectively.

The nematodes were grown on these plates at 15º, 20º or 25ºC depending on the purpose of the experiment. The *t1903* mutant is ts so it was regularly grown at 15ºC, except in those cases where we wanted to see the phenotype at 25ºC.

Worms were transferred from an old plate to a new plate either by cutting out a chunk of agar with a spatula and putting it on a new plate, or by picking up individual worms with a platinum wire mounted on a Pasteur pipette. To avoid dehydration or contamination, plates were wrapped in parafilm and stored in cardboard boxes within the incubators.

If larger quantities of worms were needed, they were grown on egg plates following the protocol described by Hochbaum et al. 2010. These egg plates were normal NGM agar seeded with a mixture of egg yolks and OP50 grown in LB. Thus, large worms populations are able to grow on the same plate without being starved. First, a pre-mixture was made by separating six egg yolks and adding them to 500 ml of LB. Afterwards, this pre-mixture was incubated for 1 h at 60ºC. After cooling, 40 ml of OP50 culture was added. Plates were seeded with 5 ml of the mixture. Any remaining liquid was discarded the next day.

3. Stock maintenance and preservation

3.1. *E. coli* glycerol stocks

E. coli stocks were stored in cryotubes at -80ºC. Single colonies from *E. coli* strains were grown on LB medium with the corresponding antibiotic (O/N; 220-240 rpm). Next, cultures were mixed with 20% glycerol total volume and frozen in liquid nitrogen for storage (-80ºC).

When a strain was required, a pipette tip was introduced into the cryotube without thawing it to withdraw a small amount to be used to seed the bacteria on an LB agar plate. If the bacterial strain was transformed with a plasmid, a single colony was checked to verify that it had maintained the plasmid and that clone was correct.

3.2. Cell line stocks

Cell lines were maintained in liquid nitrogen for long-term storage. The freezing process should be done slowly to avoid cell death due to possible crystal formation. Cells were trypsinized and spun down (1000 rpm, 5 min) at RT. Afterwards, cells were re-suspended in culture medium with 20% FBS. In addition, dimethyl sulfoxide (DMSO, Sigma), a cryoprotectant, was added to a

10% final concentration. Cells were frozen in cryotubes at -80ºC in isopropanol containers for a week. Afterwards, they were preserved permanently in liquid nitrogen.

In contrast, thawing and recovery of cells from liquid nitrogen had to be done quickly. Cryotubes were removed from liquid nitrogen and maintained in dry ice until they thawed, at which point they were immediately placed in a 37ºC water bath. Cells were washed with culture medium plus 10% FBS to eliminate DMSO. Finally, cells were re-suspended in fresh medium, transferred to a cell culture plate and placed in an incubator.

3.3. *C. elegans* stocks

3.3.1. Freezing and recovery of *C. elegans* stocks

C. elegans strains were frozen and stored indefinitely at -80°C and in liquid nitrogen. Freshly starved worms of two or three 90 mm plates (containing many L1-L2 stage larvae) were washed off plates with M9 buffer and collected in a 15 ml Falcon tube. The tube was incubated on ice until the worms had settled to the bottom of the tube. The supernatant was removed and the volume of the worm suspension was adjusted to 2 ml with M9 buffer. After adding an equal volume of freezing solution (0.3% (w/v) KH_2PO_4, 0.6% (w/v) Na_2HPO_4, 0.5% (w/v) NaCl, 1 mM $MgSO_4$, 30% glycerin (v/v)), the worm suspension was mixed and aliquoted into cryotubes (1 ml/tube). For every three tubes of a strain that were frozen, another one was kept as a freezing control (to check strain survival after freezing). To promote viability, strains were frozen slowly. The cryotubes were packed in a Styrofoam box and frozen at -80°C. Thus, the temperature decreased slowly and the worms could easily survive freezing. For recovery of a *C. elegans* stock, a frozen aliquot was thawed at RT and transferred onto a fresh NGM plate.

3.3.2. *C. elegans* synchronization or decontamination (bleaching)

C. elegans eggs are surrounded by a cuticle that protects them from harmful environmental factors such as chemicals. This feature was used to clean them when the strains were contaminated with fungi or bacteria, or to obtain synchronized populations. Worms were harvested when the plates contained many gravid hermaphrodites and washed several times with M9 buffer.

Worms were dissolved by treatment with bleaching solution (12% NaClO) for 10 min with vigorous shaking in between. To avoid an excess of the reaction and thereby damage to the embryos, destruction of the adult tissue was monitored under a dissecting microscope. The released eggs were collected by centrifugation (1500 rpm, 2 min) and then washed three times with M9 buffer in order to get rid of any hypochlorite residues.

The eggs were either directly transferred to fresh NGM plates or left incubating in M9 buffer O/N (usually at 20ºC with rotation) to get a synchronized population since the absence of food, the hatched larvae arrested at the L1 stage.

4. Nucleic acid manipulation techniques

4.1. RNA extraction

All total RNA extractions, either *C. elegans* RNA extractions or cell line RNA extractions, were performed using the *mirVana* ^{TM} *miRNA isolation kit* (Ambion) following the manufacturer's protocol for total RNA isolation. This kit enables recovery of small RNAs such as the snRNAs.

In the case of *C. elegans* samples, worms from 3 to 5 plates were washed off with M9 buffer and collected in 50 ml Falcon tubes, allowing them to settle at the bottom of the tube. Worm pellets were subsequently washed off with M9 buffer until there were no visible bacterial remains. Worm pellets were transferred to Eppendorf tubes and as much supernatant as possible was removed. Next, 300 µl of *Lysis Binding Buffer* provided by the kit was added to each sample and they were homogenized. Worm tissue homogenization was carried out with the assistance of a polytron pre-chilled with liquid nitrogen. In the case of adherent cell cultures, two cell culture plates were lysed directly on the plates by adding *Lysis Binding Buffer*, collected with a rubber spatula and pipetted into a tube. To obtain a homogenous lysate, tubes were vigorously vortexed. The following steps of RNA extraction were the same for *C. elegans* or cultured cells. Basically, 30 µl of *miRNA Homogenate Additive* (provided by the kit) was added to each lysate and mixed well by vortexing. Then, they were left on ice for 10 min followed by organic extraction, 300 µl of acid-phenol:chloroform was added to each

sample and tubes were vigorously vortexed (30-60 sec). Samples were centrifugated (10000x g, 5 min) at room temperature to separate the aqueous and organic phases. After centrifugation, the aqueous (upper) phases were carefully removed without disturbing the lower phases, and transferred to fresh tubes. Then, 1.25 volumes of 100% ethanol were added to each sample and the lysate/ethanol mixtures were passed through Cartridge Filters (10000x g, 15 sec) followed by three washing steps (washing solutions provided by the kit). Finally, each RNA sample was eluted from the filters with 35 µl of pre-heated nuclease free water. The RNA concentration was determined photometrically with a *NanoDrop™ 1000* spectrophotometer at 260 nm.

4.2. RNA deep sequencing.

RNA deep sequencing was performed in the genomics platform of the CIBIR (http://www.cibir.es/en/technology-platforms-and-services/genomics-and-bioinformatics). Sequencing libraries were prepared from 1 mg of RNA using *TruSeq® RNA* sample preparation kit according to the manufacturer's instructions (Illumina). RNA quality and integrity was evaluated using the *Experion™ Automated Electrophoresis System* (Bio-Rad). Briefly, the RiboMinus RNA sequencing libraries were prepared using *Ribo-Zero® rRNA Removal Kit* (Illumina). The poly(A) sequencing libraries were prepared by isolating polyadenylated mRNAs using poly-T oligo-attached magnetic beads. Following purification, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using RT and random primers. Second strand cDNA synthesis followed, using DNA Polymerase I and RNase H. The cDNA fragments were then passed through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were then purified and enriched with PCR to create the final cDNA library. Pools of indexed libraries were mixed (multiplexed) in equimolar ratios to yield a total oligonucleotide mix of 10 nM. The resulting libraries were sequenced on the *Illumina Genome Analyzer IIx* to generate 150 bp single-end reads.

4.3. Polymerase Chain Reaction (PCR)

Specific DNA sequences were amplified by PCR using *GeneAmp® PCR System 9700* thermal cyclers (Applied Biosystems). PCR conditions were adjusted in each reaction based on the DNA fragment to be amplified and the primer pairs used (Table 9).

NAME	SEQUENCE (5'-3')	DESCRIPTION/OBJECTIVE
18SF	CGCCGCTAGAGGTGAAATCC	*H. sapiens* 18S/ Fig. 39
18SR	CTTCGCTCTGGTCCGTCTT	*H. sapiens* 18S/ Fig. 39
109	CCAGGAATTGCTGATCGTATG	*C. elegans* actins/Fig. 44D
110	GGAGAGGGAAGCGAGGATAG	*C. elegans* actins/Fig. 44D
133	GTTTTCCCAGTCACGACGTT	L4440 sequencing/ Clonings
134	TGGATAACCGTATTACCGCC	L4440 sequencing/Clonings
370	GCTGTCGTTTCGATCTCTCG	*sls-2.8* /Fig. 44 C
373	TGTCGTGAGTAGGTGTGCAA	Y75B8A.23 /Fig. 44 C
448	TTACTAAGCTTCCATAGATCGCCGTAATCGT	*ints-8* cloning
449	TTACTCTCGAGGTGAGTGGGCCGTGAAGTAT	*ints-8* cloning
500	TAATGAATTCAATGCCCATCTTACTGTTCCTG	*ints-6* cloning in pEGP-C/N
501	TAATGGTACCTTAATTGCTATTAATATGGTTGATCTG	*ints-6* cloning in pEGFP-C
503	ACGCGTCGACACATTGCTATTAATATGGTTGATCTG	*ints-6* cloning in pEGFP-N
513	CCTGTTTCTGGGCTCTTTCTT	RNU1-109P 3' region/Fig.39
514	GCCTCGCCAACATAGTGAAAC	RNU1-109P 3' region/Fig.39
516	GAACTTTGAAGCCACTCAACC	RNU1-46P 3' region/Fig. 39
517	CCCTGCATACTCGAACACTCA	RNU1-46P 3' region/Fig. 39
518	CAAGGAGCTGAAAGGCACTGA	RNU2-69P 3' region/Fig. 39
519	GACCTGTGCTTTCTGGGGTAG	RNU2-69P 3' region/Fig. 39
520	GAGTGCAGTGGTGTGATCAAG	RNU2-18P 3' region/Fig. 39
521	GATGGAGGGGTGGTTTGAATA	RNU2-18P 3' region/Fig. 39
567	AAACCACGAGTTGGACAAGG	*ints-1* cloning
568	TCAAATCAATCGGCATTTCA	*ints-1* cloning
753	GTGTGGCAGTCTCGAGTTGA	H27M09.8/Fig. 44A
754	TTGAACCTTTTCGTCGGAAC	H27M09.5/Fig. 44A
755	TGGAACCTAGGGAAGACTCG	F08G2.9/Fig. 44B
756	TTGAACTTGTCCGGGATTCT	*ins-37*/Fig. 44B
757	ATTTTTGGAACCCAGGGAAG	W04G5.11/Fig. 44B
758	GTGGAGATTTCTGCGACACA	W04G5.8/Fig. 44B
759	TGACCTATGTGGCAGTCTCG	F08H9.10/Fig. 44A, 59
760	TCGACAATCTCATTCCGACA	F08H9.3/Fig. 44A, 59

Table 9. Primers used in this work

Generally, the *GoTaq® DNA Polymerase* (Promega) was used. The final concentrations in each PRC reaction were: 1x *GoTaq® Reaction Buffer* (1.5 mM $MgCl_2$), 0.2 mM dNTPs, 0.2 µM upstream primer and downstream primer, 2,5 units *GoTaq® DNA Polymerase* plus the required amount of DNA

template (<500 ng) in each case. In the case of the upstream primers that target the snRNAs, the concentration used was 0.8 µM.

4.4. Reverse Transcription Polymerase Chain Reaction (RT-PCR)

First, RNA was treated with DNase to eliminate any DNA contamination. In each sample, a total reaction of 10 µl contained: 500 ng RNA, 1µl *RQ1 RNase-Free DNase* (Promega), 1x *RQ1 DNase 10X Reaction Buffer* and DEPC water. Reactions were incubated for 30 min at 37ºC and stopped by adding 1 µl STOP solution (Promega) and incubating them for 10 min at 65ºC. cDNA synthesis was performed using a *Transcriptor First Strand cDNA Synthesis Kit* (Roche). First, Random Hexamers (60 µM final concentration) were added to each sample and to ensure RNA secondary structures denaturation, template-primer mixtures were incubated for 10 min at 65ºC and immediately cooled on ice.

For each tube containing the template-primer mix, the remaining components were added to a final concentration: *Transcriptor Reverse Transcriptase* Reaction Buffer 1x (8 mM $MgCl_2$), Protector RNase Inhibitor 20U, Deoxynucleotide Mix 1 mM each, and *Transcriptor Reverse Transcriptase* 10U.

Samples were incubated for 10 min at 25°C, followed by 60 min at 50°C. Afterwards, *Transcriptor Reverse Transcriptase* was inactivated by heating (5 min at 85ºC min) and then placing the tubes on ice. Tubes were stored at 4°C for immediate use or at -20°C for longer periods.

4.5. RT-qPCR

Real time quantitative PCR was used to analyze mRNA expression levels. In this case, we studied the snRNAs 3' downstream regions RPE cell line. Total RNA was obtained using the *mirVana™ miRNA isolation kit* (Ambion) as described in MM 4.1. cDNA synthesis (MM 4.4) was performed with 1 µg total RNA. All real-time PCR reactions were performed using the *iQ5 Multicolor Real-Time PCR Detection System* (Bio-Rad) and the amplifications were done using the specific primers shown in Table 9 (18SF, 18SR, 513-521) and the iQ™ SYBR® Green Supermix (Bio-Rad).

4.6. Separation and isolation of DNA fragments

Agarose gel electrophoresis was used to separate DNA fragments according to their size. The term electrophoresis refers to the movement of charged molecules in response to an electric field, resulting in their separation. When electrophoresis is performed in acrylamide or agarose gels, the gel serves as a size-selective sieve during separation. The gel porous structure allows smaller molecules to travel more rapidly than larger ones.

To prepare the agarose gel, 0.7% to 2% *Agarose D1 Low EEO* (Conda) was dissolved in 1x TBE buffer (10x TBE Fisher Scientific) and *SYBER®Safe* 10000x (Invitrogen) was added. To load the DNA molecules into gel wells, 6x loading dye (Fisher Scientific) was added to DNA samples and they were separated in agarose gels and TBE buffer using the *Sub-Cell® GT Agarose Gel Electrophoresis Systems* (Bio-Rad) at a setting of 100 Volts (V) for 30-60 min.

DNA was visualized under UV light (320 nm) and sizes of the DNA fragments were estimated by comparing them with the bands of the molecular size marker *1 kb Plus DNA ladder* (Invitrogen).

If necessary, DNA bands of interest were cut out of the gel and purified from the agarose gel with the *Jet Quick Gel Extraction Spin Kit* (Genomed) according to the manufacturer's protocol.

4.7. DNA digestion

PCR products and plasmids were digested with restriction endonucleases using buffers and appropriate incubation conditions for each enzyme according to the manufacturer's instructions.

4.8. DNA ligation

Ligation was performed in a total reaction volume of 10 µl, usually containing 10 ng vector and 3-fold molar excess of insert DNA, 1 µl of T4 DNA ligase (1 unit/µl) (Invitrogen) and 5x reaction buffer (provided by the manufacturer). After O/N incubation at 15°C, 2 µl of the reaction was transformed into *E. coli*

competent cells (*NovaBlue Singles™ Competent Cells*, Merck Millipore). Transformed bacteria were plated on agar plates with the corresponding antibiotic and incubated at 37°C O/N. Next, single colonies were checked by PCR to see if the plasmids carried the correct insert. Afterwards the correct colonies were cultured (MM 2.1) to obtain the desired plasmid (MM 4.11)

4.9. DNA sequencing

DNA molecules were sequenced on an *ABI Prism 3100 sequencer* (Applied Biosystems) by the Genomics Service of the Cancer Research Institute (Salamanca). The final volume of each sequencing reaction was 8 µl, containing 3 pM of the necessary primer and an adequate amount of DNA: 50-150 ng PCR product or 100-600 ng plasmid DNA.

4.10. Plasmid cloning:

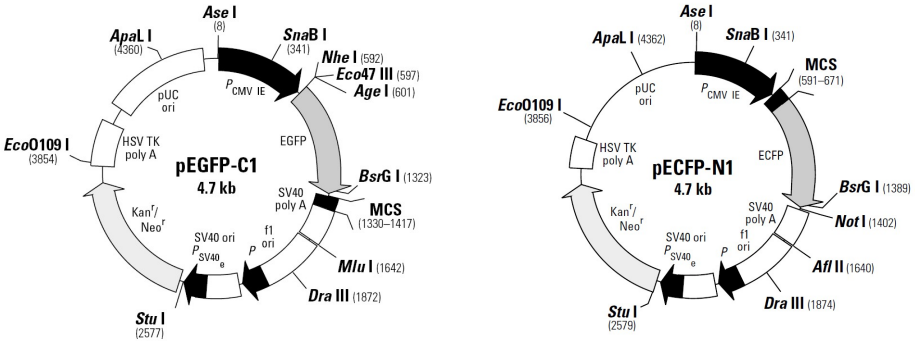
4.10.1. Mammalian plasmids:

Human Ints6 was cloned into the backbone commercial vectors pEGFP-C1 and pEGFP-N1, generating the pBS15 (eGFP-Ints6) and pBS16 (Ints6-eGFP) plasmids respectively. Backbone vectors are shown below.

The Ints6 cDNA sequence was amplified by PCR from a commercial plasmid (pCMVSport6, Addgene). Ints6 in pBS15 was amplified using the 500 (5'-TAATGAATTCAATGCCCATCTTACTGTTCCTG-3') and 501 (5'-TAATGGTACCTTAATTGCTATTAATATGGTTGATCTG-3') primers. The PCR product obtained and the pEGFP-C1 backbone vector were digested with *Eco*RI and *Kpn*I restriction endonucleases (MM 4.7). Afterwards, they were ligated (MM 4.8) and the ligation was used to transform *E. coli* DH5α competent cells (MM 6.1). Next, a screening of the colonies obtained was performed to check the positive transformants, either by PCR of the colonies or by digestion with restriction endonucleases of the previously purified plasmids from each colony.

The pBS16 plasmid was generated similarly but the 500 (5'-TAATGAATTCAATGCCCATCTTACTGTTCCTG-3') and 503 (5'-ACGCGTCGACACATTGCTATTAATATGGTTGATCTG-3') primers were used to amplify Ints6. This time, the PCR product and the pEGFP-N1 backbone vector

were digested with the restriction endonucleases *Eco*RI and *Sal*I. Backbone vectors used to generate the pBS15 and pBS16 plasmids are shown below.

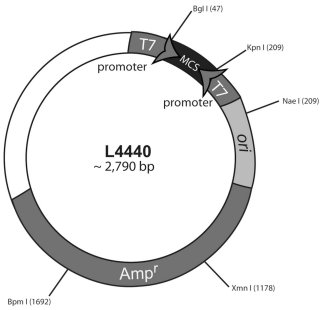


4.10.2. *C. elegans* plasmids:

4.10.2.1. Plasmids for RNAi silencing

When we wanted to silence a particular *C. elegans* gene and it was not in any RNAi clone library, (Rual et al., 2004; Kamath et al., 2003) they were clone into the empty L4440 vector shown below.

A 3860 bp fragment of *ints-1* was amplified using the 567 (5'-AAACCACGAGTTGGACAAGG-3') and 568 (5'-TCAAATCAATCGGCATTTCA-3') primers from cDNA. Then, the PCR product was inserted into the intermediate vector pJET1.2/blunt according to *cloneJET PCR Cloning Kit* manufacturer's instructions. Next, the intermediate plasmid was generated and the L4440 backbone vector was digested with *Not*I and *Xba*I restriction endonucleases (MM 4.7). The vector and the insert were ligated (MM 4.8) and the ligation product was used to transform *E. coli NovaBlueSingles ™ Competent Cells* (MM 6.1). Similarly, a fragment 1878 bp of *ints-8* was amplified from cDNA using the 448 (5'-TTACTAAGCTTCCATAGATCGCCGTAATCGT-3') and 449 (5'-TTACTCTCGAGGTGAGTGGGCCGTGAAGTAT-3') primers. Then, the PCR product and the plasmid were digested with *Xho*I and *Hind*III restriction endonucleases (MM 4.7). The vector and the insert were ligated (MM 4.8) and the ligation was used to transform *E. coli NovaBlueSingles ™ Competent Cells* (MM 6.1). Afterwards, colonies were screened by PCR (MM 4.3) using the 133 (5'-GTTTTCCCAGTCACGACGTT-3') and 134 (5'-TGGATAACCGTATTACCGCC-3') primers.



T7 promoter CCGGCAGATCTGATATCATCGATGAATTCGAGCTCCACCGCGGTGGCGGCCGCTC
Bgl II Sac I Sac II Not I Eag I

TAGATCTAGAACTAGTGGATCCACCGGTTCCATGGCTAGCCACGTGACGCGTGGATCCCCCGGGCTGCA
Xba I Spe I Age I Nco I Nhe I Mlu I Xma I Sma I Pst I

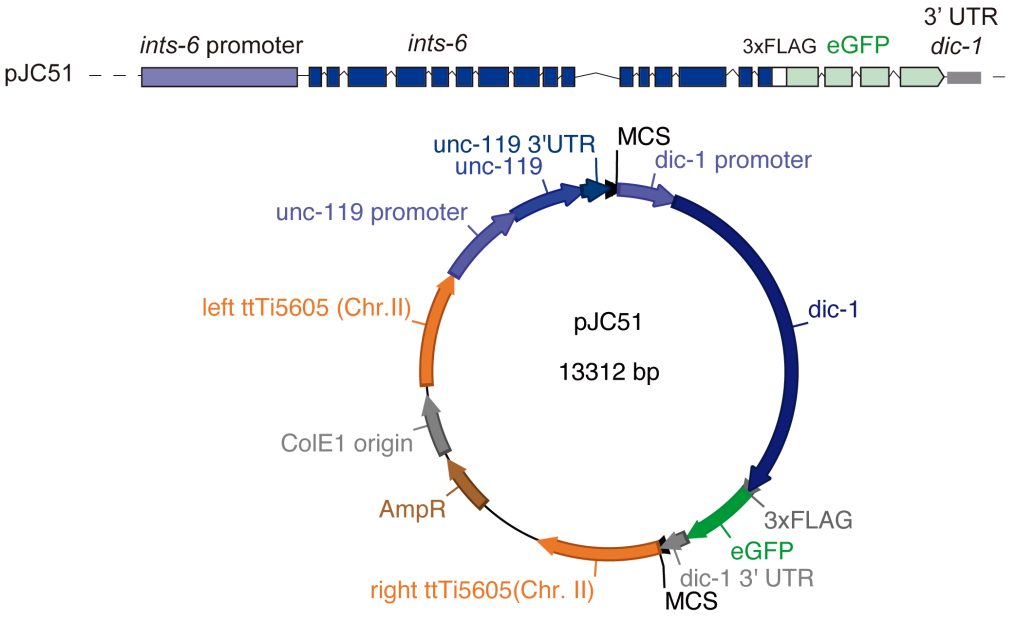
GGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATT T7 promoter
Hind III Hinc II Sal I Xho I Apa I Kpn I

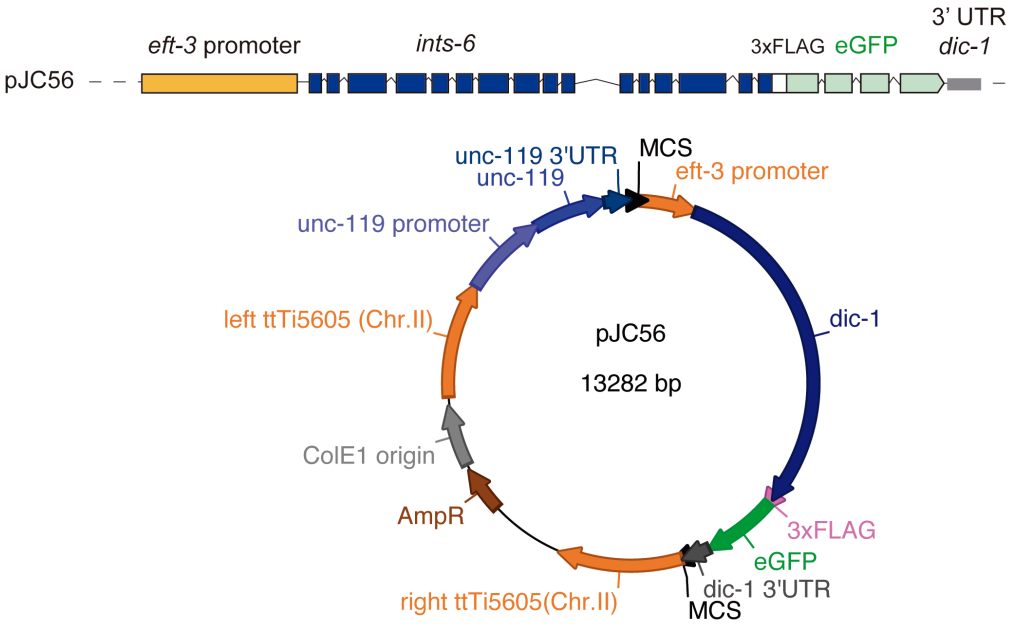
L4440 backbone vector

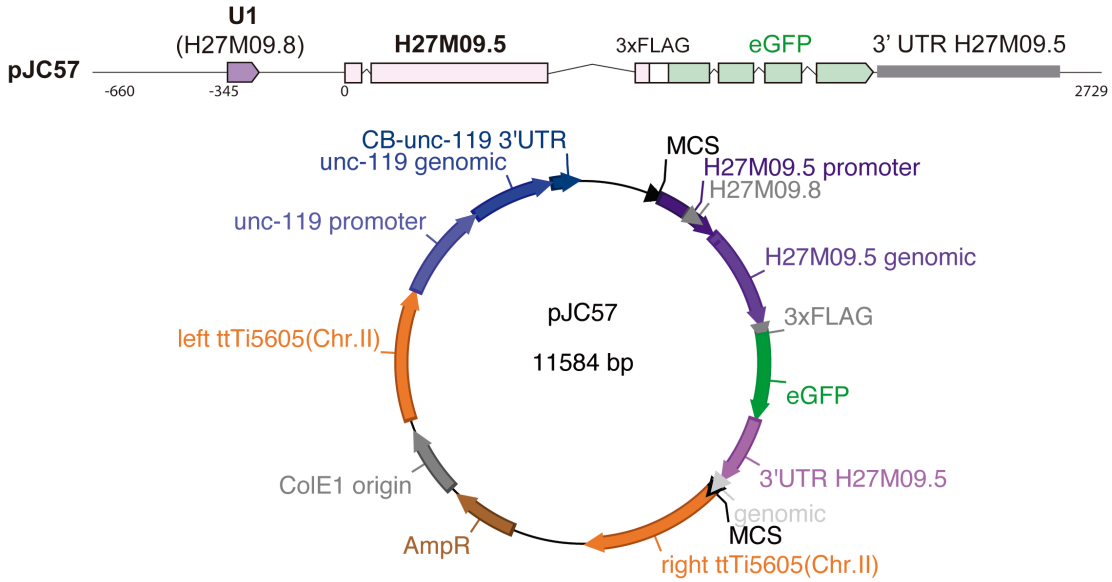
4.10.2.2. Plasmids to generate *C. elegans* transgenics

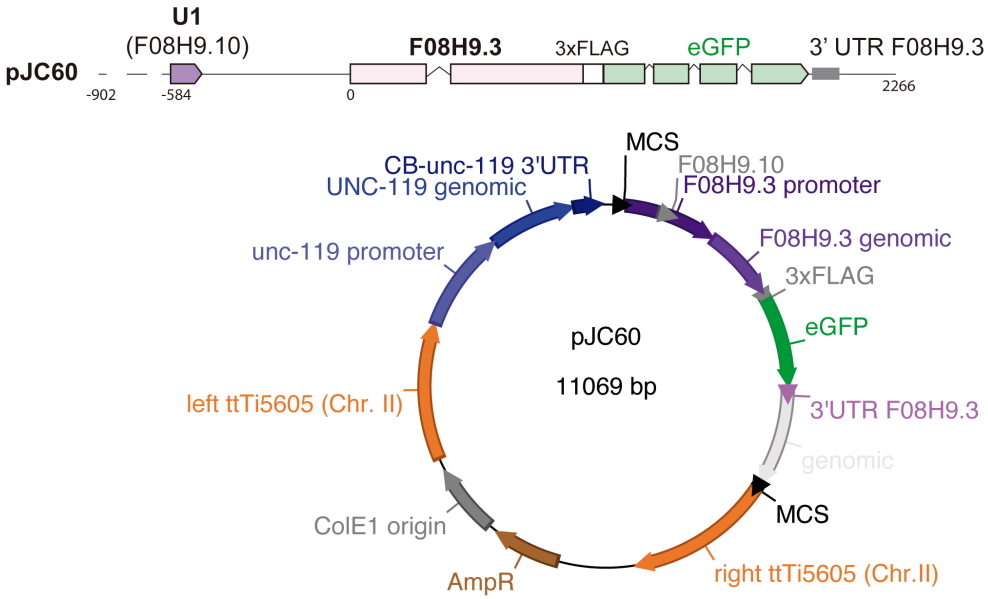
We designed the following plasmids that were created by the Knudra enterprise: pJC51, pJC56, pJC57, pJC60, pJC55, pJC50, pJC58, PJC54, pJC63 and pJC64. All of them were used to integrate the desired gene constructs into chromosome II using the mosSCI system. They have two "arms" called *ttTi5605*, which are homologous sequences (1500 bp) to chromosome II that are used for recombination and integration of the desired gene sequence at this location on chromosome II in the *C. elegans* genome (Frokjaer-Jensen et al. 2008).

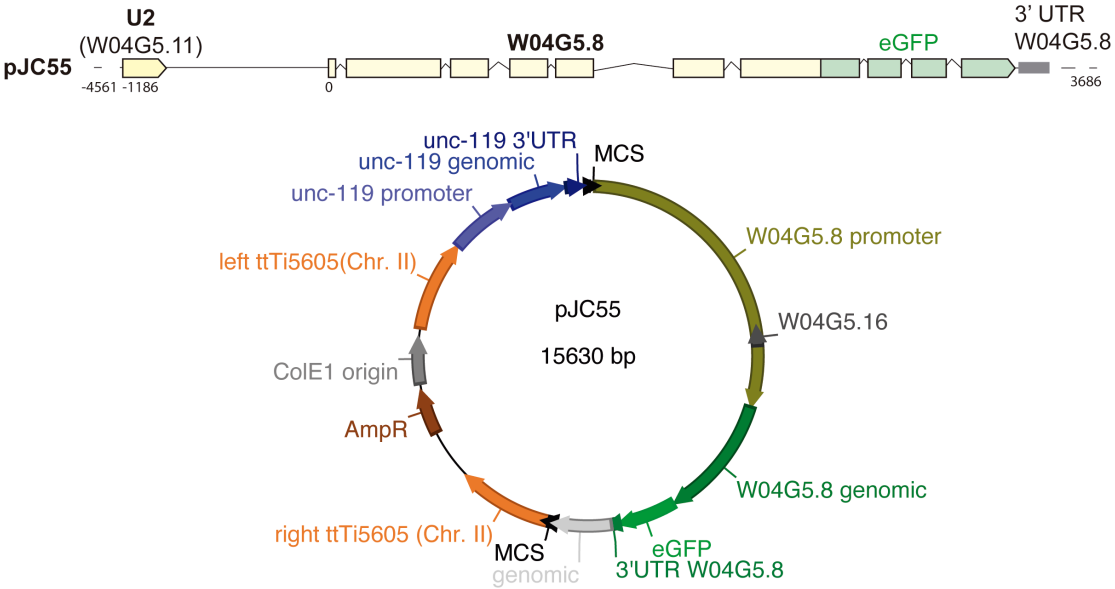
A schematic drawing of each insert designed and the full plasmid is shown below.

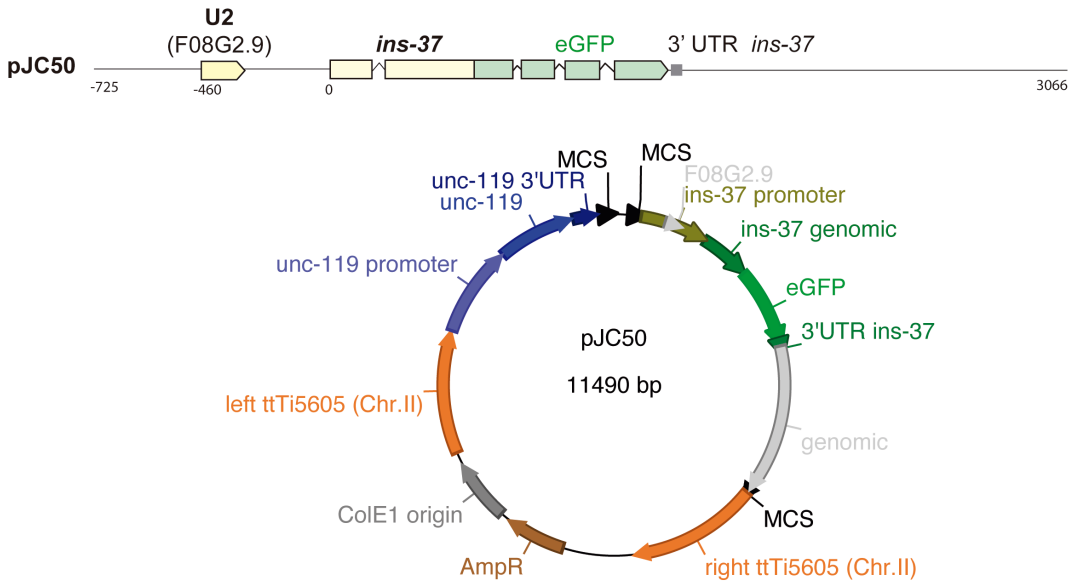

pJC51
ints-6 promoter
ints-6
3xFLAG
eGFP
3' UTR
dic-1
unc-119 3'UTR
unc-119
MCS
dic-1 promoter
unc-119 promoter
left ttTi5605 (Chr.II)
pJC51
13312 bp
dic-1
ColE1 origin
AmpR
3xFLAG
eGFP
dic-1 3' UTR
right ttTi5605(Chr. II)
MCS

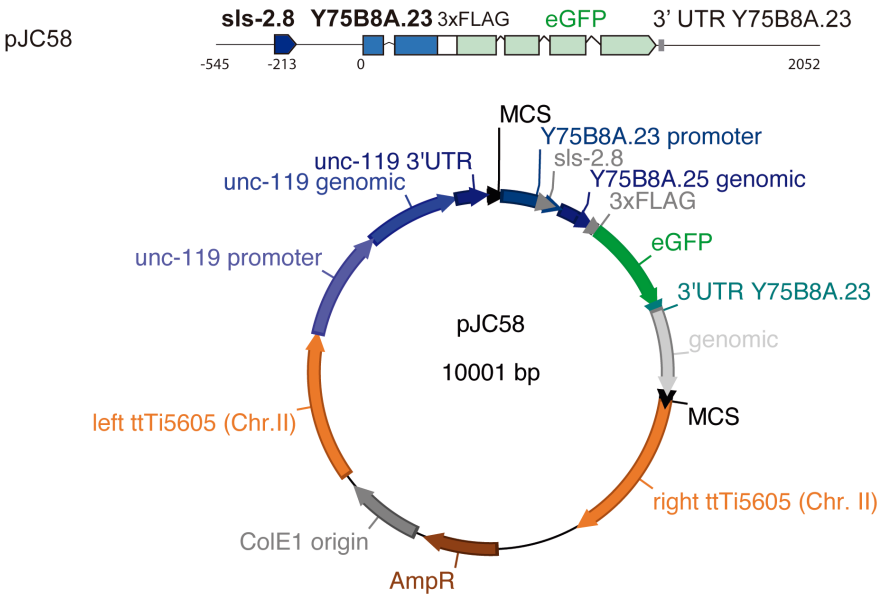

pJC56
eft-3 promoter
ints-6
3xFLAG
eGFP
3' UTR
dic-1
unc-119 3'UTR
unc-119
MCS
eft-3 promoter
unc-119 promoter
left ttTi5605 (Chr.II)
pJC56
13282 bp
dic-1
ColE1 origin
AmpR
3xFLAG
eGFP
dic-1 3'UTR
right ttTi5605(Chr.II)
MCS

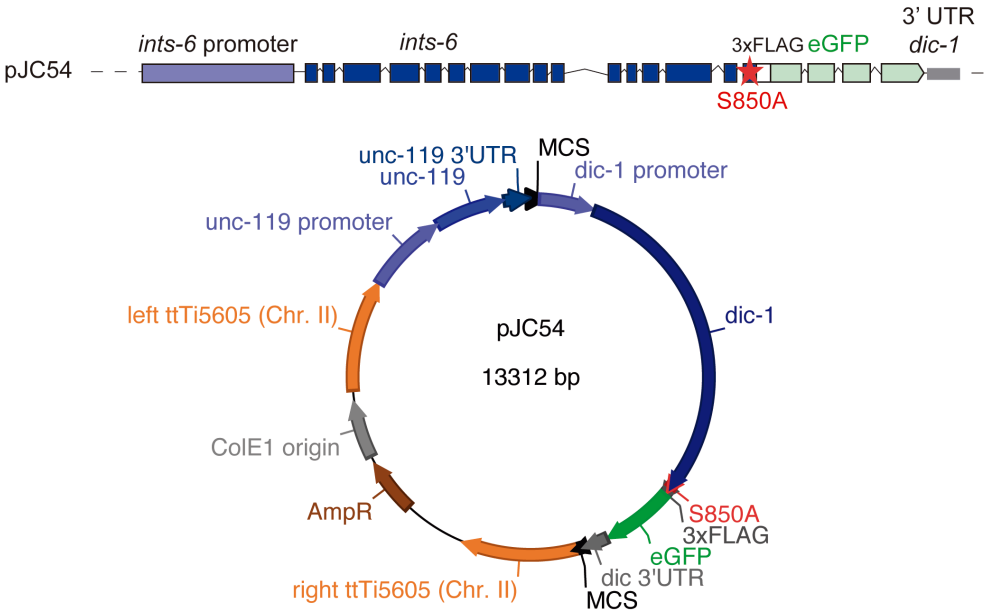

pJC57
U1
(H27M09.8)
H27M09.5
3xFLAG
eGFP
3' UTR H27M09.5
-660
-345
0
2729
CB-unc-119 3'UTR
unc-119 genomic
MCS
H27M09.5 promoter
H27M09.8
unc-119 promoter
H27M09.5 genomic
pJC57
11584 bp
3xFLAG
left ttTi5605(Chr.II)
eGFP
3'UTR H27M09.5
genomic
ColE1 origin
MCS
AmpR
right ttTi5605(Chr.II)

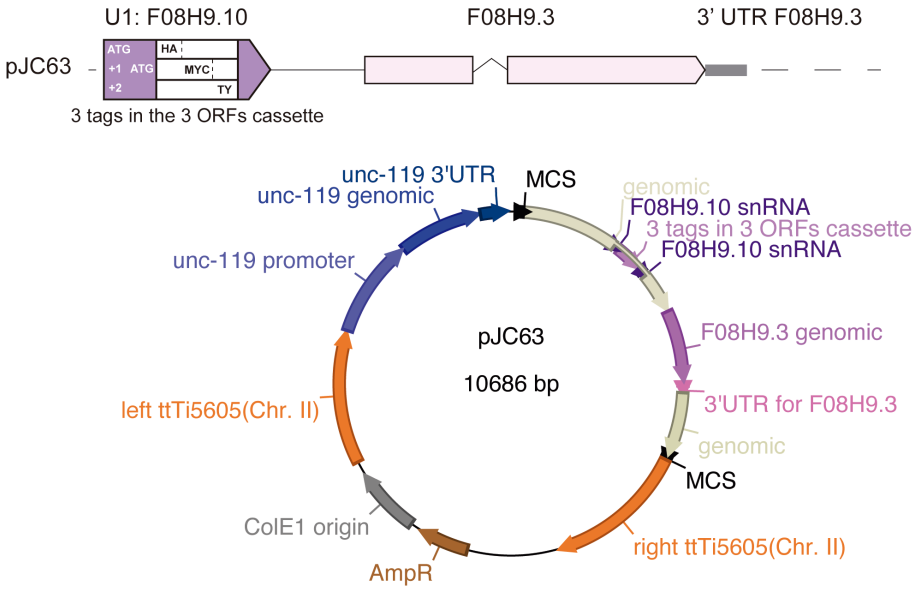

pJC60
U1
(F08H9.10)
F08H9.3
3xFLAG
eGFP
3' UTR F08H9.3
-902
-584
0
2266
MCS
CB-unc-119 3'UTR
UNC-119 genomic
F08H9.10
F08H9.3 promoter
F08H9.3 genomic
unc-119 promoter
3xFLAG
eGFP
pJC60
11069 bp
left ttTi5605 (Chr. II)
3'UTR F08H9.3
genomic
MCS
ColE1 origin
AmpR
right ttTi5605 (Chr.II)

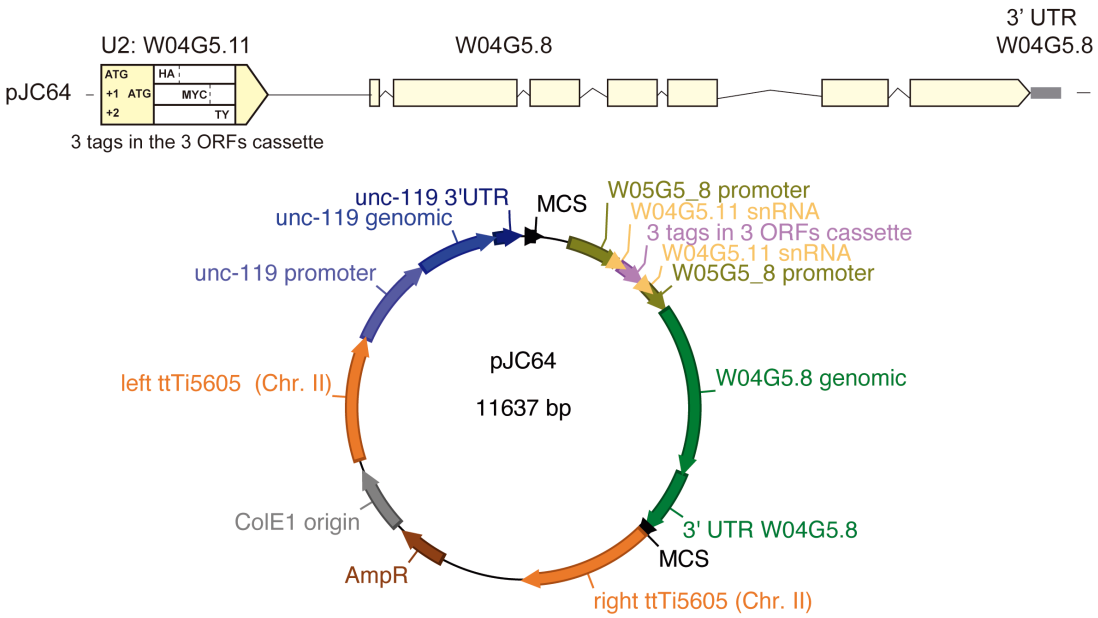

U2
(W04G5.11)
W04G5.8
eGFP
3' UTR
W04G5.8
pJC55
-4561 -1186
0
3686
unc-119 3'UTR
unc-119 genomic
MCS
unc-119 promoter
W04G5.8 promoter
left ttTi5605(Chr. II)
pJC55
15630 bp
W04G5.16
ColE1 origin
AmpR
W04G5.8 genomic
right ttTi5605 (Chr. II)
MCS
genomic
eGFP
3'UTR W04G5.8


U2
(F08G2.9)
ins-37
eGFP
3' UTR ins-37
pJC50
-725
-460
0
3066
MCS
MCS
unc-119 3'UTR
unc-119
F08G2.9
ins-37 promoter
ins-37 genomic
unc-119 promoter
eGFP
3'UTR ins-37
pJC50
11490 bp
left ttTi5605 (Chr.II)
genomic
ColE1 origin
MCS
AmpR
right ttTi5605 (Chr.II)


pJC58
sls-2.8
Y75B8A.23
3xFLAG
eGFP
3' UTR Y75B8A.23
-545
-213
0
2052
MCS
Y75B8A.23 promoter
sls-2.8
Y75B8A.25 genomic
3xFLAG
eGFP
3'UTR Y75B8A.23
genomic
MCS
right ttTi5605 (Chr. II)
AmpR
ColE1 origin
left ttTi5605 (Chr.II)
unc-119 promoter
unc-119 genomic
unc-119 3'UTR
pJC58
10001 bp


pJC54
ints-6 promoter
ints-6
3xFLAG
eGFP
3' UTR
dic-1
S850A
MCS
dic-1 promoter
unc-119 3'UTR
unc-119
unc-119 promoter
left ttTi5605 (Chr. II)
dic-1
pJC54
13312 bp
ColE1 origin
AmpR
S850A
3xFLAG
eGFP
dic 3'UTR
MCS
right ttTi5605 (Chr. II)


U1: F08H9.10
F08H9.3
3' UTR F08H9.3
pJC63
ATG
+1 ATG
+2
HA
MYC
TY
3 tags in the 3 ORFs cassette
unc-119 3'UTR
unc-119 genomic
MCS
genomic
F08H9.10 snRNA
3 tags in 3 ORFs cassette
F08H9.10 snRNA
unc-119 promoter
F08H9.3 genomic
pJC63
10686 bp
3'UTR for F08H9.3
left ttTi5605(Chr. II)
genomic
MCS
ColE1 origin
right ttTi5605(Chr. II)
AmpR


3' UTR
W04G5.8
U2: W04G5.11
W04G5.8
pJC64
ATG
+1 ATG
+2
HA
MYC
TY
3 tags in the 3 ORFs cassette
unc-119 3'UTR
unc-119 genomic
MCS
W05G5_8 promoter
W04G5.11 snRNA
3 tags in 3 ORFs cassette
W04G5.11 snRNA
W05G5_8 promoter
unc-119 promoter
pJC64
11637 bp
W04G5.8 genomic
left ttTi5605 (Chr. II)
3' UTR W04G5.8
MCS
ColE1 origin
AmpR
right ttTi5605 (Chr. II)

4.11. Plasmid DNA purification

E. coli (*NovaBlue*^{TM} *Singles Competent Cells*, Merck Millipore) transformed with the plasmid of interest were grown in 5 ml LB medium (1% (w/v) bacto-tryptone, 0.5% (w/v) bacto yeast, 0.5% (w/v) NaCl in distilled H_20) at 37°C O/N with constant shaking (220 rpm). Plasmid DNA was isolated using the *QIAprep Spin Miniprep kit* (Qiagen) according to the manufacturer's description. When higher amounts of DNA were necessary, the transformed *E. coli* where grown in 200 ml of LB and the *Plasmid Maxi kit* (Qiagen) was used instead. The plasmid DNA concentration was determined spectrophotometrically with *NanoDrop™ 1000* at 260 nm.

5. Protein biochemical methods

5.1. Immunostaining

5.1.1. Immunocytochemistry

U2OS cells were seeded and incubated on pre-treated poly-L-Lysine coated coverslips (Sigma) to improve adhesion. Cells were washed with PBS/$CA^{2+}Mg^{2+}$ (1 mM) and fixed with 4% paraformaldehyde in PBS/$Na^{+}K^{+}$ (1 mM) for 30 min at RT under gentle agitation. Next, two washes with PBS (1 mM) were performed for 5 min each and the cells were permeabilized with 0.1% Triton X-100/PBS for 10 min at RT under gentle agitation.

Next, cells were washed with 0.2% PBS-BSA for 10 min at RT (or 4ºC O/N). Subsequently, the cells were incubated with Ints6 antibody (Bethyl Laboratories, 1:50 dilution) in 0.2% PBS-BSA for 1 h at RT in a humid chamber. Afterwards, cells were washed with PBS (3 times, 7 min each) and incubated with the secondary antibody (*Cy*^{TM}*3-conjugated AffiniPure Goat Anti-Rabbit IgG*, 115-165-003, Jackson ImmunoResearch Laboratories, 1:500 dilution) in 0.2% PBS-BSA for 30 min at RT in the dark. After this incubation, cells were washed with PBS (3 times, 5 min) and incubated with DAPI (2 µg/ml) in PBS for 5 min, all in the dark. Finally, cells were washed with Milli-Q water (3 times) and the coverslips were mounted on slides using the *SlowFade Antifade kit* (Invitrogen). Coverslip edges were sealed with nail polish.

5.1.2. *C. elegans* germline isolation and immunostaining

5.1.2.1. *C. elegans* germline dissection, fixation and permeabilization

First, 8 µl of dissection buffer (1X egg buffer, 0.02% Tween-20, 0.2 mM Levamisole and Milli-Q H_20) (10X egg buffer: 1.18 M NaCl, 480 mM KCl, 20 mM $CaCl_2$, 20 mM $MgCl_2$, 250 mM HEPES pH 7.4) was placed in the center of a poly-L-lysine coated slide (*polysine® slides*, Thermo Scientific). Young adults worms (8-10) were picked with a platinum wire and transferred to the dissection buffer. They were immediately dissected by cutting off either the head (just behind the pharynx) or the tail using the sharp tip of a needle (G20). Gonads burst out and were carefully isolated from the rest of the worm. Next, 8 µl fixation buffer (1X egg buffer, 0,02% Tween-20, 4% formaldehyde, and Milli-Q H_20) was added to the drop and mixed by pipetting 3 or 4 times. Then, 8 µl was removed from the mixture. Carefully, a cover slip (24x24 mm) was placed onto the poly-L-lysine slide followed by 5 min incubation at RT. Afterwards, the slide was dipped in liquid nitrogen by using a clothespin for a handle.

Cover slips were flipped away with the tweezers and slides were incubated in Coplin jars filled with a pre-cooled (-20ºC) 1:1 acetone:methanol solution for 10 min. Next, slides were washed three times (10 min each) with 1% Triton PBS buffer followed by another 5 min wash with 0.1% Tween PBS.

5.1.2.2. *C. elegans* germ line blocking, antibody incubation and mounting

When using Alexa dyes, slides were pre-blocked for 20-30 min using *Image-iT® FX Signal Enhancer* (Invitrogen). One drop was added and gently covered with a small piece of parafilm. Next, samples were blocked for 20-30 min in a Coplin jar with 10% Fetal Bovine Serum diluted in 0.1% Tween PBS. Next, 35-45 µl of the corresponding primary antibody (Table 9) was added to each slide and covered with a small piece of parafilm. Slides were incubated O/N at 4ºC in a humid chamber. The following day the primary antibody was washed three to four times (10 min each) with 0.1% Tween PBS in a Coplin jar and 35-45 µl of the corresponding secondary antibody (Table 9) was added to

each slide before they were again covered with a small piece of parafilm. Slides were incubated for 2 h in a dark box at RT. Next, the slides were washed three to four times (10 min each) with 0.1% Tween PBS in a Coplin jar in the dark.

Finally, as much liquid as possible was wiped off and 8 µl of *VECTASHIELD® Antifade Mounting Medium with DAPI* were added into each cover slip (*18 mm x 18 mm Zeiss Thickness no. 1 ½ High performance*). The cover slips were very gently placed over the slide with the worms. After a few minutes, slides were sealed with nail polish.

PRIMARY AB	DILUTION	SOURCE
α-RAD-51	1:10000	SDIX 2948.00.02
α-pHH3 (Ser 10)	1:1000	Santa Cruz Biotechnology sc-8656R
α-FLAG	1:250	Sigma F1804
α-pTyr15 CDK-1	1:1000	CALBIOCHEM 213940
SECONDARY AB	**DILUTION**	**SOURCE**
Goat Anti-Rabbit Alexa Fluor® 488	1:1000	Thermo Fisher Scientific
Goat Anti-Mouse Alexa Fluor® 488	1:1000	Thermo Fisher Scientific
Goat Anti-Rabbit Cy3	1:500	Jackson ImmunoResearch (115-165-003)
Goat Anti-Rabbit Cy3	1:1000	Thermo Fisher Scientific

Table 9. Antibodies used in this work for immunostaining experiments.

5.2. Protein extracts

5.2.1. Culture cells protein extracts

Protein extracts were obtained directly from cultured cells that had been trypsinized, collected by centrifugation, PBS washed and subsequently processed or frozen in dry ice and kept at -80ºC until processing.

At the time of their processing, cells were resuspended in lysis buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 0.5% NP-40 and 1 mM Dithiothreitol, DTT) with protease and phosphatase inhibitors (1 mM Sodium Orthovanadate, 10 mM

NaF, 10 mM β-Glycerophosphate, 1 mM PMSF, 10 µg/ml aprotinine, 10 µg/ml leupeptine and 2 µg/ml pepstatine) and incubated in ice for 15 min with frequent vortexing in between. Then, samples were centrifuged (maximum speed, 15 min) to eliminate any non-soluble cellular remains and the supernatants were transferred to fresh Eppendorf tubes.

5.2.2. *C. elegans* protein extracts

Worms from 8 to 10 NGM plates were harvested with M9 buffer and collected in 50 ml Falcon tubes. They were washed several times, allowing them to settle to the bottom with gravity. After the last washing step, as much supernatant as possible was removed. Then, an equal volume of lysis buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100), containing 1x protease inhibitors (*Complete ™ EDTA-free Protease inhibitor*, Roche) and 1x phosphatase inhibitors (*PhosSTOP™*, Roche) was added to each worm pellet. Next, samples were ground in liquid N_2 with a pre-chilled mortar and pestle. Ground worms were thawed on ice followed by centrifugation at 4ºC (15000 rpm, 15 min) to eliminate any non-soluble tissue or cellular remains and the supernatants were transferred to fresh Eppendorf tubes.

5.3. Protein extract quantification

Protein extract concentrations were determined using *BCA Protein Assay Kit* (Fisher Scientific) according to the manufacturer's instructions. This kit is based on a colored reaction for the detection and quantitation of total protein. The purple-colored reaction product of this assay is formed by the chelation of two molecules of BCA with one cuprous ion. This water-soluble complex exhibits a strong absorbance at 562 nm that is linear with increasing protein concentrations over a broad working range of 20 µg/ml to 2,000 µg/ml. The absorbance of each sample was quantified in a *POLARstar® Omega Spectrophotometer.* Then, each protein concentration was determined by referring the values to a standard of known BSA concentrations.

5.4. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Using *Sodium-DodecylSulfate-Polyacrilamide gel electrophoresis* (SDS-PAGE) proteins are separated according to their molecular weight (Laemmli, 1970). When proteins are separated in the presence of SDS and denaturing agents, they become fully denatured and dissociate from each other. SDS binds non-covalently to proteins and since SDS is negatively charged, it masks the intrinsic charge of the protein it binds giving a similar charge-to-mass ratio for all proteins in a mixture (a stoichiometry of about one SDS molecule per two amino acids). As a result, the rate at which an SDS-bound protein migrates in a gel depends primarily on its size, enabling molecular weight estimation.

Depending on the molecular weight of the proteins to be analyzed, different percentages of commercial polyacrylamide gels (*Mini-PROTEAN® TGX™ Precast Gels*) were used (between 8% and 12%, or any KDa). Usually, 30 µg per sample were loaded onto the gels after boiling for 5 min in Laemmli buffer (80 mM Tris-HCl pH 6.8, 5 mM DTT, 2% SDS, 7.5% glycerol 7.5%, 5 mM EDTA, 0.002% bromophenol blue). Samples were run using the *Mini-PROTEAN™ Tetra Cell* or the *Criterion™ Cell* electrophoresis system (Bio-Rad) at a constant voltage of 120 V in SDS-PAGE running buffer (25 mM Tris, 200 mM glycine, 0.1% (w/v) SDS) until the tracking dye reached the bottom of the gel. *Precision Plus Protein™ Dual Color Standards* (Biorad) was used as the size reference.

Once electrophoresis was completed, gels were fixed and stained (MM 5.5.) or they were electroblotted to immunodetect proteins by WB (MM 5.6).

5.4.1. SDS polyacrylamide gel electrophoresis (SDS-PAGE) with PhosTag™

SDS-PAGE with Phos-Tag™ was performed to check for the possible phosphorylation of *C. elegans* INTS-6. The PhosTag™ (*Phos-tag Acrylamide AAl-107*, NARD Institute) is a compound of high molecular weight (Phos-tag™ ligand) that binds to divalent cations such as Zn^{2+}, forming a complex that acts as a selective phosphate-binding tag molecule. The Phos-

tag™ ligand is added into normal acrylamide gels, along with Zn^{2+}, and the phosphorylated and non-phosphorylated forms of a protein can be distinguished based on their mobility shift in a general SDS-PAGE system because the phosphate groups of the proteins are bound to the Phos-Tag Zn^{2+} complex and therefore their migration velocity is slower.

Phos-tag™ and $ZnCl_2$ are added only in the resolving gel. The gels were made at 6%. The Phos-Tag™ final concentration was 50 µM. For 5 ml of resolving gel we added 2,6 ml H_2O, 1 ml Tris pH 8 (1 M), 1,3 ml 30% acrylamide and bis-acrylamide solution, 37.5:1 (Bio-rad), 50 µl 10% SDS, 50 µl Phos-Tag™, 100 µl $ZnCl_2$ (10 mM), 50 µl 10% APS and 5 µl TEMED. For 2 ml of stacking gel we added 1.4 ml H_2O, 330 µl acrylamide and bis-acrylamide solution, 37.5:1 (Bio-rad), 250 µl Tris pH 6.8, 20 µl SDS, 20 µl APS and 2 µl TEMED.

Samples were run for 360 min at constant voltage (100 V) in the dark. Next, the gel was washed for 30 minutes with gentle shaking using fresh transfer buffer (25 mM Tris-HCl, 192 mM glycine and 20% methanol) supplemented with 100mM EDTA pH 8 to remove the $ZnCl_2$. Then we washed three times (15 min each) with transfer buffer to remove the EDTA. Afterwards, the proteins were transferred to nitrocellulose membranes (O/N at constant amperage 200 mA in fresh transfer buffer at 4ºC) and incubations with the primary and secondary antibodies, as well as protein detection were performed in the same manner as for the normal WB protocol described in MM 5.6.

5.5. Gel staining

5.5.1. Commassie Blue Staining

Polyacrylamide gels were stained with Coomassie blue using the *Colloidal Blue Staining Kit* (Invitrogen) according to the manufacture's instructions. Basically, gels were fixed for 10 min in fixing solution (40% deionized water, 50% methanol, 10% acetic acid) while shaking at RT. Then, gels were shaken in the *Staining Solution* (55% deionized water, 20% methanol, 25% Stainers

provided by the kit) for a minimum of 3 h and a maximum of 12 h. Protein bands began to appear in 2–5 minutes. Afterwards, the Staining Solution was removed and replaced with deionized water. Gels were shaken in water for at least 7 h.

5.5.2. Silver Staining

Polyacrylamide gels were silver stained using *Silver Stain for Mass Spectrometry* (Thermo Scientific) according to the manufacture's instructions. Basically, gels were washed in Milli-Q water, fixed for 30 min in Fixing solution (30% ethanol, 10% acetic acid) while shaking at RT, followed by ethanol washes and an incubation step with the *Silver stain solution* provided by the kit. Protein bands appeared within 2 to 3 min. Immediately, the staining solution was replaced with 5% acetic acid for 10 min to stop the reaction.

5.6. Western blot

For antibody-specific detection of proteins, samples were separated by SDS-PAGE and transferred to a 0.45 µm nitrocellulose membrane (*Protran BA 85*, GE Healthcare) in transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol) for 90 min at 4°C and a constant voltage of 90 V using the *Mini Trans-Blot® Electrophoretic Transfer Cell* (Bio-Rad) or *Criterion ™ Blotter*. After transfer, the membrane was blocked in TBS-T (49 mM Tris base 102 mM NaCl, 5.4 mM KCl, 0.05% (v/v) Tween®-20, pH 8) with 5% (w/v) non-fat dry milk (*Sveltesse* Nestlé) for 60 min with gentle rocking. To detect the protein of interest, the membrane was incubated with the primary antibody diluted in TBS-T milk (for dilutions see Table 10) for 120 min at RT or O/N at 4°C. Afterwards, the membrane was washed three times with TBS-T for 10 min each and then incubated with the respective horseradish peroxidase (HRP) conjugated secondary antibody in TBS-T milk (for dilutions see Table 10) for 60 min at RT. The membrane was washed twice more with TBS-T two times for 10 min each time and then once more with TBS only.

To chemiluminescently detect the protein of interest, *ECL ™ Blotting Detection Reagents* (GE Healthcare) and *Amersham Hyperfilms ™ ECL* (GE Healthcare) were used according to the manufacturer's instructions. When

necessary because the signal of the desired protein was very low, the *SuperSignal™ West Femto Maximum Sensitivity Substrate* (Thermo Fisher) detection reagents were used instead.

Films were developed manually: 1 min in developing solution (Agfa developer G153), 1 min in fixing solution (Agfa fixer G-345) and then rinsed in water.

PRIMARY AB	DILUTION	SOURCE
α- Ints6	1:500	A301-658A, Bethyl Laboratories
α-RNAP II	1:500	(H-224): sc-9001, Santa Cruz Biotechnology
α-GFP	1:1000	Living Colors® GFP Monoclonal 632381 Clontech
α-FLAG	1:1000	F1804, Sigma
α-MYC	1:1000	9B11, Cell Signaling
α-HA	1:1000	6E2, Cell Signaling
α-Ty1	1:1000	SAB4800032, Sigma
α-β-actin (*H. sapiens*)	1:5000	AC-15, Sigma
α-actin (*C. elegans*)	1:1000	sc-1616, Santa Cruz Biotechnology
SECONDARY AB	**DILUTION**	**SOURCE**
α-rabbit HRP-linked	1:3500	GE Healthcare NA934
α-mouse HRP linked	1:2500	GE Healthcare NA931
α-goat HRP-linked	1:5000	Jackson Immunoresearch 805-035-180

Table 10. Antibodies used in this work for WB/IP experiments.

5.7. Immunoprecipitation

Immunoprecipitation is the affinity “purification” of antigens using a specific antibody that is immobilized to a solid support such as magnetic particles or agarose resin.

5.7.1. IP of proteins from cellular extracts

H. sapiens Ints6 was immunoprecipitated from HEK293T cultured cells transfected with pBS15. Proteins were extracted were using lysis buffer (200 mM NaCl, 50 mM TrisCl pH 7,5, 0,1% (w/v) NP-40) with protease and phosphatase inhibitors (MM 5.2.1). *Dynabeads® Protein G* (Invitrogen) was used together with the GFP antibody (Table 10). These beads are magnetic, and prior to use, their storage buffer was removed by placing the tube that contained the beads into the *DynaMag™-Spin* magnetic separator to collect them and to eliminate the storage buffer. Then, the beads were collected in lysis buffer and the storage buffer was discarded. IPs were performed in 2ml Eppendorf tubes by adding 1 mg of protein extract, 0.5 µg GFP antibody (Table 10) and 5 µl of *Dynabeads® Protein G* (Invitrogen). The mixture was incubated for 3h-4h at 4ºC on a rotating rack with gentle mixing. The negative control consisted of incubating the protein extracts with only the beads. Subsequently the beads were recovered using the *DynaMag™-Spin* magnetic separator and three to five washes were performed using 500 µl of the lysis buffer previously used. The immunoprecipitated proteins were eluted by boiling the samples for 10 min in 20 µl SDS-PAGE Sample Buffer.

5.7.2. IP of proteins from *C. elegans* extracts

In order to immunoprecipitate INTS-6, and Co-IP their interactors, protein extracts from the JCP378 strain (jcpSi19[pJC56(*eft-3*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))]II;*unc-119(ed3)*III) were used and extracts from N2 worms were used as the negative control. IPs/Co-IPs were performed with *ANTI-FLAG®M2 Magnetic Beads* (Sigma) that are composed of the murine derived ANTI-FLAG M2 monoclonal antibodies attached to superparamagnetic iron impregnated 4% agarose beads.

As large amounts of protein extracts were needed, worms were grown in NGM egg plates (MM 2.3). Protein extracts (MM 5.2.1) were filtered through a 5.0 µm filters and subsequently through 0.45 µm filters to remove any remaining cell debris and particulates that could interfere with protein binding.

In each IP/Co-IP reaction, 150 µl of the *ANTI-FLAG®M2 Magnetic Beads* (which have a binding capacity of 45 µg) were incubated with 30 mg of protein extract (harvested from 60-100 egg plates, for a total of 3-5 g worms). Prior to using the *ANTI-FLAG®M2 Magnetic Beads,* the storage buffer was removed by placing the tube that contained the beads in the *DynaMag™-Spin* magnetic separator to collect them and eliminate the storage buffer. Immediately afterwards, the beads were equilibrated with TBS buffer (375 µl per IP/Co-IP reaction: 50 mM Tris HCl, 150 mM NaCl, pH 7.4). This step was repeated once, leaving the beads with a small amount of buffer. Then, the protein extract was incubated with the equilibrated beads for 3-4 h or O/N, always at 4ºC in a rotating rack with gentle mixing. Once the binding step was complete, the beads were collected and the supernatants were removed, followed by the washing steps. The beads were washed with TBS buffer (1500 µl per IP/Co-IP reaction) three sequential times for 10 min on a rotating rack at 4ºC to remove all non-specifically bound proteins. INTS-6::3xFLAG::eGFP fusion protein (and consequently their interacting proteins) were eluted from the magnetic beads by different methods:

1) Boiling samples for 10 min in SDS-PAGE Sample Buffer. Each IP was eluted in 100 µl SDS-PAGE Sample Buffer.

2) Under acidic conditions with 0.1 M glycine HCl, pH 3.0. Each sample was incubated for 20 min at RT with 375 µl 0.1 M glycine on a rotating rack with gentle mixing. Then the supernatants were removed and precipitated using TCA (MM 5.8). Dry protein pellets were recovered in SDS Sample Buffer and if necessary, neutralized by adding 10 µl of 0.5 M Tris HCl, pH 7.4, with 1.5 M NaCl.

3) Elution with 3xFLAG peptide. To elute each IP, 375 µl of 3xFLAG peptide (400 ng/µl 3xFLAG peptide in TBS) was incubated for 1 h at RT in a rotating rack with gentle mixing. Then supernatants were collected and the 3xFLAG peptide elution was repeated one more time. Finally, eluates were precipitated using TCA (MM 5.8). Dry protein pellets were recovered in SDS Sample Buffer and if necessary, they were neutralized by adding 10 µl of 0.5 M Tris

HCl, pH 7.4, with 1.5 M NaCl.

5.8. Trichloroacetic acid (TCA) protein precipitation

Protein extracts were precipitated using TCA. One volume 20% (w/v) TCA was added to one volume of each protein extract so that the final TCA concentration was 10%. Then, samples were incubated in ice for 30 min followed by centrifugation (maximum speed, 10 min). Afterwards, the supernatants were removed, leaving protein pellets intact. The pellets were washed with 200 µl cold acetone twice. Finally, pellets were dried by placing tubes in a 95°C thermal-block for 5-10 min to drive off the acetone.

5.9. Protein identification by mass spectrometry

Eluted IPs were run on SDS-PAGE (*Mini-PROTEAN® TGX™ Precast Gels* any kDa) by SDS-PAGE. Next, gels were stained with Coomassie Blue and bands were excised. The CIC Biogune proteomics platform (https://www.cicbiogune.es/org/plataformas/Proteomics) performed the proteomics analysis. Proteins were digested with Trypsin from each gel band and analyzed by LC-MS/MS: Liquid Chromatography-Mass Spectrometry/Mass Spectrometry.

6. Transformation techniques

6.1. Transformation of plasmids into chemically competent bacteria

Plasmid DNA was transformed into chemically competent *E. coli* (*NovaBlueSingles™ Competent Cells*) via heat shock treatment. Competent cells were thawed on ice and incubated with the DNA (1-10 ng) on ice during 5 min. Cells were heat shocked at 42ºC for 30 sec and then immediately placed on ice for 5 min. Next, 250 µl of SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) was added and the mixture was incubated at 37°C for 60 min with constant shaking (220 rpm). An aliquot of the mixture was spread on an LB-agar plate (1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast, 0.5% (w/v) NaCl, 15% (w/v)

agar in distilled H_20) containing the required antibiotic, usually 50 µg/ml ampicillin and the plate was incubated at 37°C O/N.

6.2. Cell line transformation: transient transfection of DNA

Depending on the cell line, different DNA transfection methods were used. HEK293T cells were transfected using the *Calcium Phosphate* method. This method is based on the formation of a precipitate between $CaCl_2$ and DNA in a phosphate saline solution that is recognized and engulfed by cells. For each p100 plate, 200 µl 0.25 M $CaCl_2$ and 2-8 µg DNA were mixed in an Eppendorf. Then, 200 µl HEBS (0.28 M NaCl, 0.05 M HEPES, 1.5 mM Na_2HPO_4, pH 7-7.15) was added drop by drop while bubbling. After a 2 min incubation at RT the mixture was added to the cultured cells. The culture medium was changed after 12 h of transfection.

RPE and U2OS cell lines were transfected by Lipofection. This method uses the formation of small liposomes that encompass the DNA of interest to introduce it into the cell. This transfection was done by employing *Lipofectamine® 2000* (GIBCO Invitrogen) according to the manufacturer's instructions.

Cells were seeded 12-14 h before transfection so that they would be at 50%-60% confluence at the time of transfection.

6.3. *C. elegans* transformation

6.3.1. *C. elegans* transformation by Microinyection

The transgenics JCP341, JCP378, JCP490, JCP524, JCP358, JCP250, JCP523, JCP479 and JCP504 were generated by mosSCI system following the protocol described in Frokjaer-Jensen et al. 2008 (Knudra enterprise). The mosSCI system inserts a single copy of a transgene into a defined site. Briefly, this method works by breaking a chromosome at a particular location by excising a Mos1 transposon. The excision creates a DNA DSB that is repaired by HR. In the presence of plasmids having DNA sequences homologous to the breakpoint, the repair process uses the plasmid as its DNA template.

The plasmids used (MM 4.10.2.2) had two “recombinant arms” called ttTi5605, which are DNA sequences with homology to the breakpoint in chromosome II. The DNA within the two recombinant arms is incorporated into the genome during the repair process. The corresponding plasmids were microinjected into *C. elegans* young adult gonads using a microinjector fitted to a dissecting microscope. In addition, the worms microinjected worms had the *unc-119(ed-3)* mutation and the plasmids had the *unc-119* gene within the two recombinant arms that rescued the phenotype, which thus served as a positive selection marker. Afterwards, only the integrated transgenic lines were selected.

The strains JCP462, JCP472 and JCP483 were generated by the CRISPR technique (Frokjaer-Jensen, 2013; Xu, 2015) (Knudra enterprise). The advantage of the CRISPR technique is that the genome can be modified at virtually any genomic locus through a guide RNA that recognizes the target DNA via Watson-Crick base pairing. Therefore, the CRISPR technique, which is derived from bacterial adaptive immune systems, is a powerful genome-editing tool. This system introduces site-specific DSBs that are subsequently repaired by either Non-Homologous End Joining or Homology directed repair.

The core components of the CRISPR technique include an endonuclease (Cas9) containing two catalytic nuclease domains (RuvC and HNH), and a single guide RNA (sgRNA) chimera that combines the functions of the CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides Cas9 to complementary target sequences, and tracrRNA, which binds to the crRNA and to Cas9 to form the ribonucleoprotein (RNP) complex.

The specific sequence requirement for chromosomal editing hinges on the 20 nt sequence at the 5′ end of the sgRNA, which has to be followed by a protospacer adjacent motif (PAM) of NGG in the DNA in order for efficient cleavage. Thus, the CRISPR mediated genome editing is programmable and can be easily targeted to most genomic locations of choice through the design of the sgRNA (Frokjaer-Jensen, 2013; Xu, 2015).

6.3.2. *C. elegans* transformation by Bombardment

C. elegans biolistic bombardment was performed following the protocol described by Hochbaum et al. 2010 with few modifications. This method used DNA-coated gold particles to introduce foreign DNA into the germline. Compared to DNA microinjection into the hermaphrodite germlines, multiple transgenic lines are usually obtained from a single bombardment and this technique requires less time investment in terms of practice to become successful.

In this work, the JCP505 strain (jcpEx5[pJC56(*ints-6*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR,*unc-119*(+))] II;*unc-119(ed3)* III) was created by bombardment. The pJC56 plasmid (*eft-3*p::*ints-6*::3xFLAG::eGFP::*ints-6*UTR, *unc-119*(+)) was used to transform the HT1593 strain (*unc-119(ed3)* III).

The *unc-119 (ed3)* mutant worms are difficult to grow on standard NGM plates. Because of their impaired mobility, they tend to starve on parts of a plate while other parts of the plate still contain food. Egg plates (MM 2.2) were employed to grow them because of the thick food layer on these plates.

The first step prior to performing a bombardment is to prepare the worms. Usually 5 egg plates are sufficient to grow enough worms for one bombardment. L1 synchronized worms were seeded on egg plates and grown at 20ºC for 7-10 days.

On the day of the bombardment, worms were washed off with M9 buffer, collected in 50 ml Falcon tubes, and allowed to settle with gravity. Then, the supernatant was removed and washing steps with M9 buffer were performed to cleanse worms of bacteria or egg yolk remains. Worms were kept at this step until the DNA preparation was completed.

Prior to preparing DNA-coated gold particles, the gold particle stock solution was made. 60 mg of gold particles (Bio-Rad) were weighed in a 1.5 ml Eppendorf and soaked in 70% ethanol for 15 min. Then, gold particles were spun down, the ethanol was removed and they were washed three times in sterile water. Finally, the water was removed and the gold particles were resuspended in 1 ml sterile 50% glycerol. This gold particle stock solution can be stored for months at 4ºC.

To prepare the DNA gold-coated particles, the gold particle stock solution was resuspended thoroughly and 50 µl was added to a 1.5 ml Eppendorf. Then, 50 µl DNA (10-16 µg) and 20-50 µl 100mM Spermidine prepared fresh (Sigma) was added. The mixture was incubated at RT for 10 min. Next, 50 µl 2.5 M $CaCl_2$ was added drop by drop while vortexing and the mixture was incubated for another 10 min at RT to produce DNA precipitation. The mixture was incubated on ice for 30 min and periodically mixed to keep the gold particles in suspension. Afterwards, DNA-coated gold particles were spun down (15 sec, 12000 rpm) and the supernatant was removed. They were washed with 300 µl 70% ethanol and then with 1 ml 100% ethanol. Then, they were resuspended in 170 µl 100% ethanol.

Once the DNA-coated gold particles were prepared, the worms that were collected previously were transferred to 10 cm NGM plates without food and chilled on ice. This NGM plates were poured well in advance and allowed to dry to facilitate absorption of the liquid added with the worms. The worms were dispersed on the NGM plate and kept on ice to prevent them from moving on the plate and aggregating into piles.

To bombard the worms with the DNA-coated gold particles, the *Biolistic*® *PDS-1000/He Hepta Particle Delivery System* (Bio-Rad) was used. This system uses high-pressure helium, released by a rupture disk, and partial vacuum, to propel a macrocarrier sheet loaded with millions of microscopic DNA-coated gold particles toward target worms at high velocity. The macrocarrier is halted after a short distance by a stopping screen. The DNA-coated particles continue traveling toward target worms to penetrate and transform them. For each bombardment, seven macrocarriers were rinsed in 2-propanol and placed on a tissue paper to dry at RT. Then, the DNA/gold mixture was sonicated and 20 µl were added to the center of each macrocarrier. To perform a bombardment, the pressure of the helium tank should be ≥1550 psi (pounds per square inch). It is important to perform a blank bombardment before the experiment to flush helium through the system. The rupture disk is moistened in 2-propanol and placed in the retaining cap for the hepta adapter. The adapter is screwed into the *Biolistic*® *PDS-1000 System* and tightened with the supplied torque wrench. Then, the 7 macrocarriers are placed into the holder, followed by the hepta stop screen

and then the bottom. The macrocarrier holder is placed into the *Biolistic® PDS-1000 System*. Next, the NGM plate coated with worms (lid off) is placed on the lowest shelf in the bombardment chamber. To perform a fire, the vacuum flow-rate is opened on the *PDS-1000 System* (Vac button) until chamber reaches 26" Hg. Then it is switched to the hold position to maintain the vacuum. The Fire button is pressed and held until the disk ruptures (1350 psi). The vacuum is realised from the chamber (vent position) and the plate is removed. The vacuum and the helium are turned off. It is important to fire several times until the helium gauge marks 0 psi before turning off the *PDS-1000 System*.

Afterwards, the worms were left on the bombarded plate to recover for approximately 20 min. Next, worms were collected in M9 buffer and placed on normal NGM plates seeded with OP50*. After about two weeks at 20ºC, the rescued worms (*unc-119*(+)) could be distinguished because of the wild type movement. We checked the rescued worms under UV light using the dissecting microscope and the GFP could be detected.

7. RNA mediated interference

7.1. Transient siRNA transfection of cell lines

To deplete Ints6 protein expression levels in cells lines, small interference RNAs (siRNAs) targeting the Ints6 sequence were used. These small RNA molecules are bound to their target sequence, generating dsRNA and leading to specific degradation of the corresponding mRNA.

The siRNAs used in this work were acquired from Dharmacon (ON-TARGET plus). We used a set of 4 siRNAs (J-012417). Their efficiency was checked individually and the experiments were performed with two of them. Target sequences: Ints6(5) siRNA: "GAAGAGCACUCGCAGAUUU" and Ints6(8) siRNA: "GAGCCGAUCACAUGGUUUA". A siRNA designed against the Luciferase sequence was used as the negative control (target sequence: 5'-NNACGUACGCGGAAUACUUCGAA-3'). Transfections were performed by lipofection as described in MM 6.2 using *Lipofectamine® RNAiMAX* (Total amount of siRNA used: 60-100nM).

7.2. *C. elegans* mediated RNA interference (RNAi)

RNA interference (RNAi) refers to a process by which dsRNA leads to specific degradation of the corresponding mRNA (Fire et al., 1998). There are three ways to carry out RNAi in *C. elegans*: injection (Fire et al., 1998), soaking (Tabara et al., 1998), and feeding (Timmons & Fire, 1998). In this work, RNAi was performed by feeding worms. Bacteria producing the desired dsRNA were fed to worms and they were scored. There are currently two RNAi feeding libraries that when combined can target about 94% of the *C. elegans* genome. One comes from the Ahringer Lab. This library was made by cloning gene-specific genomic fragments between two inverted T7 promoters (Kamath et al., 2003). The other is from the Vidal Lab and was made by cloning full-length open reading frame (ORF) cDNAs into a double T7 vector (Rual et al., 2004). Both libraries use the HT115 bacterial strain as a host for the plasmid RNAi clones. The HT115 *E. coli* strain has an IPTG inducible T7 polymerase and lacks a functional RNAse III (a dsRNAse). Thus, in both libraries, the expression of the respective insert is controlled by IPTG and large quantities of dsRNA can be accumulated in the cell.

For inactivation of a specific gene, the corresponding RNAi clone was selected. 5 ml LB medium containing ampicillin (100 µg/ml) was inoculated with a single bacterial colony and incubated at 37°C for 8 h with constant shaking. 400 µl of the bacterial culture was spread on 90 mm NGM RNAi feeding plates (NGM plates with 100 µg/ml ampicillin, 12.5 µg/ml tetracycline, 1 mM IPTG) and incubated O/N at RT to grow a bacterial lawn and induce dsRNA expression. The next day, synchronized L1 populations were transferred to RNAi feeding plates. The phenotypes were studied over the following days.

In each RNAi experiment, an empty RNAi clone vector (L4440), which did not alter the wild type phenotype of animals, was used as a negative control. The vector L4440 expressing the *pos-1* gene was used as a positive control for RNAi efficiency, *pos-1* RNAi treated worms did not produce viable progeny.

In general the experiments were performed at 20°C.

RNAi clone	SOURCE
C06A5.1 (*ints-1*)	Our Lab
ZC376.6 (*ints-2)*	Vidal Library
Y92H12A.4 (*ints-3*)	Our Lab
W04A4.5 (*ints-4*)	Our Lab
Y51A2D.7 (*ints-5*)	Vidal Library
F08B4.1 (*ints-6*)	Our Lab
D1043.1 (*ints-7*)	Our Lab
Y48G10A.4 (*ints-8*)	Our Lab
F19F10.12 (*ints-9*)	Our Lab
F47C12.3 (*ints-10*)	Ahringer Library
F10B5.8 (*ints-11*)	Our Lab
T23B12.1 (*ints-12*)	Vidal Library
R02D3.4 (*ints-13*)	Vidal Library
Emp. vector (L4440)	Our Lab
F52E1.1 (*pos-1*)	Vidal Library

Table 11. *C. elegans* RNAi clones used in this work.

8. Microscopic techniques

8.1. Differential interference contrast microscopy (DIC or Nomarski)

For microscope preparations, worms were monitored on NGM plates under a Leica dissecting microscope (MZ16FA model). Differential interference contrast microscopy (DIC) was performed on a motorized fluorescent Leica microscope (*DM6000B model*) equipped with a *Hamamatsu Orca-ER C10600* camera and fitted with DIC optics. Samples were prepared by mounting worms on 4.5% (w/v) agar pads (Difco Noble agar) in H_2O. Agar pads were prepared by spotting a drop of melted 4.5% agar on a slide and placing a second slide on top, then flattening the drop to create a thin film of agar. If necessary, worms were immobilized by adding a drop of 1 mM Tetramisol (Sigma). Preparations were sealed with Vaseline.

Images were captured using the open source Micro-manager software (www.micro-manager.org) and processed with XnView software and Image J software.

8.1.1. 4D microscopy

Gravid hermaphrodites were dissected and 2- to 4-cell stage embryos were mounted on 4% agar pads in water, and sealed with Vaseline. Imaging was

performed at 25ºC. The temperature was maintained constant by coupling a metal ring through which water circulated at 25ºC to the microscope objective, thus creating a thermal bath.

Multi-focal time-lapse microscopy of the samples was controlled with the open source Micro-manager software. Images on 30 focal planes (1 micron/section) were taken every 30 seconds for 12 h.

8.2. Fluorescence microscopy

Fluorescence microscopy was used to visualize the fluorescence of a sample, such as eGFP in cells or in *C. elegans*. A Leica microscope (*DM6000B*) equipped with a *Hamamatsu Orca-ER C10600* was used, and the appropriate microscope filters were selected.

Confocal microscopy imaging was performed at Hilde Nilsen's lab (Oslo University) with the *Zeiss 780* confocal microscope. Images were acquired and processed using ZEN lite open software from Zeiss.

9. *C. elegans* based specific techniques

9.1. Crossing *of C. elegans*

Crosses were set up by placing 3-4 L4 stage hermaphrodites of the corresponding strain together with 10-15 young males of the corresponding strain on the same NGM plate O/N. If mating was successful (~50% occurrence of male progeny), L4 stage hermaphrodites of the F1 generation were singly placed on new NGM plates and allowed to lay eggs. To obtain males, which are produced at low frequencies (~0.02%) in wild-type populations, about a dozen N2 males were selected and crossed with several N2 hermaphrodites. This allowed us to have our own stock of males. In addition, males were generated from L4 hermaphrodites following a mild heat shock (34ºC for 3-4 hours).

10. Cell based specific techniques

10.1. Cell culture synchronization

10.1.1. G1/S synchronization by treatment with double thymidine

Cell synchronization is a process by which cells at different stages in the cell cycle in a culture are brought to the same phase. Cells can be synchronized at G1/S using thymidine, a nucleotide that when used in excess, stops DNA replication by inhibiting nucleotides synthesis. Culture medium with 2.5 mM thymidine (Sigma) was added to growing on p100 plates (100 mm diameter). During the first treatment or block, cells were maintained at this thymidine concentration for 24 h. Afterwards, thymidine was removed by washing the cells twice with fresh PBS and then, fresh medium without thymidine was added for the next 12 h. This period is referred as liberation because cells go through the cell cycle. Then, cell cultures were again incubated with 2.5 mM thymidine for 24 h. This is the second block.

Some cells were collected at G1/S (0 h) and the rest of the them were liberated by removing thymidine as described before. Other samples were collected at different times throughout cell cycle.

10.1.2. Mytosis synchronization by nocodazole treatment

Nocodazole is a chemical agent that binds to β-tubulin and inhibits the polymerization of microtubules, preventing proper formation of the mitotic spindle and causing a stop in cells prior to the metaphase stage.
Culture medium with nocodazole (Sigma, 50 ng/ml) was added to cultured cells attached to p100 plates for 12-15 h. Following nocodazole treatment, the mitotic cells (rounded, not attached) were selected by shaking them off the plates and collecting the culture medium with the non-adherent cells.

10.2. Cell cycle analysis by FACS (Fluorescence Activated Cell Sorting) in cell cultures.

One of the possibilities that FACS (Fluorescence Activated Cell Sorting) offers is the analysis of DNA content in cells, which is informative of their cell cycle phase. The typical FACS profile of human cells growing in asynchrony shows

a high peak with 2N DNA (corresponding to G1 cells) and a lower peak with 4N DNA (corresponding to G2/M cells). Between these two peaks there is a plateau that corresponds to cells with an intermediate DNA content (between 2N and 4N) correlating to S phase cells. The peak called pre-G1 corresponds to cells with less than 2N DNA content, and those are mainly apoptotic cells.

To analyze the cell-cycle profile, cell culture samples were collected by trypsinization, washed with PBS and recovered by centrifugation (1200 rpm, 5 min). Cells were resuspended in 1 ml cold 70% ethanol (in PBS) and incubated for 1h in ice or O/N at 4ºC to fix them. Afterwards, cells were washed twice in PBS (2000rpm, 10 min) and resuspended in 0.5 ml Propidium iodide (4 µg) solution and RNase (10 µg) (Roche). Propidium iodide binds to DNA by intercalating between the bases. Cells were incubated for 1 h at 37ºC in the dark with constant shaking. Finally, they were acquired in a *BD FACScalibur* (Becton-Dickinson) flow cytometer and analyzed using the Cell Quest Pro programme.

Resumen & Conclusiones

INTRODUCCIÓN

1. *C. elegans* como organismo modelo

C. elegans es un pequeño nematodo que se encuentra distribuido por todo el mundo, predominantemente en áreas húmedas (Anderson et al., 2012; Frezal & Félix 2015). Inicialmente se caracterizó de forma errónea como un nematodo del suelo, dónde se encuentra mayormente en un estadío de resistencia llamado *dauer*. En cambio, adultos y otros estadíos larvarios se pueden encontrar más fácilmente en materia vegetal en descomposición como frutas o tallos herbáceos gruesos.

C. elegans fue descrito por primera vez en 1900 por Emilie Maupas. Sin embargo, no fue hasta 1965 cuando Sydney Brenner estableció *C. elegans* como un organismo modelo para estudio de la biología del desarrollo y la neurobiología. Desde entonces, numerosos descubrimientos clave se han realizado con este pequeño gusano. Sydney Brenner, Robert Horvitz y John Sulston fueron premiados en 2002 con el premio Nobel Fisiología o Medicina por sus descubrimientos en relación con la organogénesis y la muerte celular programada. Además, *C. elegans* ha permitido desarrollar herramientas técnicas de un gran impacto biológico como son el silenciamiento génico por el uso de RNAi (RNA de interferencia) (Fire et al. ,1998) o el desarrollo de la GFP (proteína verde fluorescente) como marcador biológico (Chalfie et al. 1994).

Hoy en día, *C. elegans* se utiliza para el estudio de una gran variedad de procesos biológicos: apoptosis, ciclo celular, señalización celular, envejecimiento, comportamiento, determinación sexual, metabolismo, etc. (Corsi et al. 2015).

La cepa de referencia *wild-type* que se utiliza comúnmente en el laboratorio la obtuvo Sydney Brenner a partir de la aislada originalmente en Bristol (Inglaterra) y la llamó N2 (Riddle et al., 1997). El cultivo de estos nematodos

es muy simple, crecen en placas de agar sembradas con *Escherichia coli* como fuente de alimentación. Los adultos tienen un tamaño aproximado de 1mm de longitud y 50 μm de diámetro (Fig. 2). Su ciclo de vida es muy rápido, tres días y medio creciendo a 20ºC, con una vida media en fase adulta de dos o tres semanas en condiciones favorables (Riddle et al., 1997; Corsi et al. 2015). Además, estos animales se pueden crecer a distintas temperaturas con rangos de entre 12 y 25ºC, por lo tanto, una de las ventajas de uso en el laboratorio es poder variar la temperatura para controlar su velocidad de desarrollo o bien usar mutantes termosensibles.

Otra de las ventajas como organismo modelo es la posibilidad de congelar las cepas en nitrógeno líquido, lo que permite mantener enormes colecciones de gusanos mutantes o transgénicos y re-vivir la cepa deseada cuando sea necesario.

El modo de reproducción de *C. elegans*, llamado androdioecia, es muy peculiar. Estos nematodos se pueden reproducir como hermafroditas o bien apareándose con machos. Ambos sexos tienen seis cromosomas, cinco cromosomas autosómicos (A) y un cromosoma determinante del sexo (X). Los hermafroditas son diploides para los seis cromosomas. En cambio los machos sólo tienen un cromosoma determinante del sexo X (XO).

La mayoría de la descendencia se produce por auto-fertilización de los hermafroditas, de esta forma un solo gusano da lugar a poblaciones enteras. Cada hermafrodita puede producir hasta 300 huevos y sólo un 0.1-0.2% de la descendencia son machos debido a un "fallo" durante la separación del cromosoma X en meiosis. La presencia de machos es rara tanto en la naturaleza como en condiciones de laboratorio. Sin embargo, cuando un hermafrodita se aparea con un macho, la capacidad de reproducción aumenta hasta unos 1000 huevos que dan lugar a una cantidad equivalente de machos y hermafroditas (Frezal & Félix, 2015).

2. El complejo Integrador

El complejo Integrador fue descubierto en 2005 por Baillat y colaboradores en las células humanas HeLa. Este complejo compuesto por catorce subunidades fue llamado Integrador por ser el responsable de integrar el extremo carboxilo terminal (CTD) de la ARN polimerasa II (RNAP II) con el procesamiento en 3' de los ARN pequeños nucleares (snRNAs) U1 y U2 (Baillat et al., 2005).

Inicialmente se identificaron 11 subunidades del complejo y se llamaron Integrador1 (Ints1) a Integrador 11 (Ints11) ordenadas de mayor a menor peso molecular. Posteriormente, un escrutinio para identificar otras subunidades del complejo realizado en células de *Drosophila S2* reveló la existencia de otras dos subunidades que se llamaron Integrador 12 (Ints12) e Integrador 13 (Ints13) (Chen et al., 2012). Las subunidades Ints11 e Ints9 del complejo Integrador presentan homología con las subunidades CPSF73 y CPSF100 de la maquinaria de corte y poliadenilación de ARNs mensajeros (mRNA) de la célula. Ambas, Ints11 e Ints9, pertenecen a un grupo de nucleasas dependientes de Zinc de la familia β-CASP. Al igual que CPSF73, Ints11 es la subunidad catalítica, contiene un dominio β-lactamasa que ha sido modificado para permitir el corte endonucleolítico de los ácidos nucleicos (Baillat et al. 2005).

El complejo Integrador está aparentemente muy bien conservado a lo largo de la evolución, análisis bioinformáticos de la secuencia proteica de sus subunidades han permitido identificar ortólogos en otras especies de metazoos. Además, algunas de sus subunidades también están presentes en organismos unicelulares como las amebas, lo que sugiere un origen evolutivo temprano (Peart et al. 2013). Sin embargo, no se han encontrado homologías en levaduras como *Saccharomyces cerevisie* donde el complejo

Nrd1/Nab3/Sen1 es el encargado de procesar los snRNAs en el extremo 3' (Steinmetz et al. 1996, Steinmetz et al. 2001).

Desde el descubrimiento del complejo Integrador, se han realizado numerosos estudios acerca de la función de sus subunidades en distintas especies. Además de ser el encargado del procesamiento en 3' de los snRNAs, este complejo, o al menos algunas de sus subunidades, está implicado en otras funciones biológicas. Por ejemplo, Ints4 e Ints11 son necesarias para la homeostasis de los cuerpos de Cajal, pequeños cuerpos nucleares caracterizados por la presencia de la proteína coilin que están implicados en procesos metabólicos del RNA como la biogénesis, maduración y reciclaje de pequeñas ribonucleoproteínas o snRNPs.

Ints6 es la única subunidad que fue descrita antes del descubrimiento del complejo Integrador, se llamó *DICE1* (Deleted In Cancer1) por su localización en el genoma, una región que en tumores de pulmón y otros órganos suele estar afectada por la pérdida de heterocigosidad. Ints6/*DICE1* se aisló como un posible supresor tumoral ya que, además de su localización, se observó una disminución de su expresión en células de carcinoma de pulmón no microcítico (Wieland et al., 1999). A parte de su actividad supresora de tumores, estudios posteriores han demostrado que Ints6 en asociación con Ints3 y las proteínas hSSB1 y C9orf80, bien formando un subcomplejo o bien en asociación con el complejo Integrator, desempeñan un papel clave en la ruta de reparación por recombinación homóloga (HR) en respuesta a daño al ADN (Huang et al., 2009; Skaar et al., 2009; Zhang et al., 2013).

Descubrimientos recientes han ampliado las funciones del complejo Integrador a un espectro más amplio del ciclo de transcripción de la RNAP II, además de su funcionamiento en el procesamiento en 3' de los snRNAs, este complejo está implicado en el inicio, pausa en promotores proximales,

elongación y la terminación de la transcripción (Gardini et al., 2014; Stadelmayer et al., 2014; Lai et al., 2015; Skaar et al. 2015).

2.1. El complejo Integrador en la transcripción

La transcripción es el proceso por el que una hebra de ADN se copia en una nueva molécula de ARN. En eucariotas, se lleva a cabo por distintos tipos de ARN polimerasas: ARN polimerasa I (RNAP I), ARN polimerasa II (RNAP II) y ARN polimerasa III (RNAP III). Ninguna de ellas se une al ADN directamente, sino que son reclutadas a través de otras proteínas. Cada una de las polimerasas se encarga de la transcripción de distintos tipos de genes (Cramer et al., 2008).

La RNAP II es la que lleva a cabo la transcripción de los mRNAs, los snRNAs de la clase Sm y otros RNAs no codificantes para proteína. Esta holoenzima está formada por doce subunidades, la subunidad mayor, Rpb1, contiene el CTD por el que se une al complejo Integrador. El CTD consiste en repeticiones seriadas de la secuencia heptapeptídica YSPTSPS (38 en *C. elegans*, 52 en humanos). Estos residuos del CTD pueden ser modificados por fosforilación (Tyr, Thr, Ser) o isomerización (Pro), permitiéndole desempeñar un papel clave en la regulación de todos los pasos de la transcripción (iniciación, elongación y terminación), además de acoplar la transcripción con el procesamiento de los RNAs nacientes de forma co-transcripcional. Aunque el proceso es considerablemente más complejo, el patrón de fosforilación en Ser5 alrededor de los sitios de inicio de la transcripción y Ser2 hacia el final de la transcripción sirve para reclutar proteínas que intervienen en los distintos momentos del ciclo de transcripción (Zaborowska et al., 2016).

La iniciación de la transcripción se lleva a cabo por el ensamblaje del complejo de pre-iniciación en los promotores, formado por factores de

transcripción generales, el complejo Mediador y la RNAP II (Fig. 9). Cuando aproximadamente 25 nucleótidos del mRNA naciente son sintetizados, se añade una 7-metilguanosina (m^7G) al extremo 5' formándose el 5'cap del mRNA (Sainsbury et al., 2015).

Una vez iniciada la transcripción, la RNAP II continua con la elongación (la progresión de la RNAP II por del locus del gen que va transcribiendo) pero antes se puede parar en los promotores, lo que se conoce como "pausa/liberación de la RNAP II". Esta pausa añade otro paso más en la regulación de la transcripción y se encuentra predominantemente en genes inducibles o regulados durante el desarrollo (Adelman & Lis, 2012).

El mecanismo de pausa y liberación es regulado por DSIF y NELF que se unen a la RNAP II inhibiéndola. Cuando p-TEFb fosforila la subunidad mayor de DSIF (Spt5) y NELF, se libera el bloqueo de la transcripción. La fosforilación de NELF produce su separación de la RNAP II y la fosforilación de DSIF hace que pase de un inhibidor transcripcional a un regulador positivo de la elongación de la transcripción (Jonkers & Lis, 2015).

En *C. elegans*, la regulación transcripcional se puede describir como la típica en eucaritas siendo la mayor diferencia la ausencia de NELF (Reinke et al., 2013).

Recientemente se ha observado que complejo Integrador está presente en los promotores de genes de expresión rápida (IEGs), dónde en respuesta a un estímulo como el EGF (factor de crecimiento epidérmico), se enriquece su cantidad en estas regiones dónde promueve el reclutamiento de p-TEFb (Gardini et al., 2014).

Una vez que la RNAP II escapa del promotor va a la región codificante del gen dónde se produce la elongación del transcrito. Durante este proceso se reclutan los factores necesarios para llevar a cabo la eliminación de los intrones (*splicing*) de forma co-transcripcional (Saldi et al., 2016).

La terminación de la transcripción tiene lugar con la liberación de la RNAP II cuando llega al final del gen que está siendo transcrito. A pesar de que los mecanismos moleculares de este proceso no son muy bien conocidos todavía, está ampliamente aceptado por la comunidad científica que el procesamiento en 3' de los transcritos juega un papel principal en la terminación de estos mismos (Porrua & Libri et al., 2015). Tres tipos de mecanismos para procesamiento en 3' se han descrito dependiendo del tipo de transcrito de la RNAP II: mRNAs poliadenilados, mRNAs de histonas dependientes de replicación o snRNAs (Fig. 11).

En el caso de los mRNAs poliadenilados el procesamiento en 3' se lleva a cabo por miembros de los complejos, CPSF y CtsF, de la maquinaria de corte y poliadenilación de transcritos, dónde la subunidad CPSF73 es la responsable del corte endonucleolítico. Los mRNAs de las histonas dependientes de replicación no están poliadenilados y su corte en 3' requiere de la proteína SBLP, un U7 snRNA y un complejo de corte formado, entre otros factores, por CPSF73. Por último, los snRNAs son procesados en 3' por el complejo Integrador (Peart et al., 2013).

El mecanismo molecular exacto por el que los snRNAs son transcritos y procesados permanece desconocido. Sin embargo, se ha propuesto un modelo para explicar la función del complejo Integrador su transcripción y procesmiento (Fig. 13). Inicialmente la RNAP II es fosforilada en Ser5 y Ser7 por la subunidad CDK7 del factor de transcripción TFIIH. La fosforilación en Ser5 es importante para la adición del 5'cap (m^7G). Después, la fosforilación en Ser7 puede ser reconocida por la proteína RPAP2 que se encarga de

eliminar la fosforilacion en Ser5, lo que conlleva el reclutamiento de, al menos, una parte de las subunidades del complejo Integrador. Mas adelante, la fosforilación en Ser2 por la subunidad CDK9 de p-TEFb permitiría el reclutamiento del resto de subunidades del complejo. Al final, cuando se transcribe una secuencia específica en el snRNA naciente (3'box), ésta es reconocida por el complejo Integrador que produce el corte por un mecanismo aún desconocido pero que implica a las subunidades Ints11 (catalítica) e Ints9 (Peart et al., 2013; Baillat & Wagner, 2015).

Una vez los snRNAs son transcritos y procesados se transportan al citoplasma dónde se ensamblan en ribonucleoproteínas, sufren la hipermetilación del cap (TMG) y un corte adicional en 3' para volver al núcleo dónde realizan su función en el splicing de pre-mRNAs (Matera et al., 2007).

Los snRNAs se denominan comúnmente "Us" por su alto contenido en Uridina. Estos snRNAs no codificantes de entre 60 y 200 nucleótidos no contienen intrones, no están poliadenilados y generalmente su expresión es alta. Se dividen en dos grupos, Sm o Lsm, en base a su secuencia y factores proteicos. Los Lsm (U6 y U6 atac) son transcritos por la RNAP III. Los Sm son transcritos por la RNAP II y están compuestos de U1, U2, U4, U4atac, U5, U7, U11 y U12. Los Us se ensamblan con otras proteínas para formar pequeñas ribonucleoproteínas nucleares (snRNPs), a excepción de la snRNP U7 implicada en el procesamiento de histonas, las demás forman parte del espliceosoma mayor y menor, ambos necesarios para el *splicing* de pre-mRNAs.

Una característica de peculiar de *C. elegans* y otros metazoos inferiores es la presencia de otro tipo de snRNAs, SL RNAs, que están implicados en el *trans-splicing*. En *C. elegans*, aproximadamente el 70% de los transcritos (a parte del *splicing* normal en *cis*) sufren *trans-splicing o SL trans-splicing*, que consiste en reemplazar la región 5' UTR del transcrito por la una secuencia

de 21- 22 nucleótidos que pertenece a la región 5' del SL RNA (Lasda & Blumental, 2011).

2.2. El complejo Integrador en la reparación del ADN

Nuestro genoma se ve constantemente atacado tanto por factores exógenos como endógenos (Fig. 14). A pesar de que las mutaciones son beneficiosas en una escala evolutiva, la reparación de éstas es vital para asegurar la estabilidad genómica. Las células son capaces de detectar cuando su ADN está dañado y activar mecanismos de reparación o tolerancia de lesiones. En caso de que no sea posible subsanar el daño, se activan mecanismos para eliminar las células dañadas (apoptosis, autofagia…).

Las lesiones más comunes son roturas en una sola hebra del ADN (SSBs) pero las más tóxicas son roturas en las dos cadenas del ADN (DSBs). En eucariotas, hay dos rutas principales de reparación de DSBs: la recombinación homóloga (HR) y la unión de extremos no homólogos (NHEJ) (Fig. 15). La elección de una u otra ruta depende del contexto de reparación, el estado de la rotura y la fase de ciclo celular en la que se encuentre la célula. La ruta de recombinación homóloga, HR, es de alta fidelidad ya que utiliza como molde para la reparación el ADN homólogo de la cromátida hermana, de tal forma que se puede completar la información que fue perdida en el sitio de rotura (actúa exclusivamente durante las fases S y G2 del ciclo celular) (Li & Heyer, 2008; Jasin & Rothstein, 2013; Schwertman et al., 2016).

El mantenimiento de la estabilidad genómica adquiere mayor relevancia durante la división celular debido a que cualquier cambio que se produzca en este momento será transmitido a la descendencia. Por eso, la integridad genómica y los elementos necesarios para la división celular son monitorizados en diversos puntos a lo largo del ciclo mitótico llamados

checkpoints o puntos de control (Lodish et al., 2013). Normalmente las células "sobrepasan" estos mecanismos de supervisión. En cambio bajo determinadas circunstancias como el daño, los *checkpoints* son activados y las células paran su ciclo para poder reparar cualquier defecto antes de que sea transmitido a la descendencia. Los *checkpoints* de daño son activados durante la Interfase en G1, G2 y M (Fig. 17). Se pueden definir como el conjunto de rutas que interactúan y se coordinan para reconocer alteraciones en el ADN y activar una respuesta que dará lugar a la inhibición de CDKs (quiniasas dependientes de ciclina) y parada del ciclo celular (Iliakis et al., 2003; Beishline & Azizkhan-Clifford, 2014).

Las quinasas ATM y ATR son los reguladores centrales de la respuesta al daño al ADN. Cuando hay DSBs, por ejemplo las causadas por irradiación, se activa la ruta ATM-Chk2. En cambio las SSBs, como las que se producen a causa de la luz ultravioleta, activan la ruta ATR-Chk1. Las quinasas Chk1 y Chk2 convergen en cascadas de señalización que inhiben CDC25 fosfatasas para bloquear la desfosforilación de complejos Cdk-ciclina y a su vez activan a p53 y éste promueve la inducción transcripcional de inhibidores de Cdk (Awasthi et al., 2015).

Un aspecto interesante del complejo Integrador es la implicación de al menos dos de sus subunidades, Ints3 e Ints6, en la respuesta a daño al ADN. Ints3, Ints6 junto con hSSB1 y C9orf80 se encuentran formando un complejo estable "hSSB1/INTS", bien como parte del complejo Integrador o bien como un sub-complejo (Zhang et al., 2013; Skaar et al., 2015). La proteína hSSB1 se había descrito previamente por su implicación en la reparación de DSBs en la ruta HR. En respuesta a daño, la quinasa ATM fosforila a hSSB1 y como consecuencia ésta se localiza formando focos en los sitios de doble rotura del ADN. Además, la activación y estabilización de hSSB1 genera una retroalimentación positiva para amplificar la cascada de señalización por

ATM (Richard et al., 2008; Richard et al., 2011). A pesar de que el mecanismo molecular del complejo hSSB1/INTS se desconoce, se sabe que juega un papel clave en la reparación del daño por la ruta HR ya que la depleción de hSSB1, Ints6 o Ints3 da lugar a un defecto en el reclutamiento de proteínas de reparación de esta ruta como Rad51 o BRCA1 a los sitios de daño (Skaar et al., 2009; Richard et al. 2011). Además, la depleción del complejo MRN (formado por Rad50, Mre11 y Nbs1), que participa en los primeros pasos de la HR, reduce significativamente los focos de Ints3 y hSSB1, lo que sugiere que el complejo hSSB1/INTS actúa en la HR por debajo de este mismo (Huang et al., 2009).

Dada la alta frecuencia de mutaciones en proteínas implicadas en la reparación del daño en cáncer, es muy posible que se encuentren mutaciones patogénicas en los miembros de este sub-complejo o otras subunidades del complejo Integrador que sean necesarias para la reparación del ADN.

OBJETIVOS

Los objetivos de esta tesis doctoral son:

1. La identificación del complejo Integrador en *C. elegans*.

2. El análisis del efecto transcripcional que se produce aguas abajo de los loci de snRNAs como consecuencia de la depleción de alguna de las subunidades del complejo Integrador.

3. El análisis de la implicación de la subunidad 6 del complejo Integrador en la respuesta a daño al ADN causado por irradiación gamma.

4. La descripción del efecto transcripcional que se produce aguas abajo de los loci de snRNAs como consecuencia de la irradiación gamma.

RESULTADOS Y DISCUSIÓN

En nuestro laboratorio empezamos a trabajar con el mutante *t1903*, aislado en el laboratorio del Dr. Ralf Schnabel (Braunschweig, Alemania). Este mutante termosensible posee una mutación puntual (C⇒T) en el gen *dic-1* que resulta en el cambio del aminoácido Ser 850 a Phe. El gen *dic-1* en *C. elegans* es ortólogo de la subunidad del complejo Integrador 6 (Ints6) de humanos, localizada en el núcleo donde interviene en el procesamiento en 3' de los snRNAs. Sin embargo, estudios previos en *C. elegans* revelaron que la proteína DIC-1 se localiza en la membrana interna de la mitocondria y su función esta relacionada con el proceso de morfogénesis mitocondrial (Han et al., 2006; Lee et al., 2009).

Las grabaciones de embriones del mutante termosensible *t1903* con microscopía 4D (a 25°C), nos revelaron que los embriones morían presentando defectos en morfogénesis y necrosis (Figs. 29, 30 y 31).

1. Localización de DIC-1

Para empezar a estudiar la función de DIC-1 decidimos corroborar su localización. Para ello, se hicieron distintas líneas de transgénicos. Se añadió la etiqueta "3xFLAG::eGFP" al extremo carboxilo-terminal de *dic-1*, tanto para chequear localización de la proteína (por visualización de GFP) como para facilitar posteriores experimentos bioquímicos. Además, se utilizaron distintos promotores, tanto el endógeno como uno que proporciona altos niveles de expresión, el *eft-3*. Estas construcciones, Fig. 32A y 33A, se integraron en el cromosoma II por el sistema mosSCI (Frokjaer-Jensen et al., 2008). Además, se hicieron arrays extracromosomales utilizando el plásmido que lleva el promotor *eft-3*. En todos los casos observamos que la localización de DIC-1 era mayoritariamente nuclear.

Esta controversia de la localización de DIC-1 entre nuestro estudio y el publicado por Han y colaboradores puede explicarse porque la aproximación para detectar DIC-1 es ligeramente distinta. Mientras que nosotros hemos fusionado la eGFP al extremo carboxilo terminal, ellos la fusionaron al extremo amino-terminal, quizá la eGFP pueda afectar a la conformación de la proteína. De todas formas, predicciones bioinformáticas indican que tanto Ints6 en humanos como su ortólogo en *C. elegans* tienen secuencias de localización a núcleo y mitocondria.

La localización de DIC-1 en núcleo, al igual que su ortólogo en humanos, nos sugirió que su función pudiera no ser divergente sino que estuviera conservada en *C. elegans*.

2. Análisis transcripcional del mutante *t1903*

Posteriormente, el análisis transcripcional del mutante *t1903*, tanto a 15ºC como a 25ºC (Fig. 35), nos permitió averigual la implicación de DIC-1 en el procesamiento en 3' de los snRNAs: U1, U2, U4, U5, U6 y SL. Por lo tanto, renombramos al gen como *ints-6*.

A pesar de que una parte de los snRNAs no estaba procesada correctamente, el *splicing* de los genes era mayoritariamente correcto. Asumimos que, o bien la porción correctamente procesada es capaz de llevar a cabo el *splicing*, o los snRNAs no procesados mantienen, al menos en parte, su actividad funcional.

3. Breve caracterización de Ints6 en humanos

En este momento de nuestro estudio, quisimos caracterizar de forma breve la subunidad Ints6 en humanos ya que queríamos comparar posibles similitudes y/o diferencias entre *H. sapiens* y *C. elegans*. Corroboramos que la localización de Ints6 en humanos es nuclear (Fig. 36) y que su depleción

conlleva a defectos en el procesamiento en 3' de los snRNAs (Fig. 39). Además, observamos que la disminución de los niveles de Ints6 en la línea celular RPE producían un aumento de la proliferación celular, de forma consistente a otros estudios previos dónde una sobre-expresión de Ints6 en otras células de cultivo inhibía la formación de colonias. También analizamos la dinámica de expresión de Ints6 a lo largo de las distintas fases del ciclo celular (G1, G1/S, G2, M) y pudimos observar que sus niveles de proteína se mantenían constantes (Fig. 37).

Tanto en el WB realizado para comprobar que la depleción de los niveles de Ints6 era correcta como en el WB para examinar los niveles de expresión de Ints6 nos llamó la atención una banda superior que, curiosamente, aumentaba tras la depleción de Ints6 y variaba a lo largo de las distintas fases del ciclo celular. En un principio consideramos que la banda era inespecífica debido a su tamaño y a que no respondía a la depleción por siRNAs. Más adelante, comprobamos que Ints6 tiene un parálogo en el genoma humano llamado DDX26B, verificamos que los siRNAs usados no mapeaban en la secuencia de DDX26B y sin embargo el anticuerpo que usamos contra Ints6 había sido diseñado contra la región Ct de Ints6 y ésta, estaba bastante conservada en su parálogo. Por tanto, sería interesante descifrar si el anticuerpo reconoce a DDX26B y si ambas proteínas tienen papeles redundantes o alguna de ellas tiene una regulación dependiente de ciclo.

4. Indentificación del complejo Integrador en *C. elegans*

Los datos obtenidos del análisis transcripcional del mutante *t1903* indicaban la existencia de un complejo Integrador en *C. elegans*, por lo que proseguimos con su identificación.

Analizamos de forma bioinformática los posibles ortólogos de los miembros del complejo Integrador (Tabla 4). A excepción de la subunidad 14, pudimos encontrar candidatos para el resto de miembros, las subunidades INTS-4, -5, -8, -10 y -12 sólo presentaron homología en una porción de la secuencia de la proteína menor al 50%. Posteriormente, inmunoprecipitamos Ints6 y analizamos el inmunoprecipitado por espectrometría de masas. Aparte de otras proteínas, detectamos péptidos para todos los ortólogos de las subunidades del complejo Integrador que habíamos obtenido en análisis bioinformático excepto para INTS-10 e INTS-12 (Tabla 5). Estas dos subunidades tienen una secuencia proteica corta (su peso molecular es muy bajo, 9 kDa y 8,5 kDA respectivamente), por lo que aumenta considerablemente la dificultad en la detección péptidos por espectrometría de masas.

Finalmente, deplecionamos cada putativa subunidad del complejo Integrador mediante RNAi y analizamos su efecto en el procesamiento en 3' de los snRNAs mediante RT-PCR (Fig. 44A, 44B y 44C). En todas los casos observamos un procesamiento incorrecto aunque en distintos grados. Los defectos más agresivos se obtuvieron con la depleción las subunidades INTS-2, -4, -5, -6 y -7, mientras que apenas se detectaron defectos con la depleción de INTS-3, INTS- 10, INTS-12 e INTS-13.

El RNA total de la depleción de cada una de las subunidades también se analizó por ultrasecuenciación (Fig. 45). No se observaron defectos en el procesamiento con la depleción de INTS-3, INTS-10, INTS-12 e INTS-13. Quizá porque en las condiciones de este experimento la técnica haya sido algo menos sensible.

Estos experimentos (la predicción bioinformática, el análisis por espectrometría de masas y análisis de la implicación de las subunidades en el

procesamiento de en 3' de los snRNAs) nos permitieron concluir que en *C. elegans* existe un complejo Integrador compuesto por, al menos, once subunidades (INTS-1, -2, -3, -4, -5, -6, -7, -8, -9, -11 e INTS-13) que es el encargado del procesamiento de los snRNAs. Cabe destacar que este es el primer estudio dónde se demuestra experimentalmente la implicación del complejo Integrador en el procesamiento en 3' de los snRNA U6 y SL.

Una de las ventajas de utilizar *C. elegans* como organismo modelo frente a células en cultivo, es la posibilidad de ver el efecto de la depleción de un gen en todo el organismo. Así, también pudimos observar los fenotipos producidos por la depleción de los distintos miembros del complejo Integrador, los cuales también variaban dependiendo de la subunidad deplecionada (Fig. 43). Los fenotipos más agresivos de arresto larvario en L2-L3 no se correlacionaban exactamente con los niveles observados en los defectos de procesamiento. Por ejemplo, la depleción de INTS-4 e INTS-7 daba lugar a altos niveles de snRNAs procesados incorrectamente. Sin embargo, los gusanos alimentados con el clon bacteriano de RNAi de *ints-4* arrestaban en L2-L3 pero los alimentados con el clon de *ints-7* llegaban a adultos y manifestaban letalidad embrionaria.

Estas diferencias en los fenotipos sugieren que el complejo Integrador (o al menos parte de sus subunidades) desempeña un papel importante durante el desarrollo de *C. elegans.*

5. Análisis de los defectos de terminación de la transcripción y la formación de "sn-mRNAs"

Tanto la depleción de alguna las subunidades del complejo Integrador como la mutación en *ints-6* del mutante *t1903* dan lugar a un defecto en el procesamiento en 3' de los snRNAs y en la terminación de la transcripción,

como consecuencia, la RNAP II continúa transcribiendo. Los genes situados a continuación de los snRNAs pueden estar en *sense,* o bien, en *antisense*. En el primer caso, la continuación de la transcripción da lugar a la formación de RNAs quiméricos formados por el snRNA y el mRNA, que hemos llamado "sn-mRNAs". En el segundo caso, se forman RNAs con un snRNA en 5' seguido de un mRNA *antisense*.

Observamos que los genes localizados en *sense* con su respectivo snRNA, eran en su mayoría correctamente procesados (los intrones eran eliminados) y poliadenilados. Además, estos mRNAs, los cuales no se expresaban en N2 en las condiciones de nuestro experimento, tenían la típica estructura del un RNA potencialmente traducible: un 5' cap, una cola poly(A) y además, el codón ATG de iniciación estaba precedido por A, algo común en los mensajeros que codifican para proteína en *C. elegans* (Riddle et al., 1997).
Quisimos comprobar si estos RNAs quiméricos se traducían a proteínas o péptidos. Para averiguarlo, elegimos dos genes localizados en *sense* y aguas abajo de un U1, otros dos de un U2 y otro de un SL. Etiquetamos eGFP o 3xFLAG::eGFP al extremo Ct de estos cinco genes (Fig. 46). Después, deplecionamos las subunidades del complejo Integrador y comprobamos por WB si los RNAs quiméricos "sn-mRNAs" se traducían. Sólo el "sn-mRNA" precedido por el snRNA *sls2.8* daba lugar a una proteína (Fig. 47). También comprobamos si los "sn-mRNAs" podían estar formando péptidos en alguna de las fases de lectura. Se hicieron transgénicos insertando el cassete "HA::MYC::TY" que poseía cada etiqueta en cada una de las pautas de lectura (Fig. 48). Tras deplecionar alguna de las subunidades del complejo Integrador no se detectó la formación de péptidos (Fig. 49).
Puede ser que la estructura secundaria del U1 y el U2 impida la traducción de los estos RNAs.

En mamíferos y levaduras se han descrito fenotipos similares en cuanto a la falta de procesamiento en 3' de los Us y la poliadenilación de los transcritos resultantes (Abou Elela & Ares, 1998; Yamamoto et al., 2014). En mamíferos la depleción de NELF o subunidades del complejo Integrador da lugar a Us poliadenilados en sitios crípticos de poliadenilación. Estos transcritos son de distinta longitud, se poliadenilan en distintos sitios, y sus colas de poliadeniladas también varían en longitud (Yamamoto et al., 2014).

En cambio, en el fenotipo de "sn-mRNAs" que nosotros hemos observado, las lecturas de ultrasecuencición muestran que los transcritos terminan al final del mRNA, por tanto, asumimos que la terminación de la transcripción está acoplada a la poliadenilación correcta de estos transcritos en el sitio PAS. Estos resultados también sugieren que los "sn-mRNAs" contienen largas colas de poliadenilación añadidas por la maquinaria canónica de poliadenilación. Además, el hecho de que los RNAs quiméricos se detecten fácilmente ante la depleción de alguna de las subunidades del complejo Integrador, indica que los transcritos son estables y no son reconocidos por los mecanismos de vigilancia y control de la calidad del RNA. De todas formas, sería intersante determinar la vida media de estos "sn-mRNAs". Su estabilidad podría deberse a la estructura secundaria del snRNA o bien a la cola poliadenilada. El hecho de que el gen Y75B8A.23 dentro del transcrito "sn-mRNA" se traduzca a proteína indica que, al menos, éste RNA quimérico es transportado al citoplasma y que los mecanismos de calidad de control del RNA no lo reconocen como aberrante.

6. Implicación de INTS-6 en la respuesta a daño al ADN

En el análisis por espectrometría de masas para detectar posibles interactores de INTS-6 observamos la presencia tanto de INTS3 como de NABP1 (Tabla 5), ortólogo de la proteína en humanos hSSB1. Estos

resultados nos sugirieron que es posible que en *C. elegans* exista un complejo NABP1-INTS y que actúe en la reparación del daño por la ruta HR al igual que en humanos.

Gracias a análisis bioinformáticos sabemos que la Ser 850 de INTS-6 tiene una alta probabilidad de ser fosforilada y además, una de las quinases probables para fosforilar ese residuo es ATM. En la ruta de reparación por HR uno de los pasos iniciales es la activación de ATM, que rápidamente inicia una cascada de señalización de respuesta a daño en la célula.

Primero comprobamos si en *C. elegans ints-6* está implicado en la respuesta al daño. Para ello, deplecionamos INTS-6 desde estadío L1 y en estadío L4 irradiamos los gusanos (90 Gy). A las 24 horas, diseccionamos las gónadas e hicimos immunohistoquímicas con el anticuerpo contra RAD-51. La proteína RAD-51 es un componente de la ruta de reparación por HR, una recombinasa que sirve para mediar la inviasión de la hebra dañada en la cromátida hermana. En respuesta a daño, RAD-51 forma focos en los sitios de DSBs.

Observamos que en condiciones de silenciamiento de *ints-6* no se producía el reclutamiento de RAD-51 en la región mitótica de la gónada, por lo que pudimos concluir que *ints-6* es un componente clave en la ruta de reparación por HR (Fig. 51). Así mismo, también observamos que la depleción de INTS-6 afectaba a la inducción de la fosforilación en Tyr15 de Cdk1 que se produce en respuesta a daño para parar el ciclo celular en G2/M (Fig. 52). En cambio, las células si que eran capaces de arrestar (Fig. 53), por lo que otros mecanimos redundantes deben estar actuando para parar el ciclo celular.

Con el objetivo de descifrar si INTS-6 esta fosforilado en la Ser 850, se hicieron trasgénicos que mimetizaban la forma fosforilada de la proteína (S850E) y otros que no podían ser fosforilados (S850A). Además se añadió la etiqueta 3xFLAG al extremo Ct de INTS-6. De forma similar, miramos el

reclutamiento de RAD-51 en respuesta a irradiación gamma. A pesar de que no se se observaron diferencias en el número total de focos de RAD-51, los transgénicos que no podían ser fosforilados (S850A) no mostrarón células con más de 16 focos, lo que indica que la fosforilación de este residuo puede ser necesaria para el reclutamiento y la estabilidad de RAD-51 en la formación de focos.

Por último, hipotetizamos que si en condiciones de daño INTS-6 se re-localiza formando focos, quizá su "ausencia" dentro del complejo Integrador en estas condiciones podría dar lugar al mismo fenotipo que observamos mediante su depleción, es decir, a defectos en el procesamiento en 3' de los snRNAs y en la terminación de la transcripción y, por consiguiente, dar lugar a la formación de "sn-mRNAs" (Fig. 60). Con el objetivo de verificar nuestra hipótesis irradiamos gusanos N2 en estadío L4 (90Gy) y comprobamos la supuesta formación de RNAs quimérios a distintos tiempos. A los 15 minutos ya se detectamos la presencia de estos "sn-mRNAs", siendo la respuesta más alta a las 24 horas.

Investigaciones recientes están desvelando la importancia del RNA en la respuesta a daño al ADN mediante diversos mecanismos (Francia et al., 2012; Wei et al., 2012; Francia et al., 2016; Capozzo et al., 2017).

Quizá estos RNAs quimérios actúen como esponjas para regulación de microRNAs o como esquetos para proteínas de reparación o de la maquinaria de remodelación de la cromatina. De todas formas, aún está por demostrar si estos "sn-mRNAs" son un artefacto o bien existe una respuesta coordinada y forman parte de la respuesta a daño al ADN y /o reparación.

CONCLUSIONES

1. El mutante termosensible *t1903* de *C. elegans* tiene mutación puntual en el gen *ints-6* (C→T) que produce un cambio en el residuo Ser 850 de la proteína a Phe. Los embriones de este mutante mueren a 25ºC presentando defectos en morfogénesis.

2. Tanto INTS-6 en *C. elegans* como su ortólogo en humanos, Ints6, muestran una localización mayoritariamente nuclear.

3. En *C. elegans*, existe un complejo Integrador formado por, al menos, once subunidades (INTS-1, INTS-2, INTS-3, INTS-4, INTS-5, INTS-6, INTS-7, INTS-8, INTS-9, INTS-11 and INTS-13), que funciona en el procesamiento en 3' de los snRNAs (U1, U2, U4, U5, U6 y SL).

4. La depleción de Ints6 en la línea celular RPE da lugar a defectos en el procesamiento en 3' de, al menos, los snRNAs U1 y U2. Además provoca una sobreproliferación celular.

5. Tanto la mutación *t1903* como la depleción de alguna de las subunidades del complejo Integrador en *C. elegans* dan lugar a defectos del procesamiento en 3' de los snRNAs y en la terminación de la transcripción. Como consecuencia, los genes localizados aguas abajo de los snRNAs localizados *en sense* se sobre-expresan. Se forman RNAs quiméricos "sn-mRNAs" que mayoritariamente están correctamente procesados y poniladenilados.

6. Los RNAs quiméricos "sn-mRNAs" formados por "U1-mRNAs" o "U2-mRNAs" no se traducen a proteínas o péptidos. Sin embargo, el "SL2-mRNA" formado por *sls-2.8* y el gen Y75B8A.23 se traduce a proteína.

7. El gen *ints-6* en *C. elegans* está implicado en la reparación del daño en el ADN. INTS-6 forma focos en respuesta al daño producido por irradiación gamma. La depleción de INTS-6 da lugar a defectos en el reclutamiento de RAD-51 a los sitios de doble rotura, así como defectos en la inducción

de fosforilación en Tyr15 Cdk1 bajo circunstancias de daño. Además, la Ser850 de INTS-6 podría fosforilarse en respuesta a daño.

8. En *C. elegans*, se forman RNAs quiméricos "sn-mRNAs" como consecuencia de la irradiación gamma. La formación de estos RNAs podría ser un mecanismo fisiológico mediado por regulación específica del complejo Integrador en la respuesta al daño al ADN.

Bibliography

Aballay, A., & Ausubel, F. M. (2001). Programmed cell death mediated by ced-3 and ced-4 protects Caenorhabditis elegans from Salmonella typhimurium-mediated killing. *Proc Natl Acad Sci U S A, 98*(5), 2735-2739. doi:10.1073/pnas.041613098

Abou Elela, S., & Ares, M., Jr. (1998). Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J, 17*(13), 3738-3746. doi:10.1093/emboj/17.13.3738

Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet, 13*(10), 720-731. doi:10.1038/nrg3293

Ahn, S. H., Kim, M., & Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell, 13*(1), 67-76.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., & Ansari, A. Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell, 34*(3), 387-393. doi:10.1016/j.molcel.2009.04.016

Albertson, D. G. (1984). Localization of the ribosomal genes in Caenorhabditis elegans chromosomes by in situ hybridization using biotin-labeled probes. *EMBO J, 3*(6), 1227-1234.

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol, 16*(3), 155-166. doi:10.1038/nrm3951

Altun, Z.F. & Hall, D.H. (2009). Introduction. *In WormAtlas.*

Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., . . . Kruglyak, L. (2012). Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nat Genet, 44*(3), 285-290. doi:10.1038/ng.1050

Anderson, M. A., Jodoin, J. N., Lee, E., Hales, K. G., Hays, T. S., & Lee, L. A. (2009). Asunder is a critical regulator of dynein-dynactin localization during Drosophila spermatogenesis. *Mol Biol Cell, 20*(11), 2709-2721. doi:10.1091/mbc.E08-12-1165

Ares, M., Jr. (1986). U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. *Cell, 47*(1), 49-59.

Avery, L., & Shtonda, B. B. (2003). Food transport in the C. elegans pharynx. *J Exp Biol, 206*(Pt 14), 2441-2457.

Avery, L., & You, Y. J. (2012). C. elegans feeding. *WormBook*, 1-23. doi:10.1895/wormbook.1.150.1

Awasthi, P., Foiani, M., & Kumar, A. (2015). ATM and ATR signaling at a glance. *J Cell Sci, 128*(23), 4255-4262. doi:10.1242/jcs.169730

Baek, H. J., Malik, S., Qin, J., & Roeder, R. G. (2002). Requirement of TRAP/mediator for both activator-independent and activator-dependent transcription in conjunction with TFIID-associated TAF(II)s. *Mol Cell Biol*, 22(8), 2842-2852.

Baillat, D., Hakimi, M. A., Naar, A. M., Shilatifard, A., Cooch, N., & Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with

the C-terminal repeat of RNA polymerase II. *Cell, 123*(2), 265-276. doi:10.1016/j.cell.2005.08.019

Baillat, D., & Wagner, E. J. (2015). Integrator: surprisingly diverse functions in gene expression. *Trends Biochem Sci, 40*(5), 257-264. doi:10.1016/j.tibs.2015.03.005

Barnum, K. J., & O'Connell, M. J. (2014). Cell cycle regulation by checkpoints. *Methods Mol Biol, 1170*, 29-40. doi:10.1007/978-1-4939-0888-2_2

Barriere, A., & Felix, M. A. (2014). Isolation of C. elegans and related nematodes. *WormBook*, 1-19. doi:10.1895/wormbook.1.115.2

Beishline, K., & Azizkhan-Clifford, J. (2014). Interplay between the cell cycle and double-strand break response in mammalian cells. *Methods Mol Biol, 1170*, 41-59. doi:10.1007/978-1-4939-0888-2_3

Bennett, C. B., Lewis, A. L., Baldwin, K. K., & Resnick, M. A. (1993). Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. *Proc Natl Acad Sci U S A, 90*(12), 5613-5617.

Bertoli, C., Skotheim, J. M., & de Bruin, R. A. (2013). Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol, 14*(8), 518-528. doi:10.1038/nrm3629

Blumenthal, T. (2005). Trans-splicing and operons. *WormBook*, 1-9. doi:10.1895/wormbook.1.5.1

Booher, R. N., Holman, P. S., & Fattaey, A. (1997). Human Myt1 is a cell cycle-regulated kinase that inhibits Cdc2 but not Cdk2 activity. *J Biol Chem, 272*(35), 22300-22306.

Bowman, E. A., & Kelly, W. G. (2014). RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus, 5*(3), 224-236. doi:10.4161/nucl.29347

Boxem, M., Srinivasan, D. G., & van den Heuvel, S. (1999). The Caenorhabditis elegans gene ncc-1 encodes a cdc2-related kinase required for M phase in meiotic and mitotic cell divisions, but not for S phase. *Development, 126*(10), 2227-2239.

Boxem, M., & van den Heuvel, S. (2001). lin-35 Rb and cki-1 Cip/Kip cooperate in developmental regulation of G1 progression in C. elegans. *Development, 128*(21), 4349-4359.

Branzei, D., & Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol, 9*(4), 297-308. doi:10.1038/nrm2351

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics, 77*(1), 71-94.

Brennwald, P., Porter, G., & Wise, J. A. (1988). U2 small nuclear RNA is remarkably conserved between Schizosaccharomyces pombe and mammals. *Mol Cell Biol, 8*(12), 5575-5580.

Busch, H., Reddy, R., Rothblum, L., & Choi, Y. C. (1982). SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem, 51*, 617-654. doi:10.1146/annurev.bi.51.070182.003153

Callebaut, I., Moshous, D., Mornon, J. P., & de Villartay, J. P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res, 30*(16), 3592-3601.

Card, C. O., Morris, G. F., Brown, D. T., & Marzluff, W. F. (1982). Sea urchin small nuclear RNA genes are organized in distinct tandemly repeating units. *Nucleic Acids Res, 10*(23), 7677-7688.

Carter, R. J., & Parsons, J. L. (2016). Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Mol Cell Biol, 36*(10), 1426-1437. doi:10.1128/MCB.00030-16

Cazalla, D., Xie, M., & Steitz, J. A. (2011). A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell, 43*(6), 982-992. doi:10.1016/j.molcel.2011.07.025

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science, 263*(5148), 802-805.

Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W., & Vogelstein, B. (1999). 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature, 401*(6753), 616-620. doi:10.1038/44188

Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T. R., Warren, W. D., Marzluff, W. F., & Wagner, E. J. (2012). An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA, 18*(12), 2148-2156. doi:10.1261/rna.035725.112

Chen, J., & Wagner, E. J. (2010). snRNA 3' end formation: the dawn of the Integrator complex. *Biochem Soc Trans, 38*(4), 1082-1087. doi:10.1042/BST0381082

Chisholm, A. D., & Hardin, J. (2005). Epidermal morphogenesis. *WormBook*, 1-22. doi:10.1895/wormbook.1.35.1

Ciarletta, P., Ben Amar, M., & Labouesse, M. (2009). Continuum model of epithelial morphogenesis during Caenorhabditis elegans embryonic elongation. *Philos Trans A Math Phys Eng Sci, 367*(1902), 3379-3400. doi:10.1098/rsta.2009.0088

Clejan, I., Boerckel, J., & Ahmed, S. (2006). Developmental modulation of nonhomologous end joining in Caenorhabditis elegans. *Genetics, 173*(3), 1301-1317. doi:10.1534/genetics.106.058628

Clemenson, C., & Marsolier-Kergoat, M. C. (2009). DNA damage checkpoint inactivation: adaptation and recovery. *DNA Repair (Amst), 8*(9), 1101-1109. doi:10.1016/j.dnarep.2009.04.008

Consortium, C. e. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science, 282*(5396), 2012-2018.

Corsi, A. K., Wightman, B., & Chalfie, M. (2015). A Transparent Window into Biology: A Primer on Caenorhabditis elegans. *Genetics, 200*(2), 387-407. doi:10.1534/genetics.115.176099

Craig, A. L., Moser, S. C., Bailly, A. P., & Gartner, A. (2012). Methods for studying the DNA damage response in the Caenorhabdatis elegans germ line. *Methods Cell Biol, 107*, 321-352. doi:10.1016/B978-0-12-394620-1.00011-4

Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., . . . Vannini, A. (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys, 37*, 337-352. doi:10.1146/annurev.biophys.37.032807.130008

Dai, Y., & Grant, S. (2010). New insights into checkpoint kinase 1 in the DNA damage response signaling network. *Clin Cancer Res, 16*(2), 376-383. doi:10.1158/1078-0432.CCR-09-1029

Davis, M. W., Hammarlund, M., Harrach, T., Hullett, P., Olsen, S., & Jorgensen, E. M. (2005). Rapid single nucleotide polymorphism mapping in C. elegans. *BMC Genomics, 6*, 118. doi:10.1186/1471-2164-6-118

Dengg, M., Garcia-Muse, T., Gill, S. G., Ashcroft, N., Boulton, S. J., & Nilsen, H. (2006). Abrogation of the CLK-2 checkpoint leads to tolerance to base-excision repair intermediates. *EMBO Rep, 7*(10), 1046-1051. doi:10.1038/sj.embor.7400782

Derheimer, F. A., O'Hagan, H. M., Krueger, H. M., Hanasoge, S., Paulsen, M. T., & Ljungman, M. (2007). RPA and ATR link transcriptional stress to p53. *Proc Natl Acad Sci U S A, 104*(31), 12778-12783. doi:10.1073/pnas.0705317104

DeSimone, J. N., Bengtsson, U., Wang, X., Lao, X. Y., Redpath, J. L., & Stanbridge, E. J. (2003). Complexity of the mechanisms of initiation and maintenance of DNA damage-induced G2-phase arrest and subsequent G1-phase arrest: TP53-dependent and TP53-independent roles. *Radiat Res, 159*(1), 72-85.

Doitsidou, M., Poole, R. J., Sarin, S., Bigelow, H., & Hobert, O. (2010). C. elegans mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS One, 5*(11), e15435. doi:10.1371/journal.pone.0015435

Dominski, Z., Yang, X. C., Purdy, M., Wagner, E. J., & Marzluff, W. F. (2005). A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol, 25*(4), 1489-1500. doi:10.1128/MCB.25.4.1489-1500.2005

Draetta, G., & Eckstein, J. (1997). Cdc25 protein phosphatases in cell proliferation. *Biochim Biophys Acta, 1332*(2), M53-63.

Egloff, S., & Murphy, S. (2008). Role of the C-terminal domain of RNA polymerase II in expression of small nuclear RNA genes. *Biochem Soc Trans, 36*(Pt 3), 537-539. doi:10.1042/BST0360537

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., . . . Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science, 318*(5857), 1777-1779. doi:10.1126/science.1145989

Egloff, S., O'Reilly, D., & Murphy, S. (2008). Expression of human snRNA genes from beginning to end. *Biochem Soc Trans, 36*(Pt 4), 590-594. doi:10.1042/BST0360590

Egloff, S., Szczepaniak, S. A., Dienstbier, M., Taylor, A., Knight, S., & Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem, 285*(27), 20564-20569. doi:10.1074/jbc.M110.132530

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., & Murphy, S. (2012). Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell, 45*(1), 111-122. doi:10.1016/j.molcel.2011.11.006

Elledge, S. J. (1996). Cell cycle checkpoints: preventing an identity crisis. *Science, 274*(5293), 1664-1672.

Emmons, S. W., & Sternberg, P. W. (1997). Male Development and Mating Behavior. In D. L. Riddle, T. Blumenthal, B. J. Meyer, & J. R. Priess (Eds.), *C. elegans II* (2nd ed.). Cold Spring Harbor (NY).

Engreitz, J. M., Ollikainen, N., & Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol, 17*(12), 756-770. doi:10.1038/nrm.2016.126

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., . . . Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol Cell, 31*(3), 337-346. doi:10.1016/j.molcel.2008.06.021

Ezzeddine, N., Chen, J., Waltenspiel, B., Burch, B., Albrecht, T., Zhuo, M., . . . Wagner, E. J. (2011). A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol Cell Biol, 31*(2), 328-341. doi:10.1128/MCB.00943-10

Falck, J., Petrini, J. H., Williams, B. R., Lukas, J., & Bartek, J. (2002). The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways. *Nat Genet, 30*(3), 290-294. doi:10.1038/ng845

Felix, M. A., & Duveau, F. (2012). Population dynamics and habitat sharing of natural populations of Caenorhabditis elegans and C. briggsae. *BMC Biol, 10*, 59. doi:10.1186/1741-7007-10-59

Fickenscher, H., & Fleckenstein, B. (2001). Herpesvirus saimiri. *Philos Trans R Soc Lond B Biol Sci, 356*(1408), 545-567. doi:10.1098/rstb.2000.0780

Filleur, S., Hirsch, J., Wille, A., Schon, M., Sell, C., Shearer, M. H., . . . Wieland, I. (2009). INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int, 9*, 28. doi:10.1186/1475-2867-9-28

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature, 391*(6669), 806-811. doi:10.1038/35888

Francia, S., Cabrini, M., Matti, V., Oldani, A., & d'Adda di Fagagna, F. (2016). DICER, DROSHA and DNA damage response RNAs are necessary for the secondary recruitment of DNA damage response factors. *J Cell Sci, 129*(7), 1468-1476. doi:10.1242/jcs.182188

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., . . . d'Adda di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature, 488*(7410), 231-235. doi:10.1038/nature11179

Frezal, L., & Felix, M. A. (2015). C. elegans outside the Petri dish. *Elife, 4*. doi:10.7554/eLife.05849

Frokjaer-Jensen, C. (2013). Exciting prospects for precise engineering of Caenorhabditis elegans genomes with CRISPR/Cas9. *Genetics, 195*(3), 635-642. doi:10.1534/genetics.113.156521

Frokjaer-Jensen, C., Davis, M. W., Hopkins, C. E., Newman, B. J., Thummel, J. M., Olesen, S. P., . . . Jorgensen, E. M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet, 40*(11), 1375-1383. doi:10.1038/ng.248

Fury, M. G., & Zieve, G. W. (1996). U6 snRNA maturation and stability. *Exp Cell Res, 228*(1), 160-163. doi:10.1006/excr.1996.0311

Gabrielli, B., Brooks, K., & Pavey, S. (2012). Defective cell cycle checkpoints as targets for anti-cancer therapies. *Front Pharmacol*, 3, 9. doi:10.3389/fphar.2012.00009

Garcia-Muse, T., & Boulton, S. J. (2005). Distinct modes of ATR activation after replication stress and DNA double-strand breaks in Caenorhabditis elegans. *EMBO J, 24*(24), 4345-4355. doi:10.1038/sj.emboj.7600896

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J. M., Wagner, E. J., . . . Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell, 56*(1), 128-139. doi:10.1016/j.molcel.2014.08.004

Gartner, A., MacQueen, A. J., & Villeneuve, A. M. (2004). Methods for analyzing checkpoint responses in Caenorhabditis elegans. *Methods Mol Biol, 280*, 257-274. doi:10.1385/1-59259-788-2:257

George, J., Castellazzi, M., & Buttin, G. (1975). Prophage induction and cell division in E. coli. III. Mutations sfiA and sfiB restore division in tif and lon strains and permit the expression of mutator properties of tif. *Mol Gen Genet, 140*(4), 309-332.

Gerstbrein, B., Stamatas, G., Kollias, N., & Driscoll, M. (2005). In vivo spectrofluorimetry reveals endogenous biomarkers that report healthspan and dietary restriction in Caenorhabditis elegans. *Aging Cell, 4*(3), 127-137. doi:10.1111/j.1474-9726.2005.00153.x

Golden J.W. & Riddle D.L. (1984). The Caenorhabditis elegans dauer larvae: developmental effects of pheromone, food and temperature. *Dev. Biol.*, 102, 368-378.

Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado, E., Chen, W., . . . Hopkins, N. (2002). Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat Genet, 31*(2), 135-140. doi:10.1038/ng896

Grant, B. D., & Sato, M. (2006). Intracellular trafficking. *WormBook*, 1-9. doi:10.1895/wormbook.1.77.1

Grallert, B., & Boye, E. (2008). The multiple facets of the intra-S checkpoint. *Cell Cycle*, 7(15), 2315-2320. doi:10.4161/cc.6389

Gullerova, M., & Proudfoot, N. J. (2010). Transcriptional interference and gene orientation in yeast: noncoding RNA connections. *Cold Spring Harb Symp Quant Biol, 75*, 299-311. doi:10.1101/sqb.2010.75.048

Hall, D. H., & Russell, R. L. (1991). The posterior nervous system of the nematode Caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions. *J Neurosci, 11*(1), 1-22.

Han, S. M., Lee, T. H., Mun, J. Y., Kim, M. J., Kritikou, E. A., Lee, S. J., . . . Koo, H. S. (2006). Deleted in cancer 1 (DICE1) is an essential protein controlling the topology of the inner mitochondrial membrane in C. elegans. *Development, 133*(18), 3597-3606. doi:10.1242/dev.02534

Hannan, M. A., & Nasim, A. (1977). Caffeine enhancement of radiation killing in different strains of Saccharomyces cerevisiae. *Mol Gen Genet*, 158(1), 111-116.

Harper, J. W., Elledge, S. J., Keyomarsi, K., Dynlacht, B., Tsai, L. H., Zhang, P., . . . et al. (1995). Inhibition of cyclin-dependent kinases by p21. *Mol Biol Cell, 6*(4), 387-400.

Hashimoto, Y., Ray Chaudhuri, A., Lopes, M., & Costanzo, V. (2010). Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat Struct Mol Biol, 17*(11), 1305-1311. doi:10.1038/nsmb.1927

Hata, T., & Nakayama, M. (2007). Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. *Biochim Biophys Acta, 1773*(7), 1039-1051. doi:10.1016/j.bbamcr.2007.04.010

Hedgecock, E. M., & White, J. G. (1985). Polyploid tissues in the nematode Caenorhabditis elegans. *Dev Biol, 107*(1), 128-133.

Hendzel, M. J., Wei, Y., Mancini, M. A., Van Hooser, A., Ranalli, T., Brinkley, B. R., . . . Allis, C. D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma, 106*(6), 348-360.

Hengartner, M. O., & Horvitz, H. R. (1994). C. elegans cell survival gene ced-9 encodes a functional homolog of the mammalian proto-oncogene bcl-2. *Cell, 76*(4), 665-676.

Hermeking, H., Lengauer, C., Polyak, K., He, T. C., Zhang, L., Thiagalingam, S., . . . Vogelstein, B. (1997). 14-3-3sigma is a p53-regulated inhibitor of G2/M progression. *Mol Cell, 1*(1), 3-11.

Hernandez, N. (1985). Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J, 4*(7), 1827-1837.

Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem, 276*(29), 26733-26736. doi:10.1074/jbc.R100032200

Hernandez, N., & Weiner, A. M. (1986). Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell, 47*(2), 249-258.

Hirao, A., Kong, Y. Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., . . . Mak, T. W. (2000). DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science, 287*(5459), 1824-1827.

Hirsh, D., Oppenheim, D., & Klass, M. (1976). Development of the reproductive system of Caenorhabditis elegans. *Dev Biol, 49*(1), 200-219.

Hitomi, M., Yang, K., Stacey, A. W., & Stacey, D. W. (2008). Phosphorylation of cyclin D1 regulated by ATM or ATR controls cell cycle progression. *Mol Cell Biol, 28*(17), 5478-5493. doi:10.1128/MCB.02047-07

Hobert, O. (2013). The neuronal genome of Caenorhabditis elegans. *WormBook*, 1-106. doi:10.1895/wormbook.1.161.1

Hochbaum, D., Ferguson, A. A., & Fisher, A. L. (2010). Generation of transgenic C. elegans by biolistic transformation. *J Vis Exp*(42). doi:10.3791/2090

Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature, 411*(6835), 366-374. doi:10.1038/35077232

Houseley, J., LaCava, J., & Tollervey, D. (2006). RNA-quality control by the exosome. *Nat Rev Mol Cell Biol, 7*(7), 529-539. doi:10.1038/nrm1964

Hsin, J. P., & Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev, 26*(19), 2119-2137. doi:10.1101/gad.200303.112

Hu, P. J. (2007). Dauer. *WormBook*, 1-19. doi:10.1895/wormbook.1.144.1

Huang, J., Gong, Z., Ghosal, G., & Chen, J. (2009). SOSS complexes participate in the maintenance of genomic stability. *Mol Cell, 35*(3), 384-393. doi:10.1016/j.molcel.2009.06.011

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., . . . Rinn, J. L. (2010). A large intergenic noncoding RNA induced by p53 mediates global

gene repression in the p53 response. *Cell*, 142(3), 409-419. doi:10.1016/j.cell.2010.06.040

Iliakis, G., Wang, Y., Guan, J., & Wang, H. (2003). DNA damage checkpoint control in cells exposed to ionizing radiation. *Oncogene, 22*(37), 5834-5847. doi:10.1038/sj.onc.1206682

Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol, 5*(11), a012740. doi:10.1101/cshperspect.a012740

Jawdekar, G. W., & Henry, R. W. (2008). Transcriptional regulation of human small nuclear RNA genes. *Biochim Biophys Acta, 1779*(5), 295-305. doi:10.1016/j.bbagrm.2008.04.001

Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G. C., Lukas, J., & Jackson, S. P. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nat Cell Biol, 8*(1), 37-45. doi:10.1038/ncb1337

Jodoin, J. N., Shboul, M., Albrecht, T. R., Lee, E., Wagner, E. J., Reversade, B., & Lee, L. A. (2013). The snRNA-processing complex, Integrator, is required for ciliogenesis and dynein recruitment to the nuclear envelope via distinct mechanisms. *Biol Open, 2*(12), 1390-1396. doi:10.1242/bio.20136981

Jodoin, J. N., Shboul, M., Sitaram, P., Zein-Sabatto, H., Reversade, B., Lee, E., & Lee, L. A. (2012). Human Asunder promotes dynein recruitment and centrosomal tethering to the nucleus at mitotic entry. *Mol Biol Cell, 23*(24), 4713-4724. doi:10.1091/mbc.E12-07-0558

Jodoin, J. N., Sitaram, P., Albrecht, T. R., May, S. B., Shboul, M., Lee, E., . . . Lee, L. A. (2013). Nuclear-localized Asunder regulates cytoplasmic dynein localization via its role in the integrator complex. *Mol Biol Cell, 24*(18), 2954-2965. doi:10.1091/mbc.E13-05-0254

Johnston, W. L., Krizus, A., & Dennis, J. W. (2006). The eggshell is required for meiotic fidelity, polar-body extrusion and polarization of the C. elegans embryo. *BMC Biol, 4*, 35. doi:10.1186/1741-7007-4-35

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol, 16*(3), 167-177. doi:10.1038/nrm3953

Kaldis, P. (1999). The cdk-activating kinase (CAK): from yeast to mammals. *Cell Mol Life Sci, 55*(2), 284-296. doi:10.1007/s000180050290

Kaletta, T., & Hengartner, M. O. (2006). Finding function in novel targets: C. elegans as a model organism. *Nat Rev Drug Discov, 5*(5), 387-398. doi:10.1038/nrd2031

Kalogeropoulos, N., Christoforou, C., Green, A. J., Gill, S., & Ashcroft, N. R. (2004). chk-1 is an essential gene and is required for an S-M checkpoint during early embryogenesis. *Cell Cycle, 3*(9), 1196-1200.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., . . . Ahringer, J. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature, 421*(6920), 231-237. doi:10.1038/nature01278

Kapp, L. D., Abrams, E. W., Marlow, F. L., & Mullins, M. C. (2013). The integrator complex subunit 6 (Ints6) confines the dorsal organizer in vertebrate embryogenesis. *PLoS Genet, 9*(10), e1003822. doi:10.1371/journal.pgen.1003822

Kar, A., Kaur, M., Ghosh, T., Khan, M. M., Sharma, A., Shekhar, R., . . . Saxena, S. (2015). RPA70 depletion induces hSSB1/2-INTS3 complex to initiate ATR signaling. *Nucleic Acids Res, 43*(10), 4962-4974. doi:10.1093/nar/gkv369

Kastan, M. B., & Bartek, J. (2004). Cell-cycle checkpoints and cancer. *Nature, 432*(7015), 316-323. doi:10.1038/nature03097

Kelleher, R. J., 3rd, Flanagan, P. M., & Kornberg, R. D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell, 61*(7), 1209-1215.

Kelly, K. O., Dernburg, A. F., Stanfield, G. M., & Villeneuve, A. M. (2000). Caenorhabditis elegans msh-5 is required for both normal and radiation-induced meiotic crossing over but not for completion of meiosis. *Genetics, 156*(2), 617-630.

Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol, 17*(4), 227-239. doi:10.1038/nrm.2015.15

Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H., & Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell, 77*(4), 599-608.

Klass, M. R. (1977). Aging in the nematode Caenorhabditis elegans: major biological and environmental factors influencing life span. *Mech Ageing Dev, 6*(6), 413-429.

Kohalmi, S. E., Roche, H. M., & Kunz, B. A. (1993). Elevated intracellular dCTP levels reduce the induction of GC-->AT transitions in yeast by ethyl methanesulfonate or N-methyl-N'-nitro-N- nitrosoguanidine but increase alkylation-induced GC-->CG transversions. *Mutagenesis, 8*(5), 457-465.

Labib, K., & De Piccoli, G. (2011). Surviving chromosome replication: the many roles of the S-phase checkpoint pathway. *Philos Trans R Soc Lond B Biol Sci*, 366(1584), 3554-3561. doi:10.1098/rstb.2011.0071

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680-685.

Lai, F., Gardini, A., Zhang, A., & Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature, 525*(7569), 399-403. doi:10.1038/nature14906

Laine, J. P., & Egly, J. M. (2006). When transcription and repair meet: a complex system. *Trends Genet, 22*(8), 430-436. doi:10.1016/j.tig.2006.06.006

Lakin, N. D., & Jackson, S. P. (1999). Regulation of p53 in response to DNA damage. *Oncogene, 18*(53), 7644-7655. doi:10.1038/sj.onc.1203015

Larrea, A. A., Lujan, S. A., & Kunkel, T. A. (2010). SnapShot: DNA mismatch repair. *Cell, 141*(4), 730 e731. doi:10.1016/j.cell.2010.05.002

Lasda, E. L., & Blumenthal, T. (2011). Trans-splicing. *Wiley Interdiscip Rev RNA, 2*(3), 417-434. doi:10.1002/wrna.71

Lavin, M. F. (2007). ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene, 26*(56), 7749-7758. doi:10.1038/sj.onc.1210880

Lee, T. H., Mun, J. Y., Han, S. M., Yoon, G., Han, S. S., & Koo, H. S. (2009). DIC-1 over-expression enhances respiratory activity in Caenorhabditis elegans by promoting mitochondrial cristae formation. *Genes Cells, 14*(3), 319-327. doi:10.1111/j.1365-2443.2008.01276.x

Lemmens, B. B., & Tijsterman, M. (2011). DNA double-strand break repair in Caenorhabditis elegans. *Chromosoma, 120*(1), 1-21. doi:10.1007/s00412-010-0296-3

Li, C., & Kim, K. (2010). Neuropeptide gene families in Caenorhabditis elegans. *Adv Exp Med Biol*, 692, 98-137.

Li, W., Notani, D., & Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet, 17*(4), 207-223. doi:10.1038/nrg.2016.4

Li, X., & Heyer, W. D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res, 18*(1), 99-113. doi:10.1038/cr.2008.1

Li, Y., Bolderson, E., Kumar, R., Muniandy, P. A., Xue, Y., Richard, D. J., . . . Wang, W. (2009). HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem, 284*(35), 23525-23531. doi:10.1074/jbc.C109.039586

Li, Z., Yu, A., & Weiner, A. M. (1998). Adenovirus type 12-induced fragility of the human RNU2 locus requires p53 function. *J Virol, 72*(5), 4183-4191.

Lindgren, V., Ares, M., Jr., Weiner, A. M., & Francke, U. (1985). Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature, 314*(6006), 115-116.

Liu, F., Stanton, J. J., Wu, Z., & Piwnica-Worms, H. (1997). The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Mol Cell Biol, 17*(2), 571-583.

Liu, H., Takeda, S., Kumar, R., Westergard, T. D., Brown, E. J., Pandita, T. K., . . . Hsieh, J. J. (2010). Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature, 467*(7313), 343-346. doi:10.1038/nature09350

Liu, P., Barkley, L. R., Day, T., Bi, X., Slater, D. M., Alexandrow, M. G., . . . Vaziri, C. (2006). The Chk1-mediated S-phase checkpoint targets initiation factor Cdc45 via a Cdc25A/Cdk2-independent mechanism. *J Biol Chem, 281*(41), 30631-30644. doi:10.1074/jbc.M602982200

Lobrich, M., & Jeggo, P. A. (2007). The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat Rev Cancer, 7*(11), 861-869. doi:10.1038/nrc2248

Lodish, H. F. (2013). *Molecular cell biology*. New York: W.H. Freeman and Co.

Logan, J., Falck-Pedersen, E., Darnell, J. E., Jr., & Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A, 84*(23), 8306-8310.

Loya, T. J., & Reines, D. (2016). Recent advances in understanding transcription termination by RNA polymerase II. *F1000Res*, 5. doi:10.12688/f1000research.8455.1

Luo, Z., Lin, C., & Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol, 13*(9), 543-547. doi:10.1038/nrm3417

Maciejewski, P. M., Peterson, F. C., Anderson, P. J., & Brooks, C. L. (1995). Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *J Biol Chem, 270*(46), 27661-27665.

MacQueen, A. J., & Villeneuve, A. M. (2001). Nuclear reorganization and homologous chromosome pairing during meiotic prophase require C. elegans chk-2. *Genes Dev, 15*(13), 1674-1687. doi:10.1101/gad.902601

Malovannaya, A., Li, Y., Bulynko, Y., Jung, S. Y., Wang, Y., Lanz, R. B., . . . Qin, J. (2010). Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci U S A, 107*(6), 2431-2436. doi:10.1073/pnas.0912599106

Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer, 9*(3), 153-166. doi:10.1038/nrc2602

Mandel, C. R., Bai, Y., & Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci, 65*(7-8), 1099-1122. doi:10.1007/s00018-007-7474-3

Mango, S. E. (2007). The C. elegans pharynx: a model for organogenesis. *WormBook*, 1-26. doi:10.1895/wormbook.1.129.1

Manke, I. A., Nguyen, A., Lim, D., Stewart, M. Q., Elia, A. E., & Yaffe, M. B. (2005). MAPKAP kinase-2 is a cell cycle checkpoint kinase that regulates the G2/M transition and S phase progression in response to UV irradiation. *Mol Cell, 17*(1), 37-48. doi:10.1016/j.molcel.2004.11.021

Marechal, A., & Zou, L. (2013). DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol, 5*(9). doi:10.1101/cshperspect.a012716

Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet, 9*(11), 843-854. doi:10.1038/nrg2438

Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol, 8*(3), 209-220. doi:10.1038/nrm2124

Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nat Rev Mol Cell Biol, 15*(2), 108-121. doi:10.1038/nrm3742

Matera, A. G., Weiner, A. M., & Schmid, C. W. (1990). Structure and evolution of the U2 small nuclear RNA multigene family in primates: gene amplification under natural selection? *Mol Cell Biol, 10*(11), 5876-5882.

Mathis, D. J., & Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. *Nature, 290*(5804), 310-315.

Maya, R., Balass, M., Kim, S. T., Shkedy, D., Leal, J. F., Shifman, O., . . . Oren, M. (2001). ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage. *Genes Dev, 15*(9), 1067-1077. doi:10.1101/gad.886901

Medlin, J., Scurry, A., Taylor, A., Zhang, F., Peterlin, B. M., & Murphy, S. (2005). P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J, 24*(23), 4154-4165. doi:10.1038/sj.emboj.7600876

Medlin, J. E., Uguen, P., Taylor, A., Bentley, D. L., & Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J, 22*(4), 925-934. doi:10.1093/emboj/cdg077

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet, 10*(3), 155-159. doi:10.1038/nrg2521

Mladenov, E., Magin, S., Soni, A., & Iliakis, G. (2016). DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: Cell cycle and proliferation-dependent regulation. *Semin Cancer Biol, 37-38*, 51-64. doi:10.1016/j.semcancer.2016.03.003

Moerman, D. G., & Fire, A. (1997). Muscle: Structure, Function, and Development. In D. L. Riddle, T. Blumenthal, B. J. Meyer, & J. R. Priess (Eds.), *C. elegans II* (2nd ed.). Cold Spring Harbor (NY).

Montecucco, A., & Biamonti, G. (2013). Pre-mRNA processing factors meet the DNA damage response. *Front Genet, 4*, 102. doi:10.3389/fgene.2013.00102

Morgan, D. O. (1995). Principles of CDK regulation. *Nature, 374*(6518), 131-134. doi:10.1038/374131a0

Mosley, A. L., Pattenden, S. G., Carey, M., Venkatesh, S., Gilmore, J. M., Florens, L., . . . Washburn, M. P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell*, 34(2), 168-178. doi:10.1016/j.molcel.2009.02.025

Mueller, P. R., Coleman, T. R., Kumagai, A., & Dunphy, W. G. (1995). Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. *Science, 270*(5233), 86-90.

Narita, T., Yung, T. M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., . . . Handa, H. (2007). NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol Cell, 26*(3), 349-365. doi:10.1016/j.molcel.2007.04.011

Nelson, F. K., Albert, P. S., & Riddle, D. L. (1983). Fine structure of the Caenorhabditis elegans secretory-excretory system. *J Ultrastruct Res, 82*(2), 156-171.

Ni, Z., Xu, C., Guo, X., Hunter, G. O., Kuznetsova, O. V., Tempel, W., . . . Greenblatt, J. F. (2014). RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol, 21*(8), 686-695. doi:10.1038/nsmb.2853

Nigon, V., & Dougherty, E. C. (1949). Reproductive patterns and attempts at reciprocal crossing of Rhabditis elegans Maupas, 1900, and Rhabditis briggsae Dougherty and Nigon, 1949 (Nematoda: Rhabditidae). *J Exp Zool*, 112(3), 485-503.

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature, 463*(7280), 457-463. doi:10.1038/nature08909

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., & Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell, 167*(4), 1001-1013 e1007. doi:10.1016/j.cell.2016.10.001

Oikonomou, G., & Shaham, S. (2011). The glia of Caenorhabditis elegans. *Glia, 59*(9), 1253-1263. doi:10.1002/glia.21084

Otani, Y., Nakatsu, Y., Sakoda, H., Fukushima, T., Fujishiro, M., Kushiyama, A., . . . Asano, T. (2013). Integrator complex plays an essential role in adipose differentiation. *Biochem Biophys Res Commun, 434*(2), 197-202. doi:10.1016/j.bbrc.2013.03.029

Page, A. P., & Johnstone, I. L. (2007). The cuticle. *WormBook*, 1-15. doi:10.1895/wormbook.1.138.1

Painter, R. B., & Young, B. R. (1980). Radiosensitivity in ataxia-telangiectasia: a new explanation. *Proc Natl Acad Sci U S A, 77*(12), 7315-7317.

Park, M., & Krause, M. W. (1999). Regulation of postembryonic G(1) cell cycle progression in Caenorhabditis elegans by a cyclin D/CDK-like complex. *Development, 126*(21), 4849-4860.

Parker, L. L., Atherton-Fessler, S., & Piwnica-Worms, H. (1992). p107wee1 is a dual-specificity kinase that phosphorylates p34cdc2 on tyrosine 15. *Proc Natl Acad Sci U S A, 89*(7), 2917-2921.

Parker, L. L., & Piwnica-Worms, H. (1992). Inactivation of the p34cdc2-cyclin B complex by the human WEE1 tyrosine kinase. *Science, 257*(5078), 1955-1957.

Peart, N., Sataluri, A., Baillat, D., & Wagner, E. J. (2013). Non-mRNA 3' end formation: how the other half lives. *Wiley Interdiscip Rev RNA, 4*(5), 491-506. doi:10.1002/wrna.1174

Piazzi, M., Williamson, A., Lee, C. F., Pearson, S., Lacaud, G., Kouskoff, V., . . . Whetton, A. D. (2015). Quantitative phosphoproteome analysis of embryonic stem cell differentiation toward blood. *Oncotarget, 6*(13), 10924-10939. doi:10.18632/oncotarget.3454

Plet, A., Eick, D., & Blanchard, J. M. (1995). Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene, 10*(2), 319-328.

Podbilewicz, B. (2006). Cell fusion. *WormBook*, 1-32. doi:10.1895/wormbook.1.52.1

Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol, 16*(3), 190-202. doi:10.1038/nrm3943

Powell, S. N., DeFrank, J. S., Connell, P., Eogan, M., Preffer, F., Dombkowski, D., . . . Friend, S. (1995). Differential sensitivity of p53(-) and p53(+) cells to caffeine-induced radiosensitization and override of G2 delay. *Cancer Res, 55*(8), 1643-1648.

Proudfoot, N. J., & Brownlee, G. G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature, 263*(5574), 211-214.

Raizen, D. M., Zimmerman, J. E., Maycock, M. H., Ta, U. D., You, Y. J., Sundaram, M. V., & Pack, A. I. (2008). Lethargus is a Caenorhabditis elegans sleep-like state. *Nature, 451*(7178), 569-572. doi:10.1038/nature06535

Ramanathan, A., Robb, G. B., & Chan, S. H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res, 44*(16), 7511-7526. doi:10.1093/nar/gkw551

Reinke, V., Krause, M., & Okkema, P. (2013). Transcriptional regulation of gene expression in C. elegans. *WormBook*, 1-34. doi:10.1895/wormbook.1.45.2

Richard, D. J., Bolderson, E., Cubeddu, L., Wadsworth, R. I., Savage, K., Sharma, G. G., . . . Khanna, K. K. (2008). Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature, 453*(7195), 677-681. doi:10.1038/nature06883

Richard, D. J., Savage, K., Bolderson, E., Cubeddu, L., So, S., Ghita, M., . . . Khanna, K. K. (2011). hSSB1 rapidly binds at the sites of DNA double-strand breaks and is required for the efficient recruitment of the MRN complex. *Nucleic Acids Res, 39*(5), 1692-1702. doi:10.1093/nar/gkq1098

Riddle, D. L., Blumenthal, T., Meyer, B. J., & Priess, J. R. (1997). Introduction to C. elegans. In D. L. Riddle, T. Blumenthal, B. J. Meyer, & J. R. Priess (Eds.), *C. elegans II* (2nd ed.). Cold Spring Harbor (NY).

Rieder, C. L., Cole, R. W., Khodjakov, A., & Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J Cell Biol, 130*(4), 941-948.

Rieder, C. L., Schultz, A., Cole, R., & Sluder, G. (1994). Anaphase onset in vertebrate somatic cells is controlled by a checkpoint that monitors sister kinetochore attachment to the spindle. *J Cell Biol, 127*(5), 1301-1310.

Ritter, S. L., & Hall, R. A. (2009). Fine-tuning of GPCR activity by receptor-interacting proteins. *Nat Rev Mol Cell Biol, 10*(12), 819-830. doi:10.1038/nrm2803

Ropke, A., Buhtz, P., Bohm, M., Seger, J., Wieland, I., Allhoff, E. P., & Wieacker, P. F. (2005). Promoter CpG hypermethylation and downregulation of DICE1 expression in prostate cancer. *Oncogene, 24*(44), 6667-6675. doi:10.1038/sj.onc.1208824

Rual, J. F., Ceron, J., Koreth, J., Hao, T., Nicot, A. S., Hirozane-Kishikawa, T., . . . Vidal, M. (2004). Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res, 14*(10B), 2162-2168. doi:10.1101/gr.2505604

Rutkowski, R. J., & Warren, W. D. (2009). Phenotypic analysis of deflated/Ints7 function in Drosophila development. *Dev Dyn, 238*(5), 1131-1139. doi:10.1002/dvdy.21922

Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol, 16*(3), 129-143. doi:10.1038/nrm3952

Saldi, T., Cortazar, M. A., Sheridan, R. M., & Bentley, D. L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol, 428*(12), 2623-2635. doi:10.1016/j.jmb.2016.04.017

Salinas, L. S., Maldonado, E., & Navarro, R. E. (2006). Stress-induced germ cell apoptosis by a p53 independent pathway in Caenorhabditis elegans. *Cell Death Differ, 13*(12), 2129-2139. doi:10.1038/sj.cdd.4401976

Sanchez, R., & Zhou, M. M. (2011). The PHD finger: a versatile epigenome reader. *Trends Biochem Sci, 36*(7), 364-372. doi:10.1016/j.tibs.2011.03.005

Scharer, O. D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol, 5*(10), a012609. doi:10.1101/cshperspect.a012609

Schatz, D. G., & Swanson, P. C. (2011). V(D)J recombination: mechanisms of initiation. *Annu Rev Genet, 45*, 167-202. doi:10.1146/annurev-genet-110410-132552

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., & Jasin, M. (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell, 145*(4), 529-542. doi:10.1016/j.cell.2011.03.041

Schmidt, K., & Butler, J. S. (2013). Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. *Wiley Interdiscip Rev RNA, 4*(2), 217-231. doi:10.1002/wrna.1155

Schumacher, B., Hofmann, K., Boulton, S., & Gartner, A. (2001). The C. elegans homolog of the p53 tumor suppressor is required for DNA damage-induced apoptosis. *Curr Biol, 11*(21), 1722-1727.

Schwertman, P., Bekker-Jensen, S., & Mailand, N. (2016). Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol, 17*(6), 379-394. doi:10.1038/nrm.2016.58

Shandilya, J., & Roberts, S. G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta, 1819*(5), 391-400. doi:10.1016/j.bbagrm.2012.01.010

Shaye, D. D., & Greenwald, I. (2011). OrthoList: a compendium of C. elegans genes with human orthologs. *PLoS One, 6*(5), e20085. doi:10.1371/journal.pone.0020085

Shiotani, B., & Zou, L. (2009). Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Mol Cell, 33*(5), 547-558. doi:10.1016/j.molcel.2009.01.024

Sitaram, P., Merkle, J. A., Lee, E., & Lee, L. A. (2014). asunder is required for dynein localization and dorsal fate determination during Drosophila oogenesis. *Dev Biol, 386*(1), 42-52. doi:10.1016/j.ydbio.2013.12.004

Skaar, J. R., Ferris, A. L., Wu, X., Saraf, A., Khanna, K. K., Florens, L., . . . Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res, 25*(3), 288-305. doi:10.1038/cr.2015.19

Skaar, J. R., Richard, D. J., Saraf, A., Toschi, A., Bolderson, E., Florens, L., . . . Pagano, M. (2009). INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol, 187*(1), 25-32. doi:10.1083/jcb.200907026

Smith, H. E., Fabritius, A. S., Jaramillo-Lambert, A., & Golden, A. (2016). Mapping Challenging Mutations by Whole-Genome Sequencing. *G3 (Bethesda), 6*(5), 1297-1304. doi:10.1534/g3.116.028316

Sperka, T., Wang, J., & Rudolph, K. L. (2012). DNA damage checkpoints in stem cells, ageing and cancer. *Nat Rev Mol Cell Biol, 13*(9), 579-590. doi:10.1038/nrm3420

Spieth, J., & Lawson, D. (2006). Overview of gene structure. *WormBook*, 1-10. doi:10.1895/wormbook.1.65.1

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., . . . Benkirane, M. (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun, 5*, 5531. doi:10.1038/ncomms6531

Steinmetz, E. J., & Brow, D. A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol, 16*(12), 6993-7003.

Steinmetz, E. J., Conrad, N. K., Brow, D. A., & Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature, 413*(6853), 327-331. doi:10.1038/35095090

Stergiou, L., Doukoumetzidis, K., Sendoel, A., & Hengartner, M. O. (2007). The nucleotide excision repair pathway is required for UV-C-induced apoptosis in Caenorhabditis elegans. *Cell Death Differ, 14*(6), 1129-1138. doi:10.1038/sj.cdd.4402115

Stricklin, S. L., Griffiths-Jones, S., & Eddy, S. R. (2005). C. elegans noncoding RNA genes. *WormBook*, 1-7. doi:10.1895/wormbook.1.1.1

Stutz, F., Liao, X. C., & Rosbash, M. (1993). U1 small nuclear ribonucleoprotein particle-protein interactions are revealed in Saccharomyces cerevisiae by in vivo competition assays. *Mol Cell Biol, 13*(4), 2126-2133.

Sulston, J. E. (1983). Neuronal cell lineages in the nematode Caenorhabditis elegans. *Cold Spring Harb Symp Quant Biol, 48 Pt 2*, 443-452.

Sulston, J. E., & Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol, 56*(1), 110-156.

Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol, 100*(1), 64-119.

Syntichaki, P., Xu, K., Driscoll, M., & Tavernarakis, N. (2002). Specific aspartyl and calpain proteases are required for neurodegeneration in C. elegans. *Nature, 419*(6910), 939-944. doi:10.1038/nature01108

Ta, H. Q., & Gioeli, D. (2014). The convergence of DNA damage checkpoint pathways and androgen receptor signaling in prostate cancer. *Endocr Relat Cancer, 21*(5), R395-407. doi:10.1530/ERC-14-0217

Tabara, H., Grishok, A., & Mello, C. C. (1998). RNAi in C. elegans: soaking in the genome sequence. *Science, 282*(5388), 430-431.

Taffoni, C., & Pujol, N. (2015). Mechanisms of innate immunity in C. elegans epidermis. *Tissue Barriers, 3*(4), e1078432. doi:10.1080/21688370.2015.1078432

Takata, H., Nishijima, H., Maeshima, K., & Shibahara, K. (2012). The integrator complex is required for integrity of Cajal bodies. *J Cell Sci, 125*(Pt 1), 166-175. doi:10.1242/jcs.090837

Tang, L., Machacek, T., Mamnun, Y. M., Penkner, A., Gloggnitzer, J., Wegrostek, C., . . . Jantsch, V. (2010). Mutations in Caenorhabditis elegans him-19 show meiotic defects that worsen with age. *Mol Biol Cell, 21*(6), 885-896. doi:10.1091/mbc.E09-09-0811

Tao, S., Cai, Y., & Sampath, K. (2009). The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. *Development, 136*(16), 2757-2765. doi:10.1242/dev.034959

The *C. elegans* Sequencing Consortium (1998) *Science* 282, 2012–2018

Thomas, J., Lea, K., Zucker-Aprison, E., & Blumenthal, T. (1990). The spliceosomal snRNAs of Caenorhabditis elegans. *Nucleic Acids Res, 18*(9), 2633-2642.

Thomas, J. D., Conrad, R. C., & Blumenthal, T. (1988). The C. elegans trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. *Cell, 54*(4), 533-539.

Timmons, L., & Fire, A. (1998). Specific interference by ingested dsRNA. *Nature, 395*(6705), 854. doi:10.1038/27579

Truebestein, L., & Leonard, T. A. (2016). Coiled-coils: The long and short of it. *Bioessays, 38*(9), 903-916. doi:10.1002/bies.201600062

Van Arsdell, S. W., & Weiner, A. M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. *Mol Cell Biol, 4*(3), 492-499.

van den Berg, D. L., Azzarelli, R., Oishi, K., Martynoga, B., Urban, N., Dekkers, D. H., . . . Guillemot, F. (2017). Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron, 93*(2), 348-361. doi:10.1016/j.neuron.2016.11.047

van den Heuvel, S. (2005). The C. elegans cell cycle: overview of molecules and mechanisms. *Methods Mol Biol, 296*, 51-67.

Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol, 16*(3), 178-189. doi:10.1038/nrm3941

Wanandi, I., Waldschmidt, R., & Seifart, K. H. (1993). Mammalian transcription factor PBP. Characterization of its binding properties to the proximal sequence element of U6 genes. *J Biol Chem, 268*(9), 6629-6640.

Ward, S., Thomson, N., White, J. G., & Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode Caenorhabditis elegans.?2UU. *J Comp Neuro*l, 160(3), 313-337. doi:10.1002/cne.901600305

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., . . . Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. *Cell, 149*(1), 101-112. doi:10.1016/j.cell.2012.03.002

Weinert, T. A., & Hartwell, L. H. (1988). The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. *Science, 241*(4863), 317-322.

West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature, 432*(7016), 522-525. doi:10.1038/nature03035

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci, 314*(1165), 1-340.

Wicks, S. R., Yeh, R. T., Gish, W. R., Waterston, R. H., & Plasterk, R. H. (2001). Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. *Nat Genet, 28*(2), 160-164. doi:10.1038/88878

Wieland, I., Arden, K. C., Michels, D., Klein-Hitpass, L., Bohm, M., Viars, C. S., & Weidle, U. H. (1999). Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene, 18*(32), 4530-4537. doi:10.1038/sj.onc.1202806

Wieland, I., Ropke, A., Stumm, M., Sell, C., Weidle, U. H., & Wieacker, P. F. (2001). Molecular characterization of the DICE1 (DDX26) tumor suppressor gene in lung carcinoma cells. *Oncol Res, 12*(11-12), 491-500.

Wieland, I., Sell, C., Weidle, U. H., & Wieacker, P. (2004). Ectopic expression of DICE1 suppresses tumor cell growth. *Oncol Rep, 12*(2), 207-211.

Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol, 3*(7). doi:10.1101/cshperspect.a003707

Wood, W. B. (1988). The nematode Caenorhabditis elegans. New York, NY: *Cold Spring Harbor Laboratory Press*.

Xiang, K., Tong, L., & Manley, J. L. (2014). Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Mol Cell Biol, 34*(11), 1894-1910. doi:10.1128/MCB.00084-14

Xie, M., Zhang, W., Shu, M. D., Xu, A., Lenis, D. A., DiMaio, D., & Steitz, J. A. (2015). The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends. *Genes Dev, 29*(14), 1552-1564. doi:10.1101/gad.266973.115

Xu, S. (2015). The application of CRISPR-Cas9 genome editing in Caenorhabditis elegans. *J Genet Genomics, 42*(8), 413-421. doi:10.1016/j.jgg.2015.06.005

Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., & Yamaguchi, Y. (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat Commun, 5*, 4263. doi:10.1038/ncomms5263

Yan, S., & Michael, W. M. (2009). TopBP1 and DNA polymerase alpha-mediated recruitment of the 9-1-1 complex to stalled replication forks: implications for a replication restart-based mechanism for ATR checkpoint activation. *Cell Cycle, 8*(18), 2877-2884. doi:10.4161/cc.8.18.9485

Yang, S. H., Zhou, R., Campbell, J., Chen, J., Ha, T., & Paull, T. T. (2013). The SOSS1 single-stranded DNA binding complex promotes DNA end resection in concert with Exo1. *EMBO J, 32*(1), 126-139. doi:10.1038/emboj.2012.314

Yoon, J. B., Murphy, S., Bai, L., Wang, Z., & Roeder, R. G. (1995). Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. *Mol Cell Biol, 15*(4), 2019-2027.

Yu, A., Bailey, A. D., & Weiner, A. M. (1998). Metaphase fragility of the human RNU1 and RNU2 loci is induced by actinomycin D through a p53-dependent pathway. *Hum Mol Genet, 7*(4), 609-617.

Yu, A., Fan, H. Y., Liao, D., Bailey, A. D., & Weiner, A. M. (2000). Activation of p53 or loss of the Cockayne syndrome group B repair protein causes metaphase fragility of human U1, U2, and 5S genes. *Mol Cell, 5*(5), 801-810.

Yudkovsky, N., Ranish, J. A., & Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature, 408*(6809), 225-229. doi:10.1038/35041603

Yuo, C. Y., Ares, M., Jr., & Weiner, A. M. (1985). Sequences required for 3' end formation of human U2 small nuclear RNA. *Cell, 42*(1), 193-202.

Zaborowska, J., Egloff, S., & Murphy, S. (2016). The pol II CTD: new twists in the tail. *Nat Struct Mol Biol, 23*(9), 771-777. doi:10.1038/nsmb.3285

Zhang, F., Ma, T., & Yu, X. (2013). A core hSSB1-INTS complex participates in the DNA damage response. *J Cell Sci, 126*(Pt 21), 4850-4855. doi:10.1242/jcs.132514

Zhang, F., Wu, J., & Yu, X. (2009). Integrator3, a partner of single-stranded DNA-binding protein 1, participates in the DNA damage response. *J Biol Chem, 284*(44), 30408-30415. doi:10.1074/jbc.M109.039404

Zhang, Y., & Xiong, Y. (2001). A p53 amino-terminal nuclear export signal inhibited by DNA damage-induced phosphorylation. *Science, 292*(5523), 1910-1915. doi:10.1126/science.1058637

Zhang, Z., & Gilmour, D. S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell, 21*(1), 65-74. doi:10.1016/j.molcel.2005.11.002

Zhao, H., Watkins, J. L., & Piwnica-Worms, H. (2002). Disruption of the checkpoint kinase 1/cell division cycle 25A pathway abrogates ionizing radiation-induced S and G2 checkpoints. *Proc Natl Acad Sci U S A, 99*(23), 14795-14800. doi:10.1073/pnas.182557299

Zhao, S., Weng, Y. C., Yuan, S. S., Lin, Y. T., Hsu, H. C., Lin, S. C., . . . Lee, E. Y. (2000). Functional link between ataxia-telangiectasia and Nijmegen breakage syndrome gene products. *Nature, 405*(6785), 473-477. doi:10.1038/35013083

Zinder, J. C., & Lima, C. D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev, 31*(2), 88-100. doi:10.1101/gad.294769.116

Zorio, D. A., Cheng, N. N., Blumenthal, T., & Spieth, J. (1994). Operons as a common form of chromosomal organization in C. elegans. *Nature*, 372(6503), 270-272. doi:10.1038/372270a0

Zou, L., & Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science, 300*(5625), 1542-1548. doi:10.1126/science.1083430

Zuryn, S., & Jarriault, S. (2013). Deep sequencing strategies for mapping and identifying mutations from genetic screens. *Worm, 2*(3), e25081. doi:10.4161/worm.25081

Agradecimientos

Esta es una historia que comenzó hace ya mucho tiempo... cuando me decidí, o más bien, me empeñé... en estudiar Biología, pues eso... era lo que gustaba, y porque... cómo no! cabezota soy un rato! Eso de que los seres vivos nacen, crecen, se desarrollan, se reproducen y mueren que aprendí en la escuela, siempre había dejado un regustillo amargo en mi interior deseando saber el cómo.

Muchas son las personas que ahora pasan por mi memoria...y que, gracias a ellas, han modelado a "la Beatriz" del hoy.

En primer lugar quiero agradecerle a Juan, mi director de tesis, y al CIBIR, haberme dado la oportunidad de realizar la tesis doctoral "en casa". Juan, gracias por aceptarme en el lab e introducirme en el mundo de la ciencia, por traer a *C. elegans* a La Rioja, por tus explicaciones y conocimientos de microscopía, por llevarme a un montón de congresos sin dudarlo, por tus consejos (tanto personales como profesionales) y por las curiosidades que nos cuentas, que siempre son de agradecer. Gracias por estos años, los recordaré con cariño.

Eva, he tenido mucha suerte de coincidir contigo, gracias por haberme enseñado y explicado las técnicas del laboratorio, las PCRs, las digestiones, las ligaciones, las minis, los screenings de colonias... qué satisfacción la de clonar! ;) Gracias por transmitirme tus conocimientos, entre otras cosas, cómo funciona el RNAi!!, por tu dedicación, por tu calma, por tus consejos, por tus conversaciones de ciencia, por estar tanto para una sonrisa como para una lágrima. Gracias. Sabes que esto también es tuyo :).

Bego, gracias por las veces que te he traído loca con pedidos que no llegan, con paquetes a enviar, con albaranes, con el servicio de mensajería... gracias por enseñarme algún que otro pequeño "truqui de lab" y porque ante la tempestad tienes la capacidad de sacar la calma y eso, es de agradecer.

Angelina, gracias por darle al lab un toque más internacional, por los cafés, por las risas, por patearte Berlín conmigo, los museos, los mercados, el muro... porque después de estar cuatro horas escuchando alemán... ahora ya sabemos que es eso de *sitzen bleiben*! me lo pasé muy bien :).

También quiero agradecer al resto de compañer@s de la segunda planta del CIBIR, tanto los que continúan como los que se fueron, que a lo largo de estos años me han mostrado su apoyo y cariño, especialmente a Jose G.P, Raquel, Rossette, Iziar, Laura, Josune y al nuevo aire fresco del CIBIR Alvarito, Rafa, María, Sergio y Chus.

Mas allá de las fronteras de las Rioja, hay muchas personas que me han apoyado en estos años y a las que les estoy muy agradecida.

Voy a empezar por María Sacristán, gracias María, porque me acogiste en el lab sin saber absolutamente nada del cultivo celular y además, como una forastera en el mundo de las fostatasas. Gracias por ser mi tutura. Gracias por enseñarme a montar figuras, a diseñar experimentos, gracias porque cada día te sentabas conmigo a discutir los resultados y a pensar posibles soluciones y, sobre todo, gracias por tu pasión por la ciencia.

Sara!! Gracias por hacer amenas esas infinitas horas en el lab! por los WB, las IPs, las dichosas "huevo frito" que no me dejaron vivir!! las risas, las charlas, tu disposición y cómo no? Gracias por presentarme a Pili!!
Charles, entre otras muchas cosas, gracias por el Graphpad y las explicaciones de los 18S, fue un placer conocerte jienense!
David, *Muito prazer em te conhecer*!
Pi, pero qué bonita es Salamanca!! gracias por entenderme y por tus consejos. Desde el principio supe que eres de las personas con las que se puede contar y que no te fallan, y así ha sido. No cambies!
Vane, mi galleguiña favorita! gracias por estar ahí, por tus conversaciones y cómo no, por tu hospitalidad. Junto con Sara y Pi, espero que nos queden por delante muchísimos más viajes por el mundo y anécdotas que contar.

Adela, gracias por tu apoyo y tus consejos. Son ya diez años... gracias porque a pesar del tiempo y la distancia, se que estás al otro lado del teléfono. Siiiii, por fin! ya he terminado la tesis!! :)

Diana, lo mismo digo, me alegra ver que pasa el tiempo y no perdemos el contacto. Recuerdo los primeros años de uni con mucho cariño.

Y bajando un poco más al sur... me gustaría destacar a Antonio Miranda, gracias Antonio, por introducirme en el mundo de la glutatión reductasa, por tus e-mails detallados y explicativos, que siempre te ponen en contexto. Gracias por recomendarme el lab de Hilde, fue una experiencia inolvidable y, mil gracias por echarme una mano sin pedírtelo.

Now moving to the North, beyond Spain borders... I would like to thank Hilde for accepting me in her lab and for her scientific support. Of course I would like to thank all people in Hilde´s lab because they were really nice and warm.

Henok, thank you very much for teaching me how to do an excellent immunostainig! Of course, thanks for your dedication and your support. It was really good working with you. Also, thanks for the non-scientific talks. Now, I can see very clearly a behavioural gradient from the South to the North :).
Sergio, fue un lujazo tener a un andaluz tan majete por tierras no españolas, espero que te vaya muy bien! *My girls!!!! Lisa and Penelope, thank you very much for your friendship, I feel I know you both for years. Thanks for the talks, coffees, laughs, walks, films and dinners!! And again... for the laughs!! Ouuuuu I miss you both!*
Kroustalaki, thanks for the ice-skating! The words in Greek and the English misunderstandings! I am waiting to try the traditional Greek Moussaka!
Lisa, thanks for living next door! I was very fortunate to have you near, thanks for the trip to Tromso, the whales and Northern lights!
Para acabar el recorrido, vuelvo a mi tierra, a mis amigas del cole, Eva y Bea, gracias por enterderme y estar ahí.

Gracias a mi familia en general. Especialmente a mis padres, Jose Luis y Rosa Mari, gracias por vuestra bondad, por la educación que me habéis dado y por la familia que habéis creado. Gracias papá, porque siempre estás dispuesto ayudarme. Gracias mamá, porque siempre estas pendiente de mí y me quieres más que nadie.

A mi hermano, Miguel Ángel, creo que... quizá tenga que darte las gracias por ser tan "chinche"... y competitivo conmigo, pues es verdad que eso ha ayudado al desarrollo de mi paciencia y afán de superación. Tato, gracias por estar ahí y por hacerme desconectar.

Gracias a los tres tanto por animarme como por aguantar mis agobios.

Y como no podía ser de otra forma...
me despido dando las gracias a
mis dos compañeras de viaje,
la música y la danza.

Gracias