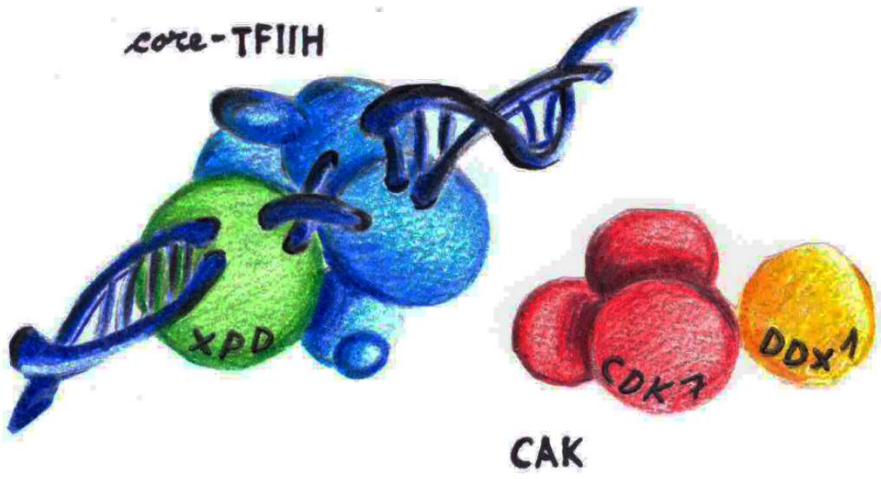
UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie

"L. SPALLANZANI"

Identification of DDX1 as a novel interactor of CDK-Activating Kinase and its implication in Trichothiodystrophy disorder



Debora Ferri

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXII– A.A. 2016-2019

UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie

"L. SPALLANZANI"

Identification of DDX1 as a novel interactor of CDK-Activating Kinase and its implication in Trichothiodystrophy disorder

Debora Ferri

Supervised by Prof. Donata Orioli

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXII – A.A. 2016-2019

Contents

Abstract

The transcription/DNA repair factor IIH (TFIIH) is made of two distinct sub-complexes, the core-TFIIH and the CDK-activating kinase (CAK), bridged together by the XPD subunit. The CDK7 kinase subunit of the CAK sub-complex targets different substrates according to the protein complex it belongs to: as part of the free-CAK complex it phosphorylates specific cyclin-dependent kinases (CDKs) and promotes cell cycle progression; as subunit of the entire TFIIH (holo-TFIIH), CDK7 phosphorylates and activates RNA polymerase II and specific transcription factors. In this context, the bridging factor XPD plays a key role in modulating the association/dissociation between the CAK and the core-TFIIH, thus linking transcription to cell cycle control and DNA repair. Mutations in the *ERCC2/XPD* gene are responsible for distinct clinical entities, including the cancer-prone xeroderma pigmentosum (XP) and the multisystemic cancer-free trichothiodystrophy (TTD).

To understand how mutations in the same gene give rise to hereditary disorders with opposite cancer proneness, we investigated whether *XPD* mutations affect the cellular composition of TFIIH complex and, in turn, the CAK-mediated signaling pathways. Therefore, by native chromatin immunoprecipitation studies, we investigated the association/dissociation dynamic of the two TFIIH sub-complexes throughout the cell cycle and found that in XPD mutated fibroblasts, isolated from either XP or TTD patients, most of the CAK sub-complex is detached from the chromatin and thus from the core-TFIIH, which is mainly bound to the DNA. Considering that the CAK substrate specificity depends on it free or core-bound state, we looked at the protein interaction profile of CDK7 kinase subunit both in physiological and pathological conditions. The CDK7 immunoprecipitation followed by mass spectrometry analysis allowed the identification of new CDK7 interactors either associated or detached from the chromatin. Some of the identified proteins display a different interaction profile in TTD

and/or XP cells, compared to control fibroblasts. We focused our attention on DDX1, an ATP-dependent RNA helicase that is involved in the resolution of RNA/DNA hybrids accumulated at DNA double strand breaks. We found that the interaction between DDX1 and the chromatin-bound CDK7 is stronger in TTD primary and immortalized fibroblasts than in XP or control fibroblasts. To gain knowledge on the functionality of the identified interaction we performed various *in vitro* assays. Upon purification of DDX1 from *E. coli* and XPD, CAK and TFIIH complex from Baculovirus and *in vitro* co-immunoprecipitation experiments, we established that DDX1 binds not only the CAK sub-complex but also the XPD subunit and the core-TFIIH. Moreover, the presence of XPD stabilizes the DDX1 interaction with the CAK but not with the core-TFIIH. We found that the C-terminal region of XPD is essential for DDX1 binding and that TTD- and XP-specific point mutations mapping in the C-terminus of *XPD* do not affect their interaction *in vitro*. Finally, by *in vitro* kinase assay we assessed that DDX1 is a substrate of the free-CAK but not of the holo-TFIIH complex. Future studies will clarify the functional relevance of the CAK-XPD interaction with DDX1.

Overall, our results point to an altered CAK distribution as consequence of XPD mutations and identify DDX1 as a novel interactor of CDK7, showing a different binding profile in TTD compared to control and XP fibroblasts.

Acknowledgements

At the end of this work, I would like to take the opportunity to express my deep sense of gratitude to all those people that have been with me during these PhD years.

First and foremost, to my supervisor Dr. Donata Orioli; you deserve my deepest gratitude for the constant support and motivation. I appreciate all your contributions of time, ideas, and funding to make my PhD experience productive and stimulating. You transmit me your passion for the research and teach me to never give up.

To Tiziana, my scientific guide and my "mum"; you helped me with your affection and your patient, you share your immense knowledge with me and most of all, you were the reference for my professional and personal worries!
To Anita, a friend more than a colleague; you have been at my side every single day of my PhD. You were there to support me with daily experiments and support me during my moments of madness!
To all the lab members: you share with me the joy for good results and the frustration for all the experiments failed. A special thanks to Martina, you rise my mood up in many scientific as well as personal situations.
To Prof. Luca Bini, who contribute to my work in terms of data and expertise and who accepted to be one of my external referees.

To Dr. Emmanuel Compe, you welcomed me, you contribute with your practical knowledge and your expertise to my personal and scientific formation. Your dynamism, vision, sincerity and motivation have deeply inspired me. Thank you for your kind supervision.

To my lab mates and friends Vale and Luca; thank you for the stimulating discussion, for sharing painful moments before deadlines and for all the fun we had in and outside the institute!

The IGBMC institute has been a source of new friendships as well as good advice and collaborations that contributed to my professional education. My time in Strasbourg was made enjoyable in large part due to friends that became part of my life.

To Fede, your encouragement and moral support during the final stage of this work has been really appreciated!

Tutti i traguardi raggiunti li dedico alla mia famiglia: senza il vostro supporto non avrei vinto tante sfide e senza il vostro amore non sarei arrivata fin qui!

Debora

Abbreviations

6-4 PP	pyrimidine(6-4)pyrimidone photoproduct
4FeS	iron-sulfur cluster binding domain
A	asynchronous
aa	aminoacids
ATP	adenosine triphosphate
BCC	basal cell carcinoma
BSA	bovine serum albumin
CAK	CDK activating kinase
Cdc	Cyclin
CDK	Cyclin-dependent kinase
CK	casein kinase
COFS	Cerebro-oculo-facial-skeletal syndrome
CPD	cyclobutane pyrimidine dimer
CRL	cullin-RING E3 ligase
CS	Cockayne syndrome
CSB	Cockayne syndrome type B protein
CTD	carboxy-terminal domain
Cyc	Cyclin
Da	Dalton
DDR	DNA damage response
DMEM	Dulbecco's modified eagle media
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DRD	damage recognition domain
DPE	downstream promoter element
dsDNA	double-stranded DNA
DSIF	DRB-sensitivity inducing factor
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EM	Electron microscopy
ES	embryonic steam

FBS	fetal bovine serum
GGR	global genome repair
GTF	general transcription factor
HCMV	human cytomegalovirus
HD	helicase domain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMGN1	high-mobility group nucleosome-binding domain-containing protein 1
ICM	inner cell mass
INR	initiator element
IP	immunoprecipitation
KD	knock-down
KMT	lysine methyltransferases
KO	Knockout
LTR	long terminal repeat
MAT1	menage à trois 1
mRNAs	messenger RNA
MMTR	MAT1-mediated transcriptional repressor
MTE	motif ten element
NELF	negative elongation factor
NER	Nucleotide excision repair
NMR	nuclear magnetic resonance
NP-40	Nonidet P40
NTD	N-terminal domain
PAGE	polyacrylamide gel electrophoresys
PBS	phosphate buffer saline
PCNA	proliferating cell nuclear antigen
PH	pleckstrin homology
PIC	Pre-initiation complex
Pol	Polymerase
PPAR	peroxisome proliferator activated receptor
pRb	Retinoblastoma protein
p-TEFb	positive transcription elongation factor b
R	Restiction

RAR	retinoic acid receptor
RCF	replication factor C
RNAP II	RNA polymerase II
RPA	replication protein A
rRNA	ribosomal RNA
RVFV	Rift Valley fever virus
SCC	squamous cell carcinoma
SDS	sodium dodecyl sulfate
SF	superfamily
snRNA	small nuclear RNA
ssDNA	single-strand DNA
TARS	threonyl-tRNA synthetase
TBP	TATA-binding protein
TCR	transcription-couple repai
TFII	Transcription factor II
ThM	thumb-like motif
TR	thyroid hormone
TRiC	TCP-1 ring complex
tRNA	transfer RNA
TSS	transcription start site
TTD	Trichothiodystrophy
TTDN1	TTD non-photosensitive 1
UDS	unscheduled DNA synthesis
USP7	ubiquitin-specific protease 7
UV	ultraviolet
UV-DDB	UV-damaged DNA-binding protein
UVSSA	UV-specific scaffold protein A
VWA	von Willebrand A
WB	western blot
Wt	wild type
XP	xeroderma pigmentosum

1. Introduction

Cell proliferation and differentiation are associated to cell cycle regulation, global gene expression, surveillance and maintenance of whole genome integrity. These processes are tightly regulated and frequently interconnected by different pathways and signal transduction events. In addition, several proteins are known to display multiple functions that act in independent molecular processes. Among these, the DNA repair and general transcription factor IIH (TFIIH) is a multi-subunit protein complex engaged in a sophisticated protein to protein interaction network crucial for transcription initiation and nucleotide excision DNA repair (NER). TFIIH is highly conserved among eukaryotes and is organized into two main functional sub-complexes: the core-TFIIH, comprising the ATP-dependent DNA translocase XPB, and the CDK activating kinase (CAK) that harbors the CDK7 kinase activity. The ATPase helicase XPD is the third enzymatic subunit of the complex that bridges the core-TFIIH to the CAK. Transcription initiation relies on the translocase activity of XPB that allows transcription bubble opening and, additionally, on the CAK complex, which phosphorylates the RNA polymerase (RNApol) II. TFIIH functions in NER through the activities of both XPD and XPB subunits to open duplex DNA and verify the presence of damage DNA. In addition, the CAK deprived of the core-TFIIH (free-CAK) is essential for regulating the cell cycle progression.

The various functional roles of TFIIH complex are regulated by different pathways and it is becoming evident that the association/disassociation of the two sub-complexes, which likely involves the bridging factor XPD, plays a role in defining the activity of the complex. Mutations in genes encoding the TFIIH subunits are responsible for the pathogenesis of trichothiodystrophy (TTD), xeroderma pigmentosum (XP), Cockayne syndrome (CS) and the combined phenotypes (XP/CS and XP/TTD).

The introduction of the following thesis deeply describes the molecular structure of TFIIH, its multiple functions and the complicate relationship between TFIIH composition and its functional activities. The various genetic diseases associated to TFIIH alterations are discussed, with a more detailed analysis of XP and TTD diseases. Finally, since this thesis work is focused on the identification and characterization of DDX1 as a novel CAK interactor, a review of the literature is introduced describing the protein structural motifs, the cellular localization and the involvement of DDX1 helicase protein in various biological processes.

1.1 The general transcription factor IIH (TFIIH)

1.1.1 Structure

The basic structure and function of TFIIH complex is conserved from yeast to humans and was firstly solved in the nineties. With an overall total molecular weight of approximately 500 kDa, the mammalian TFIIH is a ten-subunit protein complex (Giglia-Mari *et al.*, 2004). It is organized in two functional sub-complexes bridged together by the XPD subunit: the core-TFIIH composed of six subunits, including XPB, p62, p52, p44, p34 and p8; and the heterotrimeric CDK-activating kinase (CAK) complex, containing the cyclin dependent kinase (CDK) 7, Cyclin H (Cyc H) and MAT1 subunits (Compe and Egly, 2012) (Figure1). Three subunits of TFIIH exhibit catalytic activities: XPB is an ATPase/translocase, contributing to promoter opening during transcription initiation and necessary for DNA repair; XPD is a DNA helicase essential for NER but dispensable in transcription (Compe and Egly, 2016); CDK7 is the kinase subunit involved in basal transcription and transcription regulation. The kinase activity of CDK7 is also required to regulate cell cycle progression from G2 to M phase. The various enzymatic activities of TFIIH are tightly regulated through the interaction with different components of the transcription machinery, such as general and regulatory transcription factors. In addition, the TFIIH functionality appears to be regulated by contacts among the various subunits.

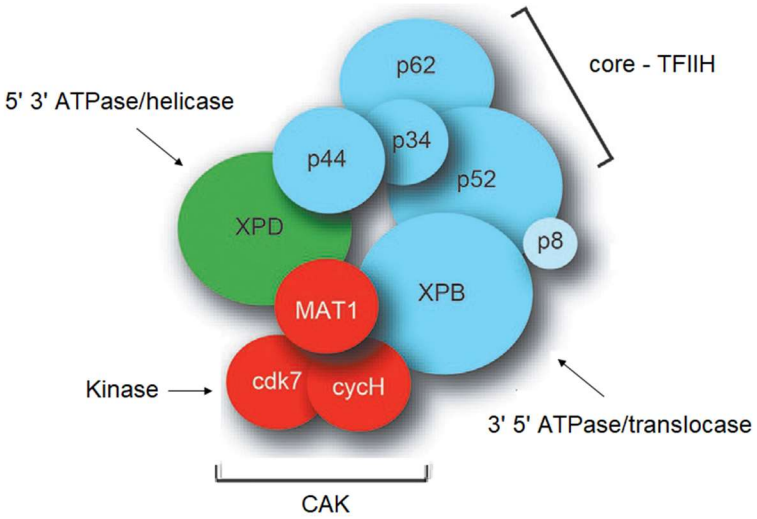


Figure 1. The TFIIH complex: Subunit composition of the Human TFIIH. Image adapted from Kolesnikova et al., 2019.

In agreement with key roles in transcription and DNA repair, the TFIIH complex is highly conserved along evolution. As expected, the three catalytic subunits (XPB, XPD and CDK7) exhibit the highest sequence conservation. Few years ago, it was reported the comparative genomic analysis of XPD and core-TFIIH subunits in more than 60 eukaryotic species (Bedez *et al.*, 2013). Although XPD, XPB, and p44 are present in all analyzed organisms, the non-enzymatic subunits of the core-TFIIH (p8/TTDA, p34, p52, and p62) are lacking in some unicellular species. The p52 subunit has not been identified in *G. lamblia* whereas p34 could not be detected in *Trypanosomatids*. Interestingly, p62 is absent in three unicellular organisms and shows only 19% sequence identity among the analyzed species. A phylogenetic comparison of the CAK complex in metazoans, fission and budding yeasts suggests a co-evolution of these components (Fisher, 2005). Kin28, the ortholog of CDK7 in *Saccharomyces cerevisiae*, is a well characterized kinase in transcription regulation, but the phosphorylation of cell cycle CDKs in this organism is mediated by the single-subunit Cak1. The fission yeast *Schizosaccharomyces pombe* has two CAKs: the essential Mcs6 complex

and the non-essential Csk1, orthologs of metazoan CDK7 and budding yeast Cak1, respectively.

Several groups have contributed to define the architecture of TFIIH and to highlight the intricate network of interactions among its subunits (Greber BJ, 2017; Greber *et al.*, 2019; Kolesnikova *et al.*, 2019). Advances in recombinant expression technologies, low resolution electron microscopy (EM) and crystal and nuclear magnetic resonance (NMR), allowed the reconstruction of the human TFIIH complex as well as some of its isolated subunits and/or functional domains. It's important to mention that the CAK sub-complex is highly flexible, explaining the lack of CDK7 and Cyc H subunits in the TFIIH solved structure. These studies revealed that XPB and XPD are located next to each other at the horseshoe-shaped structure of core-TFIIH and they directly interact with p52 and p44 subunits, respectively. The p52 subunit throughout its C-terminal region is associated with p8/TTDA to regulate the ATPase and translocase activities of XPB, whereas p44 stimulates the ATPase and helicase activities of XPD. Three different interaction sites have been described to connect p52 to p34. Moreover, MAT1 anchor the dimer CDK7-CycH to the core-TFIIH throughout interactions with both XPB and XPD. In turn, XPD is wrapped by numerous molecular bridges with XPB, p62, p44 and MAT1, indicating how its enzymatic activity is tightly regulated and triggered only when required (Figure 2).

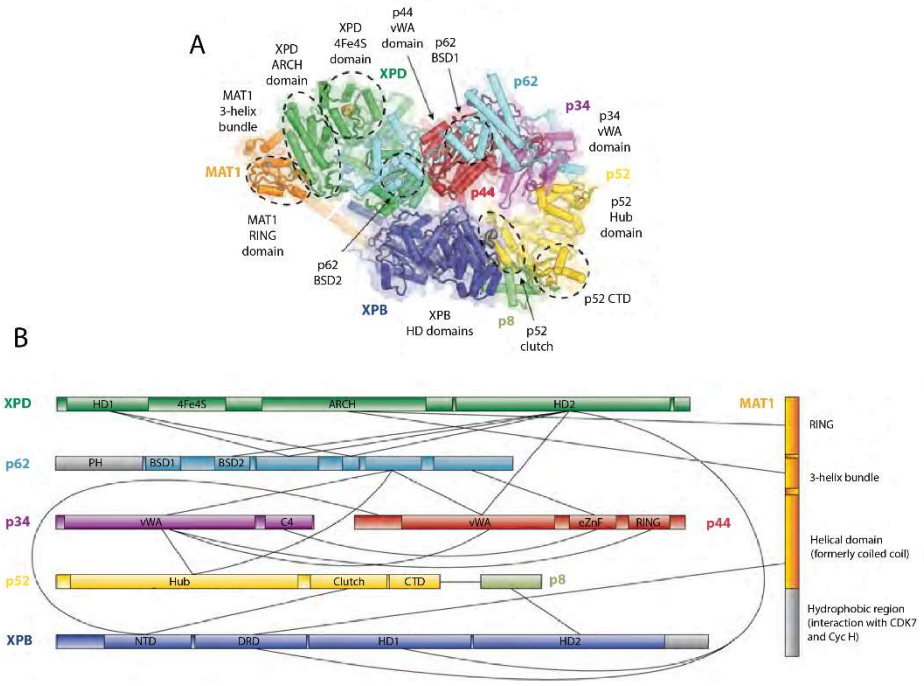


Figure 2. (A) Structure of the TFIIH core complex in association with a portion (long α-helix and helical bundle) of the CAK subunit MAT1. **(B)** Protein-protein interaction network between the components of the TFIIH core complex and MAT1 fragment. Different protein domains are indicated. Abbreviations: C4, C-terminal zinc motif; vWA, von Willebrand A; CTD, C-terminal domain; P,: pleckstrin homology; NTD, N-terminal domain; DRD, damage recognition domain. Image adapted from Basil *et al.*, 2019.

The various TFIIH subunits are summarized in Table1 and described below.

Table 1. Composition of the TFIIH complex

TFIIH	Protein	Gene	Function	Human genetic disorders
Core	XPB	*ERCC3*	3′ to 5′ ATP-dependent translocase	TTD and XP/CS
	p62	*GTF2H1*	Structural function and interaction with transcription and NER factors	
	p52	*GTF2H4*	It regulates the ATPase activity of XBP	
	p44	*GTF2H2*	An E3 ubiquitin ligase activity in yeasts	
	p34	*GTF2H3*	Structural function and strong interaction with p44	
	p8	*GTF2H5*	It regulates the ATPase activity of XBP	TTD
XPD	XPD	*ERCC2*	5′ to 3′ ATP-dependent helicase. It forms a bridge between CAK and core	TTD, XP and XP/CS XP/TTD
CAK	CDK7	*CDK7*	Kinase activity	
	Cyclin H	*CCNH*	It modulates the CDK7 kinase activity	
	MAT1	*MNAT1*	It stabilizes the CAK and regulates cullin neddylation in yeasts	

Abbreviations. CAK, cyclin-dependent kinases-activating kinase sub-complex; CDK7, cyclin dependent kinase 7; MAT1, ménage à trois 1; NER, nucleotide excision repair; XPB, xeroderma pigmentosum group B complementing protein; TTD, trichothiodystrophy; XP, xeroderma pigmentosum; CS, Cockayne syndrome.

1.1.1.1 The core-TFIIH sub-complex

XPB/Ssl2 (782 aa) is the largest (89 kDa) subunit of TFIIH complex. It is an ATPase/translocase protein, initially identified as a 3'-5' DNA helicase involved in the DNA damage response (Schaeffer *et al.*, 1993; Guzder *et al.*,

1994). The XPB protein is an unconventional helicase required for both transcription initiation and DNA repair, which contributes to DNA unwinding at gene promoters and at the lesion site,respectively (Schaeffer *et al.*, 1993). During NER, the hydrolysis of ATP induces a conformational change of XPB that contributes to separate the two DNA strands around the lesion site thus facilitating the chromatin accessibility to other repair factors (Egly and Coin, 2011). Recent studies indicated that XPB poses a dsDNA translocase activity necessary during transcription to open the double helix (Compe and Egly, 2016). It has been suggested that the ATPase but not the helicase activity of XPB determines a conformational change that contributes to anchoring TFIIH to the chromatin (Fishburn *et al.*, 2015).

As all other helicase belonging to the superfamily 2 (SF2), XPB structure comprises the HD1 and HD2 RecA-like domains. Additionally, two other domains take part in determining XPB structure: a DNA damage recognition domain (DRD) and an N-terminal domain (NTD). The three-dimensional crystal structure analysis of Archaea XPB reveal a specific RED motif (consisting of Arg, Glu and Asp residues) located in the N-terminal RecA-like domain (HD1) and a DNA polymerase thumb-like motif (ThM) in the C-terminal RecA-like domain (HD2). The N-terminal region of XPB helicase is characterized by a highly conserved 20 kDa domain involved in p52 binding. This interaction not only anchors XPB to TFIIH, but also stimulates its ATPase activity *in vitro* (Coin *et al.*, 2007). The C-terminal domain is important for the regulation of NER factors, such as ERCC1-XPF and XPG. Phosphorylation of serine 751 on human XPB prevents the 5' incision of the lesion-containing DNA fragment by ERCC1-XPF endonuclease (Coin *et al.*, 2004). Several biochemical data suggest that the N- and C-terminal ends of XPB protein are connected and constitute a structural unit, called the Lock region, proposed to be the main regulatory region of XPB via interactions with other TFIIH subunits (XPD, p52, p8 and MAT1) as well as other factors (Luo *et al.*, 2015). Another reported interactor of XPB is SUG1, an integral component of the 26S proteasome. SUG1 overexpression in normal fibroblasts results in transcription block *in vivo,* suggesting that SUG1 can inhibit transcription by binding to TFIIH via XPB (Weeda *et al.*, 1997). In yeast, the XPB-SUG1 interaction

stabilizes the DNA repair factor Rad4 and increases the cellular DNA repair efficiency (Lommel et al., 2000). These findings suggest a link between the transcription/repair and proteasome machineries. Finally, XPB, but not other core-TFIIH subunits, associates with the centrosome and mitotic spindle thus showing a potential role for XPB during cell cycle (Weber *et al.*, 2010).

The **p62**/Tfb1 (548 aa) is a core-TFIIH subunit interacting with XPD, p44, and p34. This subunit has not only the structural function to anchor different core subunits, but it also regulates the activity of different factors involved in transcription and DNA repair. The most studied domain of p62 is the pleckstrin homology (PH) domain, which mediates interactions with different members of the transcriptional machinery (Schilbach *et al.*, 2017), specific transcriptional regulators (Di Lello *et al.*, 2006) and DNA repair proteins (Okuda *et al.*, 2017). Through the PH domain, p62 binds to the *α* subunit of the general transcription factor TFIIE and allows the recruitment of TFIIH at the pre-initiation complex (PIC) (Di Lello *et al.*, 2008). Transcriptional activators such as the tumor suppressor p53, erythroid Kruppel-like factor, cell cycle controlling factor DP1 and a member of the NF-kB family p65, all target p62 N-terminal PH domain. The architecture of p62 also comprises two folding units: the BSD domain 1 (BSD1) involved in p34 and p44 binding and the BSD2 domain responsible for interaction with XPD and regulation of its activity (Greber *et al.*, 2019). Even though p62 does not directly contact the DNA-binding site of XPD, it may still physically interfere with XPD recruitment to the chromatin. Therefore, p62 should be removed in order to allow XPD binding to DNA and dsDNA unwinding. This could be achieved by conformation changes in TFIIH structure due to CAK and core dissociation and/or interaction with other NER factors (Greber *et al.*, 2019). The relevant role of p62 in NER is attested by the observation that TFIIH is recruited to the lesion site by XPC protein in global genome repair (GGR) (Okuda *et al.*, 2015) and by UVSSA in transcription coupled repair (TCR) via their interaction with the PH domain of p62 (Okuda *et al.*, 2017).

p52/Tfb2 (462 aa) is another subunit of core-TFIIH. Two important domains are located in the C-terminal region of p52: the first, known as the 'clutch', interacts with XPB (Jawhari *et al.*, 2002) while the second, a dimerization module, binds to the smallest subunit of the core-TFIIH TTDA/p8 (Kainov *et al.*, 2008). Both interaction sites are functionally important to modulate the XPB ATPase activity. Studies in *Drosophila* revealed how *p52* mutations cause strong reductions in the levels of other TFIIH core subunits, including Xpb, suggesting that when p52 is affected, Xpb becomes unstable and prone to degradation (Cruz-Becerra *et al.*, 2016). The p52 subunit also presents the so-called Hub region, which interacts extensively with the vWA domain of p34 and with the N-terminus of XPB, thus anchoring p52 subunit to the core TFIIH (Luo *et al.*, 2015).

The **p44**/Ssl1 (395 aa) subunit is crucial for TFIIH structure and function. Structural analysis of p44 shows the presence of three different domains. The N-terminal domain (von Willebrand A, vWA) binds the C-terminal region of XPD, and this direct interaction enhances the helicase activity of XPD but does not alter its ATPase activity. It has been proposed that p44 exerts its regulatory role through conformational changes of the helicase motifs, which are involved in DNA binding. The relevance of this interaction in cell viability has been demonstrated by genetics studies: a double mutation in that specific region of Ssl1/p44 is lethal in yeast (Kim *et al.*, 2015). The structure of p44 zinc finger (eZnF) domain is essential for p62 subunit incorporation within TFIIH (Kellenberger *et al.*, 2005). The RING finger domain was shown to exhibit an E3 ubiquitin ligase activity in *S. cerevisiae.* Yeast strains mutated in the this domain exhibit an increased sensitivity to DNA-damage agents due to reduced transcriptional induction of genes involved in DNA repair rather than inefficient NER activity (Takagi *et al.*, 2005). To date, no evidence concerning an enzymatic activity for p44 has been reported in higher eukaryotes. The RING domain of p44 also mediates the association with p34 subunit, via both its vWA and C4 domains. Their interactions are important for the core-TFIIH integrity as disruption of both binding sites affects the TFIIH

composition and its functionality in transcription and NER (Radu *et al.*, 2017).

The **p34**/Tfb4 (308 aa) core-TFIIH subunit contains a C-terminal zinc motif (C4) involved in binding the p44 ring domain as well as p62 and p52 subunits. The C4 region is important to stabilize the association between p34 and p44, creating a scaffold for the assembly of the other core-TFIIH subunits. The loss of the C4 domain in p34 affects both the stability of human TFIIH and its NER/transcriptional activities (Radu *et al.*, 2017). Structural analysis revealed that p34 contains a vWA like domain, which is generally involved in protein-protein interfaces and dimerization. Throughout this domain p34 establishes another important contact with the C-terminal domain of p44*. Lou and collaborators* reported an extensive network of connections between the vWA domain of p34 and p44, which allows the formation of a stable heterodimer that makes contacts with other core subunits (Luo *et al.*, 2015). No enzymatic activity is known for p34, although it has been associated to mRNA-splicing, based on phylogenetic considerations, and its depletion in the complex leads to reduced DNA repair capacity (Schmitt *et al.*, 2014).

TTDA/p8 (71 aa) is the smallest subunit of core-TFIIH. Despite its size, the TTDA protein seems to impact on different biological processes. Even if no enzymatic activity has been identified, TTDA displays an essential NER-stimulating role. Its interaction with p52 (both in yeast and human) stabilizes TFIIH structure and triggers DNA opening by stimulating XPB ATPase activity (Nonnekens *et al.*, 2013). A possible explanation is that TTDA binding to TFIIH promotes the complex association to DNA lesions and facilitates the recruitment of downstream NER factors. The lack of TTDA subunit makes TFIIH less stable and causes a reduction of DNA repair capacity *in vitro*. Indeed, the majority of the protein is not bound to the TFIIH complex *in vivo*, but its association increases after UV induced DNA damage. Differently, TTDA is dispensable for transcription, suggesting a selective involvement of this subunit in NER pathway but not in gene expression (Coin *et al.*, 2006). The lack of Ttda in mice causes embryonic lethality,

whereas the viability of *Ttda*-/- cells is not affected. Ttda deficient cells show a complete inactivation of NER pathway and are extremely sensitive to oxidative DNA damage, indicating that beyond NER TTDA (and likely the entire TFIIH) also takes part to other DNA repair pathways (Theil *et al.*, 2013).

1.1.1.2 The bridging factor XPD

XPD/Rad3 (also known as *ERCC2*) is a 761 amino acids protein with a molecular weight of 80 kDa. XPD belongs to the group of ATP-dependent SF2 helicases and displays a 5'-3' translocase activity. The relevance of XPD in transcription does not depend on its helicase activity, but relies on the structural integrity of TFIIH complex, which is XPD-dependent (Lehmann, 2001). Conversely, the helicase activity of XPD is pivotal during NER, when lesion verification and opening of the damaged DNA needs to be regulated with the recruitment and activation of the repair machinery (Mathieu *et al.*, 2013). Crystal structures of different archaeal homologs of human XPD (Fan et al. 2008; Wolski et al. 2008; Liu et al. 2008) revealed a four domains organization with HD1 and HD2 RecA-like motor domains, an iron-sulfur-cluster-binding (4Fe4S) domain and a helical domain named the ARCH domain. The C-terminal region of XPD, which is conserved among eukaryotes but absent in archaea, is responsible for its interaction with p44/Ssl1 subunit. Biochemical analysis of human XPD as well as studies on the *Chaetomium thermophilum* model system showed that the ATPase activity of XPD is enhanced by p44 binding (Kuper *et al.*, 2014). The 4Fe4S cluster domain contains four redox-sensitive cysteines involved in DNA damage recognition and essential for XPD structure and helicase activity (Fan *et al.*, 2008). The ARCH domain is a platform crucial for CAK binding (Abdulrahman *et al.*, 2013). Throughout interaction with the coiled-coil region of the MAT1 subunit and the N-terminal domain of p44, XPD bridges the CAK to the core-TFIIH complex. Different lines of evidence point to a crucial role for XPD in regulating the assembly of the two TFIIH sub-complexes. The current hypothesis suggests that XPD helicase activity is blocked in transcription when CAK is part of the entire TFIIH complex whereas it

is active in NER, following CAK dissociation from the core. In this context the ARCH domain of XPD mediates the TFIIH enzymatic activities by acting as molecular switch between transcription and DNA repair (Abdulrahman *et al.*, 2013). Additionally, biochemical data show inhibition of XPD helicase activity upon MAT1 binding (Sandrock and Egly, 2001), as well as a reduction of TFIIH ssDNA affinity in the presence of the CAK (Li et al., 2015). Conversely, the release of MAT1 from XPD might allow a greater freedom of movement especially to the ARCH domain, thereby de-repressing XPD. In *Drosophila*, Xpd is linked to cell cycle control via its interaction with the CAK module: a reduction of Xpd protein amount occurs at the beginning of mitosis. This event results in detachment of the CAK from the core and allows the free-CAK to positively regulate mitotic progression (Chen *et al.*, 2003). Human XPD is not only engaged with the CAK and core-TFIIH but can also form a complex known as CAK-XPD that plays a role in the coordination and progression of mitosis (Li *et al.*, 2010). Moreover, XPD is a component of the MMXD complex involved in chromosome segregation, where it interacts with the MMS19 and MIP18 proteins (Ito *et al.*, 2010).

1.1.1.3 The CAK sub-complex

CDK7/Kin28 (346 aa) is a 40 kDa protein located in the protruding domain of the TFIIH complex. It belongs to a large family of proteins exhibiting serine/threonine kinase activity. The structure of CDK7 has been determined in 2004 in its inactive conformation and exhibits a typical kinase fold composed of an N-terminal lobe build of β sheets and a C-terminal lobe mostly α-helical (Lolli *et al.*, 2004). ATP is bound to a protein region between the two lobes. In metazoans, CDK7 has essential roles in both cell cycle and RNApol II transcription regulation. As part of the free-CAK CDK7 controls the cell cycle progression by phosphorylating the activation segment threonine of CDK1, CDK2, CDK4, and CDK6. Differently, as subunit of the entire TFIIH complex, CDK7 phosphorylates and activates the carboxy terminal domain (CTD) of RNApol II, thus triggering the transition from transcription initiation to elongation. The CTD carries an unusual heptapeptide sequence composed of the tandemly repeated consensus sequence tyrosine –

serine – proline – threonine – serine – proline – serine; CDK7 preferentially phosphorylates serine in position 5 Ser7 (Glover-Cutter et al., 2009; Larochelle et al., 2012). The CDK7 kinase activity and substrate specificity are regulated by the phosphorylation state of the enzyme as well as by its association with regulatory subunits. For maximum activity and stability, CDK7 requires phosphorylation and association with both Cyclin H and MAT1. In order to form an active stable dimer with Cyclin H *in vivo* , CDK7 must be phosphorylated on the conserved Thr170 (in the human sequence) of the activation segment also known as T-loop (Fisher *et al.*, 1995; Martinez *et al.*, 1997). This phosphorylation event greatly stimulates the activity of the CAK complex towards the C-terminal domain of RNApol II without significantly affecting the CDKs phosphorylation. Differently from the other CDKs, CDK7 has an additional phosphorylation site (Ser164 in the human protein) that enhances its kinase activity and cyclin binding properties (Martinez *et al.*, 1997; Lolli and Johnson, 2005). Among all CDKs, only CDK1 and CDK2 are able to phosphorylate CDK7 (Lolli and Johnson, 2007). *In vivo*, CDK7 T-loop phosphorylation increases during the G1 phase, upon cell stimulation by mitogens (Schachter *et al.*, 2013). In metazoan, it has been established that *in vitro* phospho-CDK1/Cyclin B (Cyc B) and phospho-CDK2/Cyclin A (Cyc A) are able to phosphorylate CDK7 on both Thr170 and Ser164 suggesting a positive feedback loop of activating phosphorylation (Garrett *et al.*, 2001). Despite their high sequence homology, CDK7, CDK1 and CDK2 cannot auto-phosphorylate their own activation segment and they select their substrate with high specificity. The association of MAT1 protein to the CDK7/Cyc H dimer *in vitro* also results in kinase activation, even if the binding of CDK7–Cyc H to MAT1 is not fully solved within TFIIH structure. In *Drosophila*, it has been demonstrated that reduced levels of Xpd increase the free-CAK kinase activity and induce cell growth and proliferation. In contrast, an excess of Xpd titrates CAK activity, resulting in reduced T-loop phosphorylation, mitotic defects and lethality (Li *et al.*, 2010). Xpd overexpression also causes a Cdk7 mislocalization in mitotic cells of *Drosophila* embryos (Chen *et al.*, 2003). It has been postulated that a decrement of Xpd levels in mitosis may cause the dissociation of the CAK sub-complex from the core, thus

shutting-down the activity of the general transcription TFIIH and simultaneously, setting the trimeric free-CAK to act as a cell cycle regulator. Differently from *Drosophila*, in human cells no variations in TFIIH subunits amounts or kinase activity have been found during the cell cycle (Tassan *et al.*, 1994; Adamczewski *et al.*, 1996). This implies that a different process may regulate the activity of the free-CAK and its association with the core-TFIIH during cell cycle.

Cyclin H /Ccl1 (323 aa) has a molecular weight of 34 kDa and its structure has been solved in its free state (Kim *et al.*, 1996). The highly conserved core domain has a canonical cyclin-fold topology composed of two α-helical domains (each containing five helices) referred to as Repeat 1 and Repeat 2 (Andersen *et al.*, 1997). It is the regulatory subunit of the CAK complex, and its presence is required for the CAK kinase activity. Cyclin C/CDK8 has been identified as one of the kinases that phosphorylate Cyclin H at two serine residues, Ser5 and Ser304, both *in vitro* and *in vivo* (Akoulitchev *et al.*, 2000). These phosphorylation events result in inhibition of TFIIH transcriptional activity. Besides CDK8, Cyclin H is also the substrate of the protein kinase CK2. Phosphorylation on Thr315 of Cyc H by CK2 seems to be essential for the CAK activity towards both CTD and CDK2 substrates (Schneider *et al.*, 2002). As other Cyclins, also Cyc H functionality is regulated by specific proteins (see below paragraph 1.1.2.3). It has been shown that p53 leads to a down-regulation of the CAK kinase activity both *in vitro* and *in vivo* (Schneider *et al.*, 1998). Indeed, p53 induces the transcription of p21 coding gene, which in turn inhibits the cyclin dependent kinases (el-Deiry *et al.*, 1993). The Cyc H subunit of TFIIH is also associated with the core-splicing component U1 of the small nuclear ribonucleoprotein (snRNP). Recruitment of U1 snRNP complex can stimulate the helicase activity of TFIIH, thus promoting the rate of transcription initiation by RNApol II (Kwek *et al.*, 2002; O'Gorman *et al.*, 2005). It has been shown that Cyc H exhibits a critical role in maintaining embryonic stem (ES) cell identity in mice, as suggested by genetic studies. Indeed, depletion of this subunit leads to differentiation of ES cells with no evident alteration of cell cycle, (Patel and Simon, 2010).

MAT1/Tfb3 (309 aa) or ménage à trois, serves as a bridging factor that connects all three ATP-dependent enzymatic subunits (CDK7, XPB and XPD) of TFIIH complex. MAT1 interacts with the CAK complex through its hydrophobic C-terminal domain and this region is sufficient to activate CDK7 kinase activity (Fisher *et al.*, 1995; Busso *et al.*, 2000). The central portion of MAT1 contains a coiled-coil motif, which anchors the CAK to the core-TFIIH by interacting with XPB and XPD (Greber BJ, 2017). The interaction between MAT1 and XPD, observed in cryo-EM map, could lead to inhibition of the XPD helicase activity, likely by reducing the conformational freedom of the ARCH domain (Abdulrahman *et al.*, 2013). The highly dynamic interface between the MAT1 helical bundle and the XPD ARCH domain regulates the dissociation of the CAK from TFIIH during NER and its subsequent re-association to restore a transcriptionally-active TFIIH (Coin *et al.*, 2008). Through the N-terminal RING finger domain, MAT1 participates in transcription regulation favoring the CTD phosphorylation: indeed, mutations in this domain reduce the overall transcriptional activity of TFIIH *in vitro* (Busso *et al.*, 2000). Furthermore, MAT1 also aids the formation of the RNApol II pre-initiation complex by making contacts with the PIC core (Schilbach *et al.*, 2017). Mat1-deficient mice exhibit peri-implantation lethality, with homozygous mutant blastocysts failing to develop the proliferative inner cell mass (ICM) cells, which underwent apoptosis (Rossi *et al.*, 2001).

1.1.2 Functional roles

Nearly three decades ago, the finding that the multiprotein complex TFIIH was both a basal transcription factor as well as a component of the NER pathway, revealed a functional link between these two biological processes (Schaeffer *et al.*, 1993; Seroz *et al.*, 1995). Furthermore, subunits of TFIIH complex participate in other important cellular functions. The free-CAK has an essential role in cell cycle control by regulating the activity of different CDKs. The XPD subunit of TFIIH takes part to the formation of other protein complexes implicated in mitotic spindle formation and chromosome segregation. Overall, these observations imply an extensive regulation of both TFIIH composition and functions to assure proper cellular activities.

1.1.2.1 TFIIH in transcription mediated by RNApol II

TFIIH plays an important role in transcription of ribosomal RNA (rRNA) by RNApol I and in transcription of protein coding genes by RNApol II (Conaway and Conaway, 1989; Feaver *et al.*, 1991; Gerard *et al.*, 1991). In eukaryotes, the transcription of coding genes, is a tightly regulated mechanism which relays on the activity of RNApol II, firstly isolated by Roeder and colleagues (Roeder and Ruttler, 1969). With an overall molecular mass of more than 500 kDa, it is composed of 12 subunits (Rpb1 to Rpb12) organized in a ten-subunit core and the Rpb4/Rpb7 dimer (Armache *et al.*, 2003). In the core structure, the two largest subunits, Rpb1 and Rpb2, form the active center which binds DNA and RNA molecules. A flexible ‘‘linker’’ emerges from the core surface and connects to the CTD of the Rbp1. The pol II transcription mechanism is divided in different steps: promoter binding, open complex formation, initiation, elongation and termination. RNApol II transcription is an ATP-dependent mechanism that requires the assembly at the promoter region of a large pre-initiation complex (PIC) (about 2.5 MDa) containing the general transcription factors (GTFs) TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH (Compe and Egly, 2012) (Figure 3).

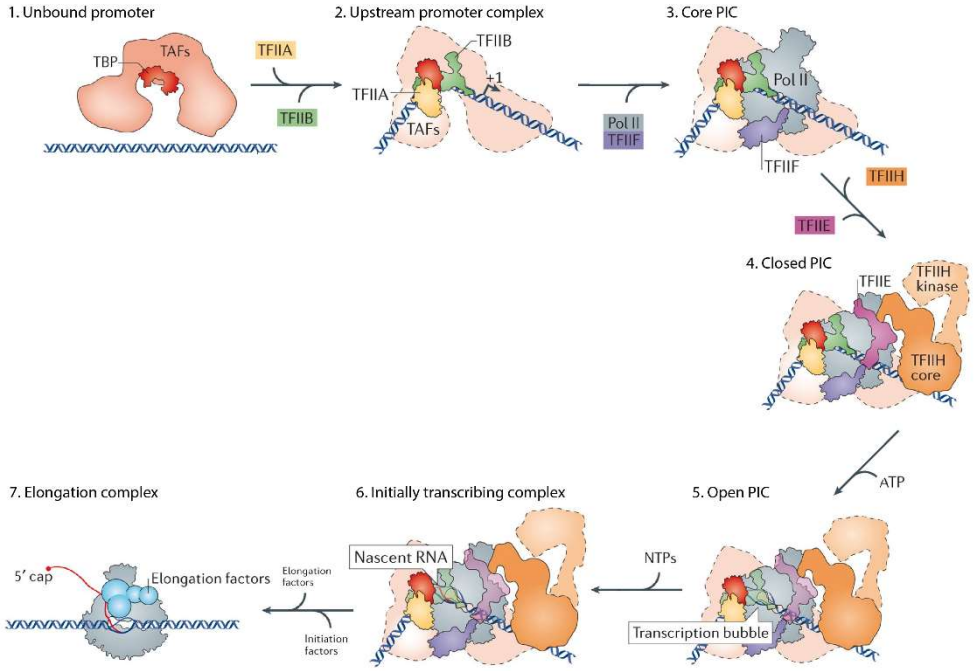


Figure 3. Schematic representation of RNApol II transcription initiation. The stepwise pre-initiation complex (PIC) assembly is depicted with the sequential recruitment of the general transcription factors (various colors) and RNApol II (gray) on promoter DNA. Abbreviations: NTP, nucleoside triphosphate; TAF, TBP-associated factor. Image adapted from Sainsbury *et al.*, 2015.

The principal purpose of the PIC is to correctly upload the RNApol II at transcription start site (TSS), a mandatory event required for transcription activation (He *et al.*, 2013). GTFs assemble at the core promoter element in a specific order. The first element recruited to the promoter is TFIID, a multi-subunit complex containing the TATA-binding protein (TBP) that recognizes the consensus sequence TATATAA (TATA box), usually located upstream of the TSS. Although most promoters do not contain a canonical TATA box (TATA-less promoters), their transcription requires TFIID (Donczew and Hahn, 2018). Some TFIID subunits are known to recognize short promoter elements such as MTE (motif ten element), INR (initiator element) and/or DPE (downstream promoter element) (Smale and Kadonaga,

2003). Since in the absence of a consensus TATA box, the PIC assembly relays on TFIID binding, the overall architecture of the initiation complex at TATA-containing and TATA-less promoters is similar (Sainsbury *et al.*, 2015). TFIID association with promoter DNA is then further stabilized by TFIIA and TFIIB. Next, TFIIF tightly associates with RNApol II and facilitates the pol II uploading to the growing PIC (*promoter binding*). Indeed, TFIIF enhances pol II affinity for the core-promoter complex, by providing additionally protein-DNA contacts (Robert *et al.*, 1998). Finally, TFIIE and TFIIH are recruited to the PIC, allowing the establishment of an open complex in which the ssDNA template, melted by TFIIH, inserts into the pol II active site. This transition mechanism leads to a transcriptional competent complex that is critically dependent on TFIIH. Experiments in both human and yeast systems have demonstrated that XPB/Ssl2 binds the DNA downstream of the transcription bubble and acts as a wrench by bending the DNA and creating a distortive melted substrate (*promoter melting*) (Kim *et al.*, 2000). Moreover, it seems that the relevance of XPB functionality during PIC formation is retained in its ATPase activity, rather than in helicase activity (Fishburn *et al.*, 2015). XPB uses energy derived from ATP hydrolysis to track the DNA in a 5'-3' direction, and ensures the correct positioning of RNApol II on the promoter and the formation of the first phosphodiester bound (Grünberg *et al.*, 2012; He *et al.*, 2013; Schilbach *et al.*, 2017). TFIIH participate in transcription initiation also via its kinase subunit. The phosphorylation of Ser5 and Ser7 in the evolutionarily conserved consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser tandemly repeated (52 repeat units in human) of the RNApol II CTD is mediated by the CDK7 kinase. These phosphorylation events allow RNApol II to engage the DNA template and start the transcription of short RNA stretches (Dvir *et al.*, 1997; Jeronimo and Robert, 2014; Wong *et al.*, 2014). When the nascent RNA strand grows to a critical length, pol II needs to be released into productive elongation (*promoter clearance*). Pol II pauses 30–60 nucleotides downstream of the TSS either to stably control gene expression or to transiently promote the recruitment of enzymes involved in mRNA processing. This mechanism depends on the presence of highly conserved DRB-sensitivity inducing factor (DSIF), a heterodimer of Spt4

and Spt5 proteins, and the negative elongation factor (NELF) recruited by DSIF itself, which promotes stable pausing. Structural and genetic studies on archaea indicate that Spt4/5 interaction with RNApol II can temporally block transcription (Grohmann *et al.*, 2011; Klein *et al.*, 2011). Overcoming the elongation block requires the positive transcription elongation factor b (P-TEFb), whose CDK9 kinase converts DSIF into a positive elongation factor, facilitates the eviction of NELF and phosphorylates the Rpb1 CTD on Ser2 (Nechaev and Adelman, 2011; Kwak and Lis, 2013). As other CDKs, also CDK9 contains an activation T-loop that must be phosphorylated for maximal activity (Larochelle *et al.*, 2012). CDK9 phosphorylation on Ser175 is principally mediated by CDK7, suggesting a possible control mechanism of pol II pausing due to CDK7, but evidences are still limited (Ebmeier *et al.*, 2017b; Mbonye *et al.*, 2018). Phosphorylated RNApol II CTD by CDK7 also acts as a scaffold for the recruitment of factors implicated in various post-initiation events (such as mRNA splicing and polyadenylation) as well as in the regulation of chromatin modification required during transcription (Eick and Geyer, 2013). The recruitment of the mRNA capping machinery through interaction of guanylyl-transferases MCE1 and CGT1 with Ser5-P, triggers the dissociation of RNApol II from the PIC that, in turn, leads to the detachment of the PIC from the promoter (Ho and Shuman, 1999; Komarnitsky, 2000). Different histone and chromatin modifying factors, important for the correct transcription cycle, have been reported to interact with the CTD (Eick and Geyer, 2013). Set1 is a H3K4 methyl-transferase that interacts with a CTD phosphorylated at Ser5 but not Ser2 (Hui Ng et al., 2003). The histone deacetylase Rpd3 is specifically recruited by the Ser5 phosphorylated CTD to actively transcribed genes (Drouin et al., 2010). The histone H3 chaperone Spt6, binds to the Ser2 phosphorylated CTD and promotes transcription elongation by altering the chromatin structure (Yoh et al., 2008). A recent proteomics analysis of phosphorylated RNApol II CTD indicates the presence of a CDK7-dependent “CTD code” capable of regulating the recruitment of transcription-associated chromatin modifiers such as the lysine methyltransferases (KMT) SET1A and SET1B (Ebmeier *et al.*, 2017a).

CDK7 is also implicated in the phosphorylation of transcriptional regulators including the CDK8 subunit of the Mediator transcriptional regulatory complex (Larochelle *et al.*, 2012), the basal transcription factor TFIIE, the p53 tumor suppressor (Lu *et al.*, 1997), nuclear receptors like the retinoic acid receptor-α1 (RARα1) and RARγ (Rochette-Egly *et al.*, 1997; Bastien *et al.*, 2000), androgen receptor (Lee *et al.*, 2001; Chymkowitch *et al.*, 2011), peroxisome proliferator activated receptor-α (PPARα), PPARβ (also known as PPARδ), PPARγ1 and PPARγ2 (Compe *et al.*, 2005) and the thyroid hormone receptor-α1 (Compe *et al.*, 2007). Studies on the role of CDK7 kinase during transcription process provided conflicting results. It was shown that *in vitro* transcription, with a highly purified enzyme system, did not required CTD phosphorylation (Serizawa *et al.*, 1993). Similarly, genetic invalidation of CDK7 in mouse affects the mRNA levels of a small number of genes, indicating that its activity is dispensable for overall transcription (Ganuza *et al.*, 2012). On the other hand, there are considerable evidences pointing to the influence of CDK7 on transcription of protein coding genes. A recent work using a covalent inhibitor reveals a role for Kin28 (the ortholog of CDK7 in *Saccharomyces cerevisiae*) in facilitating promoter release, enhancing the transition to productive elongation and priming the CTD for downstream stages of transcription cycle (Rodríguez-Molina *et al.*, 2016). Indeed, inhibition of CDK7 kinase activity (with two novel highly specific inhibitors) resulted in attenuated pol II pausing as well as productive elongation and affect mRNA capping (Kelso *et al.*, 2014; Nilson *et al.*, 2015). These observations suggest that blocking CDK7 from hyper-phosphorylating the CTD effectively inhibits promoter release and affect transcription initiation by pol II.

The CDK7 kinase activity toward the CTD is modulated by differential association with other proteins, or by other post-translational modifications. Within the trimeric complex, MAT1 is able to increase the activity of CDK7 towards the CTD at the expense of CDKs phosphorylation (Yankulov and Bentley, 1997; Komarnitsky, 2000). The CDK8-Cyclin C complex can repress general transcription by targeting the CAK complex. Indeed, CDK8 phosphorylates Cyclin H

subunit on Ser5 and Ser304 both *in vitro* and *in vivo*. These residues are in closed proximity to the N- and C-terminal domains, which are known to be essential for the CDK7 kinase activity *in vitro*. Thus, CDK8-mediated phosphorylation of Cyc H can abolish the TFIIH transcriptional activity (Akoulitchev *et al.*, 2000). Moreover, Cyc H is also the target of the CK2 kinase, which phosphorylates the residue Thr315 at the C- terminal end of the protein. This phosphorylation event has no effect on CAK structure but impacts negatively on the CAK kinase activity towards both the CTD and CDKs (Schneider *et al.*, 2002). Finally, it has been reported that the recombinant p16INK4A can specifically inhibit the CDK7-mediated CTD phosphorylation in transcription but not the phosphorylation of the CDKs in cell cycle progression (Nishiwaki *et al.*, 2000).

1.1.2.2 TFIIH in NER

DNA is continuously exposed to the deleterious action of exogenous and endogenous agents that cause genome instability. Unrepaired DNA insults can be fixed into mutations during DNA replication, leading to cellular dysfunctions and carcinogenesis (Ciccia and Elledge, 2010). Cells have developed a network of DNA damage response (DDR) mechanisms to maintain genome integrity and prevent accumulation of random DNA lesions that progressively interfere with vital biological processes, such as DNA replication and transcription. Nucleotide excision repair (NER) is an evolutionary conserved pathway that repairs DNA lesions with variable degree of distortion of the DNA helix (Mullenders, 2018). Among the different DNA repair pathways, NER is the most flexible concerning the diversity of DNA lesions. It recognizes: pyrimidine dimers (CPDs), 6-4 photoproducts (6-4PPs) caused by ultraviolet (UV) light, bulky chemical adducts, DNA intra-strand cross-links and some forms of oxidative DNA damage (Apostolou *et al.*, 2019; Kusakabe *et al.*, 2019). Eukaryotic NER can be divided into two mechanistically distinct sub-pathways: global genome repair (GG-NER) and transcription-coupled repair (TC-NER). While GG-NER functions anywhere in the genome, TC-NER is responsible for the repair of DNA lesions blocking the progression of RNApol II transcription. Although

the two sub-pathways use different players to recognize the DNA lesion, thereafter they converge in a common molecular pathway (Figure 4).

In **GG-NER**, the XPC protein is the prime essential NER factor that accumulates at the lesion site (Alekseev and Coin, 2015). Its recruitment is indispensable for the accumulation of the rest of the NER factors. *In vivo*, XPC is part of a stable protein complex with HR23B - CETN2 (Masutani *et al.*, 1994; Araki *et al.*, 2001), which detects double helix distortions and bends the DNA by destabilizing one or more base pairs. To ensure that GG-NER removes a wide spectrum of structurally different lesions, XPC displays a low binding specificity. Recent studies suggest that XPC scans the DNA filament for the presence of helical distortions. This leads to a longer retention of XPC at the suspicious site and decreases the energy required to form a stable DNA-protein interaction. XPC can also recognize and associate to some DNA sites that are prone to be naturally unwound, like it occurs during DNA replication and transcription. A considerable fraction of the DNA lesion search initiates in this way but if no real damage exists, the repair process is aborted. Therefore, it has been proposed the existence of specialized molecular mechanisms, that help XPC to efficiently discriminate between its specific targets and undamaged DNA. One of such mechanisms involves the UV-damaged DNA-binding protein complex (UV-DDB). This factor was identified as a heterodimer complex consisting of DDB1 and DDB2, which exhibits an incredible high binding affinity and specificity for UV-damaged DNA, especially for CPDs. The UV-DDB complex not only promotes the recruitment of XPC to the damage sites, but also contributes to chromatin structure remodeling thus allowing DNA lesion access to XPC and to the following repair proteins (Kusakabe *et al.*, 2019). DDB2 is involved in chromatin poly ADP-ribosylation (PARylation) and recruitment of the chromatin-remodeling enzyme ALC1, which locally modulates the chromatin structure through nucleosome sliding and thereby favoring the recruitment of XPC (Pines *et al.*, 2012)

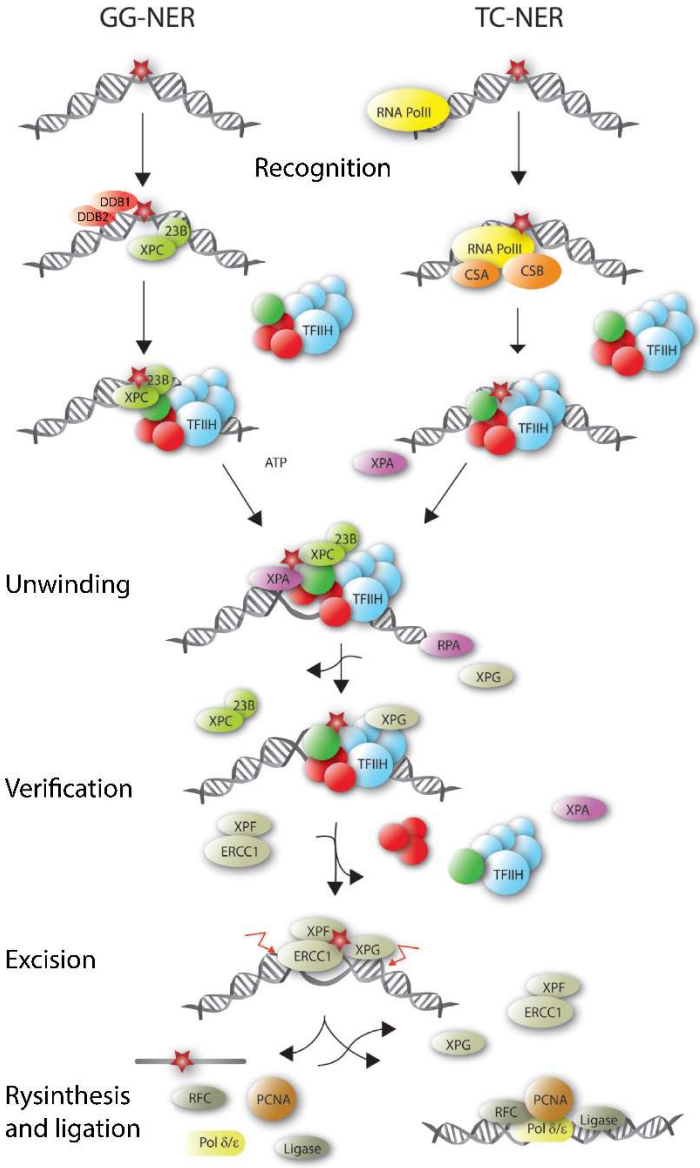


Figure 4. Nucleotide excision repair (NER). Molecular mechanisms of global genome repair (GG-NER) and transcription coupled repair (TC-NER). The NER pathway consists of different steps: recognition of DNA lesion, DNA opening, lesion verification, incision of damaged DNA and resynthesis of DNA. Image adapted from Kolesnikova et al., 2019.

Moreover, UV-DDB is part of the CUL4A-RBX1 ubiquitin ligase complex that can ubiquitinate itself and XPC protein upon UV irradiation. Polyubiquitylated XPC exhibits a higher DNA binding activity and displaces the UV-DDB complex from the site of lesion (Sugasawa *et al.*, 2005). The recruitment of the NER complex at the lesion site is strictly dependent on XPC, indeed the assembly of NER factors to photolesions is abolished in XPC-deficient cells (Mullenders, 2018). After lesion identification, XPC recruits TFIIH through the interaction with XPB and p62 subunits of the core (Coin *et al.*, 2008).

The **TC-NER** pathway relies on the transcription machinery for the initial recognition of DNA damage. The cytotoxic damage on the transcribed strand is detected by the RNApol II that bumps into the damaged DNA. Therefore, the stalled RNApol II constitutes the first step to activate TC-NER (Spivak, 2016). Thus, TCR-initiation complex, comprising CS complementation group proteins CSB (ERCC6) and CSA (ERCC8), UV-specific scaffold protein A (UVSSA) and ubiquitin-specific protease 7 (USP7) is recruited at the lesion site and subsequently displaces the polymerase. The CSB gene encodes a 168 kDa protein, member of the SWI2/SNF2 family of DNA-dependent ATPases that also displays DNA binding and nucleosome remodeling activities (Citterio *et al.*, 2000). The transcription elongation factor CSB is the first protein to be loaded onto arrested pol II and is considered the master coordinating factor in TC-NER (Sameer and Nissar, 2018). It binds the DNA upstream of RNApol II, promoting the forward movement of RNApol II and making the lesion accessible to repair proteins (Xu *et al.*, 2017). CSB than recruits CSA protein, which contains WD-40 repeats and is part of an E3-ubiquitin ligase (CRL) complex consisting of DDB1, Cullin 4A and ROC1 proteins (Groisman *et al.*, 2003). One of the functions of CRL is the ubiquitination and subsequent degradation of RNApol II under conditions of slow or no repair of the blocking lesions (Svejstrup, 2003). In normal conditions, the ligase activity of CRL is inhibited by interaction with the COP9 signalosome. Upon DNA damage, the COP9 signalosome dissociates, thus resulting in CRL activation (Groisman, 2006). Recently a stable interaction between CSA and the chaperonin TCP-1 ring complex (TRiC) has been reported, whose

function is to ensure the proper folding and stability of CSA (Pines *et al.*, 2018). CSA is also involved in the recruitment of UVSSA and USP7 proteins, which regulate the proteasome-dependent degradation of CSB. UVSSA-mediated positioning of USP7 in close proximity to CSB, allows the de-ubiquitinating activity of USP7 to prevent CSB degradation (Schwertman *et al.*, 2013). Additionally, CSB, together with CSA, recruits other factors of the TC-NER complex including the pre-mRNA splicing involved protein XAB2, the nucleosome binding protein HMGN1 and the p300 histone-acetyl transferase (Fousteri *et al.*, 2006). TFIIH is subsequently recruited to the damaged site via UVSSA (Okuda *et al.*, 2017). TFIIH is loaded downstream of the pol II-CSB complex and it uses its XPB and XPD subunits to verify whether a DNA lesion is present on the template strand (Wang *et al.*, 2018). With the recruitment of TFIIH to the repair site, both branches of the NER pathway converge.

TFIIH uses the XPD helicase to verify the presence of the lesion by scanning the DNA in 5′-3′ direction, while both XPB and XPD are required for DNA unwinding. The ATPase activity of XPB is necessary to stabilize TFIIH to the damaged site whereas its helicase activity seems to be dispensable. In opposition, XPD helicase activity is necessary to mediate the initial opening between the two DNA strands. The activity of XPB and XPD in NER has to be precisely regulated to ensure the correct DNA bubble opening. Both p52 and TTDA subunits are able to modulate the ATPase activity of XPB. The relevance of p52 was first demonstrated in *Drosophila* cells, where mutations in Dmp52 coding gene (the homolog of p52) resulted in reduced interaction between p52 and Xpb, decreased ATPase activity of Xpb and instability of the entire TFIIH complex (Fregoso *et al.*, 2007). In agreement, human mutations destabilizing the XPB–p52 interaction lead to a reduction of XPB ATPase activity and reduced DNA unwinding by TFIIH at the lesion site (Coin *et al.*, 2007). Another regulatory subunit of XPB is p8/TTDA that acts by stimulating the ATPase activity required for DNA opening (Coin et al., 2006). Finally, the p44 subunit of the core-TFIIH is critical for XPD activity in NER. Mutations that disrupt the interaction between these two subunits, perturb the correct DNA unwinding mediated by XPD

helicase activity (Coin et al., 1998; Kim et al., 2015). Once the damaged DNA has been unwound by TFIIH, the replication protein A (RPA), XPA and XPG bind to the undamaged strand forming the pre-incision complex that contains a ~30 nucleotides region of separated DNA strands around the lesion. RPA protein is recruited to coat the undamaged strand and protect the ssDNA ends from hydrolysis. XPA, which directly interacts with TFIIH, enhances the TFIIH helicase activity, triggers the release of the CAK module and promotes the recruitment of XPG and XPF endonucleases (Kokic et al., 2019). Interestingly, the CAK sub-complex was shown to inhibit the XPD helicase activity in vitro (Abdulrahman et al., 2013). A complete release of the CAK during NER may not only account for de-repression of XPD helicase activity but also to provide a more flexibility of TFIIH structure that could possibly facilitate the recruitment of downstream NER factors. Specifically, the interaction of MAT1 with XPD and XPB may inhibit global conformational changes associated with TFIIH activation in the context of DNA repair (Kokic *et al.*, 2019). At the late step of NER the re-association of the CAK with the core-TFIIH correlates with resumption of transcription (Coin *et al.*, 2008). Therefore, the interaction of TFIIH with XPA may modulate the structural and/or enzymatic features of XPD helicase and thus affect its mobility across the lesion sites. The two endonucleases XPG and ERCC1-XPF are then loaded to make dual incisions at the 3′ and 5′ of the lesion, respectively. Following the release of the lesion containing DNA fragment (about 30-oligonucleotide), the RPA protein remains associated to the undamaged strand where it recruits the proliferating cell nuclear antigen (PCNA) and replication factor C (RFC). These two factors allow the correct positioning of DNA polymerases δ and ε. The gap is filled with the novo DNA synthesis complementary to the undamaged DNA strand. Finally, the 3’ end of the newly synthesized DNA is covalently ligated to the parental strand by DNA ligases I and 3 (Kolesnikova *et al.*, 2019).

1.1.2.3 TFIIH in cell cycle

Some of the TFIIH subunits also function in cellular processes other than transcription and DNA repair (Figure 5). Indeed, the CAK was initially identified as a complex involved in cell proliferation whose function is to tightly regulate the cell cycle progression (Fesquet *et al.*, 1993).

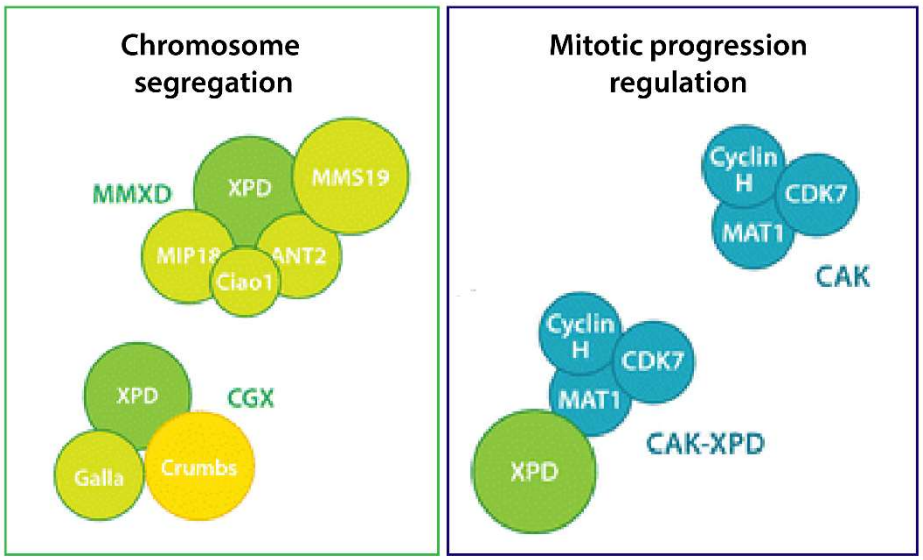


Figure 5. Involvement of TFIIH subunits/sub-complexes in cell cycle progression and chromosome segregation. The free-CAK, is involved in cell cycle regulation; CAK-XPD might have a role in the progression of mitotic divisions; XPD within the MMXD or CGX complexes is implicated in chromosome segregation. Image adapted from Compe and Egly, 2016.

Cell cycle consists of consecutive processes mainly characterized by DNA replication and subsequent chromosomes segregation into daughter cells. Originally, cell division was divided into two steps: the process of nuclear division (mitosis, M), and the interphase that lasts between two M phases. The mitosis is further characterized by four different stages, namely the prophase, metaphase, anaphase and

telophase. Afterward, different steps have been described also for the interphase, which includes the G1, S and G2 phases. During G1 phase, cells are actively preparing for the DNA replication that occurs in S phase; the following G2 period is necessary to verify the correct duplication of the genome and allow entry in mitosis. The transition from one phase to another must be tightly regulated to ensure the proper progression of cellular events. Key regulatory players are the cyclin-dependent kinases (CDK), a family of protein serine/threonine kinases whose activity depends on the association with a non-catalytic subunit called cyclin. Till now, 21 genes encoding CDKs have been identified in the human genome (Malumbres *et al.*, 2009) and, among these, five (CDK1, CDK2, CDK4, CDK6 and CDK7) are active during the cell cycle. CDK4, CDK6 and CDK2 function in G1 phase, CDK2 also in S and CDK1 in G2 and M phases. To be fully activated, CDKs require both the binding to a specific cyclin and phosphorylation on a conserved threonine residue within the activation T-loop domain. Although threonine phosphorylation does not play a direct role in catalysis, it appears to stabilize the overall structure of the CDK–cyclin complex. When activated, CDK promotes downstream signaling events through the phosphorylation of specific target proteins (Morgan, 1995; Pines, 1995). The coordinated synthesis and degradation of cyclins during cell cycle progression as well as the composition of specific CDK-cyclin partnerships allows a timely controlled activation of the different CDKs. The three D type Cyclins (Cyclin D1, Cyclin D2, Cyclin D3) are partners of CDK4 and CDK6 that allow cells to move through G1 (Sherr, 1994). CDK4/6 - Cyc D dimer is necessary to indirectly activate the transcription of *cyclin A*, *cyclin E* and *Cdc25* genes, required for progression into the S phase. Another G1 Cyclin is Cyc E ,which associates with CDK2 to regulate the progression from G1 to S phase (Ohtsubo *et al.*, 1995). CDK2 - Cyc E phosphorylates histone H1 and this may be important for chromosome decondensation required during DNA replication. Cyc A - CDK2 is required during S phase; the complex phosphorylates the DNA polymerase α primase, promoting initiation of DNA replication (Girard *et al.*, 1991; Walker and Maller, 1991). In late G2, CDK1 binds Cyc A to promote entry into M phase. Then A-type Cyclins are degraded by ubiquitin-mediated proteolysis and B-type

Cyclins are synthetized. As a consequence, CDK1 changes its partner and CDK1 - Cyc B complex promotes the beginning of mitosis (King *et al.*, 1994; Arellano and Moreno, 1997).

CDK7, together with Cyc H and the assembly factor MAT1, is responsible for the correct timing of activation and cyclin-binding specificity of different cell cycle CDKs, including CDK1, CDK2, CDK4, and CDK6 (Compe and Egly, 2016). CDK7 is itself phosphorylated on the activation segment (Thr170) but, in contrast to other CDKs, this event is not essential for the kinase activity. Indeed, the presence of MAT1 can substitute Thr170 phosphorylation and is sufficient for CDK7/Cyc H activation (Devault *et al.*, 1995; Fisher *et al.*, 1995; Lolli and Johnson, 2005). In addition, CDK7 has a second phosphorylation site in the activation segment (Ser164) whose phosphorylation enhances activity and cyclin binding (Martinez *et al.*, 1997; Lolli and Johnson, 2005). Although still debated, to date the CDK7/Cyclin H/MAT1 complex is the only CDK-activating kinase known in metazoans (Ganuza *et al.*, 2012). Inactivating CDK7, either by chemical inhibition (Bisteau *et al.*, 2013) or gene disruption (Ganuza *et al.*, 2012), abolishes CAK activity in whole-cell extracts and causes general activating phosphorylation failure that affects G1, S and mitotic CDKs (Larochelle *et al.*, 2012). Cdk7 genetic ablation in mice causes reduced T-loop phosphorylation of cell cycle CDKs, leading to cell cycle arrest and premature aging *in vitro* and early embryonic lethality *in vivo* (Ganuza *et al.*, 2012). In agreement, mice lacking the Mat1 subunit die early in embryogenesis, confirming the essential role of CAK in cell proliferation and development (Rossi *et al.*, 2001). In *Drosophila*, Cdk7 is an essential gene and a temperature-sensitive Cdk7 mutant prevents cell proliferation in the germ line and activation of CDK1 (Larochelle *et al.*, 1998; Fisher, 2005). Chemical inhibition of CDK7 kinase in human cancer cells blocks both S and M phases by abolishing CDK2 and CDK1 activation, respectively (Larochelle *et al.*, 2007).

The main mechanism used by CAK to identify its best substrate during different cell cycle phases relies on the presence of cyclins. CDK2 activation occurs in a two-step process: first, monomeric CDK2 is

phosphorylated by CAK on the threonine residue in the activation segment (Thr160). This event optimizes the substrate binding by favoring the association with cyclin (Cyc E in G1 and subsequently Cyc A) (Wood and Endicott, 2018). CDK1 follows a distinct activation pathway. Indeed, CDK1 cannot be phosphorylated in the absence of cyclin but it must be phosphorylated to form stable complexes with either Cyclins A or B, implying that the two events must occur in concert (Merrick and Fisher, 2010). CDK4 and CDK6 are implicated in progression acting prior to the Restriction (R) point, the regulatory point for cell division commitment that allows cells to proceed in S phase. It has been demonstrated that CDK7 can activate CDK4 and CDK6 *in vitro* and when associated with Cyc D, both kinases depend on CDK7 for their *in vivo* activation. In mouse fibroblasts, the gene disruption of Cdk7 causes a reduction of Cdk4-associated kinase activity (Ganuza *et al.*, 2012). Structural analysis has revealed significant differences between CDK4 and CDK2, suggesting that their mechanisms of activation and regulation might be different (Takaki *et al.*, 2009). CDK4 and CDK6 appear to follow the canonical pathway in which cyclin-binding is a key step to allow T-loop phosphorylation. Even after binding with Cyc D and phosphorylation on Thr172, CDK4 requires additional events to become active, such as binding to a substrate or to an assembly factor. The p27 molecule, for example, can bind the CDK4/Cyc D complex and behave as inhibitory or non-inhibitory factor depending on its own phosphorylation state (Chu *et al.*, 2007; James *et al.*, 2008). Differently from CDK2 and CDK1 complexes that remain phosphorylated for hours after CDK7 inactivation, the T-loop phosphorylation of CDK4 and CDK6 is rapidly lost. Therefore, CDK7 activity is required to maintain the active state of CDK4 and CDK6 while it only promotes the activation of CDK1 and CDK2.

Considering the involvement of the free-CAK in cell cycle progression and that of holo-TFIIH (CAK plus core) in basal and activated transcription, it's important to understand how the CAK discriminates among its targets and therefore activates specific signaling pathways. Studies on *Drosophila* Cdk7 highlighted how CAK's substrate specificity

may be regulated without large fluctuations in levels or activities. In 2003, *Suter and coworkers* used a *Drosophila* model of rapid embryo development to demonstrate that Xpd amounts vary throughout the cell cycle and drop during metaphase, anaphase and telophase (Chen *et al.*, 2003). Reduced Xpd levels resulted in increased amount of free-CAK, increased Cdk7 kinase activity and cell growth and proliferation. In contrast, an excess of Xpd titrated CAK activity, resulting in reduced T-loop phosphorylation, mitotic defects and embryo lethality. *Xpd* overexpression in wild-type *Drosophila* embryos also caused a redistribution of Cdk7 from cytoplasm to the nucleus. The lack of Xpd caused changes in the subcellular and temporal distribution of Cdk7 as well as in the local mitotic kinase activity (Li *et al.*, 2010). It has been postulated that the reduced Xpd levels in mitosis causes the dissociation of the CAK sub-complex from the core thus shutting-down the activity of the general transcription TFIIH. In parallel, the detached free-CAK can act as a cell cycle regulator. Finally, it has been shown that the lack of Xpd caused decreased degradation and abnormal cellular distribution of Cyc B, whose degradation normally occurs to exit the mitosis. According to this model, XPD presence becomes a critical regulator of both cell cycle and basal transcription by defining the distribution of CDK7 between free-CAK and entire TFIIH. The identification of a quaternary complex made of the CAK subunits and XPD both in mammalian and *Drosophila* cell extracts, also supports this idea.

In mammalian cells, the amount of all TFIIH subunits remains constant during the cell cycle, suggesting that a different process may regulate the CDK7 kinase activity and its substrate specificity during cell cycle (Tassan *et al.*, 1994; Adamczewski *et al.*, 1996). Although previous studies reported that the overall T-loop phosphorylation of CDK7 does not drastically change during the cell cycle in human cells (Garrett *et al.*, 2001), *Schachter and collaborators*, demonstrated that in HCT116 and in RPE-hTERT cells, the Thr170 phosphorylation of CDK7 increases during mitosis, whereas the total levels of CDK7, Cyc H and MAT1 remain roughly constant (Schachter *et al.*, 2013). Thus, the regulation of CAK activity by T-loop phosphorylation is becoming a real possibility.

The Thr170 phosphorylation is induced by mitogenic signaling and could be important for the switch from transcription to cell cycle programs. Phosphorylation at Ser164, apparently increases in mitotic cells treated with nocodazole, and correlates with TFIIH inability to activate transcription in mitosis (Akoulitchev and Reinberg, 1998). Indeed, CDK7 kinase activity associated to TFIIH has been identified as a target of mitotic repression by CDK1/Cyc B phosphorylation (Long et al., 1998).

Another level of CAK activity regulation is mediated by inhibitors and/or activation factors. As an example, the MAT1-mediated transcriptional repressor (MMTR) is an intrinsic negative cell cycle regulator that operates through inhibition of CAK activity. MMTR, via MAT1 binding, blocks the CAK-mediated phosphorylation of CDK1 both *in vitro* and *in vivo*. MMTR over-expression delayed the G1/S and G2/M transitions, whereas co-expression of MAT1 and MMTR rescue the cell growth and proliferation rate. Moreover, the expression level of MMTR is modulated during cell cycle progression and its down regulation correlates with G1/S and G2/M transitions. These findings suggest that the regulatory role of MMTR on cell cycle is strictly dependent on its expression level and acts as an anti-proliferative factor in a dose-dependent manner (Shin *et al.*, 2010).
The discovery that a CAK-XPD sub-complex exist *in vivo* and is implicated in coordinating the progression of mitosis during the late nuclear division steps in *Drosophila* embryos (Chen *et al.*, 2003; Li *et al.*, 2010), could represent an additionally regulatory mechanism for CDK7 activity.

In addition to the well described role of CAK in mitotic progression, another TFIIH subunit is involved in cell cycle. XPD, but no one of the other TFIIH subunits, is also part of the MMXD multi-protein complex (Ito et al., 2010). Composed of XPD, MMS19, MIP18, Ciao1, and ANT2, the MMXD complex localizes at the mitotic spindles where it regulates the correct chromosomes segregation and the nuclear shape formation. This finding suggests that XPD may have distinct roles during mitosis, including spindle pole formation and chromosome segregation. Even

though no other TFIIH subunits take part to the MMXD complex, MMS19 was previously shown to interact and regulate the TFIIH activity during DNA repair (Seroz, 2000). In the yeast *Saccharomyces cerevisiae,* deletion of *Mms19* correlates with a reduction of Rad3 (XPD homologue) and Ssl2 (XPB homologue) protein levels. Indeed, the NER pathway is deficient in yeast cell extracts depleted of Mms19 and can be restored upon overexpression of Rad3, but not Ssl2. Therefore, Mms19 is involved in the NER pathway by sustaining an adequate amount of the Rad3 subunit (Kou *et al.*, 2008). In addition, the iron-sulfur cluster assembly (CIA) complex, composed of MMS19, CIAO1 and FAM96B proteins, physically associates with XPD to facilitate its Fe-S cluster assembly. This finding highlights the presence of a control mechanism that prevents the incorporation of an immature XPD in the TFIIH complex (Vashisht *et al.*, 2015).
Drosophila XPD was identified to associate with Crb and Galla-2 proteins during mitosis to form the Crb-Galla-Xpd (CGX) complex. Genetic evidence indicates that this complex is required in nuclear division cycles for proper chromosome segregation (Yeom et al., 2015).

XPD silencing in tumor cells significantly promotes cell proliferation, reduces apoptosis and promotes cell invasion. Specifically, XPD reduction results in the down-regulation of p53 and p21 proteins that can promote cell proliferation by increasing the expression of Cyc D1 and Bcl-2 and C-sis oncogenes (Ding *et al.*, 2018). Conversely, the overexpression of XPD in vascular smooth muscle cells results in cell cycle arrest and suppression of CDK4 and Cyc D1 expression, both required for G1 to S transition. Thus, XPD has a tumor suppression effect and serves as a proliferation inhibitor possibly via CDK7 activity down-regulation (Li *et al.*, 2018a).

1.2 Hereditary disorders associated to mutations in TFIIH subunit encoding genes

The relevance of TFIIH complex in human health is highlighted by the dramatic consequences of alterations in some of its subunits (Table 1). Bi-allelic mutations in TFIIH genes result in different rare diseases characterized by a varied range of clinical features. Specifically, mutations in *XPD* or *XPB* genes are associated with xeroderma pigmentosum (XP), trichothiodystrophy (TTD) or Cockayne syndrome (CS), whereas mutations in *TTDA/P8* gene are associated only to TTD. Although these pathologies share a common photosensitive phenotype associated with an altered cellular response to UV light and NER defects, they represent distinct clinical entities. Besides the finding that different genes are responsible for the same pathology, the scenario gets further complicated by the observation that different alterations in the same gene give rise to different disorders (Ferri *et al.*, 2019) (Figure 6).

XP patients usually present with sunburn on minimal sun exposure, pigmentary changes at exposed skin regions and they develop skin cancers (both melanoma and non-melanoma) with an incidence 2000 times higher than that observed in the general population. This increased skin cancer predisposition can be explained by defects that directly or indirectly impair the NER activity. Progressive neuronal degeneration is also observed in approximately one-third of cases, generally after the appearance of cutaneous signs (Lehmann *et al.*, 2011). Although XP is typical of childhood, late presentations with inconspicuous skin manifestations and neurological signs have also been reported and included in XP spectrum (Garcia-Moreno *et al.*, 2018).

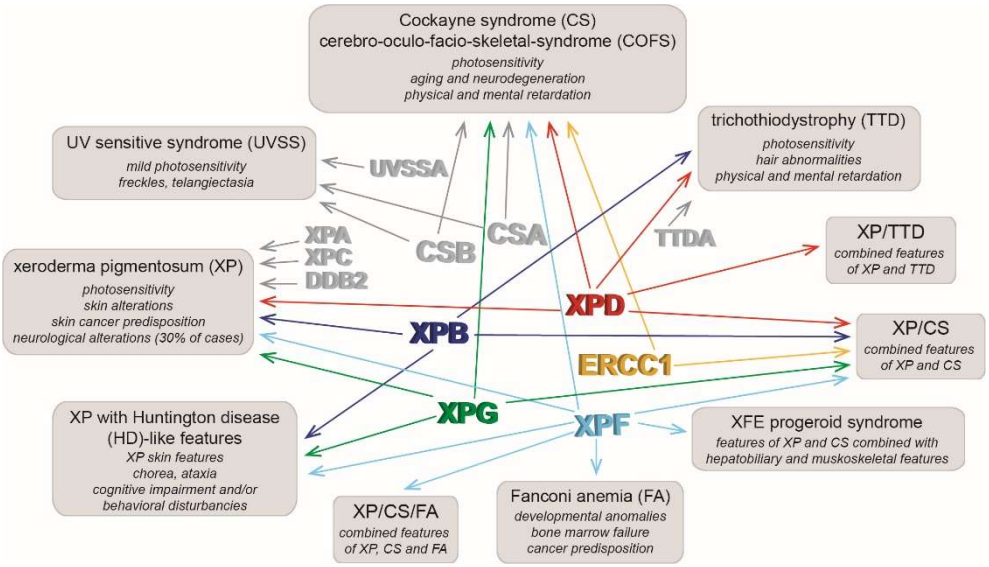


Figure 6: Relationship between NER factors and human diseases. Ten clinical conditions with the corresponding main symptoms are represented. Each disease may be caused by alterations in different NER factors and, conversely, alterations in one NER factor may result in different diseases. The pleiotropic core-NER factors are in colors. Image from Ferri et al., 2019.

Conversely, TTD and CS are invalidating multisystem disorders. TTD defining feature is the sulfur-deficient brittle hair and nails caused by the reduced level of cysteine-rich matrix proteins. Patients also display ichthyotic skin, developmental delay, neurological disability, signs of premature aging and recurrent infections, the last one being the primary cause of early death (Stefanini *et al.*, 2010). Apart from photosensitivity, the other symptoms cannot be simply explained by a DNA repair impairment but they likely reflect transcriptional deficiencies associated to TFIIH malfunctioning (Lehmann, 2003; Compe and Egly, 2016). CS patients show growth failure, progressive neurological degeneration, ocular and skeletal abnormalities and signs of premature aging (Calmels *et al.*, 2018). Various degrees of severity have been described in CS: the mildest form of CS, called UVSS, shows only mild photosensitive signs; a classical form of CS (CSI) exhibits typical CS features within the first few years of life; a severe form (CSII)

shows symptoms already at birth, and a mild form (CSIII) is characterized by late onset and slow progression of the disease. Clinical features of the cerebrooculofacioskeletal (COFS) syndrome overlap with the most severe form of CS and is characterized by fetal abnormalities and early lethality. In addition, COFS infants are born with arthrogryposis, extreme microcephaly, facial dysmorphism, progressive pigmentary retinopathy and/or cataract.

Besides these clinically distinct pathologies, "overlapping syndromes" with combined hallmarks can also occur. Careful assessment of clinical features has led to the identification of rare subsets of patients in which TTD or CS features are present in combination with cutaneous changes typical of XP, thus defining the combined phenotypes XP/TTD and XP/CS (Broughton, 2001).

For the propose of this thesis work, a more detailed description of TTD and XP diseases is presented below.

1.2.1 Trichothiodystrophy

TTD is a rare autosomal recessive disorder firstly described in 1980 (Price, 1980). It is characterized by the presence of fragile brittle hair due to reduced cysteine content. The clinical hallmark of this pathology is the presence of the so called "tiger tail" pattern: sulfur deficient brittle hairs that under the light of a polarizing microscope appear characterized by a specific dark and light banding. All patients exhibit hair abnormalities and a wide spectrum of clinical manifestations including physical and mental retardation, signs of premature aging, facial dysmorphism, ichthyotic skin and, in about half of the reported cases, cutaneous photosensitivity. Notably, TTD is also associated with increased infections frequency, especially respiratory infections, which contribute to the early mortality. Even if many patients exhibit neurological defect like speech delays, reduced learning ability and difficulties in motor control (ataxia), they show a friendly personality. The clinical features of TTD patients display a wide spectrum of severity ranging from only hair involvement to severe neurological and

skeletal abnormalities. Despite the sun sensitivity, TTD patients with mutations in TFIIH subunit coding genes do not develop pigmentary abnormalities neither increased skin cancer frequency that otherwise characterize XP subjects. Even though the accumulation of unrepaired DNA lesions could account for the aging feature in TTD patients, this defect cannot explain the developmental delay and mental retardation frequently observed in TTD.

Most cases of photosensitive TTD result from mutations in *ERCC3/XPB*, *ERCC2/XPD* or *GTF2H5/TTDA* genes that encode three of TFIIH subunits. Most of photosensitive TTD patients are mutated in the *XPD* gene, whereas alterations in *XPB* have been found rarely, probably reflecting the essential role of this subunit in TFIIH functionality. Different evidences show that also the non-photosensitive form of TTD is genetically heterogeneous with different genes involved. Even though most of these patients remained genetically uncharacterized, four different genes (*MPLKIP*, *GTF2E2*, *RNF113A* and *TARS*) have been associated with the non-photosensitive phenotype. Less than 20% of the cases are mutated in the *MPLKIP* gene encoding the TTDN1 protein, whose function is still unclear. It has been suggested that TTDN1 may have a role in transcription regulation but not in DNA repair. Alterations responsible for the photosensitive form of TTD are known to destabilize the TFIIH structure and, in turn, to cause a reduction in the intracellular concentration of the complex, leading to basal transcriptional defects (Botta *et al.*, 2002; Giglia-Mari *et al.*, 2004). Thus, impairment of TFIIH activities contributes to the TTD clinical phenotype. Structural data on the p8/TTD-A subunit revealed how p8 stabilizes TFIIH, in particular why the TTD-A causing mutations reduce the association with p52 (Kainov *et al.*, 2008) and impair the stimulation of XPB ATPase activity (Coin *et al.*, 2006; Schilbach *et al.*, 2017), leading to NER defects. Several XPD mutations have been shown to weaken the anchoring of CAK to core-TFIIH, thus reducing the phosphorylation activity of CDK7 towards specific substrates such as nuclear hormone receptors and therefore impacting on hormone-responsive gene expression. These specific mutations in XPD could account for the transcriptional defects resulting in skin and

development abnormalities. A transcriptional impairment has also been reported for PPAR target genes, as observed in an XPD-mutated TTD mouse model which develops hypoplasia of adipose tissue (de Boer *et al.*, 1998). The altered genes expression is ascribable to a weaker CDK7-mediated phosphorylation of PPARs (Compe *et al.*, 2005). More recently, it has been shown that TFIIH favors the removal of SREBP-1 regulatory element from COL6A1 promoter, thus up-regulating the expression of the gene. This mechanism failed in XPD-affected TTD patients' cells leading to reduced synthesis of COL6A1 in the extracellular matrix, which might contribute at least in part to the skin deficiencies (Orioli *et al.*, 2013). Furthermore, a common feature of TTD is the overexpression of the matrix metalloproteinase-1 (MMP-1) gene. MMP-1 up-regulation increases secretion of the active enzyme in the extracellular matrix, leading to altered tissue remodeling (Arseni *et al.*, 2015). To date it is common opinion that the reduced amounts of mutated TFIIH and the consequent altered transcriptional program typically detected in TTD cells could prevent neoplastic transformation, interfering with the expression of critical genes involved in tumor formation and/or progression. In fact, despite TTD patients are defective in the NER pathway and thus unable to correctly repair UV-induced DNA damage, they do not display increase cancer predisposition. In addition, many TTD patients develop neurological symptoms like microcephaly and demyelination, that might be connected to transcriptional impairment, including the deregulation of thyroid hormone (TR) responsive genes. Effectively, these hormones play an important role during myelinogenesis, axon formation and synaptogenesis. This suggests that, the reduced levels of TFIIH found in TTD individuals can contribute to the dysregulation of TR receptor activity and thus explain some neurological manifestations observed in TTD individuals (Compe *et al.*, 2007). The recent finding that mutations in the *GTF2E2* gene, encoding the β subunit of the TFIIE general transcription factor, or in the *TARS* gene, that encode the threonyl-tRNA synthetase, are responsible for the non-photosensitive form of TTD, supports the theory that gene expression impairments contributes to TTD etiopathogenesis (Kuschal *et al.*, 2016; Theil *et al.*, 2017, 2019).

1.2.2 Xeroderma Pigmentosum

XP is an autosomal recessive disorder caused by defective NER activities that result in skin abnormalities associated in one third of cases to neurodegeneration. First described by Kaposi in 1870, XP is diffused worldwide, with no significant difference between the sexes. The incidence in Europe and North America is of 1 in 1 000 000 , but it becomes higher in region where consanguinity is more frequent, as observed in Japan (Hirai *et al.*, 2006). The clinical marker for this disorder, especially in children under 2 years of age, is the skin photosensitivity and the presence of pigmentary abnormalities after sun exposure (Kraemer *et al.*, 2007). Although not all XP patients exhibit acute abnormal reaction due to sunlight exposure (Fassihi *et al.*, 2016), they all show freckling-like skin manifestations. The extreme UV sensitivity associated with XP greatly enhanced susceptibility to develop skin cancer, including melanoma, squamous (SCC) and basal cell carcinoma (BCC) (Bradford *et al.*, 2011). Moreover, an increased frequency (20-50 times higher) in risk of internal neoplasms, such as, breast, uterine, brain, renal, gastric, testicular and lung tumors, has been reported. Approximately 30% of XP patients also develop progressive neuronal degeneration, including intellectual disability, hearing loss, difficulty walking and swallowing. Autopsies on XP patients revealed brain atrophy and diffuse neuronal loss in the cerebellum and cerebrum. XP has no cure, but skin manifestations can be controlled and prevented with adequate prevention and sun protection. Regular skin cancer screening is fundamental to detect early malignancies. Surgery and cryosurgery are the most suitable methods to treat cutaneous tumors in children, in accordance with guidelines used for non-XP patients (Naik *et al.*, 2013).

XP is related to defect in different genes whose products are involved in the NER pathway (XPA, XPB, XPC, XPD, XPE, XPF, XPG and pol η), allowing XP to be divided into eight clinically heterogeneous complementation groups (Bowden *et al.*, 2015). The XPA to XPG proteins are involved in different steps of the NER pathway (Figure 4) whereas patients with the XP variant (XP-V) harbor mutations in the

DNA polymerase η, whose function is to bypass the UV-induced DNA lesions during DNA synthesis (Lehmann *et al.*, 1975; Masutani *et al.*, 1999). XP-V patients exhibit normal DNA repair activity, even if their cells are sensible to UV radiation in the presence of caffeine. The neurological manifestations are frequently found in XP-A, -B, -D and -G complementation groups and their severity are not related to the degree of skin damage.

XP patients can be classified in three clinical forms (severe, moderate and mild) depending on the prognosis. The most severe form is characterized by a premature onset, before 1 year of age, with cutaneous erythema associated to a strong photophobia. Skin tumors develop during childhood and patient typically die before 15 year of age. This form is usually associated with mutations in *XPA* and *XPC* genes. The moderate phenotype is characterized by cancer development after age 10 and less severe cutaneous manifestations. The mildest form exhibits a late onset (after 3 years of age) and cancer development around age 20. These correspond usually to XPE, XPF, XPG and XPV classes. The unscheduled DNA synthesis (UDS), that measures the cellular capacity to repair UV-induced DNA lesions, and cell survival after UV irradiation are the technical approaches used as reference tests for the diagnosis. More recently, the accessibility of advanced sequencing approaches allows the quick identification of causative mutations in most cases and make safer the prenatal diagnosis.

All the reported notions could explain how mutations in the same gene *ERCC2* or *ERCC3* give rise to distinct clinical entities, with the assumption that XP is primarily caused by NER defects while TTD is the result of both NER and transcriptional impairments (Lehmann, 2003; Hashimoto and Egly, 2009).

The heterogeneity of the phenotypes observed in patients with TFIIH mutations suggests that DNA repair as well as transcription and/or other cellular processes might be differently affected according to the nature of the mutations. Notably XPD is also part of MMXD complex and thus mutations in *ERCC2* gene could also affect chromosome

segregation and contribute to the observed phenotype (such as cancer predisposition in XP patients). This variety of functions and clinical presentations makes NER disorders, although rare, important model systems to investigate basic biological mechanisms and, at the same time, their implication in cancer, neurodegeneration and aging.

1.3 The ATP-dependent helicase DDX1

The work described in this thesis manuscript concerns the identification of a novel CAK interactor, the ATP-dependent helicase DDX1 that belongs to the DEAD-box protein family.

DEAD-box encoding genes are a family of RNA helicases which use the energy derived from ATP hydrolysis to modify the secondary structure of RNA molecules. DEAD box proteins are found in most species, from complex multicellular organisms to bacteria and viruses. A total of 37 DEAD box proteins, sharing the signature domain D(asp)-E(glu)-A(ala)-D(asp), have been identified in humans (Linder and Jankowsky, 2011). Despite increasing interest in this protein family, the biological role of many DEAD-box proteins remains poorly characterized. It has been proposed that they serve as multifunctional RNA chaperones that facilitate RNA folding and transport from one cellular compartment to another (Lorsch, 2002). Most biological processes related to RNA metabolism, including synthesis, splicing, translation and degradation, are modulated by DEAD-box proteins (Schmid and Linder, 1992; de la Cruz *et al.*, 1999).

DEAD box 1 (DDX1) is a 740 amino acids protein able to unwind RNA-RNA and RNA-DNA duplexes in an ATP-dependent manner. DDX1 belongs to the superfamily 2 (SF2) group of helicases, characterized by the presence of two RecA-like globular domains. In addition to the nine motifs (I, Ia, Ib, II-VI, Q), conserved in all DEAD box proteins, DDX1 contains a unique SPRY domain (Figure 7).

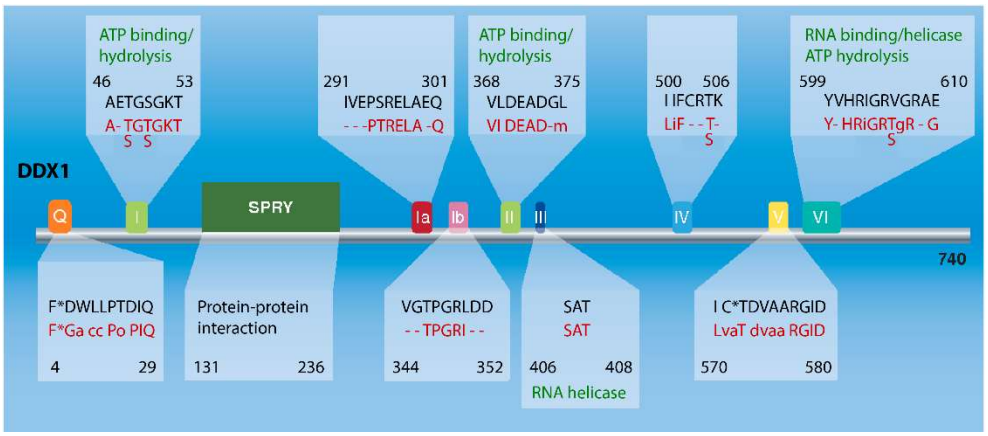


Figure 7. DEAD box motifs and SPRY domain in DDX1. Schematic representation of the nine conserved motifs found in all DEAD box proteins (Q, Ia, Ib, II, III, IV, V, VI) and the SPRY domain. The amino acid position and the functional activity associated to each motif are indicated. Capital and small case letters represent conserved (>80%) or less conserved (50-79%) amino acids in DEAD box proteins, respectively. The Asterix (*) represents a region of 16 amino acids. Image adapted from Godbout *et al.*, 2007.

The various motifs are involved in ATP binding/hydrolysis (I and II), in linking ATP hydrolysis to helicase activity (VI) or play roles in RNA binding and unwinding (Ia, Ib, IV and V)(Godbout *et al.*, 2007). Finally, the SPRY domain is a putative platform for either protein-protein or protein-RNA interactions (Kellner and Meinhart, 2015).

DDX1 was initially identified as a gene co-amplified with *MYCN* and overexpressed in a subset of retinoblastoma and neuroblastoma tumors (Godbout and Squire, 1993). Indeed, *DDX1* is widely expressed but shows its highest expression levels in proliferating cells, including cancer and neuroectodermal-derived cells (Godbout *et al.*, 2002). DDX1 is a crucial cell growth factor in diverse tumors. It sustain the tumorigenic activity of human testicular germ-cells, acting as a transcriptional activator for *Cyclin D2* and enhancing the expression of other stem cell genes (Tanaka et al., 2009). In colorectal cancer stem cells (CSCs), DDX1 promotes the expression of the *LGR5* gene by directly interactioning with its promoter (Tanaka *et al.*, 2018). High expression levels of DDX1 mRNA was reported to be a prognostic

marker for early recurrence in primary breast cancer (Germain et al., 2011). *Ddx1* knockout leads to early embryonic lethality in mice and severely reduced fertility in flies (Germain *et al.*, 2015; Hildebrandt *et al.*, 2015).

In human fibroblasts and in various cell lines, DDX1 localizes predominantly in the nucleus, suggesting that its primary function is carryed out in this cellular compartment. In contrast, in DDX1-amplified cells the protein was detected both in nucleus and cytoplasm (Godbout et al., 1998). Within the nucleus, the protein has a punctuate distribution and accumulates in a few discrete foci, known as DDX1 nuclear bodies (Bléoo et al., 2001; Li et al., 2006). Most of these foci co-localize with cleavage bodies which are involved in mRNA 3' cleavage and polyadenylation (Bléoo et al., 2001).

Although the precise function of DDX1 is still unknown, different evidence point to DDX1 involvement in pre-mRNA processing, RNA transport, transcription and translation. DDX1 has been found in a complex with hCLE protein (Trowitzsch, 2008), a component of the spliceosome, an interactor of the capping machinery and a positive modulator of RNApol II transcription (Rappsilber et al., 2003; Pérez-González et al., 2006; Jeronimo et al., 2007). Moreover, DDX1 was found to interact with hnRNP K protein, a factor of the heterogeneous nuclear ribonucleoprotein complex (hnRNP), which participates in various steps of mRNA processing, from transcription to trusduction (Chen et al., 2002). More recently, it was reported that DDX1 is recruited to the site of DNA double strand breaks (DSBs) in cells exposed to ionizing radiation (IR). Here DDX1 removes single stranded RNAs to facilitate the repair reaction of transcriptionally active regions of the genome (Li et al., 2008, 2016, 2017). In the context of metabolism regulation, it was shown that DDX1 directly blocks the translation of insulin mRNA in pancreatic β cells in response to fatty acid overload (Li *et al.*, 2018b). Furthermore, DDX1 is an essential host factor for the replication and transciption of different viruses, including human immuodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV) and JC virus (JCV). Regarding HIV-1, the endogenous DDX1 is a co-factor for the functional

activity of the Rev protein that promotes the export from the nucleus to the cytoplasm of viral mRNAs, contributing to virions formation (Fang *et al.*, 2004, 2005). It associates with the 3' untranslated region (UTR) of HCV genome thus modulating the initiation of viral RNA replication (Tingting *et al.*, 2006). Even thought DDX1 has no effect on JCV replication, its overpression significantly increased promoter activation of early and late genes important for virus infection (Sunden *et al.*, 2007).

2. Aim of the research

The transcription initiation factor IIH (TFIIH) is a multiprotein complex involved in many biological processes. It is made of ten different subunits organized in two main functional sub-complexes: the core-TFIIH (containing XPB, p62, p52, p44, p34 and p8) and the CDK-activating kinase complex (CAK, made of CDK7, Cyc H and MAT1 subunits) (Compe and Egly, 2012). The core and CAK sub-complexes are bridged together by the XPD subunit, which interacts with p44 and MAT1 of the core and the CAK sub-complexes, respectively. The multifaceted roles of TFIIH are likely regulated by multiple cellular signaling pathways, nevertheless emerging evidences underline the possibility that the structural composition of TFIIH plays a role in defining the substrate specificity of CDK7 kinase activity. As subunit of the entire TFIIH (holo-TFIIH), CDK7 targets the carboxy terminal domain (CTD) of the largest subunit of the RNA polymerase (RNApol) II, thus triggering the transition from transcription initiation to elongation. As part of the free-CAK (the fraction of the CAK sub-complex deprived of the core-TFIIH) CDK7 phosphorylates specific CDKs and controls cell cycle progression. A sub-fraction of the free-CAK is found associated to the XPD subunit but the function of this tetrameric complex and therefore its specific targets are still largely debated. The association/dissociation dynamic of TFIIH sub-complexes, which appears fundamental to determine the biological function of the complex, seems to be mediated by the bridging factor XPD, whose alterations are responsible for distinct clinical entities, including the cancer-prone xeroderma pigmentosum (XP) and the multisystemic cancer-free trichothiodystrophy (TTD).

The main goal of this thesis work is to investigate how mutations in the same gene *XPD* can results in distinct pathological phenotypes, XP or TTD, characterized by different cancer predisposition. To address this aim we explore the impact that TTD- or XP-specific mutations in the *ERCC2/XPD* gene may have on TFIIH composition and its interaction

with the chromatin. We use both primary and immortalized dermal fibroblasts isolated from healthy individuals or from XP and TTD patients with mutations in the *XPD* gene (XP-D). Since the free-CAK sub-complex is a key player in cell cycle regulation, the analysis is performed in mitotic or interphase cells. By this approach we identify the consequence of *XPD* mutations on the tightly regulated equilibrium between core-associated and free-CAK states that could suggest altered CAK functionalities in TTD and XP cells. Then, we evaluate the effect that XP- or TTD-specific mutations may have on the substrate specificity of CDK7 kinase. Co-immunoprecipitation experiments followed by mass spectrometry analysis in control and *XPD* mutated fibroblasts allow the identification of the CDK7 protein interaction framework, either associated or detached from the chromatin and in mitosis or at the interphase. Among the identified proteins we select the DDX1 helicase that shows an altered CDK7 interaction pattern in TTD and XP cells. We deeply investigate DDX1 binding capacity towards the free-CAK or holo-TFIIH, its interaction with the chromatin, its possible phosphorylation by the free-CAK or holo-TFIIH and the impact that the CDK7-mediated phosphorylation may have on the interactor functionality.

3. Materials and methods

3.1 Primary and stable cell lines

The study was performed on primary dermal fibroblasts (Table 2A) established from biopsies taken from sun-unexposed areas of the skin or fibroblasts transformed with SV40 (Table 2B). Cells were routinely cultured at 37°C in humidified atmosphere conditioned with 5% CO2 and grown in DMEM High Glucose (EuroClone) supplemented with 10% fetal bovine serum (FBS, Gibco), 0.1 mg/ml Streptomycin (EuroClone), 2 mM L-glutamine (EuroClone) and 100 U/ml Penicillin (EuroClone).

3.2 Subcellular Fractionation

SV-40 immortalized fibroblasts were subjected to mitotic shake-off, to isolate a fraction enriched in mitotic cells, separate from the others in interphase, and subsequently undergo cellular fractionation to separate a chromatin-enriched fraction from all the soluble cellular components. For primary dermal fibroblasts we used a similar approach, that involves only the cellular fractionation step.

Mitotic shake-off has been performed via vigorous shaking of the flask, that allows the release into the culture medium of mitotic cells, which are less firmly attached to the culture plastic surface. The isolated cell population enriched of mitotic cells (M) is then collected. All the cells still attached to the cell culture surface were harvested and collect as the asynchronous cell population (A). Subsequently, each cellular fraction was lysated in Triton 1% Lysis Buffer (Tris HCl pH 7.4 10 mM, Triton 1%, MgCl2 2.5 mM, Na3VO4 0.025 mM, NaF 12.5 mM, B-glycerophosphate 6.25 mM) to separate the Triton-insoluble cellular structures (Ch), enriched in chromatin, from the soluble cellular components (S). The Ch-fraction was resuspended in WCE Buffer (Hepes pH 7.9 25 mM, NaCl 130 mM, MgCl2 1.5 mM, EDTA 0.2 mM DTT 0.5 mM, Triton 0,1%) and then subjected to digestion with Mnase S7 (at

Table 2. Clinical, cellular and molecular features of the patients analyzed in this study

A) Primary dermal fibroblasts

Subject Code	Phenotype	Sex	Mutated gene	Mutated alleles	Refs[a]
C3PV	none	M	none	none	
TTD8PV mother	none	F	none	Arg112His none	
TTD8PV father	none	M	none	Arg112His none	
TTD8PV	TTD	M	*ERCC2/ XPD*	Arg112His Arg112His	(Stefanini *et al.*, 1992; Botta *et al.*, 1998)
TTD12PV	TTD	M	*ERCC2/ XPD*	Arg722Trp Cys259Tyr	(Botta *et al.*, 1998, 2002)
TTD23PV	TTD	F	*ERCC2/ XPD*	Arg112His not expressed	(Orioli *et al.*, 2013; Arseni *et al.*, 2015)
XP15PV	XP	F	*ERCC2/ XPD*	Arg683Gln Arg683Gln	(Taylor *et al.*, 1997; Arseni *et al.*, 2015)
XP49PV	XP	M	*ERCC2/ XPD*	Arg683Trp [Leu461Val;Val716 Arg730del]	(Arseni *et al.*, 2015)
XP22VI	XP	nr	*ERCC2/ XPD*	Arg683Trp Arg683Trp	(Taylor *et al.*, 1997)
XP26VI	XP	nr	*ERCC2/ XPD*	Arg616Pro Arg683Trp	(Taylor *et al.*, 1997)

B) SV-40 immortalized fibroblasts

Subject Code	Phenotype	Sex	Mutated gene	Mutated alleles	Refs[a]
MRC5	none	M	none	none	
TTD2GL	TTD	F	*ERCC2/ XPD*	Arg112His Thr460_Met493del	(Broughton *et al.*, 1994)
XP6BE	XP	F	*ERCC2/ XPD*	Val488_Met493del Arg683Trp	(Takayama *et al.*, 1995; Kobayashi *et al.*, 2002; Boyle *et al.*, 2008)

Abbreviations. TTD: trichothiodystrophy; XP: xeroderma pigmentosum; F: female; M: male; nr: not reported.

a final concentration of 0.216 U/ µl; Roche) in the presence of 1mM of CaCl2 and 4mM of MgCl2. All the described steps were performed on ice and all the buffers were supplemented with Protease Inhibitor Cocktail and PhosSTOP (Roche) to avoid protein degradation and phosphatase activity.

3.3 Protein analysis

3.3.1 Whole cell extracts

All the procedures described in this section were carried out at 4°C and all the buffers were supplemented with Protease Inhibitor Cocktail and PhosSTOP (Roche). Adherent cells were lysed 30 min at 4°C in Ripa Buffer (Tris HCl pH 7.4 50 Mm, NaCl 150 mM, 1% Triton X-100, 0.25% Na-Deoxycholate,0.1% SDS), centrifuged at 14000xg for 30 min and the supernatant stored at -20ºC.

3.3.2 Co-immunoprecipitation

The co-immunoprecipitation analysis has been performed on cell fractions derived from the subcellular fractionation described before. As first step, protein A/G PLUS Agarose Beads (Santa Cruz) were washed three times in PBS containing 0.1% Tween 20 and saturated overnight with the addition of 0.1 mg/ml BSA. Ch-enriched and Soluble protein fractions were pre-cleared by incubation with saturated beads for 1-2 hours in constant rotation. The suspension was then centrifugated, the beads were pulled apart and the supernatant was collected in a new eppendorf. After beads removal, CDK7 primary antibody conjugated to agarose beads (Santa Cruz) was added to the supernatant and incubated overnight in constant rotation. The following day, the beads were washed two times for 5 min in 1x IP Low Buffer (Active Motif) containing 1 mg/ml BSA, two times in 1x IP High Buffer (Active Motif) and two times in 1x IP Low Buffer alone in constant rotation. Immunoprecipitated proteins were eluted in 30 µl 2X NuPage-LDS Sample buffer (Life Technologies) supplemented with 50 mM DTT, 10 minutes at room temperature and further investigated by immunoblotting.

3.3.3 Immunoblotting

Whole cell extracts, cellular fractions or immunoprecipitated sample were boiled for 10 minutes at 95°C and analyzed by SDS-PAGE. Proteins were separated on 4-20% Mini-Protean TGX Gels (Bio-Rad) or 12% SDS-polyacrylamide at a constant voltage of 120 V and 90 V, respectively. Protein were transferred onto a nitrocellulose membrane by using the Trans-Blot Turbo Transfer System (BioRad), for 7 min at 2.5 A constant, up to 25 V. Alternatively, proteins were transferred onto a 0.2 µm nitrocellulose membrane (Whatman) in phosphate transfer buffer (12.2 mM Na2HPO4.2H2O, 7.8 mM NaH2PO4.H2O) by applying an electrical current of 140 V at 4°C for 1 h and 40 minutes. After the transfer, the membrane was incubated at RT for 1 hour in blocking solution made of PBS-T (1X PBS and 0.05% Tween 20) supplemented with 5% skim milk (Difco). Then, the membrane was incubated overnight at 4°C in blocking solution containing the primary antibody. The dilution of each antibody is indicated in Table 3. After three washes of 6 min each in PBS-T, the membrane was incubated at RT for 1 hour with horseradish-peroxidase-(HRP)-conjugated secondary antibodies diluted 1:10000 in blocking solution. Following this second hybridization, the membrane was washed three times as previously performed and developed using the SuperSignal West Pico PLUS (ThermoScientific) or Westar Supernova (Cyanagen), according to the manufacturer instructions. The chemiluminescent signals were then detected using the ChemiDoc XRS system (BioRad) and quantified using the Quantity one (BioRad) software.

3.3.4 Blue Coomassie staining

To stain proteins directly on polyacrilamide gels, InstantBlue™ Ready to Use Coomassie (Expedeon) was used according to the manufacturer instructions. After electrophoretic run, the gel has directly covered with the Coomassie solution and left in constant agitation for 15-30 minutes. Finally, the gel has been washed with ultrapure water.

Table 3. Antibodies used in this study

Antibody	Dilution by application		Origin	Reference
Anti-H3	WB	1:10000	Rabbit Polyclonal	Ab1791, Abcam
Anti–H3S10p	WB	1:10000	Rabbit Polyclonal	GT-128116, GeneTex
Anti-γtubulin	WB	1:10000	Mouse Monoclonal	T 6557, Sigma
Anti-Orc2	WB	1:1500	Rat Monoclonal	sc-32734, Santa Cruz
Anti-Mek2	WB	1:2000	Mouse Monoclonal	BD610235,
Anti-XPB	WB	1:2000	Mouse Monoclonal	A gift from JM Egly
Anti-p62	WB	1:2000	Mouse Monoclonal	A gift from JM Egly
Anti-XPD	WB	1:2000	Mouse Monoclonal	A gift from JM Egly
Anti-CDK7	WB	1:2000	Mouse Monoclonal	A gift from JM Egly
Anti-CycH	WB	1:2000	Mouse Monoclonal	A gift from JM Egly
Anti-DDX1	WB	1:2000	Rabbit Polyclonal	NB200-349, NovusBio

Abbreviation. WB: western blotting

3.3.5 Mass Spectrometry

Cellular fractions were subjected to CDK7 IP, as described. All sample were loaded and resolved on 4-16% polyacrylamide linear gradient gels at 40mA/gel constant current, at 9°C. Gels were stained according to MS compatible silver staining protocol. Each visible band was manually excised, destained in 2.5 mM ammonium bicarbonate and 50% (v/v) acetonitrile, and 100% (v/v) acetonitrile dehydrated. Before protein digestion, protein bands were reduced with 10 mM DTE in 25 mM ammonium bicarbonate (1 h at 56°C) and then alkylated with 55 mM iodoacetamide in 25 mM ammonium bicarbonate at RT for 45 minutes, in darkness. After 10 minutes of incubation with 50 mM ammonium bicarbonate, protein bands were dehydrated with 100% (v/v) acetonitrile. Dehydrated protein bands were then rehydrated in trypsin solution (Sigma Aldrich) and in-gel protein digestion was performed by overnight incubation at 37°C. Protein identification was carried out by Peptide Mass Fingerprinting (PMF) on an ultrafleXtreme™ MALDIToF/ ToF instrument (Brucker Corporation). After acquiring the mass of the peptides, a mass fingerprinting search was carried out in Swiss-Prot/TrEMBL and NCBInr databases using MASCOT search engine available on-line (Matrix Science Ltd).

3.4 Protein expression and purification

3.4.1 Transformation

HT96 BL21(DE3) *Escherichia coli* cells were transformed in collaboration with the laboratory of Prof. Maga (IGM-CNR, Pavia) with heat shock at 42°C for 30 seconds using pET-30a(+) plasmid (Novagen) containing DDX1 cDNA. The bacteria were plated on LB + agar plates containing kanamycin (100 µg/ml). Plates were then incubated overnight at 37°C. The obtained colonies were grown in LB medium with kanamycin under agitation at 37°C overnight. DNA was extracted from each colony using Miniprep Promega Kit and the positivity of the sample was checked by enzymatic digestion with specific restriction enzymes. The positive sample were conserved at -80°C with additional 20% of glycerol.

3.4.2 Protein expression and purification

Few µL of bacterial glycerol stock were added to 5 ml of LB + kanamycin (100 µg/ml) and grown overnight at 37°C. The following day, the pre-inoculation was transferred into 500 mL of LB + kanamycin (100 µg/ml) and grown at 37°C up to OD600 of 0.6-0.8. The expression of DDX1 protein was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) 1 mM (Sigma) and let it grow at 14°C overnight at 200 rpm.
Cells were harvested by centrifugation 4000 rpm for 20 minutes at 4 °C. Obtained pellets were washed once in 1X PBS and then resuspended in 30 ml lysis buffer (20 mM Tris HCl pH 6.8, 20% glycerol, 150 mM KCl, 0.1% NP40) in the presence of 1X PIC, freezed and defreezed in liquid nitrogen for two times to lysate and sonicate 3 times for 2 minutes. The lysed samples were ultracentifuged at 30 000 rpm for 45 minutes at 4°C and the supernatant collect. His-DDX1 present in the supernatant was purified using Talon Superflow metal affinity resin (Takara Biotechnology). Briefly, 150 µl of cobalt resin pre-washed three times in lysis buffer supplemented with 10 mM Imidazole, and 10mM of Imidazole were added to the supernatant and leave at 4°C in constant agitation for at least 2 hours. After centrifugation at 1500 rpm for 5

minutes at 4°C, the flow through was collected and the resin washed three times in wash buffer 1 (20 mM Tris HCl pH 6.8, 20% glycerol, 500 mM KCl, 0.1% NP40, 10 mM imidazole, 1X PIC) and three times in wash buffer 2 (20 mM Tris HCl pH 6.8, 20% glycerol, 150 mM KCl, 0.1% NP40, 10 mM imidazole, 1X PIC). Finally, his-DDX1 was eluted in 200 µl of elution buffer (20 mM Tris HCl pH 6.8, 20% glycerol, 150 mM KCl, 0.01% NP40, 200 mM imidazole, 1X PIC). The elution step was repeated several times to collect the maximum amount of recombinant protein. DDX1 purified protein was dialyzed for 2 hours at 4°C in dialysis buffer (50 mM Tris HCl pH 8, 20% glycerol, 0.1 mM EDTA, 50 mM KCl and 0.5 mM DTT). The quality of DDX1 recombinant protein purification was assessed by Blue Coomassie staining and western blot analysis (Figure 8).

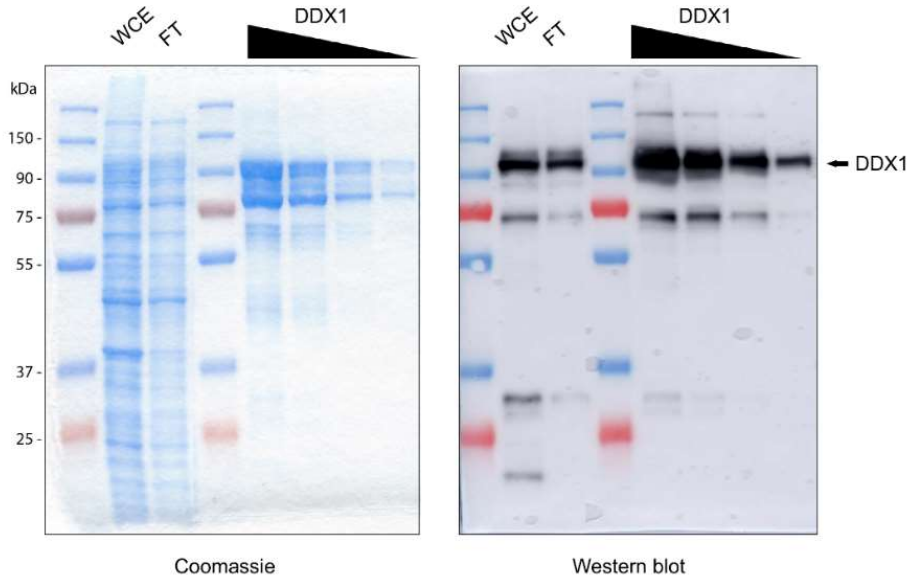


Figure 8. Coomassie and Western blot of purified recombinant DDX1 protein.
WCE, whole cell extract; FT, flow through.

3.5 *In vitro* interaction assay

Purified DDX1 was used to investigate *in vitro* its interaction with recombinant CAK and core sub-complexes as well as WT or mutated form of XPD subunit purified in the laboratory of J.M. Egly and E. Compe. 5µl of Protein G Dynabeads (Invitrogen) were used per immunoprecipitation reaction and washed three times in wash buffer 1 (1X PBS, 0.01% NP-40 and 5 mg/ml BSA). 2.5 µl of DDX1 primary antibody (Bethyl) was used for each immunoprecipitation reaction, added to the magnetic beads and left overnight on a rotary shaker at 4°C. To allow protein complex formation, the different recombinant proteins were added to 750 µl of Interaction buffer (20 mM Hepes pH 7.8, 100 mM KCl, 10% glycerol, 0.2 mM EDTA, 0.05% NP-40, 0.2 mM DTT and 1X PIC) and left overnight on a rotary shaker at 4°C. The primary-secondary antibody complex was washed one time in the wash buffer and one time in the interaction buffer and subsequently co-incubated with the purified protein solution for 2 hours at 4°C on a rotary shaker. The resulting complexes was washed four times in wash buffer 2 (20 mM Hepes pH 7.8, 300 mM KCl, 10% glycerol, 0.2 mM EDTA, 0.05% NP-40, 0.2 mM DTT and 1X PIC) and finally each co-IP reaction was eluted in 20 µl on wash buffer 2 + leammli buffer 4X + 5% beta-mercaptoethanol. Samples were stored at -20°C.

3.6 Helicase assay based on Fluorescence Resonance Energy Transfer (FRET)

After protein purification, the RNA helicase activity of DDX1 was verify using the fluorescence assay described by Tani *et al.*, 2010. Briefly, a double strand RNA prepared by hybridizing a Fluo-FAM ssRNA carries a 6-carboxyfluorescein fluorophore at its 3' end and a Qu-BHQ1 ssRNA carries a Black Hole quencher group at its 5' end, was used. The sequences of the two ssRNA oligonucleotides are following:

Fluo-FAM 5’ UUUUUUUUUUUUUUAGUACCGCCACCCUCAGAACC 3’
Qu-BHQ1 5’ GGUUCUGAGGGUGGCGGUACUA 3’

When the DDX1 protein, in the presence of ATP and MgCl2, unwinds the RNA double strand substrate, FAM molecule become free from BHQ-1 and fluorescence emission was detectable. To avoid a reannealing between complementary strands, in the reaction mix was also put a single strand DNA filament (capture) with a sequence complementary to the BHQ-1 filament.

DNA capture 5' TAGTACCGCCACCCTCAGAACC 3'

DNA-RNA hybrids are more stable than RNA-RNA double strands, so capture molecules favor the hybrid formation. Helicase assay using the dsRNA substrate was performed in 20 mM Tris HCl (pH 8), 70 mM KCl, 2mM MgCl2, 2mM dithiothreitol, 12 units RNasin (Promega), 2 mM ATP, 50 nM dsRNA and 100 nM capture strand in 20 µl of reaction volume. The unwinding reaction was started by adding 30 pmols of DDX1 recombinant protein and carried out at 37°C for 40 min using a LightCycler 480 (Roche). The fluorescence intensity was detected every 30 s. Output data were analyzed with Excel program to calculate the enzymatic activity at different substrate concentrations (Figure 9A).

3.7 Unwinding Assay

The helicase activity of DDX1 was evaluated by measuring the conversion of a double stranded (ds) molecule (labelled at the 5'-end of one strand with a 6-FAM fluorescent group) into single stranded (ss) nucleic acid. 50 nM of double strand filament (DNA/RNA hybrid or dsRNA as indicated) was used as substrate for 3ug of DDX1 recombinant protein. Reactions were performed in 20 mM Tris HCl pH 8, 2 mM DTT, 70 mM KCl, and 1 mM ATP, 2 mM MgCl2 at 37°C degrees for 30 minutes and stopped by adding EDTA 50 mM pH 8. Products were separated through non-denaturating 7% PAGE in TBE buffer (89 mM Tris, 89 mM boric acid and 2 mM EDTA). Substrates and products were visualized by laser scanning densitometry (Thyphoon-TRIO, GE Healthcare) (Figure 9B).

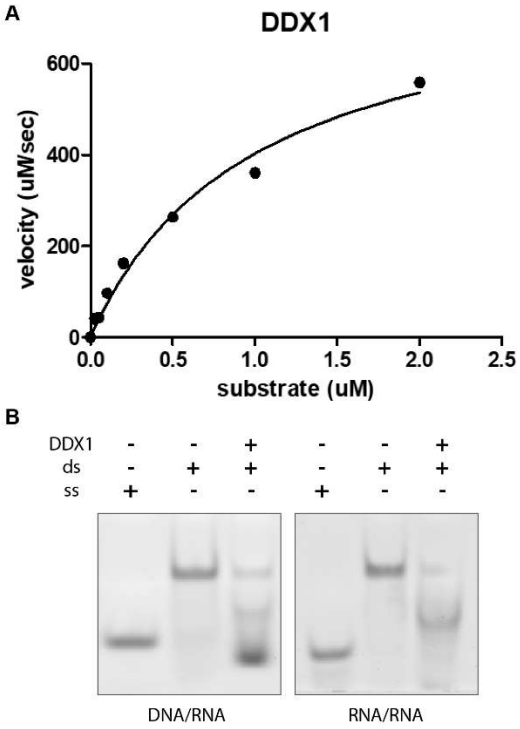


Figure 9. Enzymatic activities of DDX1 recombinant protein. Graphical representation of unwinding DDX1 activity in helicase assay in which the protein concentration (μM) is put in correlation with the velocity of the reaction (μM/sec). FRET-based assay was used to measure DDX1 capability to unwind dsRNA molecules. on increasing concentration of double strand RNA (dsRNA). **(B)** Two different double strand (ds) substrate molecules where incubated (+) with DDX1 recombinant protein to test its activity. ss: single strand.

3.8 Kinase Assay

To determine the kinase activity of CAK complex, DDX1 recombinant protein was incubated 30 min at 30°C with purified CAK sub-complex or highly purified TFIIH in the presence of [γ-32P] ATP (0.14 μM). When indicated a preincubation with THZ1, a specific inhibitor of CDK7 kinase activity, has been performed 30 minutes at 37°C. Reaction was

performed in 100 mM Hepes pH 7.9, 100 mM Tris HCl pH 7.9, 35 mM MgCl2, 20% glycerol, 2.5 mg/ml BSA and 150 mM KCl.

3.9 Statistical analysis

All the experiments were repeated at least two-three times. P-values were obtained by the unpaired two-tailed student t-test. Fisher F-ratio at a probability level (p-value) of 0.05 was used to compare variances among the analyzed groups. Data are reported as mean ± standard error (SE). All comparisons were considered statistically significant when P-values <0.05.

4. Results

4.1 Distribution of TFIIH sub-complexes in control and XP-D cell lines

In mammalian cells the substrate specificity of the CDK7 kinase varies depending on whether it acts as part of the entire TFIIH complex (holo-TFIIH) or as a subunit of the free-CAK (the fraction of CAK complex physiologically dissociated from the core-TFIIH). Relevant of note, the free-CAK is involved in cell cycle regulation through Cyclin Dependent Kinases (CDKs) phosphorylation. Therefore, the assembly/disassembly dynamic of CAK and core-TFIIH is fundamental to coordinate the various cellular functions of the CDK7-containing complexes. Considering the relevance of the XPD subunit as bridging factor between the CAK and core-TFIIH, we compared the cellular distribution of the two sub-complexes, in immortalized fibroblasts isolated from a healthy individual or patients carrying mutations in the *ERCC2/XPD* gene (XP-D) typical of the cancer-prone xeroderma pigmentosum (XP) or the cancer-free trichothiodystrophy (TTD). To identify possible differences between the healthy and the two pathological conditions that could occur in various moment of the cell cycle, we decided to separate mitotic cells from the remaining cell population and to isolate the chromatin with all the associated proteins from the other cellular components. For the separation of mitotic cells, we excluded cell cycle synchronization protocols since they are based on chemical treatments, that could interfere with the assembly/disassembly of TFIIH and its interaction with the chromatin. Thus, we established the experimental procedure for mitotic shake-off and cellular fractionation. By vigorously shaking the culture flasks, we were able to separate a sub-population of fibroblasts enriched in mitosis (M) from the cells in different phases of the cell cycle (asynchronous, A). This procedure takes advantage of the physical phenomenon according to which, cells approaching cell division (early mitosis) round up and get less firmly attached to the

culture plastic surface. Next, we lysed cells in 1% Triton X-100-containing buffer to separate a Triton unsoluble fraction that contains the chromatin (Ch) from a Triton soluble fraction that contains most of the cellular components (S) (Figure 10A). The quality of Triton-based cell fractionation and mitotic shake-off was investigated by immunoblot analysis. The phosphorylation on serine 10 of histone H3 (H3S10p) occurs during mitosis and it is tightly correlated with chromosome condensation. Histone H3 as well as Orc2 protein were used as markers for the Ch-enriched fractions, whereas Mek2 is a cytosolic protein used as a marker for the soluble fractions. As shown by the immunoblot in Figure 10B, Mek2 is mainly present in the soluble fractions whereas H3 and Orc2 are more abundantly found in the Ch-enriched fractions. The stronger signal of H3S10p in whole cell extract of mitotic cells compared to asynchronous population, clearly demonstrates the efficacy of the shake-off protocol.

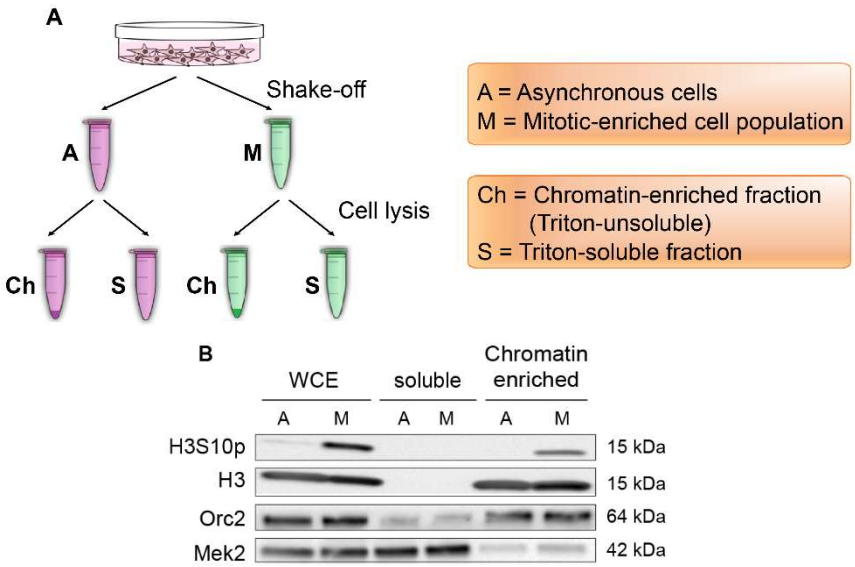


Figure 10. Shake-off and cellular fractionation. **(A)** Schematic representation of the experimental procedure for shake off and cell lyses. **(B)** Immunoblot analysis of H3S10p, H3, Orc2 and Mek2 proteins in whole cell extracts (WCE), soluble and Ch-enriched fractions of asynchronous (A) and mitotic (M) cell fractions.

Based on these data, we established that the adopted protocol is suitable to separate the Ch-enriched fractions from the remaining cellular components in mitotic or interphase cells.

Next, we investigated the cellular distribution of CAK and core-TFIIH subunits in immortalized control (MRC5) and XP-D mutated fibroblasts isolated from either XP (XP6BE) or TTD (TTD2GL) patients. These two cell lines carry the most frequent mutations representative of XP and TTD phenotype (Table 2B). In control cells, the XPB and p62 subunits of the core-TFIIH are more abundantly localized in the Ch-enriched fraction (about 90%) whereas only 10% of the total amounts are found in the soluble fraction (Figure 11A). Conversely, the CDK7 and Cyc H subunits of the CAK sub-complex are equally distributed (50%) between the Ch-enriched and soluble fractions. This distribution reveals that inside the cells not all the CAK is bound to the core-TFIIH, but a consistent amount of the sub-complex is present in a free state. The bridging factor XPD displays an intermediate situation being found in the soluble and Ch-enriched fractions for about 20% and 80%, respectively. For all the investigated TFIIH subunits no major differences were observed between the asynchronous and mitotic cell populations. In TTD2GL and XP6BE (Figure 11B and C, respectively), most of the core-TFIIH remains associated to the chromatin (86% and 85%, respectively) but it becomes slightly more soluble in mitosis (the fraction associated to the chromatin in both cell lines is reduced to about 70%). Differently from the control, in XP-D cells the majority of the CAK (80%) is detected in the soluble fraction and only 20% is attached to the chromatin. This occurs both in TTD2GL and XP6BE cell lines in asynchronous as well as mitotic cell populations. The different CAK distribution indicates that *XPD* alterations either affect the stability of holo-TFIIH complex, without impacting on the amount of the free-CAK, or they result in a loosed association of the CAK to the core-TFIIH regardless of the type of mutation. Concerning the XPD subunit, we could not detect the protein band in TTD cells likely because of the reduced stability caused by the TTD-specific mutations (Vermeulen *et al.*, 2000; Botta *et al.*, 2002).

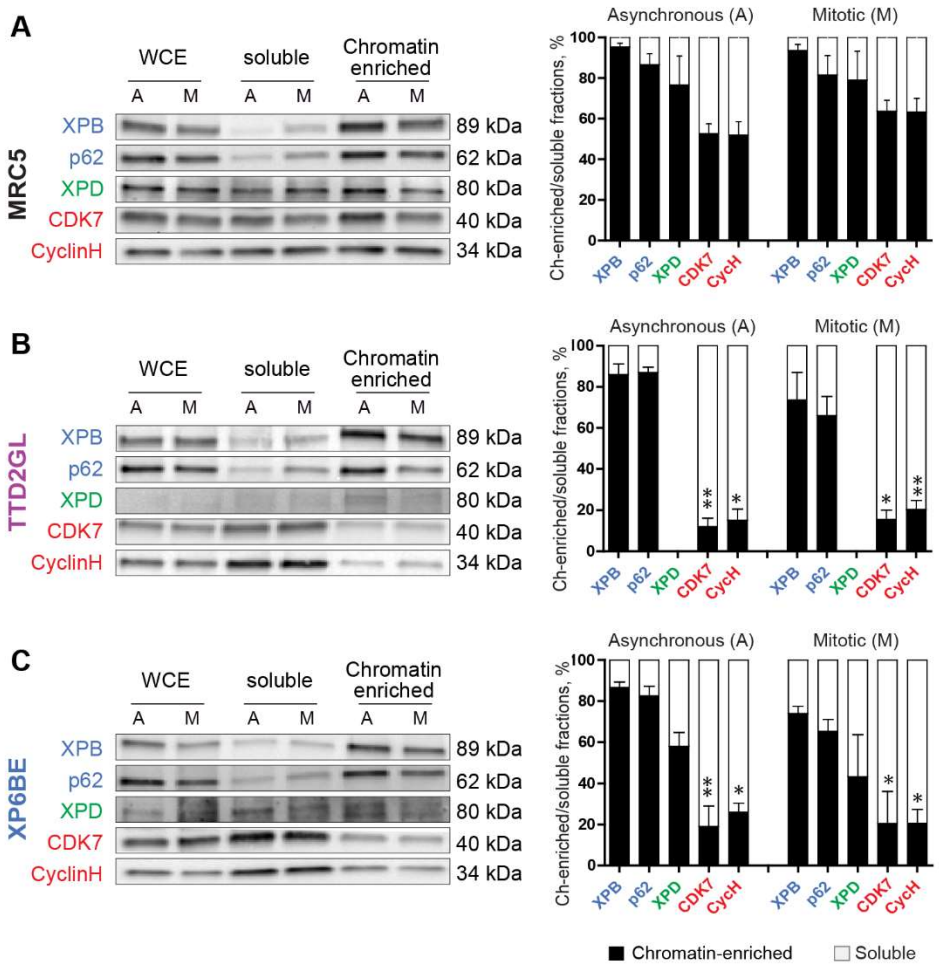


Figure 11. Shake-off and cellular fractionation of control and XP-D cell lines. (A) MRC5, **(B)** TTD2GL and **(C)** XP6BE cells were separated in asynchronous (A) and mitotic (M) cell populations. Quantification of immunoblot analysis with antibodies raised against the different TFIIH subunits (XPB and p62 of the core-TFIIH, the bridging factor XPD and the CDK7 and Cyc H subunits of the CAK) in Ch-enriched and soluble fractions. All quantifications were normalized on MEK2 and Orc2 protein amount for the soluble and Ch-enriched fractions, respectively (WB not shown). Bars indicate standard errors *(*P <0.05, **P <0.01; Student's t-test)*.

In contrast, about 50% of XPD is released from the chromatin of XP cells and detected in the soluble fraction indicative of the findings that XPD can also bind the CAK in the absence of the core-TFIIH and that takes part to other protein complexes (Chen *et al.*, 2003; Ito *et al.*, 2010).

The association/dissociation kinetic of TFIIH complex has been then investigated in primary dermal fibroblasts isolated from TTD and XP patients as well as from healthy donors. Primary fibroblasts are slow replicating cells; therefore, it was not possible to isolate an adequate number of mitotic cells, from the asynchronous population. Upon Triton-based cell fractionation of the whole cell culture, we found that in control fibroblasts (C3PV and the healthy mother of the TTD8PV patient), the core-TFIIH, the CAK sub-complex and the bridging factor XPD are predominantly attached to the chromatin (Figure 12). This cellular distribution is quite different from the situation previously observed in immortalized control cells where half of the CAK complex was found detached from the chromatin. The different behavior of the CAK could be ascribed to the high proliferation rate that characterizes the SV-40 transformed cell lines, suggesting that more free-CAK is required to sustain the cell cycle progression. In TTD and XP dermal fibroblasts the core-TFIIH remains associated to the chromatin, as observed in control cells, whereas most of the CAK results detached from the DNA as a consequence of *XPD* mutations. This finding agrees with the CAK distribution previously observed in immortalized XP-D cells. The amount of soluble CAK is even higher in TTD rather than XP fibroblasts. Moreover, in all the analyzed cell lines, the XPD factor mimics the CAK distribution.

Overall, our results highlight a different CAK distribution in XP-D fibroblasts, that is more pronounced in TTD compared to XP cells. Since the substrate specificity of CDK7 kinase activity varies between the free-CAK and holo-TFIIH, our data implies that XPD mutated cells may be characterized by an unbalanced CAK functionality that might result in an altered regulation of CDK7-dependent pathways.

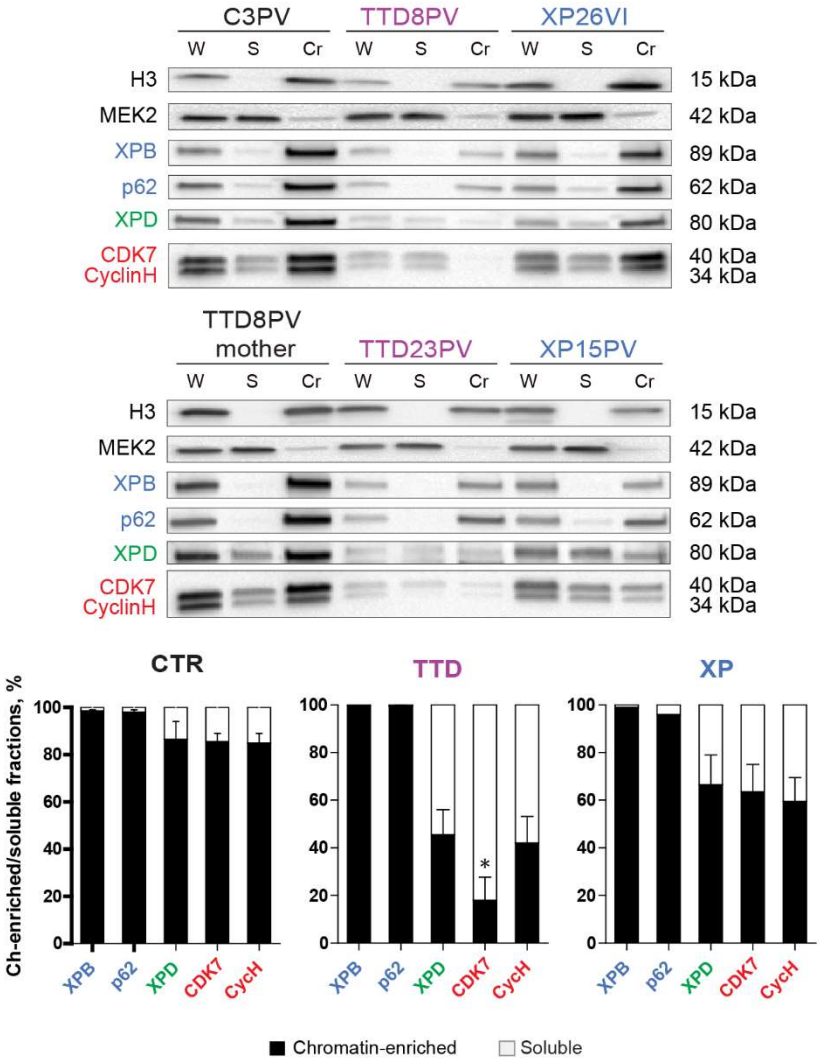


Figure 12. Shake-off and cellular fractionation of control and XP-D primary fibroblasts. Primary dermal fibroblasts derived from TTD and XP patients and from healthy donor were subjected to Triton-based fractionation to separate Ch-enriched fractions from soluble cellular components. Quantification of immunoblot analysis with antibodies raised against the different TFIIH subunits (XPB and p62 of the core-TFIIH, the bridging factor XPD and the CDK7 and Cyc H subunits of the CAK) in Ch-enriched and soluble fractions. Bars indicate standard errors *(*P <0.05; Student's t-test)*.

4.2 Identification of CDK7 interactors and their alterations in XP-D patient cells by mass spectrometry

To gain knowledge on CDK7-dependent signaling pathways and their putative alterations in XP-D cells, we searched for CDK7 interactors in control, TTD and XP cell lines. To address this issue, we established the experimental conditions to perform immunoprecipitation (IP) experiments using anti-CDK7 antibody, in the Ch-enriched and soluble fractions of mitotic or asynchronous MRC5, TTD2GL and XP6BE cells. Upon cell fractionation, samples were treated with Nuclease S7 to ensure the total nucleic acids degradation from all samples and to avoid the isolation of DNA- or RNA-mediated protein interactions. Then, cell suspensions were IP with anti-CDK7 antibody and the co-immunoprecipitated proteins were visualized by silver nitrate staining of SDS-PAGE gels (Figure 13). Each band was isolated and subjected to Peptide Mass Fingerprinting (mass spectroscopy analysis) in the laboratory of Prof. Luca Bini at the University of Siena. The identified peptides belong to several open reading frames (ORFs), the majority of which are related to RNA metabolism, including transcription regulation, pre-mRNA processing, RNA splicing, transcript stabilization and mRNA translation. Among all the identified interactors, we focused our attention on the ATP-dependent RNA helicase DDX1, which was shown to play a role in RNA clearance at DNA double-strand breaks (DSBs) mapping in the transcriptionally active genomic regions (Li *et al.*, 2008, 2016, 2017). The co-IP pattern of DDX1 protein observed by gel Silver staining suggests an altered interaction profile in *XPD* mutated cells compared to control MRC5. According to the literature data, DDX1 is capable of unwinding both RNA-RNA and RNA-DNA duplexes and exhibits an ATP-independent 5' single-stranded RNA overhang nuclease activity. The finding that CDK7 interacts with DDX1 opens the possibility of a crosstalk between the transcription/NER factor TFIIH and the DNA repair pathways operating on DSBs. Alternatively, it may indicate additional functions involving DDX1 with the CAK and/or TFIIH complex outside DSBs repair.

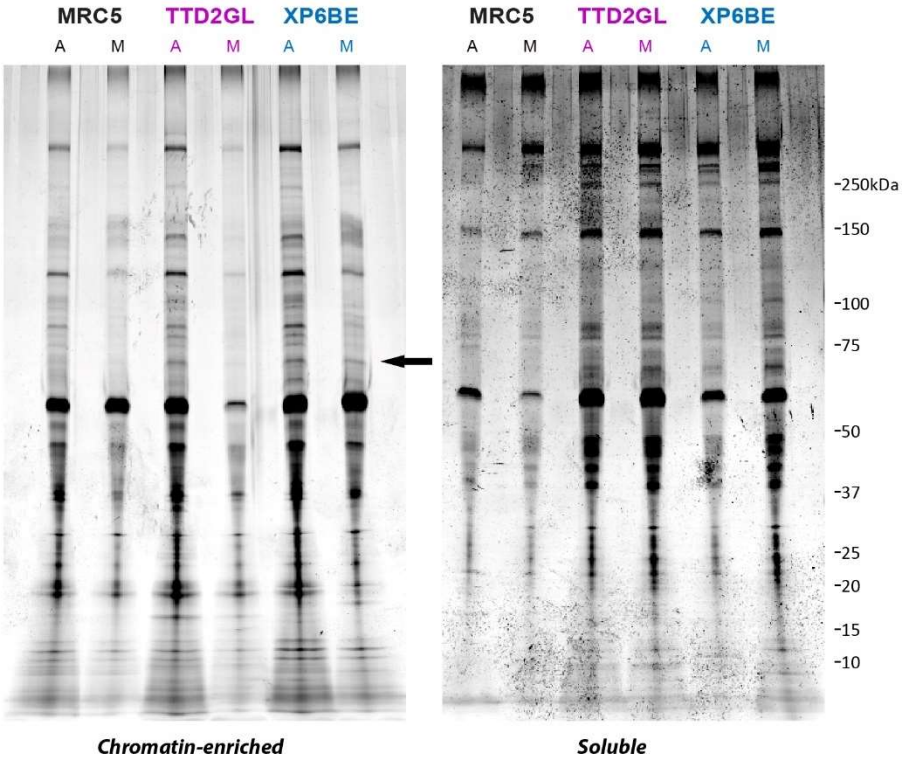


Figure 13. Gel Silver Staining on CDK7 co-immunoprecipitated proteins in control and XP-D cell lines. Gel silver staining of elutes obtained by immunoprecipitating CDK7 in the Ch-enriched and Triton-soluble fractions from control and XP-D immortalized fibroblasts. The black arrow highlights the DDX1 protein bend.

Next, the interaction between CDK7 and DDX1 was further investigated by IP and immunoblot analysis in the different cellular fractionations of mitotic or asynchronous MRC5, TTD2GL and XP6BE cells (Figure 14). As subunit of the CAK sub-complex, we observed the co-immunoprecipitation of Cyclin H with CDK7 in all the cell lines. Compared to control MRC5, the amount of immunoprecipitated CAK in TTD and XP cells is higher in the Triton-soluble than in the Ch-enriched fractions. This finding agrees with the previous results obtained in the cellular fractionation analysis (Figure 11).

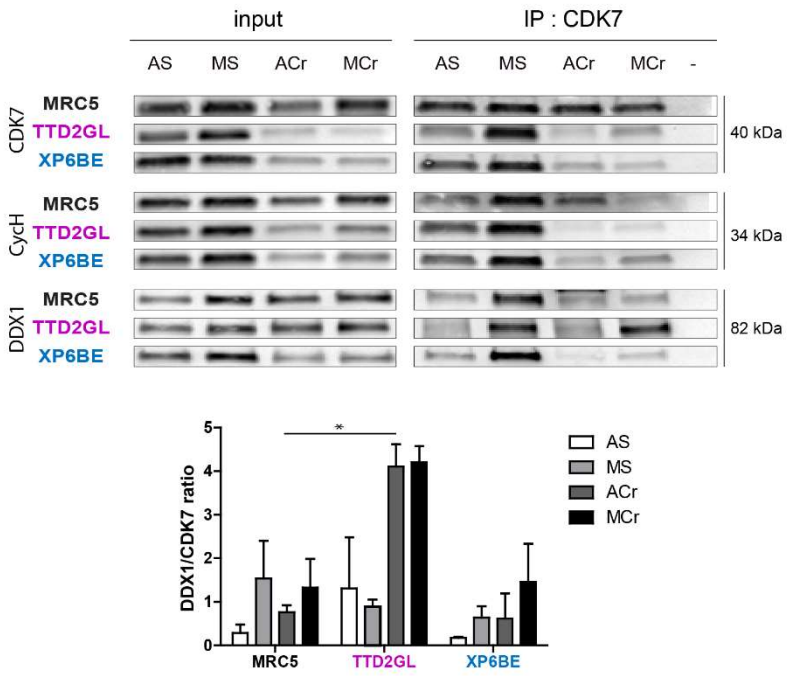


Figure 14. Interaction of DDX1 with CDK7 protein in control and XP-D cell lines. Immunoblot analysis of total and co-immunoprecipitated proteins in the different cellular fractions of MRC5, TTD2GL and XP6BE cells. The levels of co-immunoprecipitated DDX1 showed in the upper panel were normalized and quantified to the amount of CDK7. Bars indicate standard errors *(*P <0.05; Student's t-test)*.

DDX1 protein co-immunoprecipitated with both the Ch-associated and the soluble CDK7 in control as well as in patient cells. Relevant of note, the interaction between DDX1 and CDK7 appeared particularly abundant in the Ch-enriched fractions of TTD cells.

In a parallel experiment we immunoprecipitated CDK7 in the Ch-enriched and soluble fractions of control and patient primary dermal fibroblasts (Figure 15). Despite the reduced levels of CDK7 in TTD Ch-enriched fraction, we found that the amount of co-IP DDX1 is similar to that in control and XP cells. This finding can be explained by either a stronger interaction of DDX1 to CDK7, eventually due to higher cellular

concentration of DDX1, or to an increased stability of DDX1-associated CDK7. In particular, we could assume that the CDK7 protein bound to DDX1 is preserved by the protein degradation, that usually affects all TFIIH subunits in TTD cells.

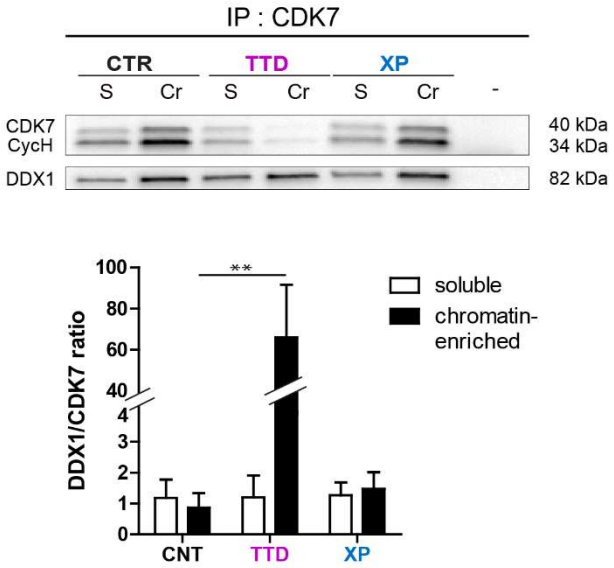


Figure 15. Interaction of DDX1 with CDK7 protein in control and XP-D primary fibroblasts. Immunoblot analysis of total and co-immunoprecipitated proteins in the Ch-enriched and soluble cellular fractions of control, TTD and XP primary dermal fibroblasts. The levels of co-immunoprecipitated DDX1 showed in the upper panel were normalized and quantified to the amount of CDK7. Bars indicate standard errors *(**P <0.01; Student's t-test)*.

4.3 DDX1 protein levels in control and *XPD*-mutated fibroblasts

To check whether the strong interaction between DDX1 and CDK7 in TTD cells does not result from a higher cellular amount of the RNA helicase protein, we evaluated by immunoblot the endogenous protein

levels of DDX1, as well as TFIIH subunits, in the three immortalized cell lines (Figure 16A) and in primary skin fibroblasts (Figure 16B). In agreement with previously published data, all TTD cell lines showed reduced TFIIH content, whereas XP cells contain normal TFIIH levels (Vermeulen *et al.*, 2000; Botta *et al.*, 2002). An exception is observed in XP6BE cells, where XPD and the core-TFIIH subunits were quantitatively reduced compared to control MRC5. In all TTD cells we observed a slight reduction of DDX1 protein content, thus denying the possibility that the increased stability of CDK7-DDX1 interaction is due to increased endogenous protein levels.

4.4 Characterization of DDX1 interaction with various subunits of TFIIH by *in vitro* assays

To establish whether the CDK7-DDX1 interaction occurs in the context of holo-TFIIH or involves only the free-CAK, we performed *in vitro* assays using purified recombinant proteins. For the TFIIH complex, baculovirus CAK and core-TFIIH sub-complexes as well as the XPD subunit were kindly provided by the laboratory of JM Egly and E Compe at the IGBMC in Strasbourg where I went to perform the *in vitro* experiments. Regarding DDX1, we successfully purified DDX1 recombinant protein (as described in Material and Methods) from *E. coli* and visualized by Gel Coomassie blue staining and immunoblot analysis (Figure 8). Since, DDX1 is an ATP-dependent RNA helicase which also possesses a nuclease activity, we verified that the purified recombinant DDX1 protein exhibits its enzymatic activities by performing a FRET-based assay on increasing concentration of double strand RNA (dsRNA) molecules and unwinding assays (see Materials and methods for details).

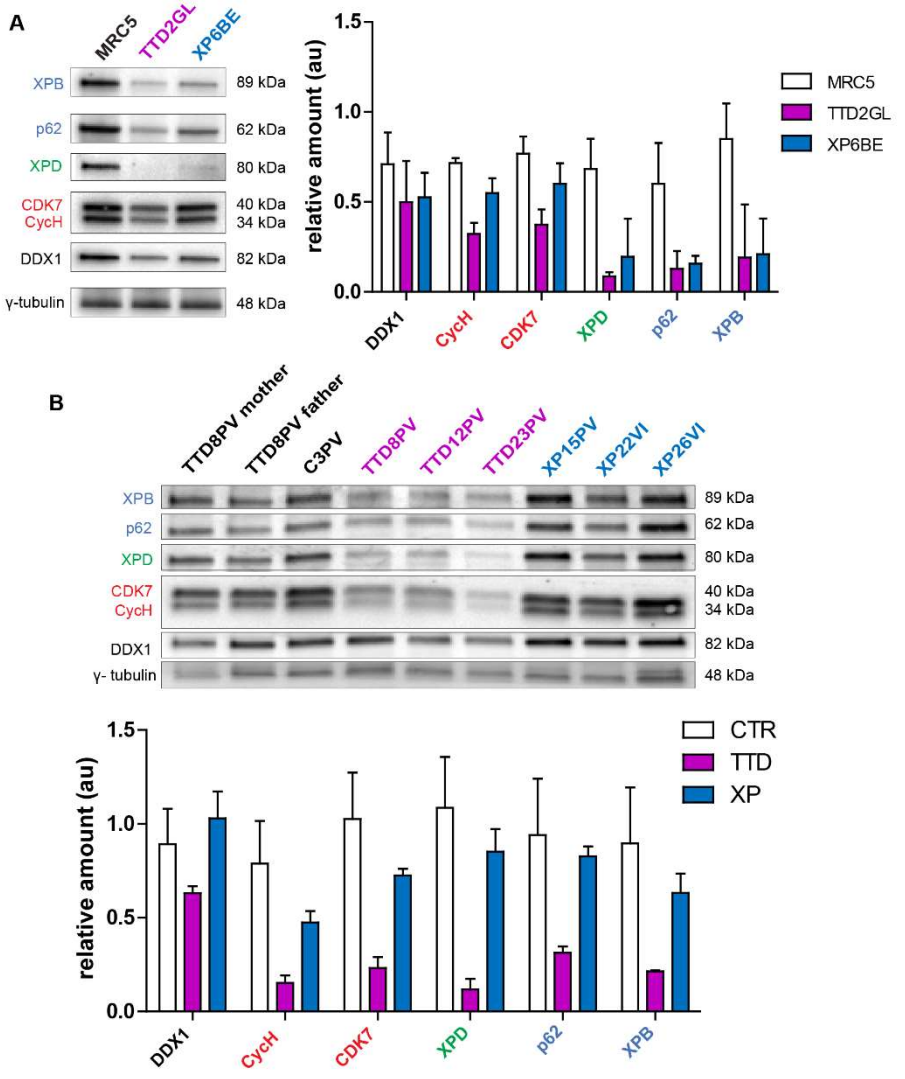


Figure 16. DDX1 and TFIIH protein levels in total cell lysates from control and XP-D fibroblasts. Immunoblot analysis and relative quantification of TFIIH subunits (XPB, p62, XPD, CDK7 and CycH) and DDX1 protein normalized to the γ-tubulin amount in **(A)** MRC5, TTD2GL and XP6BE immortalized fibroblasts or **(B)** control, TTD and XP primary dermal fibroblasts. γ-tubulin is the loading control. Au, arbitrary unit.

To confirm DDX1 interaction with CDK7 and to assess whether other TFIIH subunits and/or sub-complexes are able to bind DDX1, we performed *in vitro* interaction assays firstly by incubating DDX1 with the indicated TFIIH subunits and subsequently by immunoprecipitating DDX1 (Figure 18). Western Blot analysis allowed the visualization of the co-immunoprecipitated proteins thus demonstrating the *in vitro* interaction of DDX1 with the CAK, even if the amount of the co-immunoprecipitated CAK (lane 5) was very low compared to the amount originally incubated (input, lane 2). To our surprise, a strong interaction was observed between DDX1 and the XPD subunit alone (lane 6) or even with the core-TFIIH (lane 7) in the absence of CDK7. Apparently, the presence of XPD stabilized the interaction of DDX1 with CAK as observed in lane 8. To this aspect, we should remember that the *in vivo* existence of the CAK-XPD subcomplex lacking the core-TFIIH has been previously demonstrated but its functional role has not been clearly elucidated (Chen *et al.*, 2003; Li *et al.*, 2010). Finally, the incubation of DDX1 with all the TFIIH subunits resulted in a strong interaction of DDX1 with the core-TFIIH and with XPD but a very faint interaction with the CAK sub-complex (lane 9). This result suggests that DDX1 does not only bind the CAK, but it also interacts with the XPD and core-TFIIH. Moreover, the strength of DDX1-CAK interaction is influenced by the presence/absence of XPD and the core-TFIIH.

Because of the strong interaction between DDX1 and XPD, we searched for the DDX1-binding domains within the XPD protein. Thus, we performed *in vitro* DDX1 interaction assays upon incubation of the protein with various truncated form of the recombinant XPD. Specifically, we used an XPD protein lacking the ARCH domain, which is known to be fundamental for the CAK binding. We used the peptide corresponding to the ARCH domain alone, the N-terminal portion containing the first 245 amino acids of the protein or C- terminal fragment spanning from amino acids 444 and 762 of XPD. A parallel reaction was performed incubating DDX1 with the wild type (WT) XPD protein used as positive control.

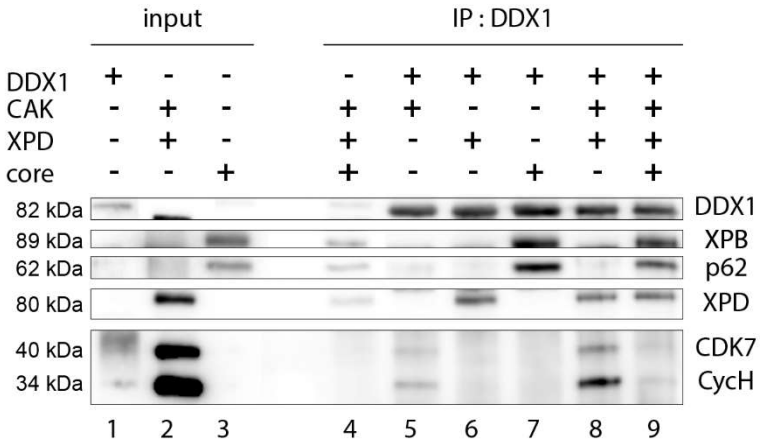


Figure 18. DDX1 *in vitro* interaction with TFIIH subunits. DDX1 was immunoprecipitated and incubated (+) with core-TFIIH, XPD and/or CAK. Immunoprecipitated proteins were analyzed by Western blotting using antibodies against different subunit of TFIIH (XPB, p62, XPD, CDK7 and Cyc H) and DDX1. The input lanes represent 10% of the total protein amount used in each immunoprecipitation reaction.

We found that the C-terminal domain of XPD is sufficient for the binding to DDX1, whereas both the N- portion and the ARCH domain were dispensable (Figure 19A).

Next, we investigated whether *XPD* mutations associated to either XP or TTD phenotype may impair the DDX1-XPD interaction. Thus, we selected the two most frequent TTD- and XP- specific alterations mapping at the C-terminal portion of the protein (R722W and R683W, respectively) or the TTD-specific alteration (R112H) located in the N-terminal portion of XPD. The DDX1-binding affinity of mutated XPD was investigated by incubating the proteins with increasing salts concentration (150mM and 300mM KCl). The analysis revealed that no one of the selected *XPD* alterations seem to affect the DDX1 binding (Figure 19B).

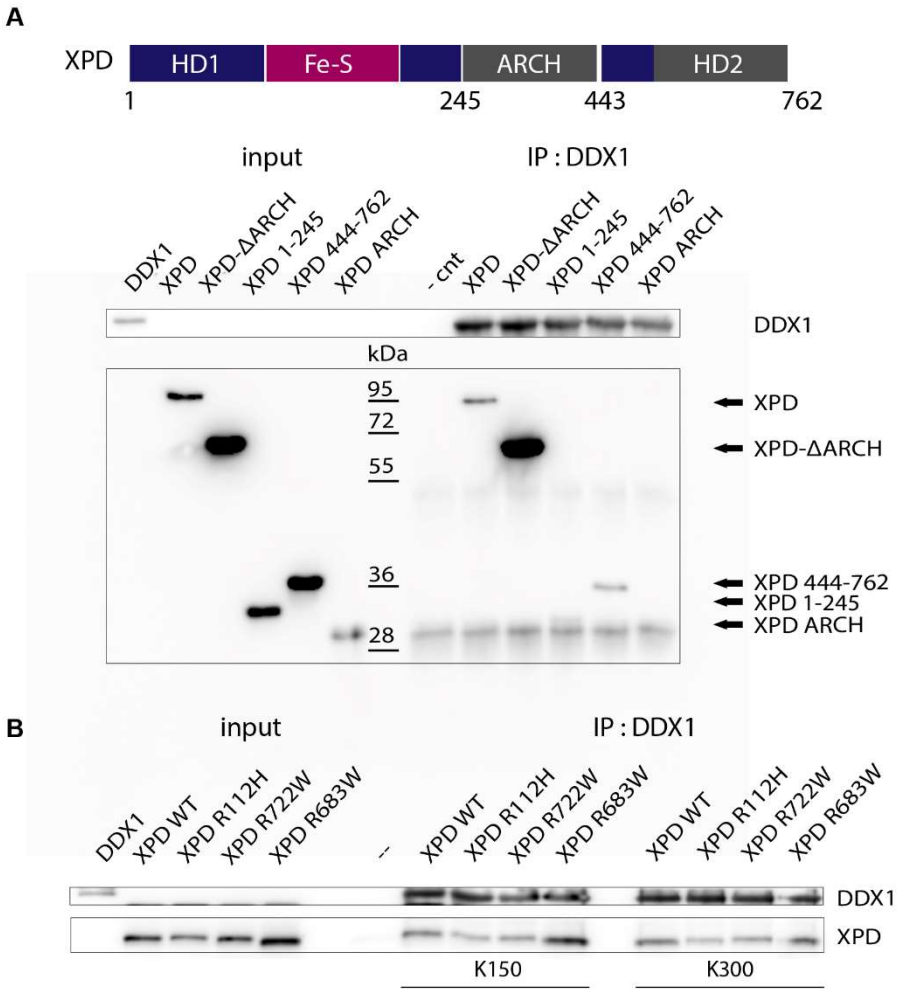


Figure 19. DDX1 *in vitro* interaction with XPD subunit. (A) Schematic representation of XPD structure and its functional domains. In the lower part of the box, DDX1 was immunoprecipitated and incubated (+) with WT or truncated form of XPD as indicated. Immunoprecipitated proteins were analyzed by Western blotting using antibodies against XPD and DDX1. The input lanes represent 10% of the total protein amount used in each immunoprecipitation. **(B)** DDX1 was immunoprecipitated and incubated with WT or mutated form of XPD as indicated. Immunoprecipitated proteins were analyzed by Western blotting using antibodies against XPD and DDX1. The input lanes represent 10% of the total protein amount used in each immunoprecipitation. K150 and K300 indicate the KCl salt concentrations used during washing steps.

4.5 Relevance of DDX1-CDK7 interaction

Based on the above discoveries that DDX1 interacts with the CAK alone or in association with XPD and/or core-TFIIH, we questioned whether DDX1 could be a target of CDK7 kinase activity as part of the free-CAK or holo-TFIIH. Thus, we performed *in vitro* kinase assays using the indicated recombinant proteins and radiolabeled ATP (Figure 20). We observed that CDK7 phosphorylates DDX1 protein when it is part of the free-CAK sub-complex whereas the presence of the entire TFIIH complex does not allow the DDX1 phosphorylation by CDK7. Similarly, the pre-incubation of the CAK with a specific CDK7 inhibitor, THZ1, efficiently prevented the phosphorylation event. These results indicate that DDX1 is a substrate of free-CAK and not of the entire TFIIH.

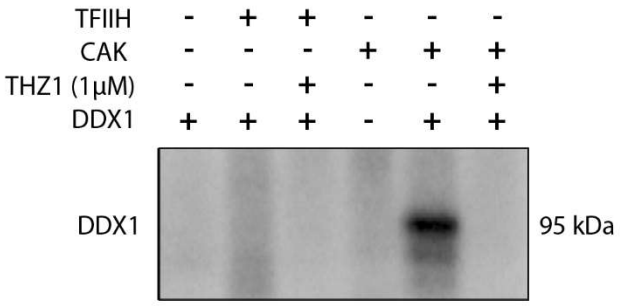


Figure 20. DDX1 phosphorylation by CAK sub-complex. DDX1 was incubated with TFIIH or CAK sub-complex as indicated, in presence of [32P] ATP and CDK7 inhibitor THZ1 (+, when indicated). Autoradiography of the incubated fractions are shown.

5 Conclusions and perspectives

TFIIH is a transcription/DNA repair factor composed of the core-TFIIH and the CAK sub-complexes, bridged together by the XPD subunit. As a multi-subunit protein complex, TFIIH is involved in different biological processes, such as transcription initiation, gene expression regulation, nucleotide excision repair (NER) and cell cycle modulation. The CDK7 kinase subunit of the CAK sub-complex phosphorylates different substrates: as subunit of the TFIIH complex, CDK7 phosphorylates the RNA polymerase (RNApol) II and activates transcription whereas, as part of the free-CAK, it targets specific Cyclin-Dependent Kinases (CDKs) regulating cell cycle progression. In this context, the bridging factor XPD plays a key role in the association/dissociation between the CAK and the core-TFIIH, thus modulating the different TFIIH functionalities. Mutations in the human *ERCC2/XPD* gene cause distinct clinical entities, including xeroderma pigmentosum (XP) and trichothiodystrophy (TTD). Whereas XP is characterized by increased cancer predisposition, TTD patients develop cutaneous photosensitivity with normal skin cancer risk, associated to other clinical features including hair abnormalities, physical and mental retardation and signs of premature aging. It has been hypothesized that mutations in *XPD* could differentially interfere with the various cellular functions of TFIIH leading to distinct phenotypes. In particular, mutations responsible for XP likely affect the repair activity of TFIIH, whereas those typical of TTD also impair transcription. Nevertheless, this issue is still an open question.

To understand why mutations in the same gene give rise to clinical entities with opposite skin cancer predisposition, we investigated the cellular distribution of the core-TFIIH and CAK sub-complexes and their association to the chromatin in mitotic and interphase cells. The study revealed that *XPD* mutations alter the relative distribution of chromatin-bound and soluble CAK. In TTD primary and immortalized fibroblasts, the CAK was mostly present in the soluble fraction (Figures

11 and 12). Since *XPD* mutations typical of TTD cause a degradation of all TFIIH subunits, without affecting the amount of the free-CAK, it is possible that in our analysis we detected mainly the distribution of free-CAK. In XP cells, the distribution of CAK and core-TFIIH appeared different between primary and immortalized fibroblasts. XP6BE cells displayed a TFIIH subunits distribution similar to TTD2GL (Figure 11) whereas XP primary fibroblasts showed an intermediate situation between control and TTD fibroblasts (Figure 12). Since XP6BE cells contained reduced TFIIH content (Figure 16A), the higher CAK solubility may also reflect the distribution of the stable free-CAK as observed in TTD2GL. Differently, XP primary fibroblasts contained normal TFIIH amounts (Figure 16B), hence the reduction of the chromatin-bound CAK paralleled by an increased level of soluble CAK may revealed the detachment of the CAK from the chromatin and likely from the core-TFIIH that, conversely, maintains its association to the DNA. Overall, we cannot exclude that both mechanisms contribute to the different CAK distribution in XP-D cells.

It has been reported that the mutation site on *XPD* could differentially interfere with TFIIH multiple functions and ultimately be relevant to explain the molecular and cellular defects observed in patients. In particular, most of XP mutations map along the ATP-binding pocket and DNA-binding channel, impairing the helicase activity of XPD protein essential for NER. On the contrary, TTD-specific alterations map in all four protein domains and are predicted to cause framework defects, impacting on TFIIH integrity (Fan *et al.*, 2008). These observations can explain the reduced TFIIH levels specifically observed in TTD cells (Vermeulen *et al.*, 2000; Botta *et al.*, 2002). It is plausible that each individual mutation differentially affects the XPD interaction with the other TFIIH subunits and causes specific alterations of TFIIH functionality (Lehmann, 2008). Our analysis reveals that *XPD* mutations may destabilize CAK association with the DNA as well as interaction with core-TFIIH.

Since the substrate specificity of CDK7 kinase subunit of the CAK sub-complex depends also on the CAK association/dissociation from core-

TFIIH, our findings suggest possible alterations of CAK-dependent signaling pathways in XP-D patient fibroblasts. Thus, we searched for CDK7 interacting proteins either associated or not associated to the chromatin. The identification of proteins interacting with CDK7 is a promising strategy to unravel the still unidentified targets of CDK7 kinase, and thus novel partners that could mediate the various functionality of the CAK or of the holo-TFIIH. Upon a complete chromatin digestion to exclude the involvement of the DNA in mediating protein-protein interactions, a mass spectrometry analysis on CDK7 co-immunoprecipitated proteins allowed the identification of CAK interactors in control, TTD and XP fibroblasts. The CDK7 interaction pattern appeared similar among the different cell lines, meaning that the majority of interactions were not affected by an altered soluble/chromatin-bound CAK ratio. This finding agrees with the viability of these specific mutation-bearing patients. Despite most of the CDK7 interactions are retained in *XPD* cells, we achieved to identify few proteins that differently associated with CDK7 in TTD or XP fibroblasts compared to control cells. Most of these interactors are involved in different steps of the RNA metabolism including transcription, pre-mRNA processing, RNA export, translation initiation and RNA degradation. Among them, we found the DEAD/box RNA helicase 1 (DDX1), which seems to display a different CDK7 binding affinity in control compared to *XPD* mutated fibroblasts. As the others DEAD box genes, *DDX1* encode for a RNA helicase containing the D(asp)-E(glu)-A(ala)-D(asp) signature domain involved in ATP binding/hydrolysis. DDX1 has been proposed to play multiple functions in RNA metabolisms (Bléoo *et al.*, 2001), tRNA splicing (Popow *et al.*, 2014), microRNA maturation (Han *et al.*, 2014), repair of transcriptionally active regions of the genome (Li *et al.*, 2008) and G-quadruplex resolution (Ribeiro de Almeida *et al.*, 2018). *DDX1* has been shown to be upregulated in various tumors (Godbout *et al.*, 2007; Tanaka *et al.*, 2009; Germain *et al.*, 2011) and implicated in mouse erly embryonic development (Germain *et al.*, 2015; Hildebrandt *et al.*, 2015), suggesting a role for DDX1 in cell proliferation. In addition to the nine conserved amino acid motifs, DDX1 also contains a SPRY domain

suggested to play a role in either protein-protein or protein-RNA interactions. This domain has been identified in more than 70 proteins, including the heterogeneous nuclear ribonucleoprotein (hnRNP) U. The SPRY domain of hnRNP U mediates the binding of TFIIH to RNApol II holoenzyme (Kim and Nikodem, 1999). Thus, in future studies it will be interesting to understand whether also the SPRY domain of DDX1 is implicated in TFIIH binding. By performing mutagenesis followed by protein purification, it will be possible to produce various DDX1 recombinant proteins lacking specific domains/motifs. The purified proteins will be tasted *in vitro* for their capacity to interact with various TFIIH subunits and sub-complexes. By so doing, it will be possible to identify the DDX1 portion mediating the association with TFIIH complex.

We have shown that DDX1 not only interacts with the CAK sub-complex, but also binds XPD and the core-TFIIH (Figure 18). The last 320 amino acids of XPD are sufficient for the DDX1 binding, whereas the N-terminal region and the ARCH domain, the latter known to mediate the interaction with the MAT1 subunit of the CAK, are both dispensable for this interaction (Figure 19A). Surprisingly, the presence of TTD- or XP- specific mutations at the C-terminus of XPD (R722W and R683W, respectively) of XPD seems not to affect the binding to DDX1 *in vitro* (Figure 19B). It is possible that the *in vitro* experimental conditions do not allow visualizing the effects of *XPD* point mutations on this interaction but it is also conceivable that the mutated XPD protein still binds DDX1 while it losses affinity for the CAK sub-complex, as demonstrated by the fractionation experiments (Figure 11 and 12). Notably, the presence of wild type XPD in *in vitro* assays appears to stabilize the interaction between DDX1 and the CAK indicating that the TFIIH-DDX1 interaction is strictly mediated by at least three players: the CAK, DDX1 and XPD.

By co-immunoprecipitation experiments we observed that the reduced levels of CAK in the chromatin-enriched fraction of TTD primary fibroblasts are sufficient to pull down the same amount of DDX1 as in control or XP cells (Figure 15). This implies that, despite the TFIIH

instability characterizing TTD cells, the amount of CAK engaged to DDX1 is not affected in TTD. A possible explanation to this finding is that DDX1 binds the free-CAK complex. Indeed, the same experiment performed in immortalized cells that contain more free-CAK than primary fibroblasts, revealed that the amount of CDK7-co-immunoprecipitated DDX1 is higher in TTD cells compared to both control and XP (Figure 14). Nevertheless, also the control and XP immortalized fibroblasts contain increased levels of free-CAK (50% in MRC5 and 80% in XP6BE, Figure 11). Therefore, we conclude that the DDX1-CAK interaction is altered in TTD cells.

Finally, by *in vitro* kinase assays we found that DDX1 recombinant protein is phosphorylated by the CAK but not by the holo-TFIIH (Figure 20). In future studies, it will be relevant to analyze the biological implication of DDX1 phosphorylation by the free-CAK as well as to understand the effect of XPD subunit on this phosphorylation event. DDX1 phosphorylation by the CAK may promote or repress DDX1 helicase activity, modulate its substrate specificity or, alternatively, could results in a feedback loop altering CAK functionality or substrate specificity. FRET-based analysis and unwinding assays on phosphorylated and unphosphorylated DDX1 will clarify whether the CAK-mediated phosphorylation may alter DDX1 helicase activity towards various double strand substrates (DNA or RNA double strands and DNA-RNA hybrids). Similarly, it was shown that another member of the DEAD box protein family, DDX3, is phosphorylated by CDK1-Cyc B complex during mitosis and this phosphorylation alter its activity (Sekiguchi *et al.*, 2007). The effect of XPD in this scenario will also be investigate.

Accumulating evidence indicates that deregulated RNA-DNA hybrids formation during transcription (R-loop), can result in the accumulation of DNA damage and genome instability. Improper processing of R-loops by transcription-coupled NER also results in DNA damage (Sollier *et al.*, 2014). Future studies will be relevant to understand whether the TFIIH-DDX1 interaction is somehow implicated in solving the R-loop structures associated to transcription and/or DNA repair. This

possibility is supported by the observation that XPB and XPD helicases target the G-quadruple structures blocking the progression of RNApol II (Gray *et al.*, 2014). Moreover, we will compare the CDK7 *in vitro* substrate affinity towards CDKs or RNApol II CTD *in vitro* in the presence or absence of DDX1 protein. DDX3 was shown to be a regulatory factor of the serine-threonine kinase CK1 since their interaction stimulates CK1 kinase activity and promotes the phosphorylation of CK1 target proteins (Cruciat *et al.*, 2013).

In summary, this study demonstrates that *XPD* mutations cause an imbalance between free- and chromatin-bound CAK. The presence of more soluble CAK, not associated to the chromatin and thus to the core-TFIIH could affect CAK functionality leading to activation and/or repression of specific cellular pathway that can ultimately contribute to the TTD or XP clinical phonotype. In addition, we identified the helicase DDX1 as a novel CDK7 interactor, showing a stronger or delayed association with the chromatin-bound CDK7 in TTD fibroblasts. It will be interesting to further investigate the biological relevance of DDX1-CDK7 interaction, to identify the residues on DDX1 protein phosphorylated by the CAK, to analyze the consequence of XPD presence on CDK7 kinase activity towards DDX1 and to evaluate the effect of DDX1 phosphorylation on its nuclease and helicase activities. Moreover, we intend to isolate *in viv*o the DDX1-CAK complex and analyze the presence/absence of XPD as well as core-TFFIH to better understand the composition of these newly identified protein complex.

6 References

Abdulrahman, W., Iltis, I., Radu, L., Braun, C., Maglott-Roth, A., Giraudon, C., Egly, J.-M. and Poterszman, A. (2013) 'ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities', *Proceedings of the National Academy of Sciences*, 110, pp. 633–642. doi: 10.1073/pnas.1213981110.

Adamczewski, J. P., Rossignol, M., Tassan, J. P., Nigg, E. A., Moncollin, V. and Egly, J. M. (1996) 'MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH.', *The EMBO Journal*, 15(8), pp. 1877–1884. doi: 10.1002/j.1460-2075.1996.tb00538.x.

Akoulitchev, S., Chuikov, S. and Reinberg, D. (2000) 'TFIIH is negatively regulated by cdk8-containing mediator complexes', *Nature*, 407(6800), pp. 102–106. doi: 10.1038/35024111.

Akoulitchev, S. and Reinberg, D. (1998) 'The molecular mechanism of mitotic inhibition of TFIIH is mediated by phosphorylation of CDK7', *Genes and Development*, 12(22), pp. 3541–3550. doi: 10.1101/gad.12.22.3541.

Alekseev, S. and Coin, F. (2015) 'Orchestral maneuvers at the damaged sites in nucleotide excision repair', *Cellular and Molecular Life Sciences*, 72(11), pp. 2177–2186. doi: 10.1007/s00018-015-1859-5.

Andersen, G., Busso, D., Poterszman, A., Hwang, J. R., Wurtz, J. M., Ripp, R., Thierry, J. C., Egly, J. M. and Moras, D. (1997) 'The structure of cyclin H: Common mode of kinase activation and specific features', *EMBO Journal*, 16(5), pp. 958–967. doi: 10.1093/emboj/16.5.958.

Apostolou, Z., Chatzinikolaou, G., Stratigi, K. and Garinis, G. A. (2019) 'Nucleotide Excision Repair and Transcription-Associated Genome Instability', *BioEssays : news and reviews in molecular, cellular and developmental biology*, 41(4), p. e1800201. doi: 10.1002/bies.201800201.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y. and Hanaoka, F. (2001) 'Centrosome Protein Centrin

2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair', *Journal of Biological Chemistry*, 276(22), pp. 18665–18672. doi: 10.1074/jbc.M100855200.

Arellano, M. and Moreno, S. (1997) 'Regulation of CDK/cyclin complexes during the cell cycle', *The International Journal of Biochemistry & Cell Biology*, 29(4), pp. 559–573. doi: 10.1016/S1357-2725(96)00178-1.

Armache, K. J., Kettenberger, H. and Cramer, P. (2003) 'Architecture of initiation-competent 12-subunit RNA polymerase II', *Proceedings of the National Academy of Sciences of the United States of America*, 100(12), pp. 6964–6968. doi: 10.1073/pnas.1030608100.

Arseni, L., Lanzafame, M., Compe, E., Fortugno, P., Afonso-Barroso, A., Peverali, F. A., Lehmann, A. R., Zambruno, G., Egly, J. M., Stefanini, M. and Orioli, D. (2015) 'TFIIH-dependent MMP-1 overexpression in trichothiodystrophy leads to extracellular matrix alterations in patient skin', *Proceedings of the National Academy of Sciences of the United States of America*, 112(5), pp. 1499–1504. doi: 10.1073/pnas.1416181112.

Bastien, J., Adam-Stitah, S., Riedl, T., Egly, J.-M., Chambon, P. and Rochette-Egly, C. (2000) 'TFIIH Interacts with the Retinoic Acid Receptor γ and Phosphorylates Its AF-1-activating Domain through cdk7', *Journal of Biological Chemistry*, 275(29), pp. 21896–21904. doi: 10.1074/jbc.M001985200.

Bedez, F., Linard, B., Brochet, X., Ripp, R., Thompson, J. D., Moras, D., Lecompte, O. and Poch, O. (2013) 'Functional insights into the core-TFIIH from a comparative survey', *Genomics*. Elsevier Inc., 101(3), pp. 178–186. doi: 10.1016/j.ygeno.2012.11.003.

Bisteau, X., Paternot, S., Colleoni, B., Ecker, K., Coulonval, K., De Groote, P., Declercq, W., Hengst, L. and Roger, P. P. (2013) 'CDK4 T172 Phosphorylation Is Central in a CDK7-Dependent Bidirectional CDK4/CDK2 Interplay Mediated by p21 Phosphorylation at the Restriction Point', *PLoS Genetics*. Edited by B. E. Clurman, 9(5), p. e1003546. doi: 10.1371/journal.pgen.1003546.

Bléoo, S., Sun, X., Hendzel, M. J., Rowe, J. M., Packer, M. and Godbout, R. (2001) 'Association of human DEAD box protein DDX1 with a cleavage stimulation factor involved in 3′-end processing of pre-mRNA', *Molecular Biology of the Cell*, 12(10), pp. 3046–3059. doi: 10.1091/mbc.12.10.3046.

de Boer, J., de Wit, J., van Steeg, H., Berg, R. J. W., Morreau, H., Visser, P., Lehmann, A. R., Duran, M., Hoeijmakers, J. H. J. and Weeda, G. (1998) 'A Mouse Model for the Basal Transcription/DNA Repair Syndrome Trichothiodystrophy', *Molecular Cell*, 1(7), pp. 981–990. doi: 10.1016/S1097-2765(00)80098-2.

Botta, E., Nardo, T., Broughton, B. C., Marinoni, S., Lehmann, A. R. and Stefanini, M. (1998) 'Analysis of Mutations in the XPD Gene in Italian Patients with Trichothiodystrophy: Site of Mutation Correlates with Repair Deficiency, but Gene Dosage Appears to Determine Clinical Severity', *The American Journal of Human Genetics*, 63(4), pp. 1036–1048. doi: 10.1086/302063.

Botta, E., Nardo, T., Lehmann, A., Egly, J., Pedrini, A. and Stefanini, M. (2002) 'Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy', *Human Molecular Genetics*, 11(23), pp. 2919–2928. doi: 10.1093/hmg/11.23.2919.

Bowden, N., Beveridge, N., Ashton, K., Baines, K. and Scott, R. (2015) 'Understanding Xeroderma Pigmentosum Complementation Groups Using Gene Expression Profiling after UV-Light Exposure', *International Journal of Molecular Sciences*, 16(7), pp. 15985–15996. doi: 10.3390/ijms160715985.

Boyle, J., Ueda, T., Oh, K.-S., Imoto, K., Tamura, D., Jagdeo, J., Khan, S. G., Nadem, C., DiGiovanna, J. J. and Kraemer, K. H. (2008) 'Persistence of repair proteins at unrepaired DNA damage distinguishes diseases with ERCC2 (XPD) mutations: cancer-prone xeroderma pigmentosum vs. non-cancer-prone trichothiodystrophy', *Human Mutation*, 29(10), pp. 1194–1208. doi: 10.1002/humu.20768.

Bradford, P. *et al.* (2011) 'Cancer and Neurologic Degeneration in Xeroderma', *J Med Gene*, 48(3), pp. 168–176. doi: 10.1136/jmg.2010.083022.

Broughton, B. C. (2001) 'Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene', *Human Molecular Genetics*, 10(22), pp. 2539–2547. doi: 10.1093/hmg/10.22.2539.

Broughton, B. C., Steingrimsdottir, H., Weber, C. A. and Lehmann, A. R. (1994) 'Mutations in the xeroderma pigmentosum group D DNA repair/transcription gene in patients with trichothiodystrophy', *Nature Genetics*, 7(2), pp. 189–194. doi: 10.1038/ng0694-189.

Busso, D., Keriel, A., Sandrock, B., Poterszman, A., Gileadi, O. and Egly, J. M. (2000) 'Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities', *Journal of Biological Chemistry*, 275(30), pp. 22815–22823. doi: 10.1074/jbc.M002578200.

Calmels, N. *et al.* (2018) 'Functional and clinical relevance of novel mutations in a large cohort of patients with Cockayne syndrome', *Journal of medical genetics*, 55(5), pp. 329–343. doi: 10.1136/jmedgenet-2017-104877.

Chen, H. C., Lin, W. C., Tsay, Y. G., Lee, S. C. and Chang, C. J. (2002) 'An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K', *Journal of Biological Chemistry*, 277(43), pp. 40403–40409. doi: 10.1074/jbc.M206981200.

Chen, J., Larochelle, S., Li, X. and Suter, B. (2003) 'Xpd/Ercc2 regulates CAK activity and mitotic progression', *Nature*, 424(6945), pp. 228–232. doi: 10.1038/nature01746.

Chu, I., Sun, J., Arnaout, A., Kahn, H., Hanna, W., Narod, S., Sun, P., Tan, C.-K., Hengst, L. and Slingerland, J. (2007) 'p27 Phosphorylation by Src Regulates Inhibition of Cyclin E-Cdk2', *Cell*, 128(2), pp. 281–294. doi: 10.1016/j.cell.2006.11.049.

Chymkowitch, P., Le May, N., Charneau, P., Compe, E. and Egly, J. M. (2011) 'The phosphorylation of the androgen receptor by TFIIH directs the ubiquitin/proteasome process', *EMBO Journal*. Nature Publishing Group, 30(3), pp. 468–479. doi: 10.1038/emboj.2010.337.

Ciccia, A. and Elledge, S. J. (2010) 'The DNA Damage Response: Making It

Safe to Play with Knives', *Molecular Cell*. Elsevier Inc., 40(2), pp. 179–204. doi: 10.1016/j.molcel.2010.09.019.

Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R. E., Hoeijmakers, J. H. J. and Vermeulen, W. (2000) 'ATP-Dependent Chromatin Remodeling by the Cockayne Syndrome B DNA Repair-Transcription-Coupling Factor', *Molecular and Cellular Biology*, 20(20), pp. 7643–7653. doi: 10.1128/MCB.20.20.7643-7653.2000.

Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W. and Egly, J. M. (2004) 'Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity', *EMBO Journal*, 23(24), pp. 4835–4846. doi: 10.1038/sj.emboj.7600480.

Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M. and Egly, J. M. (1998) 'Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH', *Nature Genetics*, 20(2), pp. 184–188. doi: 10.1038/2491.

Coin, F., Oksenych, V. and Egly, J. M. (2007) 'Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair', *Molecular Cell*, 26(2), pp. 245–256. doi: 10.1016/j.molcel.2007.03.009.

Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C. and Egly, J. M. (2008) 'Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH', *Molecular Cell*, 31(1), pp. 9–20. doi: 10.1016/j.molcel.2008.04.024.

Coin, F., De Santis, L. P., Nardo, T., Zlobinskaya, O., Stefanini, M. and Egly, J. M. (2006) 'p8/TTD-A as a repair-specific TFIIH subunit', *Molecular Cell*, 21(2), pp. 215–226. doi: 10.1016/j.molcel.2005.10.024.

Compe, E., Drane, P., Laurent, C., Diderich, K., Braun, C., Hoeijmakers, J. H. J. and Egly, J.-M. (2005) 'Dysregulation of the Peroxisome Proliferator-Activated Receptor Target Genes by XPD Mutations', *Molecular and Cellular Biology*, 25(14), pp. 6065–6076. doi: 10.1128/MCB.25.14.6065-6076.2005.

Compe, E. and Egly, J.-M. (2016) 'Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond', *Annual Review of*

Biochemistry, 85(1), pp. 265–290. doi: 10.1146/annurev-biochem-060815-014857.

Compe, E. and Egly, J. M. (2012) 'TFIIH: When transcription met DNA repair', *Nature Reviews Molecular Cell Biology*, 13(6), pp. 343–354. doi: 10.1038/nrm3350.

Compe, E., Malerba, M., Soler, L., Marescaux, J., Borrelli, E. and Egly, J.-M. (2007) 'Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH', *Nature Neuroscience*, 10(11), pp. 1414–1422. doi: 10.1038/nn1990.

Conaway, J. W. and Conaway, R. C. (1989) 'A multisubunit transcription factor essential for accurate initiation by RNA polymerase II.', *The Journal of biological chemistry*, 264(4), pp. 2357–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2914911.

Cruciat, C. M., Dolde, C., De Groot, R. E. A., Ohkawara, B., Reinhard, C., Korswagen, H. C. and Niehrs, C. (2013) 'RNA helicase DDX3 is a regulatory subunit of casein kinase 1 in Wnt-β-catenin signaling', *Science*, 339(6126), pp. 1436–1441. doi: 10.1126/science.1231499.

Cruz-Becerra, G., Juárez, M., Valadez-Graham, V. and Zurita, M. (2016) 'Analysis of Drosophila p8 and p52 mutants reveals distinct roles for the maintenance of TFIIH stability and male germ cell differentiation', *Open Biology*, 6(10), p. 160222. doi: 10.1098/rsob.160222.

Devault, A., Martinez, A. M., Fesquet, D., Labbé, J. C., Morin, N., Tassan, J. P., Nigg, E. A., Cavadore, J. C. and Dorée, M. (1995) 'MAT1 ('menage à trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK.', *The EMBO journal*, 14(20), pp. 5027–36.

Ding, H., Wen, Z. and Sun, G. (2018) 'Silencing of Xeroderma Pigmentosum Group D Gene Promotes Hepatoma Cell Growth by Reducing P53 Expression', *Medical Science Monitor*, 24, pp. 8015–8021. doi: 10.12659/MSM.910944.

Donczew, R. and Hahn, S. (2018) 'crossm ABSTRACT', *Molecular and Cellular*

Biology, pp. 1–14.

Drouin, S., Laramée, L., Jacques, P. É., Forest, A., Bergeron, M. and Robert, F. (2010) 'DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes', *PLoS Genetics*, 6(10), pp. 1–12. doi: 10.1371/journal.pgen.1001173.

Dvir, A., Tan, S., Conaway, J. W. and Conaway, R. C. (1997) 'Promoter escape by RNA polymerase II. Formation of an escape-competent transcriptional intermediate is a prerequisite for exit of polymerase from the promoter.', *The Journal of biological chemistry*, 272(45), pp. 28175–8. doi: 10.1074/jbc.272.45.28175.

Ebmeier, C. C., Erickson, B., Allen, B. L., Allen, M. A., Kim, H., Fong, N., Jacobsen, J. R., Liang, K., Shilatifard, A., Dowell, R. D., Old, W. M., Bentley, D. L. and Taatjes, D. J. (2017a) 'Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications', *Cell Reports*, 20(5), pp. 1173–1186. doi: 10.1097/CCM.0b013e31823da96d.Hydrogen.

Ebmeier, C. C., Erickson, B., Allen, B. L., Allen, M. A., Kim, H., Fong, N., Jacobsen, J. R., Liang, K., Shilatifard, A., Dowell, R. D., Old, W. M., Bentley, D. L. and Taatjes, D. J. (2017b) 'Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications', *Cell Reports*. ElsevierCompany., 20(5), pp. 1173–1186. doi: 10.1016/j.celrep.2017.07.021.

Egly, J. M. and Coin, F. (2011) 'A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor', *DNA Repair*. Elsevier B.V., 10(7), pp. 714–721. doi: 10.1016/j.dnarep.2011.04.021.

Eick, D. and Geyer, M. (2013) 'The RNA polymerase II carboxy-terminal domain (CTD) code', *Chemical Reviews*, 113(11), pp. 8456–8490. doi: 10.1021/cr400071f.

el-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., Lin, D., Mercer, W. E., Kinzler, K. W. and Vogelstein, B. (1993) 'WAF1, a potential mediator of p53 tumor suppression.', *Cell*, 75(4), pp. 817–25. doi: 10.1016/0092-8674(93)90500-p.

Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K. and Tainer, J. A. (2008) 'XPD Helicase Structures and Activities: Insights into the Cancer and Aging Phenotypes from XPD Mutations', *Cell*, 133(5), pp. 789–800. doi: 10.1016/j.cell.2008.04.030.

Fang, J., Acheampong, E., Dave, R., Wang, F., Mukhtar, M. and Pomerantz, R. J. (2005) 'The RNA helicase DDX1 is involved in restricted HIV-1 Rev function in human astrocytes', *Virology*, 336(2), pp. 299–307. doi: 10.1016/j.virol.2005.03.017.

Fang, J., Kubota, S., Yang, B., Zhou, N., Zhang, H., Godbout, R. and Pomerantz, R. J. (2004) 'A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev', *Virology*, 330(2), pp. 471–480. doi: 10.1016/j.virol.2004.09.039.

Fassihi, H. *et al.* (2016) 'Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect', *Proceedings of the National Academy of Sciences*, 113(9), pp. E1236–E1245. doi: 10.1073/pnas.1519444113.

Feaver, W. J., Gileadi, O., Li, Y. and Kornberg, R. D. (1991) 'CTD kinase associated with yeast RNA polymerase II initiation factor b.', *Cell*, 67(6), pp. 1223–30. doi: 10.1016/0092-8674(91)90298-d.

Ferri, D., Orioli, D. and Botta, E. (2019) 'Heterogeneity and overlaps in nucleotide excision repair disorders', *Clinical Genetics*, p. cge.13545. doi: 10.1111/cge.13545.

Fesquet, D., Labbé, J. C., Derancourt, J., Capony, J. P., Galas, S., Girard, F., Lorca, T., Shuttleworth, J., Dorée, M. and Cavadore, J. C. (1993) 'The MO15 gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues.', *The EMBO journal*, 12(8), pp. 3111–21.

Fishburn, J., Tomko, E., Galburt, E. and Hahn, S. (2015) 'Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation', *Proceedings of the National Academy of Sciences*, 112(13), pp. 3961–3966. doi: 10.1073/pnas.1417709112.

Fisher, R. P. (2005) 'Secrets of a double agent: CDK7 in cell-cycle control and transcription', *Journal of Cell Science*, 118(22), pp. 5171–5180. doi: 10.1242/jcs.02718.

Fisher, R. P., Jin, P., Chamberlin, H. M. and Morgan, D. O. (1995) 'Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase.', *Cell*, 83(1), pp. 47–57. doi: 10.1016/0092-8674(95)90233-3.

Fousteri, M., Vermeulen, W., van Zeeland, A. A. and Mullenders, L. H. F. (2006) 'Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo', *Molecular Cell*, 23(4), pp. 471–482. doi: 10.1016/j.molcel.2006.06.029.

Fregoso, M., Laine, J.-P., Aguilar-Fuentes, J., Mocquet, V., Reynaud, E., Coin, F., Egly, J.-M. and Zurita, M. (2007) 'DNA Repair and Transcriptional Deficiencies Caused by Mutations in the Drosophila p52 Subunit of TFIIH Generate Developmental Defects and Chromosome Fragility', *Molecular and Cellular Biology*, 27(10), pp. 3640–3650. doi: 10.1128/MCB.00030-07.

Ganuza, M., Sáiz-Ladera, C., Cañamero, M., Gómez, G., Schneider, R., Blasco, M. A., Pisano, D., Paramio, J. M., Santamaría, D. and Barbacid, M. (2012) 'Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion', *EMBO Journal*, 31(11), pp. 2498–2510. doi: 10.1038/emboj.2012.94.

Garcia-Moreno, H., Fassihi, H., Sarkany, R. P. E., Phukan, J., Warner, T., Lehmann, A. R. and Giunti, P. (2018) 'Xeroderma pigmentosum is a definite cause of Huntington's disease-like syndrome', *Annals of Clinical and Translational Neurology*, 5(1), pp. 102–108. doi: 10.1002/acn3.511.

Garrett, S., Barton, W. A., Knights, R., Jin, P., Morgan, D. O. and Fisher, R. P. (2001) 'Reciprocal activation by cyclin-dependent kinases 2 and 7 is directed by substrate specificity determinants outside the T loop.', *Molecular and cellular biology*, 21(1), pp. 88–99. doi: 10.1128/MCB.21.1.88-99.2001.

Gerard, M., Fischer, L., Moncollin, V., Chipoulet, J. M., Chambon, P. and

Egly, J. M. (1991) 'Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2.', *The Journal of biological chemistry*, 266(31), pp. 20940–5.

Germain, D. R., Graham, K., Glubrecht, D. D., Hugh, J. C., MacKey, J. R. and Godbout, R. (2011) 'DEAD box 1: A novel and independent prognostic marker for early recurrence in breast cancer', *Breast Cancer Research and Treatment*, 127(1), pp. 53–63. doi: 10.1007/s10549-010-0943-7.

Germain, D. R., Li, L., Hildebrandt, M. R., Simmonds, A. J., Hughes, S. C. and Godbout, R. (2015) 'Loss of the Drosophila melanogaster DEAD box protein Ddx1 leads to reduced size and aberrant gametogenesis', *Developmental Biology*. Elsevier, 407(2), pp. 232–245. doi: 10.1016/j.ydbio.2015.09.012.

Giglia-Mari, G. *et al.* (2004) 'A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A', *Nature Genetics*, 36(7), pp. 714–719. doi: 10.1038/ng1387.

Girard, F., Strausfeld, U., Fernandez, A. and Lamb, N. J. C. (1991) 'Cyclin a is required for the onset of DNA replication in mammalian fibroblasts', *Cell*, 67(6), pp. 1169–1179. doi: 10.1016/0092-8674(91)90293-8.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R. P. and Bentley, D. L. (2009) 'TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II', *Molecular and Cellular Biology*, 29(20), pp. 5455–5464. doi: 10.1128/mcb.00637-09.

Godbout, R., Li, L., Liu, R.-Z. and Roy, K. (2007) 'Role of DEAD box 1 in retinoblastoma and neuroblastoma', *Future Oncology*, 3(5), pp. 575–587. doi: 10.2217/14796694.3.5.575.

Godbout, R., Packer, M. and Bie, W. (1998) 'Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines', *Journal of Biological Chemistry*, 273(33), pp. 21161–21168. doi: 10.1074/jbc.273.33.21161.

Godbout, R., Packer, M., Katyal, S. and Bléoo, S. (2002) 'Cloning and

expression analysis of the chicken DEAD box gene DDX1', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1574(1), pp. 63–71. doi: 10.1016/S0167-4781(01)00346-3.

Godbout, R. and Squire, J. (1993) 'Amplification of a DEAD box protein gene in retinoblastoma cell lines.', *Proceedings of the National Academy of Sciences*, 90(16), pp. 7578–7582. doi: 10.1073/pnas.90.16.7578.

Gray, L. T., Vallur, A. C., Eddy, J. and Maizels, N. (2014) 'G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD', *Nature Chemical Biology*, 10(4), pp. 313–318. doi: 10.1038/nchembio.1475.

Greber, B. J., Toso, D. B., Fang, J. and Nogales, E. (2019) 'The complete structure of the human TFIIH core complex', *eLife*, 8. doi: 10.7554/eLife.44771.

Greber BJ (2017) 'The cryo-electron microscopy structure of human transcription factor IIH', *Nature*. doi: 10.1097/CCM.0b013e31823da96d.Hydrogen.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J. and Werner, F. (2011) 'The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation', *Molecular Cell*. Elsevier Inc., 43(2), pp. 263–274. doi: 10.1016/j.molcel.2011.05.030.

Groisman, R. (2006) 'CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome', *Genes & Development*, 20(11), pp. 1429–1434. doi: 10.1101/gad.378206.

Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J., Saijo, M., Drapkin, R., Kisselev, A. F., Tanaka, K. and Nakatani, Y. (2003) 'The Ubiquitin Ligase Activity in the DDB2 and CSA Complexes Is Differentially Regulated by the COP9 Signalosome in Response to DNA Damage', *Cell*, 113(3), pp. 357–367. doi: 10.1016/S0092-8674(03)00316-7.

Grünberg, S., Warfield, L. and Hahn, S. (2012) 'Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent

promoter opening.', *Nature structural & molecular biology*, 19(8), pp. 788–96. doi: 10.1038/nsmb.2334.

Guzder, S. N., Sung, P., Bailly, V., Prakash, L. and Prakash, S. (1994) 'RAD25 is a DMA helicase required for DNA repair and RNA polymerase II transcription', *Nature*, 369(6481), pp. 578–581. doi: 10.1038/369578a0.

Han, C., Liu, Y., Wan, G., Choi, H. J., Zhao, L., Ivan, C., He, X., Sood, A. K., Zhang, X. and Lu, X. (2014) 'The RNA-Binding Protein DDX1 Promotes Primary MicroRNA Maturation and Inhibits Ovarian Tumor Progression', *Cell Reports*, 8(5), pp. 1447–1460. doi: 10.1016/j.celrep.2014.07.058.

Hashimoto, S. and Egly, J. M. (2009) 'Trichothiodystrophy view from the molecular basis of DNA repair/transcription factor TFIIH', *Human Molecular Genetics*, 18(R2), pp. 224–230. doi: 10.1093/hmg/ddp390.

He, Y., Fang, J., Taatjes, D. J. and Nogales, E. (2013) 'Structural visualization of key steps in human transcription initiation', *Nature*, 495(7442), pp. 481–486. doi: 10.1038/nature11991.

Hildebrandt, M. R., Germain, D. R., Monckton, E. A., Brun, M. and Godbout, R. (2015) 'Ddx1 knockout results in transgenerational wild-type lethality in mice', *Scientific Reports*, 5, pp. 1–8. doi: 10.1038/srep09829.

Hirai, Y., Kodama, Y., Moriwaki, S.-I., Noda, A., Cullings, H. M., MacPhee, D. G., Kodama, K., Mabuchi, K., Kraemer, K. H., Land, C. E. and Nakamura, N. (2006) 'Heterozygous individuals bearing a founder mutation in the XPA DNA repair gene comprise nearly 1% of the Japanese population', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 601(1–2), pp. 171–178. doi: 10.1016/j.mrfmmm.2006.06.010.

Ho, C. K. and Shuman, S. (1999) 'Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme', *Molecular Cell*, 3(3), pp. 405–411. doi: 10.1016/S1097-2765(00)80468-2.

Hui Ng, H., ois Robert, F., Young, R. A. and Struhl, K. (2003) 'Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity genome-

wide fashion (Kuo et al that is linked to subsequent events of mRNA production including capping, splicing', *Orphanides and Reinberg*, 11, pp. 709–719.

Ito, S., Tan, L. J., Andoh, D., Narita, T., Seki, M., Hirano, Y., Narita, K., Kuraoka, I., Hiraoka, Y. and Tanaka, K. (2010) 'MMXD, a TFIIH-Independent XPD-MMS19 Protein Complex Involved in Chromosome Segregation', *Molecular Cell*, 39(4), pp. 632–640. doi: 10.1016/j.molcel.2010.07.029.

James, M. K., Ray, A., Leznova, D. and Blain, S. W. (2008) 'Differential Modification of p27Kip1 Controls Its Cyclin D-cdk4 Inhibitory Activity', *Molecular and Cellular Biology*, 28(1), pp. 498–510. doi: 10.1128/MCB.02171-06.

Jawhari, A., Lainé, J. P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D. and Egly, J. M. (2002) 'p52 mediates XPB function within the transcription/repair factor TFIIH', *Journal of Biological Chemistry*, 277(35), pp. 31761–31767. doi: 10.1074/jbc.M203792200.

Jeronimo, C. *et al.* (2007) 'Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme', *Molecular Cell*, 27(2), pp. 262–274. doi: 10.1016/j.molcel.2007.06.027.

Jeronimo, C. and Robert, F. (2014) 'Kin28 regulates the transient association of Mediator with core promoters.', *Nature structural & molecular biology*, 21(5), pp. 449–55. doi: 10.1038/nsmb.2810.

Kainov, D. E., Vitorino, M., Cavarelli, J., Poterszman, A. and Egly, J.-M. (2008) 'Structural basis for group A trichothiodystrophy.', *Nature structural & molecular biology*, 15(9), pp. 980–4.

Kellenberger, E., Dominguez, C., Fribourg, S., Wasielewski, E., Moras, D., Poterszman, A., Boelens, R. and Kieffer, B. (2005) 'Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions', *Journal of Biological Chemistry*, 280(21), pp. 20785–20792. doi: 10.1074/jbc.M412999200.

Kellner, J. N. and Meinhart, A. (2015) 'Structure of the SPRY domain of the human RNA helicase DDX1, a putative interaction platform within a DEAD-box protein', *Acta Crystallographica Section:F Structural Biology Communications*. International Union of Crystallography, 71, pp. 1176–1188. doi: 10.1107/S2053230X15013709.

Kelso, T. W. R., Baumgart, K., Eickhoff, J., Albert, T., Antrecht, C., Lemcke, S., Klebl, B. and Meisterernst, M. (2014) 'Cyclin-Dependent Kinase 7 Controls mRNA Synthesis by Affecting Stability of Preinitiation Complexes, Leading to Altered Gene Expression, Cell Cycle Progression, and Survival of Tumor Cells', *Molecular and Cellular Biology*, 34(19), pp. 3675–3688. doi: 10.1128/MCB.00595-14.

Kim, Ebright, R. H. and Reinberg, D. (2000) 'Mechanism of ATP-dependent promoter melting by transcription factor IIH.', *Science (New York, N.Y.)*, 288(5470), pp. 1418–22. doi: 10.1126/science.288.5470.1418.

Kim, J. S., Saint-André, C., Lim, H. S., Hwang, C. S., Egly, J. M. and Cho, Y. (2015) 'Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44', *Journal of Biological Chemistry*, 290(13), pp. 8321–8330. doi: 10.1074/jbc.M115.636514.

Kim, K. K., Chamberlin, H. M., Morgan, D. O. and Kim, S. H. (1996) 'Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase.', *Nature structural biology*, 3(10), pp. 849–55.

Kim, M. K. and Nikodem, V. M. (1999) 'hnRNP U Inhibits Carboxy-Terminal Domain Phosphorylation by TFIIH and Represses RNA Polymerase II Elongation', *Molecular and Cellular Biology*, 19(10), pp. 6833–6844. doi: 10.1128/mcb.19.10.6833.

King, R. W., Jackson, P. K. and Kirschner, M. W. (1994) 'Mitosis in transition', *Cell*, 79(4), pp. 563–571. doi: 10.1016/0092-8674(94)90542-8.

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X. and Murakami, K. S. (2011) 'RNA polymerase and transcription elongation factor Spt4/5 complex structure', *Proceedings of the National Academy of Sciences*, 108(2), pp. 546–550. doi: 10.1073/pnas.1013828108.

Kobayashi, T., Uchiyama, M., Fukuro, S. and Tanaka, K. (2002) 'Mutations in theXPD gene in xeroderma pigmentosum group D cell strains: Confirmation of genotype-phenotype correlation', *American Journal of Medical Genetics*, 110(3), pp. 248–252. doi: 10.1002/ajmg.10465.

Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H. and Cramer, P. (2019) 'Structural basis of TFIIH activation for nucleotide excision repair', *Nature Communications*. Springer US, 10(1), p. 2885. doi: 10.1038/s41467-019-10745-5.

Kolesnikova, O., Radu, L. and Poterszman, A. (2019) 'TFIIH: A multi-subunit complex at the cross-roads of transcription and DNA repair', *Advances in Protein Chemistry and Structural Biology*. Elsevier Inc., 115, pp. 21–67. doi: 10.1016/bs.apcsb.2019.01.003.

Komarnitsky, P. (2000) 'Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription', *Genes & Development*, 14(19), pp. 2452–2460. doi: 10.1101/gad.824700.

Kou, H., Zhou, Y., Gorospe, R. M. C. and Wang, Z. (2008) 'Mms19 protein functions in nucleotide excision repair by sustaining an adequate cellular concentration of the TFIIH component Rad3', *Proceedings of the National Academy of Sciences*, 105(41), pp. 15714–15719. doi: 10.1073/pnas.0710736105.

Kraemer, K. H., Patronas, N. J., Schiffmann, R., Brooks, B. P., Tamura, D. and DiGiovanna, J. J. (2007) 'Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: A complex genotype-phenotype relationship', *Neuroscience*, 145(4), pp. 1388–1396. doi: 10.1016/j.neuroscience.2006.12.020.

Kuper, J., Braun, C., Elias, A., Michels, G., Sauer, F., Schmitt, D. R., Poterszman, A., Egly, J. M. and Kisker, C. (2014) 'In TFIIH, XPD Helicase Is Exclusively Devoted to DNA Repair', *PLoS Biology*, 12(9). doi: 10.1371/journal.pbio.1001954.

Kusakabe, M., Onishi, Y., Tada, H., Kurihara, F., Kusao, K., Furukawa, M., Iwai, S., Yokoi, M., Sakai, W. and Sugasawa, K. (2019) 'Mechanism and regulation of DNA damage recognition in nucleotide excision repair', *Genes*

and Environment. Genes and Environment, 41(1), pp. 1–6. doi: 10.1186/s41021-019-0119-6.

Kuschal, C. *et al.* (2016) 'GTF2E2 Mutations Destabilize the General Transcription Factor Complex TFIIE in Individuals with DNA Repair-Proficient Trichothiodystrophy', *The American Journal of Human Genetics*, 98(4), pp. 627–642. doi: 10.1016/j.ajhg.2016.02.008.

Kwak, H. and Lis, J. T. (2013) 'Control of Transcriptional Elongation', *Annual Review of Genetics*, 47(1), pp. 483–508. doi: 10.1146/annurev-genet-110711-155440.

Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N. J. and Akoulitchev, A. (2002) 'U1 snRNA associates with tfiih and regulates transcriptional initiation', *Nature Structural Biology*, 9(11), pp. 800–805. doi: 10.1038/nsb862.

de la Cruz, J., Kressler, D. and Linder, P. (1999) 'Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families', *Trends in Biochemical Sciences*, 24(5), pp. 192–198. doi: 10.1016/S0968-0004(99)01376-6.

Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J. J., Shokat, K. M., Bentley, D. L. and Fisher, R. P. (2012) 'Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II', *Nature Structural & Molecular Biology*, 19(11), pp. 1108–1115. doi: 10.1038/nsmb.2399.

Larochelle, S., Merrick, K. A., Terret, M.-E., Wohlbold, L., Barboza, N. M., Zhang, C., Shokat, K. M., Jallepalli, P. V. and Fisher, R. P. (2007) 'Requirements for Cdk7 in the Assembly of Cdk1/Cyclin B and Activation of Cdk2 Revealed by Chemical Genetics in Human Cells', *Molecular Cell*, 25(6), pp. 839–850. doi: 10.1016/j.molcel.2007.02.003.

Larochelle, S., Pandur, J., Fisher, R. P., Salz, H. K. and Suter, B. (1998) 'Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity', *Genes and Development*, 12(3), pp. 370–381. doi: 10.1101/gad.12.3.370.

Lee, D. K., Duan, H. O. and Chang, C. (2001) 'Androgen Receptor Interacts

with the Positive Elongation Factor P-TEFb and Enhances the Efficiency of Transcriptional Elongation', *Journal of Biological Chemistry*, 276(13), pp. 9978–9984. doi: 10.1074/jbc.M002285200.

Lehmann, A. (2003) 'DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy', *Biochimie*, 85(11), pp. 1101–1111. doi: 10.1016/j.biochi.2003.09.010.

Lehmann, A. R. (2001) 'The xeroderma pigmentosum group D (XPD) gene: One gene, two functions, three diseases', *Genes and Development*, 15(1), pp. 15–23. doi: 10.1101/gad.859501.

Lehmann, A. R. (2008) 'XPD structure reveals its secrets', *DNA Repair*, 7(11), pp. 1912–1915. doi: 10.1016/j.dnarep.2008.07.008.

Lehmann, A. R., Kirk-Bell, S., Arlett, C. F., Paterson, M. C., Lohman, P. H., de Weerd-Kastelein, E. A. and Bootsma, D. (1975) 'Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation.', *Proceedings of the National Academy of Sciences*, 72(1), pp. 219–223. doi: 10.1073/pnas.72.1.219.

Lehmann, A. R., McGibbon, D. and Stefanini, M. (2011) 'Xeroderma pigmentosum', *Orphanet Journal of Rare Diseases*, 6(1), p. 70. doi: 10.1186/1750-1172-6-70.

Di Lello, P., Jenkins, L. M. M., Jones, T. N., Nguyen, B. D., Hara, T., Yamaguchi, H., Dikeakos, J. D., Appella, E., Legault, P. and Omichinski, J. G. (2006) 'Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53.', *Molecular cell*, 22(6), pp. 731–40. doi: 10.1016/j.molcel.2006.05.007.

Di Lello, P., Miller Jenkins, L. M., Mas, C., Langlois, C., Malitskaya, E., Fradet-Turcotte, A., Archambault, J., Legault, P. and Omichinski, J. G. (2008) 'p53 and TFIIEalpha share a common binding site on the Tfb1/p62 subunit of TFIIH.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(1), pp. 106–11. doi: 10.1073/pnas.0707892105.

Li et al. (2015) 'Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair', *Mol cell*, 59(6), pp. 1025–1034. doi: 10.1016/j.physbeh.2017.03.040.

Li, L., Germain, D. R., Poon, H.-Y., Hildebrandt, M. R., Monckton, E. A., McDonald, D., Hendzel, M. J. and Godbout, R. (2016) 'DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks', *Molecular and Cellular Biology*, 36(22), pp. 2794–2810. doi: 10.1128/mcb.00415-16.

Li, L., Monckton, E. A. and Godbout, R. (2008) 'A Role for DEAD Box 1 at DNA Double-Strand Breaks', *Molecular and Cellular Biology*, 28(20), pp. 6413–6425. doi: 10.1128/mcb.01053-08.

Li, L., Poon, H. Y., Hildebrandt, M. R., Monckton, E. A., Germain, D. R., Fahlman, R. P. and Godbout, R. (2017) 'Role for RIF1-interacting partner DDX1 in BLM recruitment to DNA double-strand breaks', *DNA Repair*. Elsevier, 55(February), pp. 47–63. doi: 10.1016/j.dnarep.2017.05.001.

Li, L., Roy, K., Katyal, S., Sun, X., Bléoo, S. and Godbout, R. (2006) 'Dynamic nature of cleavage bodies and their spatial relationship to DDX1 bodies, Cajal bodies and Gems', *Molecular Biology of the Cell*, 16(September), pp. 4341–4349. doi: 10.1091/mbc.E05.

Li, Q., Liao, C., Xu, W., Li, G., Hong, K., Cheng, X. and Li, J. (2018a) 'Xeroderma Pigmentosum Group D (XPD) Inhibits the Proliferation Cycle of Vascular Smooth Muscle Cell (VSMC) by Activating Glycogen Synthase Kinase 3β (GSK3β)', *Medical Science Monitor*, 24, pp. 5951–5959. doi: 10.12659/msm.909614.

Li, X., Urwyler, O. and Suter, B. (2010) 'Drosophila Xpd regulates Cdk7 localization, mitotic kinase activity, spindle dynamics, and chromosome segregation', *PLoS Genetics*, 6(3). doi: 10.1371/journal.pgen.1000876.

Li, Z., Zhou, M., Cai, Z., Liu, H., Zhong, W., Hao, Q., Cheng, D., Hu, X., Hou, J., Xu, P., Xue, Y., Zhou, Y. and Xu, T. (2018b) 'RNA-binding protein DDX1 is responsible for fatty acid-mediated repression of insulin translation', *Nucleic Acids Research*, 46(22), pp. 12052–12066. doi: 10.1093/nar/gky867.

Linder, P. and Jankowsky, E. (2011) 'From unwinding to clamping ĝ€" the DEAD box RNA helicase family', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 12(8), pp. 505–516. doi: 10.1038/nrm3154.

Liu, H., Rudolf, J., Johnson, K. A., Mcmahon, S. A., Oke, M., Mcrobbie, A., Brown, S. E., Naismith, J. H. and White, M. F. (2008) 'Structure of the DNA repair helicase XPD', *Cell*, 133(5), pp. 801–812. doi: 10.1016/j.cell.2008.04.029.Structure.

Lolli, G. and Johnson, L. N. (2005) 'A Key Kinase in Cell Cycle Control and a Target for Drugs ?', *Cell Cycle*, 4(4)(April), pp. 572–577.

Lolli, G. and Johnson, L. N. (2007) 'Recognition of Cdk2 by Cdk7.', *Proteins*, 67(4), pp. 1048–59. doi: 10.1002/prot.21370.

Lolli, G., Lowe, E. D., Brown, N. R. and Johnson, L. N. (2004) 'The crystal structure of human CDK7 and its protein recognition properties', *Structure*, 12(11), pp. 2067–2079. doi: 10.1016/j.str.2004.08.013.

Lommel, L. (2000) 'The 26S proteasome negatively regulates the level of overall genomic nucleotide excision repair', *Nucleic Acids Research*, 28(24), pp. 4839–4845. doi: 10.1093/nar/28.24.4839.

Lorsch, J. R. (2002) 'RNA chaperones exist and DEAD box proteins get a life', *Cell*, 109(7), pp. 797–800. doi: 10.1016/S0092-8674(02)00804-8.

Lu, H., Fisher, R. P., Bailey, P. and Levine, A. J. (1997) 'The CDK7-cycH-p36 complex of transcription factor IIH phosphorylates p53, enhancing its sequence-specific DNA binding activity in vitro.', *Molecular and Cellular Biology*, 17(10), pp. 5923–5934. doi: 10.1128/MCB.17.10.5923.

Luo, J., Cimermancic, P., Viswanath, S., Ebmeier, C. C., Kim, B., Dehecq, M., Raman, V., Greenberg, C. H., Pellarin, R., Sali, A., Taatjes, D. J., Hahn, S. and Ranish, J. (2015) 'Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH', *Molecular Cell*. Elsevier Inc., 59(5), pp. 794–806. doi: 10.1016/j.molcel.2015.07.016.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J. M., Manning, G., Morgan, D. O., Tsai, L.-H. and Wolgemuth, D. J. (2009) 'Cyclin-dependent

kinases: a family portrait', *Nature Cell Biology*, 11(11), pp. 1275–1276. doi: 10.1038/ncb1109-1275.

Martinez, A. M., Afshar, M., Martin, F., Cavadore, J. C., Labbé, J. C. and Dorée, M. (1997) 'Dual phosphorylation of the T-loop in cdk7: its role in controlling cyclin H binding and CAK activity.', *The EMBO journal*, 16(2), pp. 343–54. doi: 10.1093/emboj/16.2.343.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K. and Hanaoka, F. (1999) 'The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η', *Nature*, 399(6737), pp. 700–704. doi: 10.1038/21447.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J. and Bootsma, D. (1994) 'Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23.', *The EMBO Journal*, 13(8), pp. 1831–1843. doi: 10.1002/j.1460-2075.1994.tb06452.x.

Mathieu, N., Kaczmarek, N., Rüthemann, P., Luch, A. and Naegeli, H. (2013) 'DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH', *Current Biology*, 23(3), pp. 204–212. doi: 10.1016/j.cub.2012.12.032.

Mbonye, U., Wang, B., Gokulrangan, G., Shi, W., Yang, S. and Karn, J. (2018) 'Cyclin-dependent kinase 7 (CDK7)-mediated phosphorylation of the CDK9 activation loop promotes P-TEFb assembly with Tat and proviral HIV reactivation', *Journal of Biological Chemistry*, 293(26), pp. 10009–10025. doi: 10.1074/jbc.RA117.001347.

Merrick, K. A. and Fisher, R. P. (2010) 'Putting one step before the other: Distinct activation pathways for Cdk1 and Cdk2 bring order to the mammalian cell cycle', *Cell Cycle*, 9(4), pp. 706–714. doi: 10.4161/cc.9.4.10732.

Morgan, D. O. (1995) 'Principles of CDK regulation', *Nature*, 374(6518), pp. 131–134. doi: 10.1038/374131a0.

Mullenders, L. H. F. (2018) 'Solar UV damage to cellular DNA: from mechanisms to biological effects', *Photochemical & Photobiological Sciences*, 17(12), pp. 1842–1852. doi: 10.1039/C8PP00182K.

Naik, S. M., Shenoy, A. M., Nanjundappa, A., Halkud, R., Chavan, P., Sidappa, K. and Gupta, S. (2013) 'Cutaneous Malignancies in Xeroderma Pigmentosum: Earlier Management Improves Survival', *Indian Journal of Otolaryngology and Head & Neck Surgery*, 65(2), pp. 162–167. doi: 10.1007/s12070-012-0614-6.

Nechaev, S. and Adelman, K. (2011) 'Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1809(1), pp. 34–45. doi: 10.1016/j.bbagrm.2010.11.001.

Nilson, K. A., Guo, J., Turek, M. E., Brogie, J. E., Delaney, E., Luse, D. S. and Price, D. H. (2015) 'THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing', *Molecular Cell*. Elsevier Inc., 59(4), pp. 576–587. doi: 10.1016/j.molcel.2015.06.032.

Nishiwaki, E., Turner, S. L., Harju, S., Miyazaki, S., Kashiwagi, M., Koh, J. and Serizawa, H. (2000) 'Regulation of CDK7-Carboxyl-Terminal Domain Kinase Activity by the Tumor Suppressor p16INK4A Contributes to Cell Cycle Regulation', *Molecular and Cellular Biology*, 20(20), pp. 7726–7734. doi: 10.1128/mcb.20.20.7726-7734.2000.

Nonnekens, J., Cabantous, S., Slingerland, J., Mari, P.-O. and Giglia-Mari, G. (2013) 'In vivo interactions of TTDA mutant proteins within TFIIH.', *Journal of cell science*, 126(Pt 15), pp. 3278–83. doi: 10.1242/jcs.126839.

O'Gorman, W., Thomas, B., Kwek, K. Y., Furger, A. and Akoulitchev, A. (2005) 'Analysis of U1 small nuclear RNA interaction with cyclin H', *Journal of Biological Chemistry*, 280(44), pp. 36920–36925. doi: 10.1074/jbc.M505791200.

Ohtsubo, M., Theodoras, A. M., Schumacher, J., Roberts, J. M. and Pagano, M. (1995) 'Human cyclin E, a nuclear protein essential for the G1-to-S phase transition.', *Molecular and Cellular Biology*, 15(5), pp. 2612–2624. doi: 10.1128/MCB.15.5.2612.

Okuda, M., Kinoshita, M., Kakumu, E., Sugasawa, K. and Nishimura, Y. (2015) 'Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC', *Structure*, 23(10), pp. 1827–1837. doi: 10.1016/j.str.2015.07.009.

Okuda, M., Nakazawa, Y., Guo, C., Ogi, T. and Nishimura, Y. (2017) 'Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways', *Nucleic Acids Research*. Oxford University Press, 45(22), pp. 13043–13055. doi: 10.1093/nar/gkx970.

Orioli, D., Compe, E., Nardo, T., Mura, M., Giraudon, C., Botta, E., Arrigoni, L., Peverali, F. A., Egly, J. M. and Stefanini, M. (2013) 'XPD mutations in trichothiodystrophy hamper collagen VI expression and reveal a role of TFIIH in transcription derepression', *Human Molecular Genetics*, 22(6), pp. 1061–1073. doi: 10.1093/hmg/dds508.

Patel, S. A. and Simon, M. C. (2010) 'Functional analysis of the Cdk7·cyclin H·Mat1 complex in mouse embryonic stem cells and embryos', *Journal of Biological Chemistry*, 285(20), pp. 15587–15598. doi: 10.1074/jbc.M109.081687.

Pérez-González, A., Rodriguez, A., Huarte, M., Salanueva, I. J. and Nieto, A. (2006) 'hCLE/CGI-99, a Human Protein that Interacts with the Influenza Virus Polymerase, Is a mRNA Transcription Modulator', *Journal of Molecular Biology*, 362(5), pp. 887–900. doi: 10.1016/j.jmb.2006.07.085.

Pines, A. *et al.* (2012) 'PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1', *Journal of Cell Biology*, 199(2), pp. 235–249. doi: 10.1083/jcb.201112132.

Pines, A. *et al.* (2018) 'TRiC controls transcription resumption after UV damage by regulating Cockayne syndrome protein A', *Nature Communications*, 9(1), p. 1040. doi: 10.1038/s41467-018-03484-6.

Pines, J. (1995) 'Confirmational change', *Nature*, 376(6538), pp. 294–295. doi: 10.1038/376294a0.

Popow, J., Jurkin, J., Schleiffer, A. and Martinez, J. (2014) 'Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors',

Nature, 511(7507), pp. 104–107. doi: 10.1038/nature13284.

Price, V. H. (1980) 'Trichothiodystrophy: sulfur-deficient brittle hair as a marker for a neuroectodermal symptom complex', *Archives of Dermatology*, 116(12), pp. 1375–1384. doi: 10.1001/archderm.116.12.1375.

Radu, L., Schoenwetter, E., Braun, C., Marcoux, J., Koelmel, W., Schmitt, D. R., Kuper, J., Cianférani, S., Egly, J. M., Poterszman, A. and Kisker, C. (2017) 'The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH', *Nucleic Acids Research*, 45(18), pp. 10872–10883. doi: 10.1093/nar/gkx743.

Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M. (2003) 'Large-scale proteomic analysis of the human spliceosome', *Genome Research*, 13(1), pp. 1231–1245. doi: 10.1101/gr.473902.

Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A. E., Sattentau, Q., Meinhart, A. and Proudfoot, N. J. (2018) 'RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination', *Molecular Cell*. Elsevier Inc., 70(4), pp. 650-662.e8. doi: 10.1016/j.molcel.2018.04.001.

Robert, F., Douziech, M., Forget, D., Egly, J.-M., Greenblatt, J., Burton, Z. F. and Coulombe, B. (1998) 'Wrapping of Promoter DNA around the RNA Polymerase II Initiation Complex Induced by TFIIF', *Molecular Cell*, 2(3), pp. 341–351. doi: 10.1016/S1097-2765(00)80278-6.

Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J.-M. and Chambon, P. (1997) 'Stimulation of RARα Activation Function AF-1 through Binding to the General Transcription Factor TFIIH and Phosphorylation by CDK7', *Cell*, 90(1), pp. 97–107. doi: 10.1016/S0092-8674(00)80317-7.

Rodríguez-Molina, J. B., Tseng, S. C., Simonett, S. P., Taunton, J. and Ansari, A. Z. (2016) 'Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization', *Molecular Cell*, 63(3), pp. 433–444. doi: 10.1016/j.molcel.2016.06.036.

Roeder, R. G. and Ruttler, W. J. (1969) 'Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms', *Nature*, 224(5216), pp. 234–

237. doi: 10.1038/224234a0.

Rossi, D. J., Londesborough, A., Korsisaari, N., Pihlak, A., Lehtonen, E., Henkemeyer, M. and Mäkelä, T. P. (2001) 'Inability to enter S phase and defective RNA polymerase II CTD phosphorylation in mice lacking Mat1', *EMBO Journal*, 20(11), pp. 2844–2856. doi: 10.1093/emboj/20.11.2844.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 129–143. doi: 10.1038/nrm3952.

Sameer, A. S. and Nissar, S. (2018) 'XPD–The Lynchpin of NER: Molecule, Gene, Polymorphisms, and Role in Colorectal Carcinogenesis', *Frontiers in Molecular Biosciences*, 5(March), pp. 1–10. doi: 10.3389/fmolb.2018.00023.

Sandrock, B. and Egly, J. M. (2001) 'A Yeast Four-hybrid System Identifies Cdk-activating Kinase as a Regulator of the XPD Helicase, a Subunit of Transcription Factor IIH', *Journal of Biological Chemistry*, 276(38), pp. 35328–35333. doi: 10.1074/jbc.M105570200.

Schachter, M. M., Merrick, K. A., Larochelle, S., Hirschi, A., Zhang, C., Shokat, K. M., Rubin, S. M. and Fisher, R. P. (2013) 'A Cdk7-Cdk4 T-Loop Phosphorylation Cascade Promotes G1 Progression', *Mol cell*, 50(2), pp. 250–260. doi: 10.1016/j.molcel.2013.04.003.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J., Chambon, P. and Egly, J. (1993) 'DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor', *Science*, 260(5104), pp. 58–63. doi: 10.1126/science.8465201.

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., C., W., Urlaub, H. and Cramer, P. (2017) 'Structures of transcription pre-initiation complex with TFIIH and Mediator', *Nature*, 551(7679), pp. 204–209. doi: 10.1038/nature24282.

Schmid, S. R. and Linder, P. (1992) 'D-E-A-D protein family of putative RNA helicases', *Molecular Microbiology*, 6(3), pp. 283–291. doi: 10.1111/j.1365-2958.1992.tb01470.x.

Schmitt, D. R., Kuper, J., Elias, A. and Kisker, C. (2014) 'The Structure of the TFIIH p34 Subunit Reveals a Von Willebrand Factor A Like Fold', *PLoS ONE*. Edited by T. Miyata, 9(7), p. e102389. doi: 10.1371/journal.pone.0102389.

Schneider, E., Kartarius, S., Schuster, N. and Montenarh, M. (2002) 'The cyclin h/cdk7/mat1 kinase activity is regulated by CK2 phosphorylation of cyclin H', *Oncogene*, 21(33), pp. 5031–5037. doi: 10.1038/sj.onc.1205690.

Schneider, E., Montenarh, M. and Wagner, P. (1998) 'Regulation of CAK kinase activity by p53.', *Oncogene*, 17(21), pp. 2733–41. doi: 10.1038/sj.onc.1202504.

Schwertman, P., Vermeulen, W. and Marteijn, J. A. (2013) 'UVSSA and USP7, a new couple in transcription-coupled DNA repair', *Chromosoma*, 122(4), pp. 275–284. doi: 10.1007/s00412-013-0420-2.

Sekiguchi, T., Kurihara, Y. and Fukumura, J. (2007) 'Phosphorylation of threonine 204 of DEAD-box RNA helicase DDX3 by cyclin B/cdc2 in vitro', *Biochemical and Biophysical Research Communications*, 356(3), pp. 668–673. doi: 10.1016/j.bbrc.2007.03.038.

Serizawa, H., Conaway, J. W. and Conaway, R. C. (1993) 'Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription', *Nature*, 363(6427), pp. 371–374. doi: 10.1038/363371a0.

Seroz, T. (2000) 'Cloning of a human homolog of the yeast nucleotide excision repair gene MMS19 and interaction with transcription repair factor TFIIH via the XPB and XPD helicases', *Nucleic Acids Research*, 28(22), pp. 4506–4513. doi: 10.1093/nar/28.22.4506.

Seroz, T., Hwang, J. R., Moncollin, V. and Egly, J. M. (1995) 'TFIIH: a link between transcription, DNA repair and cell cycle regulation', *Current Opinion in Genetics and Development*, 5(2), pp. 217–221. doi: 10.1016/0959-437X(95)80011-5.

Sherr, C. J. (1994) 'G1 phase progression: Cycling on cue', *Cell*, 79(4), pp. 551–555. doi: 10.1016/0092-8674(94)90540-1.

Shin, J. H., Kang, H. C., Park, Y.-Y., Ha, D. H., Choi, Y. H., Eum, H. Y., Kang, B.

G., Chae, J. H., Shin, I., Lee, J.-H. and Kim, C. G. (2010) 'Corepressor MMTR/DMAP1 is an intrinsic negative regulator of CAK kinase to regulate cell cycle progression', *Biochemical and Biophysical Research Communications*, 402(1), pp. 110–115. doi: 10.1016/j.bbrc.2010.09.126.

Smale, S. T. and Kadonaga, J. T. (2003) 'The RNA Polymerase II Core Promoter', *Annual Review of Biochemistry*, 72(1), pp. 449–479. doi: 10.1146/annurev.biochem.72.121801.161520.

Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A. and Cimprich, K. A. (2014) 'Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability', *Molecular Cell*, 56(6), pp. 777–785. doi: 10.1016/j.molcel.2014.10.020.

Spivak, G. (2016) 'Transcription-coupled repair: an update', *Archives of Toxicology*. Springer Berlin Heidelberg, 90(11), pp. 2583–2594. doi: 10.1007/s00204-016-1820-x.

Stefanini, M., Botta, E., Lanzafame, M. and Orioli, D. (2010) 'Trichothiodystrophy: From basic mechanisms to clinical implications', *DNA Repair*, 9(1), pp. 2–10. doi: 10.1016/j.dnarep.2009.10.005.

Stefanini, M., Giliani, S., Nardo, T., Marinoni, S., Nazzaro, V., Rizzo, R. and Trevisan, G. (1992) 'DNA repair investigations in nine Italian patients affected by trichothiodystrophy', *Mutation Research/DNA Repair*, 273(2), pp. 119–125. doi: 10.1016/0921-8777(92)90073-C.

Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K. and Hanaoka, F. (2005) 'UV-Induced Ubiquitylation of XPC Protein Mediated by UV-DDB-Ubiquitin Ligase Complex', *Cell*, 121(3), pp. 387–400. doi: 10.1016/j.cell.2005.02.035.

Sunden, Y., Semba, S., Suzuki, T., Okada, Y., Orba, Y., Nagashima, K., Umemura, T. and Sawa, H. (2007) 'DDX1 promotes proliferation of the JC virus through transactivation of its promoter', *Microbiology and Immunology*, 51(3), pp. 339–347. doi: 10.1111/j.1348-0421.2007.tb03907.x.

Svejstrup, J. Q. (2003) 'Rescue of arrested RNA polymerase II complexes',

Journal of Cell Science, 116(3), pp. 447–451. doi: 10.1242/jcs.00271.

Takagi, Y., Masuda, C. A., Chang, W. H., Komori, H., Wang, D., Hunter, T., Joazeiro, C. A. P. and Kornberg, R. D. (2005) 'Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage', *Molecular Cell*, 18(2), pp. 237–243. doi: 10.1016/j.molcel.2005.03.007.

Takaki, T., Echalier, A., Brown, N. R., Hunt, T., Endicott, J. A. and Noble, M. E. M. (2009) 'The structure of CDK4/cyclin D3 has implications for models of CDK activation', *Proceedings of the National Academy of Sciences*, 106(11), pp. 4171–4176. doi: 10.1073/pnas.0809674106.

Takayama, K., Salazar, E. P., Lehmann, A., Stefanini, M., Thompson, L. H. and Weber, C. A. (1995) 'Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D.', *Cancer research*, 55(23), pp. 5656–63.

Tanaka, K., Ikeda, N., Miyashita, K., Nuriya, H. and Hara, T. (2018) 'DEAD box protein DDX1 promotes colorectal tumorigenesis through transcriptional activation of the LGR5 gene', *Cancer Science*, 109(8), pp. 2479–2489. doi: 10.1111/cas.13661.

Tanaka, K., Okamoto, S., Ishikawa, Y., Tamura, H. and Hara, T. (2009) 'DDX1 is required for testicular tumorigenesis, partially through the transcriptional activation of 12p stem cell genes', *Oncogene*. Nature Publishing Group, 28(21), pp. 2142–2151. doi: 10.1038/onc.2009.89.

Tani, H., Fujita, O., Furuta, A., Matsuda, Y., Miyata, R., Akimitsu, N., Tanaka, J., Tsuneda, S., Sekiguchi, Y. and Noda, N. (2010) 'Real-time monitoring of RNA helicase activity using fluorescence resonance energy transfer in vitro', *Biochemical and Biophysical Research Communications*. Elsevier Inc., 393(1), pp. 131–136. doi: 10.1016/j.bbrc.2010.01.100.

Tassan, J. P., Schultz, S. J., Bartek, J. and Nigg, E. A. (1994) 'Cell cycle analysis of the activity, subcellular localization, and subunit composition of human CAK (CDK-activating kinase)', *Journal of Cell Biology*, 127(2), pp. 467–478. doi: 10.1083/jcb.127.2.467.

Taylor, E. M., Broughton, B. C., Botta, E., Stefanini, M., Sarasin, A., Jaspers,

N. G. J., Fawcett, H., Harcourt, S. A., Arlett, C. F. and Lehmann, A. R. (1997) 'Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene', *Proceedings of the National Academy of Sciences*, 94(16), pp. 8658–8663. doi: 10.1073/pnas.94.16.8658.

Theil, A. F. *et al.* (2017) 'Trichothiodystrophy causative TFIIEβ mutation affects transcription in highly differentiated tissue', *Human Molecular Genetics*, 26(23), pp. 4689–4698. doi: 10.1093/hmg/ddx351.

Theil, A. F. *et al.* (2019) 'Bi-allelic TARS Mutations Are Associated with Brittle Hair Phenotype', *The American Journal of Human Genetics*. ElsevierCompany., 105(2), pp. 434–440. doi: 10.1016/j.ajhg.2019.06.017.

Theil, A. F., Nonnekens, J., Steurer, B., Mari, P. O., de Wit, J., Lemaitre, C., Marteijn, J. A., Raams, A., Maas, A., Vermeij, M., Essers, J., Hoeijmakers, J. H. J., Giglia-Mari, G. and Vermeulen, W. (2013) 'Disruption of TTDA Results in Complete Nucleotide Excision Repair Deficiency and Embryonic Lethality', *PLoS Genetics*, 9(4). doi: 10.1371/journal.pgen.1003431.

Tingting, P., Caiyun, F., Zhigang, Y., Pengyuan, Y. and Zhenghong, Y. (2006) 'Subproteomic analysis of the cellular proteins associated with the 3′ untranslated region of the hepatitis C virus genome in human liver cells', *Biochemical and Biophysical Research Communications*, 347(3), pp. 683–691. doi: 10.1016/j.bbrc.2006.06.144.

Vashisht, A. A., Yu, C. C., Sharma, T., Ro, K. and Wohlschlegel, J. A. (2015) 'The association of the xeroderma pigmentosum group D DNA helicase (XPD) with transcription factor IIH is regulated by the cytosolic iron-sulfur cluster assembly pathway', *Journal of Biological Chemistry*, 290(22), pp. 14218–14225. doi: 10.1074/jbc.M115.650762.

Vermeulen, W., Bergmann, E., Auriol, J., Rademakers, S., Frit, P., Appeldoorn, E., Hoeijmakers, J. H. and Egly, J. M. (2000) 'Sublimiting concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder.', *Nature genetics*, 26(3), pp. 307–13. doi: 10.1038/81603.

Walker, D. H. and Maller, J. L. (1991) 'Role for cyclin A in the dependence of

mitosis on completion of DMA replication', *Nature*, 354(6351), pp. 314–317. doi: 10.1038/354314a0.

Wang, W., Xu, J., Chong, J. and Wang, D. (2018) 'Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair', *DNA Repair*. Elsevier, 71(August), pp. 43–55. doi: 10.1016/j.dnarep.2018.08.006.

Weber, A., Chung, H. J., Springer, E., Heitzmann, D. and Warth, R. (2010) 'The TFIIH subunit p89 (XPB) localizes to the centrosome during mitosis', *Cellular Oncology*, 32(1–2), pp. 121–130. doi: 10.3233/CLO-2009-0509.

Weeda, G., Rossignol, M., Fraser, R. A., Winkler, G. S., Vermeulen, W., Van't Veer, L. J., Ma, L., Hoeijmakers, J. H. J. and Egly, J. M. (1997) 'The XPB subunit of repair/transcription factor TFIIH directly interacts with SUG1, a subunit of the 26S proteasome and putative transcription factor', *Nucleic Acids Research*, 25(12), pp. 2274–2283. doi: 10.1093/nar/25.12.2274.

Wolski, S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, B. and Kisker, C. (2008) 'Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD', *PLoS Biology*, 6(6), pp. 1332–1342. doi: 10.1371/journal.pbio.0060149.

Wong, K. H., Jin, Y. and Struhl, K. (2014) 'TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.', *Molecular cell*, 54(4), pp. 601–12. doi: 10.1016/j.molcel.2014.03.024.

Wood, D. J. and Endicott, J. A. (2018) 'Structural insights into the functional diversity of the CDK–cyclin family', *Open Biology*, 8(9). doi: 10.1098/rsob.180112.

Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M. A., Chong, J., Hare, A. A., Dervan, P. B., DiMaio, F., Leschziner, A. E. and Wang, D. (2017) 'Structural basis for the initiation of eukaryotic transcription-coupled DNA repair', *Nature*, 551(7682), pp. 653–657. doi: 10.1038/nature24658.

Yankulov, K. Y. and Bentley, D. (1997) 'Regulation of CDK7 substrate specificity by MAT1 and TFIIH', *The EMBO Journal*, 16(7), pp. 1638–1646.

doi: 10.1093/emboj/16.7.1638.

Yeom, E., Hong, S. T. and Choi, K. W. (2015) 'Crumbs interacts with Xpd for nuclear division control in Drosophila', *Oncogene*. Nature Publishing Group, 34(21), pp. 2777–2789. doi: 10.1038/onc.2014.202.

Yoh, S. M., Lucas, J. S. and Jones, K. A. (2008) 'The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation', *Genes and Development*, 22(24), pp. 3422–3434. doi: 10.1101/gad.1720008.

List of original manuscripts

- **Ferri D**, Orioli D and Botta E. Clin Genet. 2019 Mar 28. doi: 10.1111/cge.13545. [Epub ahead of print] Review.
  Appendix I

- Lanzafame M, Landi C, Uggè M, Vaz B, Nardo T, **Ferri D**, Stefanini M, Peverali FA, Bini L and Orioli D. Cockayne syndrome and erythropoietic protoporphyria come together at the nucleoli. *Submitted*

- Lombardi A, Arseni L, Carriero R, Compe E, Botta E, **Ferri D**, Uggè M, Biamonti G, Peverali AF, Bione S and Orioli D. Reduced levels of Prostaglandin I2 Synthase specify the cancer-free trichothiodystrophy. *Submitted*

Appendix I

Heterogeneity and overlaps in nucleotide excision repair disorders

Abstract

Nucleotide excision repair (NER) is an essential DNA repair pathway devoted to the removal of bulky lesions such as photoproducts induced by the ultraviolet (UV) component of solar radiation. Deficiencies in NER typically result in a group of heterogeneous distinct disorders ranging from the mild UV sensitive syndrome to the cancer-prone xeroderma pigmentosum and the neurodevelopmental/progeroid conditions trichothiodystrophy, Cockayne syndrome and cerebro-oculo-facio-skeletalsyndrome. A complicated genetic scenario underlines these disorders with the same gene linked to different clinical entities as well as different genes associated with the same disease. Overlap syndromes with combined hallmark features of different NER disorders can occur and sporadic presentations showing extra features of the hematological disorder Fanconi Anemia or neurological manifestations mimicking Hungtinton disease-like syndromes have been described. Here, we discuss the multiple functions of the five major pleiotropic NER genes (ERCC3/XPB, ERCC2/XPD, ERCC5/ XPG, ERCC1 and ERCC4/XPF) and their relevance in phenotypic complexity. We provide an update of mutational spectra and examine genotype-phenotype relationships. Finally, the molecular defects that could explain the puzzling overlap syndromes are discussed.

Received: 18 December 2018 | Revised: 27 February 2019 | Accepted: 26 March 2019
DOI: 10.1111/cge.13545

INVITED REVIEW

CLINICAL GENETICS WILEY

Heterogeneity and overlaps in nucleotide excision repair disorders

Debora Ferri | Donata Orioli | Elena Botta

Istituto di Genetica Molecolare (IGM), Consiglio Nazionale delle Ricerche, Pavia, Italy

Correspondence
Elena Botta, IGM Via Abbiategrasso 207, 27100 Pavia, Italy.
Email: botta@igm.cnr.it

Funding information
Associazione Italiana per la Ricerca sul Cancro, Grant/Award Number: Grant IG 17710

Abstract

Nucleotide excision repair (NER) is an essential DNA repair pathway devoted to the removal of bulky lesions such as photoproducts induced by the ultraviolet (UV) component of solar radiation. Deficiencies in NER typically result in a group of heterogeneous distinct disorders ranging from the mild UV sensitive syndrome to the cancer-prone xeroderma pigmentosum and the neurodevelopmental/progeroid conditions trichothiodystrophy, Cockayne syndrome and cerebro-oculo-facio-skeletal-syndrome. A complicated genetic scenario underlines these disorders with the same gene linked to different clinical entities as well as different genes associated with the same disease. Overlap syndromes with combined hallmark features of different NER disorders can occur and sporadic presentations showing extra features of the hematological disorder Fanconi Anemia or neurological manifestations mimicking Hungtinton disease-like syndromes have been described. Here, we discuss the multiple functions of the five major pleiotropic NER genes (*ERCC3/XPB*, *ERCC2/XPD*, *ERCC5/XPG*, *ERCC1* and *ERCC4/XPF*) and their relevance in phenotypic complexity. We provide an update of mutational spectra and examine genotype-phenotype relationships. Finally, the molecular defects that could explain the puzzling overlap syndromes are discussed.

KEYWORDS
genotype-phenotype relationships, NER disorders, nucleotide excision repair (NER), overlap syndromes

1 | INTRODUCTION

Nucleotide excision repair (NER) is the major DNA repair pathway devoted to the removal of bulky adducts induced by the ultraviolet (UV) component of sun light or other environmental carcinogens. Depending on the site of the damage, the NER process is referred as global genome NER (GG-NER) or transcription-coupled NER (TC-NER), the first sensing lesions anywhere in the genome and the last acting on lesions located in the transcribed strand of active genes. Specific GG-NER and TC-NER factors carry out the initial step of lesion recognition and, thereafter, a common set of core activities completes the repair reaction by (a) unwinding of the DNA double helix (DNA bubble) in the region containing the lesion, (b) incision of the damaged DNA strand at both sites of the lesion and removal (excision) of the lesion-containing fragment, (c) synthesis and ligation of a novel DNA stretch that fills the gap and restores integrity of the DNA molecule.[1,2] In human cells, the entire NER process requires the sequential action of a variety of proteins encoded by more than 40 genes (Table 1).

The relevance of NER factors for human health is showed by the dramatic consequences of their alteration. Bi-allelic mutations in NER genes result in several rare diseases characterized by a wide and varied spectrum of clinical features. NER disorders also include combined presentations that can sporadically present symptoms associated with

TABLE 1 Human NER genes and their functions

Gene[a] (alternative names)	NCBI gene ID	Locus	Exons (isoforms)	Reference transcript[b] (CDS)	Reference protein[c] (symbol; length)	Function in NER
GG-NER						
XPC	7508	3p25.1	16	NM_004628	NP_004619	DNA damage recognition factor
XP3; RAD4; XPCC; P125			(5)	(105-2927)	(XPC; 940 aa)	
RAD23B	5887	9q31.2	12	NM_002874	NP_002865	It associates with XPC
HR23B; HHR23B; P58			(3)	(352-1581)	(HR23B; 409 aa)	
RAD23A	5886	19p13.13	9	NM_005053	NP_005044	It associates with XPC in substitution of HR23B
HR23A; HHR23A			(3)	(88-1179)	(HR23A; 363 aa)	
CETN2	1069	Xq28	5	NM_004344	NP_004335	It associates with XPC
CALT; CEN2			(1)	(48-566)	(Centrin-2; 172 aa)	
DDB1	1642	11q12.2	27	NM_001923	NP_001914	Subunit of the DNA damage-binding (DDB) complex
DDBA; XAP1; UV-DDB1			(1)	(245-3667)	(DDB p127; 1140 aa)	
DDB2	1643	11p11.2	10	NM_000107	NP_000098	Subunit of the DDB complex
XPE; DDBB; UV-DDB2			(2)	(196-1479)	(DDB p48; 427 aa)	
CUL4A	8451	13q34	25	NM_0010088	NP_001008895	Component of the DDB1-containing ubiquitin ligase multi-protein complex
			(11)	95 (16-2295)	(Cullin 4A; 759 aa)	
RBX1	9978	22q13.2	5	NM_014248	NP_055063	Component of the DDB1-containing ubiquitin ligase multi-protein complex
ROC1; RNF75; BA554C12.1			(1)	(53-379)	(RBX; 108 aa)	
TC-NER						
ERCC8	1161	5q12.1	13	NM_000082	NP_000073	Component of the DDB1-containing ubiquitin ligase complex acting at sites of stalled RNA Polymerase II
CSA; CKN1; UVSS2			(4)	(71-1261)	(CSA; 396 aa)	
ERCC6	2074	10q11.23	23	NM_000124	NP_000115	It operates at sites of stalled RNA polymerase II
CSB; CKN2; COFS; ARMD5; COFS1; POF11; RAD26; UVSS1; CSB-PGBD3			(4)	(175-4656)	(CSB; 1493 aa)	
UVSSA	57654	4p16.3	27	NM_020894	NP_065945	It stabilizes CSB
UVSS3; KIAA1530			(3)	(448-2577)	(UVSSA; 709 aa)	
USP7	7874	16p13.2	35	NM_003470	NP_003461	Ubiquitin-specific protease interacting with UVSSA
TEF1; HAUSP			(4)	(622-3930)	(USP7; 1102 aa)	
XAB2	56949	19p13.2	19	NM_020196	NP_064581	Chromatin remodeling factor
HCNP; HCRN; SYF1; NTC90			(1)	(19-2586)	(SYF1; 855 aa)	
HMGN1	3150	21q22.2	10	NM_004965	NP_004956	Chromatin remodeling factor
HMG14			(1)	(172-474)	(HMG-14; 100 aa)	
Common pathway						
XPA	7507	9q22.33	9	NM_000380	NP_000371	DNA damage verification factor
XP1; XPAC			(2)	(118-939)	(XPA; 273 aa)	
TFIIH (transcription factor IIH)						
ERCC3	2071	2q14.3	15	NM_000122	NP_000113	ATPase/helicase, subunit of the core-TFIIH
XPB; BTF2; TTD2; GTF2H; RAD25			(3)	(96-2444)	(XPB; 782 aa)	
ERCC2	2068	19q13.32	24	NM_000400	NP_000391	ATPase/helicase, subunit of the core-TFIIH
EM9; XPD; TTD1; COFS2			(2)	(48-2330)	(XPD; 760 aa)	
GTF2H5	404672	6q25.3	3	NM_207118	NP_997001	Subunit of the core-TFIIH. It stabilizes p52

TABLE 1 (Continued)

Gene[a] (alternative names)	NCBI gene ID	Locus	Exons (isoforms)	Reference transcript[b] (CDS)	Reference protein[c] (symbol; length)	Function in NER
TFB5; TTD3; TTDA; TGF2H5; C6orf175; bA120J8.2			(1)	(84-299)	(TTDA; 71 aa)	
GTF2H3	2967	12q24.31	13	NM_001516	NP_001507	Subunit of the core-TFIIH
P34; BTF2; TFB4			(4)	(32-958)	(p34; 308 aa)	
GTF2H2	2966	5q13.2	19	NM_001515	NP_001506	Subunit of the core-TFIIH. It stimulates XPD activity
P44; BTF2; BTF2P44; T-BTF2P44			(8)	(211-1398)	(p44; 395 aa)	
GTF2H4	2968	6p21.33	14	NM_001517	NP_001508	Subunit of the core-TFIIH. It stimulates XPB activity
P52; TFB2			(1)	(201-1589)	(p52; 462)	
GTF2H1	2965	11p15.1	18	NM_0011423	NP_001135779	Subunit of the core-TFIIH
P62; BTF2; TFB1			(2)	07 (695-2341)	(p62; 584 aa)	
MNAT1	4331	14q23.1	12	NM_002431	NP_002422	Subunit of the cyclin-dependent kinase (CDK) activating kinase (CAK) sub-complex
MAT1; TFB3; CAP35; RNF66			(2)	(103-1032)	(MAT1; 309 aa)	
CDK7	1022	5q13.2	13	NM_001799	NP_001790	Subunit of the CAK sub-complex
CAK1; HCAK; MO15; STK1; CDKN7; p39MO15			(9)	(91-1131)	(CDK7; 346 aa)	
CCNH	902	5q14.3	11	NM_0013635	NP_001350468	Subunit of the CAK sub-complex
p34; p37; CycH			(5)	39 (240-1217)	(Cyclin H; 325 aa)	
ERCC5	2073	13q33.1	15	NM_000123	NP_000114	Endonuclease cleaving at 3′ of the DNA lesion
XPG; UVDR; XPGC; COFS3; ERCM2; ERCC5-201			(1)	(427-3987)	(XPG; 1186 aa)	
ERCC4	2072	16p13.12	13	NM_005236	NP_005227	Endonuclease cleaving at 5′ of the DNA lesion
XPF; RAD1; FANCQ XFEPS; ERCC11			(1)	(10-2760)	(XPF; 916 aa)	
ERCC1	2067	19q13.32	14	NM_001983	NP_001974	It associates with XPF
UV20; COFS4; RAD10			(3)	(147-1040)	(ERCC1; 297 aa)	
RPA1	6117	17p13.3	17	NM_002945	NP_002936	Largest subunit of the heterotrimeric replication protein A (RPA)
HSSB; RF-A; RP-A; REPA1 RPA70; MST075			(3)	(124-1974)	(RP-A p70; 616 aa)	
RPA2	6118	1p35.3	9	NM_0012975	NP_001284487	Subunit of the RPA complex
REPA2; RPA32; RP-A p32; RP-A p34			(5)	58 (138-974)	(RP-A p32; 278 aa)	
RPA3	6119	7p21.3	8	NM_002947	NP_002938	Subunit of the RPA complex
REPA3; RP-A p14			(1)	(1173-1538)	(RP-A p14; 121 aa)	
PCNA *ATLD2*	5111	20p12.3	7	NM_002592	NP_002583	Cofactor of DNA polymerase δ
			(2)	(240-1025)	(PCNA; 261 aa)	
RFC1	5981	4p14	25	NM_0012047	NP_001191676	Largest subunit of the DNA polymerase accessory replication Factor C (RFC)
A1; RFC; PO-GA; RECC1 MHCBFB; RFC140			(4)	47 (141-3587)	(RFC140; 1148 aa)	
RFC2	5982	7q11.2 3	11	NM_181471	NP_852136	Subunit of the RFC complex
RFC40			(5)	(17-1081)	(RFC40; 354 aa)	
RFC3	5983	13q13.2	15	NM_002915	NP_002906	Subunit of the RFC complex
RFC38			(2)	(31-1101)	(RFC38; 356 aa)	
RFC4	5984	3q27.3	11	NM_002916	NP_002907	Subunit of the RFC complex

TABLE 1 (Continued)

Gene[a] (alternative names)	NCBI gene ID	Locus	Exons (isoforms)	Reference transcript[b] (CDS)	Reference protein[c] (symbol; length)	Function in NER
A1; RFC37			(2)	(224-1315)	(RFC37; 363 aa)	
RFC5	5985	12q24.23	14	NM_007370	NP_031396	Subunit of the RFC complex
RFC36			(6)	(126-1148)	(RFC36; 340 aa)	
POLD1	5424	19q13.33	30	NM_0013086	NP_001295561	Catalytic subunit of DNA Polymerase δ
CDC2; MDPL; POLD; CRCS10			(3)	32(1-3402)	(Pol δ p125; 1133 aa)	
POLD2	5425	7p13	14	NM_006230	NP_006221	Accessory subunit of DNA Polymerase δ
			(3)	(12-1526)	(Pol δ p50; 504 aa)	
POLD3	10714	11q13.4	14	NM_006591	NP_006582	Accessory subunit of DNA Polymerase δ
P66; P68; PPP1R128			(2)	(130-1530)	(Pol δ p68; 466 aa)	
POLE	5426	12q24.33	51	NM_006231	NP_006222	Catalytic subunit of DNA Polymerase ε
FILS; POLE1; CRCS12			(1)	(210-7070)	(Pol ε subunit A; 2286 aa)	
POLE2	5427	14q21.3	20	NM_002692	NP_002683	Accessory subunit of DNA Polymerase ε
DPE2			(5)	(12-1595)	(Pol ε subunit B; 527 aa)	
POLK	51426	5q13.3	21	NM_016218	NP_057302	DNA Polymerase
DINP; POLQ; DINB1			(3)	(179-2791)	(Pol k; 870 aa)	
XRCC1	7515	19q13.31	17	NM_006297	NP_006288	Accessory protein of polymerases and ligases
RCC; SCAR26			(1)	(121-2022)	(XRCC1; 633 aa)	
LIG1	3978	19q13.33	29	NM_000234	NP_000225	DNA ligase acting in proliferating cells
			(5)	(162-2921)	(DNA ligase I; 919 aa)	
LIG3	3980	17q12	22	NM_013975	NP_039269	DNA ligase acting in proliferating and non-proliferating cells
LIG2; LIG3alpha			(2)	(130-3159)	(DNA ligase III; 1009 aa)	

Abbreviations: aa, amino acid; CDS, coding sequence.
[a]Official gene symbol in bold.
[b]It refers to the longest transcript.
[c]It refers to the longest isoform.

different disease categories (Figure 1). Typical NER disorders are xeroderma pigmentosum (XP, MIM 278700, 610651, 278720, 278730, 278740, 278760, 278780), UV sensitive syndrome (UVSS, MIM 600630), trichothiodystrophy (TTD, MIM 601675, 211390, 275550 Cockayne syndrome (CS, MIM 21640), 133540 and cerebro-oculo-facio-skeletal-syndrome (COFS, MIM 214150). Although displaying cutaneous sun sensitivity as a common trait and sometimes overlapping neurological symptoms, NER disorders are distinct and heterogeneous clinical entities. XP patients usually present with sunburn on minimal sun exposure, pigmentary changes at exposed skin regions and multiple early age skin cancers. Progressive neuronal degeneration is also observed in approximately one-third of cases, generally after the appearance of cutaneous signs.[3] Although XP is typical of childhood, late presentations with inconspicuous skin manifestations and neurological signs mimicking conditions such as Huntington disease (HD)-like syndromes have been recently included in the XP spectrum.[4,5] No cancer predisposition or neurodegeneration is instead found in the UVSS patients who only show mild photosensitivity with freckling and telangiectasia.[6] Very few UVSS cases are known, possibly because the symptoms are so mild that patients may be overlooked. Conversely, TTD and CS are invalidating multisystem disorders. TTD defining feature is the sulfur-deficient brittle hair associated with ichthyotic skin, developmental delay, intellectual disability, signs of premature aging and recurrent infections, the last ones representing a frequent cause of early death.[7] CS patients show growth failure, progressive neurological degeneration, ocular and skeletal abnormalities, and premature aging.[8] Various degrees of severity have been described in CS, including the devastating COFS variant that is characterized by fetal abnormalities. Infants are born with arthrogryposis, extreme microcephaly, congenital cataract and facial dysmorphism.[9]

Besides the distinct NER disorders, "overlapping syndromes" with combined clinical hallmarks can also occur. Careful assessment of clinical features has led to the identification of rare subsets of patients in

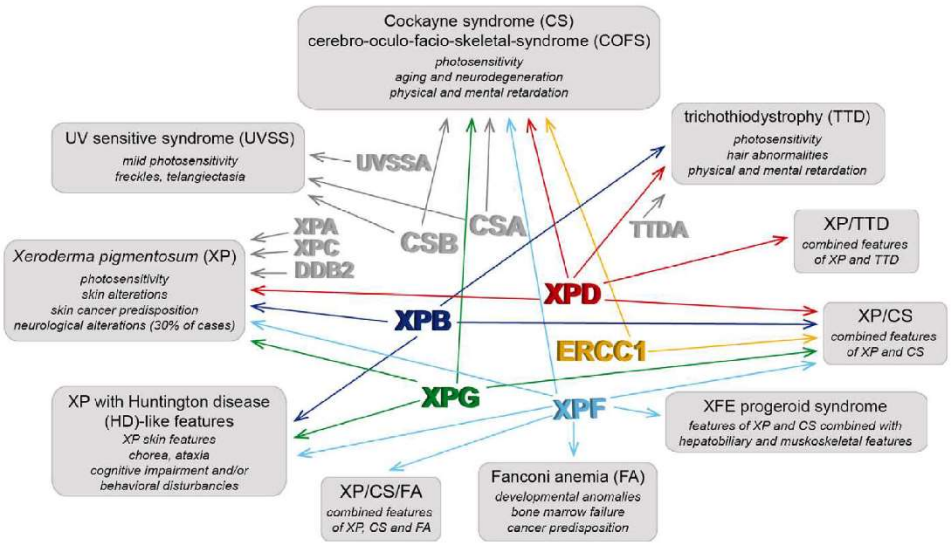


FIGURE 1 Relationship between NER factors and human diseases. Ten clinical conditions with the corresponding main symptoms are represented. Each disease may be caused by alterations in different NER factors and, conversely, alterations in one NER factor may result in different diseases. The pleiotropic core-NER factors are in colors [Colour figure can be viewed at wileyonlinelibrary.com]

which TTD or CS features are present in combination with cutaneous changes typical of XP, thus defining the combined phenotypes XP/TTD and XP/CS. Very few cases with XP/TTD are known and they all show attenuated features of both diseases.[10,11] Several XP/CS cases have instead been described showing remarkable variability in the type and severity of symptoms. Some XP/CS patients show mild neurological/developmental abnormalities associated with late appearing XP skin features. Others present at birth with severe CS or COFS features and minimal skin alterations.[12] Few exceptional cases have been also reported. In addition to combined XP and CS symptoms, unique hepatobiliary and musculoskeletal features were observed in a patient classified as XFE progeroid syndrome (MIM 610965)[13] whereas clinical features of Fanconi Anemia (FA, MIM 227650) were found in an English case.[14] FA is a myelodysplastic disorder characterized by progressive bone marrow failure and blood cancer predisposition resulting from defective repair of DNA interstrand crosslinks (ICLs). Of note, partial NER deficiency has been found in few FA cases.[15,16]

The remarkable phenotypic complexity of NER disorders is underlined by an intricate genetic scenario. Twelve NER genes have been associated to five distinct disorders and five different combined phenotypes. As highlighted in Figure 1, the same NER gene may give rise to multiple clinical phenotypes as well as distinct NER genes may result in the same phenotype. In particular, the genes encoding XPB, XPD, XPG, ERCC1 and XPF proteins result in a constellation of distinct disorders and overlapping syndromes. Such extensive pleiotropy can be explained by the multifunctional nature of these factors. Besides participating to the core repair reaction common to GG-NER and TC-NER, they are also involved in other DNA repair mechanisms and/or DNA-associated processes such as transcription. This variety of functions and clinical presentations makes NER disorders, although rare, important model systems to investigate basic biological mechanisms and, at the same time, their implication in cancer, neurodegeneration and aging.

In this article, we briefly overview the multifaceted roles of pleiotropic NER factors. We report on patient cohorts with the different clinical outcomes and we provide the updated mutational spectra discussing genotype-phenotype relationships. Finally, we examine the so far proposed models for puzzling out the complex overlap phenotypes.

2 | MULTITASKING CORE-NER FACTORS

2.1 | XPB and XPD

XPB and XPD (xeroderma pigmentosum group B and group D complementing proteins, respectively) are both ATPase/helicases belonging to the multiprotein transcription factor IIH (TFIIH). Besides its role in basal and activated transcription, TFIIH serves an essential function in NER.[17,18] It is composed by 10 subunits organized in the core-TFIIH and in the CDK-activating enzyme kinase (CAK) module. XPB is part of the core whereas XPD is a bridging factor that links the core to the

CAK. In NER, TFIIH major contribution is to open the double-stranded DNA around the lesion site. XPB and XPD co-operate during this process. It has been shown that the XPD helicase unwinds the DNA around the damage, whereas the XPB ATPase contributes to anchoring TFIIH to the chromatin.[19] In transcription, TFIIH plays various roles. It is a general transcription factor for the RNA polymerase II (RNAPII) transcription initiation machine. In this context, the XPB helicase seems to be implicated in promoter opening whereas the kinase activity of the CDK7 subunit stimulates promoter-proximal pausing, transcription elongation and termination. CDK7 kinase also phosphorylates several transcription factors thereby contributing to the role of TFIIH in transcription regulation.[20] By anchoring the CAK to the core, the bridging subunit XPD influences CDK7 activity and, in turn, proper gene expression ([17] and references therein). A further layer of complexity is added by the fact that XPD can be also found in non-TFIIH complexes. A fraction of XPD is associated with CAK in the absence of the core subunits and this association might be important in modulating mitotic progression.[21] XPD is also part of the MMXD (MMS19-MIP18-XPD) and CGX (Crumbs-Galla-XPD) complexes, both involved in chromosome segregation.[22,23] Finally, XPD localizes inside mitochondria where it might be implicated in protection of mitochondrial genome from oxidative DNA damage.[24]

XP-B and XP-D cohorts: Alterations in XPB have been found rarely, probably reflecting the essential role of this subunit in TFIIH activities. Only 12 cases (from eight unrelated families) have been reported so far (Table S1). Alterations in XPD have instead been observed with relatively high frequency, 142 patients (from 117 families) are known (Table S2). Both XP-B and XP-D cohorts include pure XP cases with minimal neurological abnormalities as well as combined XP/CS presentations of varying severity. Two couples of young siblings with a suspected clinical diagnosis of UVSS have been also recently reported to carry alterations in either XPB or XPD.[25] Diagnosis is however difficult in young patients because skin cancers may develop later in life. Long-term follow-up is therefore needed to test the hypothesis that the clinical spectrum associated with XPB or XPD extends to the UVSS phenotype. Several TTD cases have been found associated to defects in XPD (46 patients from 42 families) or XPB (only two siblings) whereas all the rare combined XP/TTD cases (four patients) are defective in XPD. The XP-D cohort also includes three severe COFS cases and the XP-B group comprises two XP siblings presenting adult-onset chorea and neuropsychiatric symptoms consistent with HD-like phenotype. The high similarity of clinical presentations associated with defects in XPB or XPD is consistent with the fact that these proteins are subunit of the same complex.

Mutational spectra and relationship to clinical outcomes: Only nine mutations have been so far identified in the *ERCC3/XPB* gene: two resulting in missense changes and seven in protein truncations because of either stop codons or splice alterations (Figure 2) (Tables S1, S6 and S7). Genotype-phenotype correlation studies in this small cohort have indicated that mild XP and TTD phenotypes are associated with missense mutations on at least one *ERCC3* allele whereas the severely affected XP/CS patients have truncating mutations on both alleles.[26] Compared to *ERCC3/XPB*, the *ERCC2/XPD* gene is more tolerant to variations (as reported in gnomAD, observed/expected [o/e] metric for loss of function variations in *ERCC2* is higher than in *ERCC3*. O/e metrics correspond to 0.85 and 0.57, respectively). Indeed, a total of 72 genetic variants predicted to result in 75 distinct protein alterations have been identified in *ERCC2/XPD* (Figure 2) (Tables S2, S6, and S7). Forty-one of the variations are missense changes, eight in frame-deletions, three insertions or deletion/insertions and the rest are truncations from stop, frameshifts or splice alterations. A complex *ERCC2* allele with two associated variants has been also identified. In addition, null mutations leading to unexpressed transcripts have been reported in seven patients. The vast majority of the mutations are clustered in the C-terminal third of the XPD protein even though no specific disease-related domains can be highlighted. Indeed, adjacent mutations can result in different phenotypes (Figure 2). Interpretation of mutational data is nevertheless complicated by the fact that most of the XP-D patients are compound heterozygotes with a combination of different alterations on the *ERCC2* alleles. In addition, some *ERCC2* mutated alleles have been found associated with more than one disorder. Three of them were shown to be lethal in yeast and thus considered as null, suggesting that these alleles probably do not contribute to the clinical outcome which would be determined by the second allele.[27] Taken this into account, each combination of mutated *ERCC2* alleles appears to be specific for a particular disease. The only exceptions so far reported are two cases with TTD and XP/CS carrying identical variants. Additional genetic variations or different allelic expression because of SNPs have been proposed to contribute to the different phenotypes in these patients.[28]

A relevant contribution to our understanding of genotype-phenotype relationships in the XP-D cohort was provided by the definition of archaeal XPD protein structures.[29–31] Based on crystal models, XP-causing mutations seem to be located either on the faces of the helicase domains or close to the ATP-binding site. These mutations impact on the XPD helicase activity that is essential for NER, without disrupting the general protein structure. TTD-causing mutations instead appear to impact on the overall structural framework of XPD, thus affecting TFIIH activities both in NER and transcription. Consistently, all mutations linked to the NER deficient form of TTD result in markedly reduced TFIIH cellular levels.[32] Differently, XP/CS mutations are predicted to affect the flexibility of the XPD structure and in turn disturb its bridging role between the CAK and core-TFIIH sub-complexes.

2.2 | XPG

XPG (xeroderma pigmentosum group G protein) is a single-stranded structure-specific DNA 3′-endonuclease belonging to the FEN1 protein family that possesses the N and I nuclease domains (Figure 2). During NER, XPG cuts the damaged DNA strand at the 3′ junction of the DNA bubble opened by TFIIH.[1] Besides NER, XPG has multiple non-enzymatic functions. It interacts directly with RNAPII and the CSB protein, the last one being implicated in the early steps of TC-NER.[33] It associates with the TFIIH complex thus stabilizing the CAK

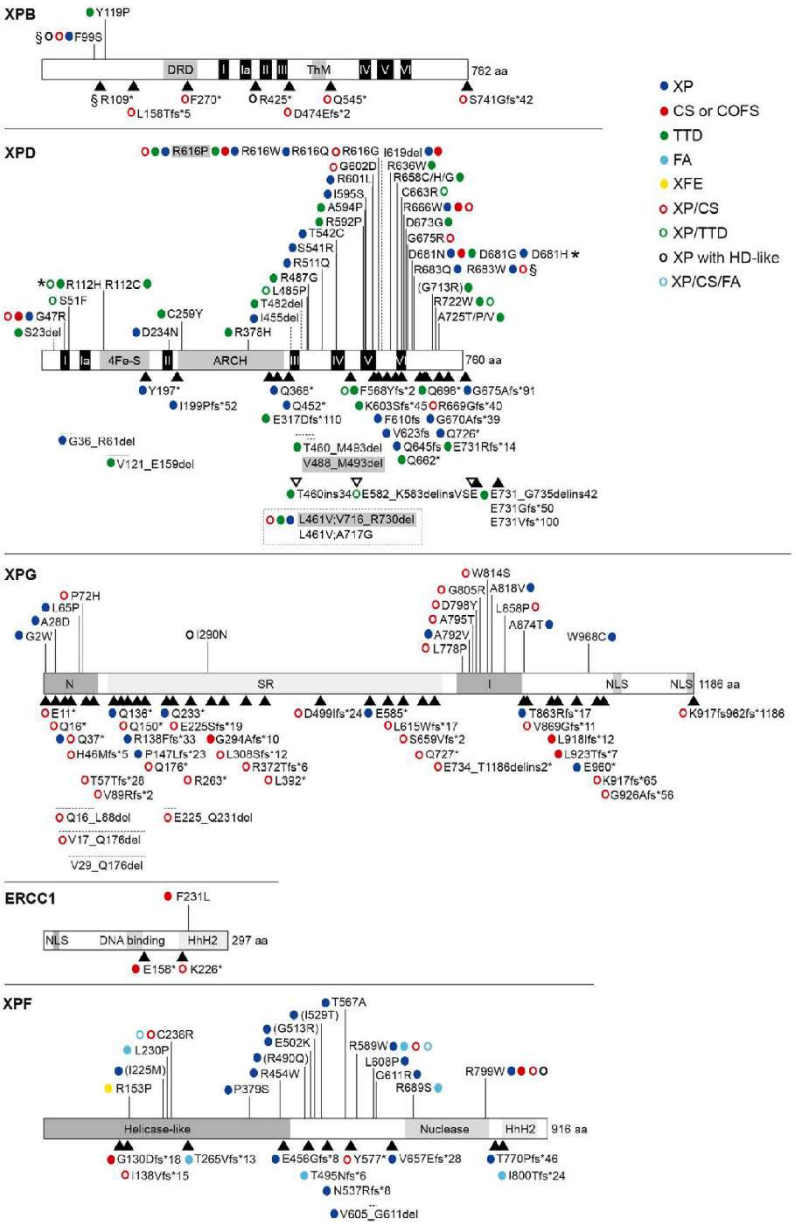


FIGURE 2 Legend on next coloumn.

module and favoring receptor-mediated transcription.[34] The XPG-TFIIH complex also serves as a transcription elongation factor for RNAPII[35] whereas, together with XPF, the XPG protein appears to be involved in chromatin looping.[36] In addition, XPG interacts in vitro and stimulates the NTH1 glycosylase involved in base excision repair (BER) of oxidative DNA damage.[37] Finally, XPG has been recently demonstrated to contribute to genomic stability preservation also by participating to the homologous recombination repair pathway.[38]

XP-G cohort: XP-G cases are quite rare. Forty-seven cases (from 36 unrelated families) have been identified so far and they span from pure XP to combined XP/CS presentations and few individuals classified as severe CS or COFS. An XP case manifesting as HD-like phenotype has been also reported (Table S3).

Mutational spectrum and relationship to clinical outcomes: XP-G cases are linked to mutations in the *ERCC5/XPG* gene. A total of 47 alterations resulting in 51 distinct affected protein products have been reported (Figure 2): they include 15 missense changes, four in-frame deletions and 32 truncations caused by either nonsense changes or frameshifts. A complex genomic rearrangement has been also identified (Tables S3, S6 and S7).

In general, XPG missense changes on at least one *ERCC5* allele give rise to XP phenotype whereas truncations on both *ERCC5* alleles lead to severe neurodevelopmental XP/CS or CS phenotypes.

Most of the missense changes are located in the nuclease domains of the XPG protein with a particularly high concentration in the I domain. Missense mutations only impair endonuclease activity required for NER.[39–41] In contrast, truncating mutations also impact on oxidative damage repair[41] and destabilize the XPG interaction with TFIIH.[34]

2.3 | ERCC1 and XPF

In mammalian cells, Excision Repair Cross-Complementation group 1 protein (ERCC1) and xeroderma pigmentosum group F protein (XPF) form an obligate heterodimer with single-stranded structure-specific DNA 5′-endonuclease activity. During the NER process, the ERCC1-XPF complex acts in cooperation with the XPG 3′-endonuclease to perform dual incision around the lesion-containing open DNA bubble. ERCC1-XPF incises the damaged DNA strand 5′ to the lesion thus providing a free 3′-OH end for the polymerase to initiate repair synthesis and replace the excised DNA fragment. XPF contains the catalytic activity whereas ERCC1 is required for DNA binding.[1] Besides NER, ERCC1-XPF has additional functions in other DNA repair processes. It participates in ICLs repair where it acts in different steps from the initial nicking at the site of the crosslink lesion to downstream processing and resolution of intermediates.[42] ERCC1-XPF is also involved in the repair pathways activated by the presence of double-strand breaks (ie, homologous recombination and end-joining) and serves a backup role in processing of oxidative DNA damage during BER.[43] In addition, ERCC1-XPF is involved in telomere length regulation.[44]

ERCC-1 and XP-F cohorts: Three cases, classified as CS, COFS and XP/CS, belong to the ERCC-1 group (Table S4) whereas 29 patients (from 27 unrelated families) have been assigned to the XP-F group (Table S5). The vast majority of the XP-F patients have XP phenotype (20 cases from 18 families) usually characterized by mild cutaneous manifestations and, in two cases, associated with late neurological involvement consistent with HD-like syndrome. On the severe end, the XP-F cohort includes three combined XP/CS presentations, one early-onset CS case and the XFE patient. In addition, three individuals showing FA features and the combined XP/CS/FA case have been also classified in the XP-F group.

Mutational spectra and relationship to clinical outcomes: In the *ERCC1* gene, one missense change, two truncations and one uncharacterized splice mutation have been identified (Figure 2), (Tables S4, S6 and S7). The CS case CS20LO is homozygous for the p.Phe231Leu change. The severe COFS individual 165TOR is instead a compound heterozygote carrying the same p.Phe231Leu change and the premature termination p.Gln158*, whose transcript is probably subjected to nonsense-mediated mRNA decay.[14,45] It has been hypothesized that the relatively milder clinical features observed in the CS compared to the COFS patient, might be because of bi-allelic expression of the mutated *ERCC1* allele that produces twice the level of ERCC1 protein. The third ERCC1-deficient individual displays relatively mild XP/CS and harbors a splice-site mutation that most

FIGURE 2 Alterations in the XPB, XPD, XPG, ERCC1 and XPF proteins. The diagrams show the different proteins with the functional domains in gray or black. In XPB and XPD, I to VI: helicase motifs. In XPB, DRD: DNA damage recognition domain, ThM: Thumb motif. In XPD, 4Fe-S: iron-sulfur-containing domain, ARCH: archaeal domain. In XPG, N and I: regions for nuclease activity, SR: spacer region. In XPG and ERCC1, NLS: nuclear localization signals. In ERCC1 and XPF, HhH2: helix-hairpin-helix motif. Missense changes are indicated by solid lines, with those classified as benign by in silico prediction tools (Table S6) in parenthesis. Truncations are indicated by triangles whereas in-frame deletions are represented by dotted horizontal lines, all below the protein schemes. In XPD, single amino acid deletions (vertical dotted lines above the protein scheme) and insertions or deletion/insertions (white triangles below the protein scheme) are also reported. The complex allele carrying two combined alterations and expressing both p.[L461 V;V716_R730del] and p.[L461 V;A717G][77] is shown boxed. Alterations described as lethal in yeast[27] are highlighted in gray. Circles flanking the alterations indicate the associated clinical phenotypes: XP (close blue), CS or COFS (close red), TTD (close green), FA (close light blue), XFE (yellow), XP/CS (open red), XP/TTD (open green), XP with HD-like (open black), XP/CS/FA (open light blue). The symbol § refers to patients recently reported with suspected UVSS[25] for whom, because of young age, long-term follow-up is needed to confirm clinical diagnosis. In XPD, the asterisks denote the identical combination of mutated alleles (p.[R112H];[D681H]) exceptionally found in two patients with TTD and XP/CS.[28] Each combination of mutated alleles is usually specific for a particular disease (details in the text). Mutation nomenclature follows the format indicated at www.hgvs.org/mutnomen. Protein sequences refer to GenBank NP_000113.1 (XPB), NP_000391.1 (XPD), NP_000114.2 (XPG), NP_001974.1 (ERCC1) and NP_005227.1 (XPF) [Colour figure can be viewed at wileyonlinelibrary.com]

probably allows production of a significant amount of normal protein.[46]

Mutations in the *ERCC4/XPF* gene are responsible for the XP-F patients. A total of 27 variants have been identified including 16 missense changes, one in-frame deletion and 10 truncations (Figure 2). A null mutation leading to unexpressed transcript has been also reported in one case (Tables S5, S6 and S7). Critical determinants in the clinical manifestations are the impact of the variants on protein expression, function and subcellular localization. Based on these features, *ERCC4* variants can be categorized into three groups. The first set of *ERCC4* variants results in reduced level of nuclear protein because the altered XPF has a tendency to aggregate in the cytoplasm. The reduced level of nuclear XPF is insufficient to perform complete NER, but it is still sufficient for ICLs repair. All these mutations are associated with the XP phenotype.[47] The second set of *ERCC4* variants results in the nuclear localization of the mutated proteins that exert a sufficient, although not optimal, NER activity but are fully deficient in ICLs repair. These mutations result in the FA phenotype.[15,16] The last category of *ERCC4* variants is characterized by very low levels of nuclear XPF, apparently insufficient to support either NER or ICLs repair. These mutations are associated with severe CS, XFE and combined presentations.[13,14]

3 | MOLECULAR BASIS OF OVERLAP SYNDROMES

3.1 | XP/CS overlap

The molecular basis of XP is explained by defects in GG-NER that cause a genome-wide accumulation of DNA lesions. Trans-lesion DNA polymerases can efficiently bypass the damage allowing completion of DNA replication in an error-prone manner. Lesion bypass therefore favors cell survival but increases mutagenesis, thus explaining the cancer predisposition typical of XP. Differently, the molecular basis of the CS phenotype has been explained by defects in TC-NER that cause the accumulation of transcription complexes stalled by DNA lesions. Transcription is a vital function for the cell and TC-NER deficiency can be expected to result in cell death, neurodegeneration and premature aging. Despite the observation that defects in the TC-NER factors CSA and CSB (Table 1) typically result in neurodevelopmental and cancer-free CS phenotype, this rationale cannot fully explain the XP-only phenotype observed in XP-A, XP-B, XP-D, XP-G and XP-F patients, who lack of functional GG-NER as well as TC-NER. As previously discussed, only some specific alterations in XPB, XPD, XPG and XPF are associated with the XP/CS overlap. The causative relationship between mutations and the CS clinical features in XP/CS cases is complex and must not only involve the NER defects but also the other functions of the NER proteins. As described below, several studies have demonstrated that transcription impairment, oxidative repair and energy metabolism alteration as well as genotoxic stress may explain the combined XP/CS phenotype. Patients with defects in the XPA protein, that has no known role outside NER, may exhibit accelerated neurodegeneration in the second decade of life which might suggest a link between TC-NER defects and neuronal loss.[48,49]

Transcription impairment: The hypothesis of a transcription defect underlying CS features was introduced more than 20 years ago, based on that time observation that combined XP/CS presentations were often related to alterations in the XPB and XPD subunits of the transcription factor TFIIH.[50] Since then, studies have indeed shown transcriptional deregulations in CS[51–53] and different molecular mechanisms have been put forward to explain the XP/CS phenotype. In 2007, the CAK sub-complex dissociation from TFIIH was proposed as a molecular marker of XP/CS by the group of Tanaka.[34] This model was based on the observation that cells from XP/CS patients mutated in *ERCC5/XPG* (XP-G/CS) or *ERCC2/XPD* (XP-D/CS) are characterized by an altered assembly state of the TFIIH complex that results in free CAK. XPG interaction with TFIIH facilitates the anchoring of the CAK to the core-TFIIH thus favoring the phosphorylation and the transactivation of nuclear receptors such as estrogen receptor α (ERα) and peroxisome proliferator-activated receptor γ (PPARγ2).[34] The lack of activation of hormone-specific signaling pathways may explain multisystem clinical features typical of CS. No CAK dissociation from TFIIH has however been observed in XP-B/CS cells.[34]

In addition to TFIIH, other members of the NER machinery appear to play a role in transcription. Several NER factors, including ERCC1-XPF and XPG endonucleases, are recruited on promoters of active genes in basal conditions. Their presence at the promoters is necessary to achieve optimal DNA demethylation and histone post-translational modifications, thus contributing to efficient transcription.[36] It was shown that NER factors are not essential for transcription initiation but they probably contribute to fine-tuning the transcription process. Their action might become relevant in specific tissues at specific stages of development and therefore potentially explain the neurological and developmental abnormalities associated with the XP/CS phenotype or the neurological problems observed in some XP patients.

Notably, compared to XP-D/XP cells, the XP-D/CS cells show a gross transcriptional dysregulation following UV irradiation. Using housekeeping genes as a model system, it has been demonstrated that XP-D/CS cells are unable to reassemble the promoters of these genes and restart transcription after UV irradiation. Promoter repression does not appear to be a consequence of deficient repair but rather an active repression process because of dysregulation of the histone deacetylase Sirt1 that leads to promoter heterochromatinization.[54] Many of the genes whose expression is impaired play a role in multiple pathways, including genomic stability and neuronal development. This may help to explain the severe XP-D/CS phenotype. Transcriptional dysregulation upon UV irradiation is also observed in XP-B/CS and XP-G/CS cells, although apparently not related to Sirt1 impairment.[54] Interestingly, it has been recently shown that UV irradiation persistently arrests transcription of about 70% of genes in CS patient cells defective in either CSA or CSB, because of the constitutive presence of the ATF3 repressor.[55]

An additional model suggests that the molecular alterations underlying XP/CS originate from the persistent binding of TFIIH to the DNA bubble during the NER process, which results in a long-lasting unfinished NER intermediate.[56] According to this model, in XP/CS cells, TFIIH is recruited to the lesion site where it opens the DNA double helix but the subsequent incision by XPG and ERCC1-XPF is somehow prevented leaving an aberrant repair intermediate without further processing. This would sequester TFIIH away from the transcription sites and consequently impair transcription resumption. The "aberrant NER intermediate" hypothesis is supported by the observation that a specific *ERCC3/XPB* mutation associated with the XP/CS phenotype impairs ERCC1-XPF incision and repair DNA synthesis, thus resulting in a long-lasting NER intermediate.[57] Furthermore, *ERCC5/XPG* mutations leading to XP/CS are able to promote TFIIH assembly and the formation of the DNA bubble, but the subsequent recruitment of DNA synthesis factors is prevented.[58] Persistent single strand-displacement at the sites of UV lesions has been also observed in XP-D/CS cells.[59] Finally, an XP/CS specific mutation in *ERCC4/XPF* was shown to cause persistent recruitment of XPF together with the upstream core NER machinery to DNA damage[60] suggesting that the persistence of NER factors at lesion sites might be a contributing factor in the XP/CS molecular pathogenesis.

Oxidative and energy metabolism alterations: Besides processing the UV-induced DNA lesions, some NER proteins are also involved in the repair of endogenous damage caused by the reactive oxygen species (ROS) carried out by the BER pathway. The TC-NER factors CSA and CSB, as well as the XPG and ERCC1-XPF core-NER endonucleases, have been shown to promote the BER activity and protect genomic DNA from ROS.[37,61,62] Interestingly, investigations on XP-G patient cells have shown that XP/CS cases are hypersensitive to UV light and also to oxidative stress whereas XP-only cases are only sensitive to UV.[41] These observations support a causal relationship between oxidative alterations and CS clinical features in overlapping phenotypes.

Mitochondria function as the "powerplant" of the cells and supply ATP by oxidative phosphorylation, a process that induces the production of high levels of ROS. As a consequence, mitochondrial DNA (mtDNA) is constantly exposed to oxidative stress leading to damage accumulation. To protect from ROS-induced damage, mitochondria hold an independent BER machinery whose components are coded by nuclear genes. Increasing evidence is demonstrating the contribution of NER factors to mitochondrial BER.[63] XPD, CSA and CSB have been found to localize in mitochondria where the last two proteins have been shown to interact with the BER enzymes.[24,64,65] Interestingly, primary fibroblasts from CS patients are characterized by increased mitochondrial content, altered membrane potential and increased ROS production ([66]and references therein). Of note, recent data indicate that mitochondrial alteration resulting from the loss of CSA or CSB might be related to stalled ribosomal DNA transcription.[67] Even more, it has been shown that in CS fibroblasts a nuclear transcription deregulation results in increased levels of the HTRA3 serine protease which degrades mitochondrial DNA polymerase gamma, thus, impairing mitochondrial function.[68]

Mitochondrial DNA damage is thought to be central to the aging process and mtDNA mutations are known to cause a variety of diseases with significant similarity to the neurodegenerative phenotype of CS.[69] Mitochondrial dysfunction and mtDNA damage are also found in the neurological condition HD.[70] Interestingly, a few XP cases fulfilling the criteria of HD-like disorders have been recently reported in the XP-B, XP-F and XP-G groups,[4,5] indicating that the NER pathway might be of general importance for mitochondrial maintenance.

Genotoxic stress: A connection between genotoxic stress because of generalized unrepaired DNA damage and premature aging features typical of CS is highlighted by studies in the mouse model for ERCC1-XPF deficiency. The *Ercc1*$^{\Delta/-}$ mutant containing one knockout allele and one hypomorphic *Ercc1* allele, in which most of the DNA repair functions are severely compromised but not completely absent, exhibits a remarkably wide range of pathological and physiological features related to accelerated aging, including progressive neurodegeneration, osteoporosis, cachexia (all commonly observed in CS patients), liver and kidney dysfunction (observed in XFE), and early hematopoietic stem cell exhaustion (as in FA). Expression data from *Ercc1*$^{\Delta/-}$ mice indicate increased cell death paralleled by suppression of the growth hormone/insulin-like growth factor 1 (GH/IGF1) axis, a known regulator of lifespan.[13] In addition, suppression of the GH/IGF1 axis is also observed in the mouse model defective for both Csb and Xpa functions, whose phenotype largely mimics the CS clinical features.[71] Of note, a recent study in the *Ercc1*$^{\Delta/-}$ mice has shown that these animals accumulate oxidative DNA damage more rapidly than age-matched wild-type mice and that oxidative stress appears to be due, at least in part, to mitochondrial-derived ROS.[72]

3.2 | XP/TTD overlap

Only four cases with the XP/TTD overlapping phenotype are known.[10,11] These patients show the typical TTD hair defects, even though it is less prominent than in TTD patients. In addition, XP/TTD patients are predisposed to skin cancer formation but with lower frequency than XP patients. As the vast majority of the "pure" TTD cases, XP/TTD patients are linked to mutations in the *ERCC2/XPD* gene. They are all compound heterozygotes: one individual carries peculiar mutations on both *ERCC2* alleles whereas the remaining three have a specific mutation combined with a typical TTD mutation (Table S2).

As a subunit of the TFIIH complex, XPD is involved in both NER and transcription. Growing evidence supports the notion that subtle transcription defects explain the TTD typical clinical features, such as hypoplasia of the adipose tissue, neurodevelopmental defects, bone and cartilage alterations.[20,73–75] The NER defect explains skin photosensitivity that, however, does not result in increased carcinogenesis. The prevailing explanation for the lack of skin cancers in TTD is that transcriptional deficiencies may prevent some crucial steps of carcinogenesis. However, the XP/TTD individuals do have features of both TTD and XP. A completely speculative explanation for this perplexing phenotype is that, a mild transcription defect not interfering with the

carcinogenic process could result in the partial TTD features combined with XP.[10] Concerning DNA repair ability, marked differences have been observed between XP and TTD cells in the recruitment and distribution of NER factors at sites of UV-induced DNA damage. Of interest, XP/TTD cells appear to show partial XP and TTD features.[11]

4 | CONCLUDING REMARKS

XP, UVSS, TTD, CS and COFS are distinct clinical entities whose etiopathogenesis involves alterations of the NER pathway and, in a number of cases, resulting from mutations in the same gene(s). In very rare cases, patients have features of more than one disorder. Symptoms resembling other disease categories, such as FA and HD, are also sporadically found in combined presentations.

It is now quite well established that such a complex phenotypic spectrum is because of the additional functions that the underlying genes have besides NER. The clinical outcome seems to depend on the balance between the degree to which NER and non-NER functional roles are disturbed. Much is known about the critical role of transcription impairment behind the heterogeneity of NER disorders and increasing evidence supports the relevance of oxidative damage repair alterations. The potential role of new functions, such as energy metabolism, is also emerging. Drawing a direct correspondence between clinical entities and given molecular dysfunctions is nevertheless complicated because of the blurred phenotypic boundaries of these syndromes. Heterogeneity and overlaps make the clinical diagnosis of NER disorders challenging. The advent of high throughput technologies has enabled rapid identification of disease-causing variations in patients with typical presentations but also in non-suspected conditions. The screening of the 46 NER genes in Table 1 must be encouraged for those cases with an NER-like phenotype. In addition, coupling with functional tests and detailed clinical evaluation is essential to provide reliable observations and hence establish a causal relationship between molecular defects and phenotype. Deep phenotyping has recently identified previously unrecognized features in some XP groups, stressing how integration of basic science with clinic is important to elucidate the remaining puzzles of NER disorders.[76]

ACKNOWLEDGEMENTS

We are grateful to clinicians and patients for their cooperation. Our work was supported by the Associazione Italiana per la Ricerca sul Cancro Grant IG 17710.

CONFLICT OF INTEREST

Nothing to declare.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Elena Botta https://orcid.org/0000-0003-2060-0110

REFERENCES

1. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*. 2014;15:465-481.
2. Spivak G. Nucleotide excision repair in humans. *DNA Repair*. 2015;36:13-18.
3. Lehmann AR, McGibbon D, Stefanini M. Xeroderma pigmentosum. *Orphanet J Rare Dis*. 2011;6:70.
4. Carré G, Marelli C, Anheim M, et al. Xeroderma pigmentosum complementation group F: a rare cause of cerebellar ataxia with chorea. *J Neurol Sci*. 2017;376:198-201.
5. Garcia-Moreno H, Fassihi H, Sarkany RPE, et al. Xeroderma pigmentosum is a definite cause of Huntington's disease-like syndrome. *Ann Clin Transl Neurol*. 2018;5:102-108.
6. Spivak G. UV-sensitive syndrome. *Mutat Res*. 2005;577:162-169.
7. Stefanini M, Botta E, Lanzafame M, Orioli D. Trichothiodystrophy: from basic mechanisms to clinical implications. *DNA Repair*. 2010;9:2-10.
8. Calmels N, Botta E, Jia N, et al. Functional and clinical relevance of novel mutations in a large cohort of patients with Cockayne syndrome. *J Med Genet*. 2018;55:329-343.
9. Laugel V, Dalloz C, Tobias ES, et al. Cerebro-oculo-facio-skeletal syndrome: three additional cases with CSB mutations, new diagnostic criteria and an approach to investigation. *J Med Genet*. 2008;45:564-571.
10. Broughton BC, Berneburg M, Fawcett H, et al. Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene. *Hum Mol Genet*. 2001;10:2539-2547.
11. Boyle J, Ueda T, Oh K-S, et al. Persistence of repair proteins at unrepaired DNA damage distinguishes diseases with ERCC2 (XPD) mutations: cancer-prone xeroderma pigmentosum vs. non-cancer-prone trichothiodystrophy. *Hum Mutat*. 2008;29:1194-1208.
12. Natale V, Raquer H. Xeroderma pigmentosum-Cockayne syndrome complex. *Orphanet J Rare Dis*. 2017;12:65.
13. Niedernhofer LJ, Garinis GA, Raams A, et al. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature*. 2006;444:1038-1043.
14. Kashiyama K, Nakazawa Y, Pilz DT, et al. Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *Am J Hum Genet*. 2013;92:807-819.
15. Bogliolo M, Schuster B, Stoepker C, et al. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet*. 2013;92:800-806.
16. Popp I, Punekar M, Telford N, et al. Fanconi anemia with sun-sensitivity caused by a Xeroderma pigmentosum-associated missense mutation in XPF. *BMC Med Genet*. 2018;19:7.
17. Compe E, Egly JM. Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annu Rev Biochem*. 2016;85:265-290.
18. Rimel JK, Taatjes DJ. The essential and multifunctional TFIIH complex. *Protein Sci*. 2018;27:1018-1037.

19. Oksenych V, de Jesus BB, Zhovmer A, Egly JM, Coin F. Molecular insights into the recruitment of TFIIH to sites of DNA damage. *EMBO J*. 2009;28:2971-2980.
20. Orioli D, Compe E, Nardo T, et al. XPD mutations in trichothiodystrophy hamper collagen VI expression and reveal a role of TFIIH in transcription derepression. *Hum Mol Genet*. 2013;22:1061-1073.
21. Chen J, Larochelle S, Li X, Suter B. Xpd/Ercc2 regulates CAK activity and mitotic progression. *Nature*. 2003;424:228-232.
22. Ito S, Tan LJ, Andoh D, et al. MMXD, a TFIIH-independent XPD-MMS19 protein complex involved in chromosome segregation. *Mol Cell*. 2010;39:632-640.
23. Yeom E, Hong S-T, Choi K-W. Crumbs and Xpd in mitosis. *Oncoscience*. 2015;2:821-822.
24. Liu J, Fang H, Chi Z, et al. XPD localizes in mitochondria and protects the mitochondrial genome from oxidative DNA damage. *Nucleic Acids Res*. 2015;43:5476-5488.
25. Calmels N, Greff G, Obringer C, et al. Uncommon nucleotide excision repair phenotypes revealed by targeted high-throughput sequencing. *Orphanet J Rare Dis*. 2016;11:26.
26. Oh K-S, Khan SG, Jaspers NGJ, et al. Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Hum Mutat*. 2006;27:1092-1103.
27. Taylor EM, Broughton BC, Botta E, et al. Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene. *Proc Natl Acad Sci U S A*. 1997;94:8658-8663.
28. Schäfer A, Gratchev A, Seebode C, et al. Functional and molecular genetic analyses of nine newly identified XPD-deficient patients reveal a novel mutation resulting in TTD as well as in XP/CS complex phenotypes. *Exp Dermatol*. 2013;22:486-489.
29. Fan L, Fuss JO, Cheng QJ, et al. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*. 2008;133:789-800.
30. Liu H, Rudolf J, Johnson KA, et al. Structure of the DNA repair helicase XPD. *Cell*. 2008;133:801-812.
31. Wolski SC, Kuper J, Hänzelmann P, et al. Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol*. 2008;6:e149.
32. Botta E, Nardo T, Lehmann AR, Egly JM, Pedrini AM, Stefanini M. Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy. *Hum Mol Genet*. 2002;11:2919-2928.
33. Sarker AH, Tsutakawa SE, Kostek S, et al. Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne syndrome. *Mol Cell*. 2005;20:187-198.
34. Ito S, Kuraoka I, Chymkowitch P, et al. XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Mol Cell*. 2007;26:231-243.
35. Narita T, Narita K, Takedachi A, Saijo M, Tanaka K. Regulation of transcription elongation by the XPG-TFIIH complex is implicated in Cockayne syndrome. *Mol Cell Biol*. 2015;35:3178-3188.
36. Le May N, Mota-Fernandes D, Vélez-Cruz R, et al. NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell*. 2010;38:54-66.
37. Bessho T. Nucleotide excision repair 3′ endonuclease XPG stimulates the activity of base excision repairenzyme thymine glycol DNA glycosylase. *Nucleic Acids Res*. 1999;27:979-983.
38. Trego KS, Groesser T, Davalos AR, et al. Non-catalytic roles for XPG with BRCA1 and BRCA2 in homologous recombination and genome stability. *Mol Cell*. 2016;61:535-546.
39. Emmert S, Slor H, Busch DB, et al. Relationship of neurologic degeneration to genotype in three xeroderma pigmentosum group G patients. *J Invest Dermatol*. 2002;118:972-982.
40. Moriwaki S, Takigawa M, Igarashi N, et al. Xeroderma pigmentosum complementation group G patient with a novel homozygous missense mutation and no neurological abnormalities. *Exp Dermatol*. 2012;21:304-307.
41. Soltys DT, Rocha CRR, Lerner LK, et al. Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. *Hum Mutat*. 2013;34:481-489.
42. Dehé P-M, Gaillard P-HL. Control of structure-specific endonucleases to maintain genome stability. *Nat Rev Mol Cell Biol*. 2017;18:315-330.
43. Manandhar M, Boulware KS, Wood RD. The ERCC1 and ERCC4 (XPF) genes and gene products. *Gene*. 2015;569:153-161.
44. Wan B, Yin J, Horvath K, et al. SLX4 assembles a telomere maintenance toolkit by bridging multiple endonucleases with telomeres. *Cell Rep*. 2013;4:861-869.
45. Jaspers NGJ, Raams A, Silengo MC, et al. First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am J Hum Genet*. 2007;80:457-466.
46. Imoto K, Boyle J, Oh K, et al. Patients with defects in the interacting nucleotide excision repair proteins ERCC1 or XPF show xeroderma pigmentosum with late onset severe neurological degeneration. *J Investig Dermatol*. 2007;127:S92-S92.
47. Ahmad A, Enzlin JH, Bhagwat NR, et al. Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet*. 2010;6:e1000871.
48. Andrews AD, Barrett SF, Robbins JH. Xeroderma pigmentosum neurological abnormalities correlate with colony-forming ability after ultraviolet radiation. *Proc Natl Acad Sci U S A*. 1978;75:1984-1988.
49. Robbins JH, Brumback RA, Mendiones M, et al. Neurological disease in Xeroderma-Pigmentosum-documentation of a late onset type of the juvenile onset form. *Brain*. 1991;114:1335-1361.
50. Bootsma D, Hoeijmakers JHJ. DNA repair. Engagement with transcription. *Nature*. 1993;363:114-115.
51. Brooks PJ. Blinded by the UV light: how the focus on transcription-coupled NER has distracted from understanding the mechanisms of Cockayne syndrome neurologic disease. *DNA Repair*. 2013;12:656-671.
52. Koch S, Garcia Gonzalez O, Assfalg R, et al. Cockayne syndrome protein a is a transcription factor of RNA polymerase I and stimulates ribosomal biogenesis and growth. *Cell Cycle*. 2014;13:2029-2037.
53. Alupei MC, Maity P, Esser PR, et al. Loss of Proteostasis is a Pathomechanism in Cockayne syndrome. *Cell Rep*. 2018;23:1612-1619.
54. Vélez-Cruz R, Zadorin AS, Coin F, et al. Sirt1 suppresses RNA synthesis after UV irradiation in combined xeroderma pigmentosum group D/Cockayne syndrome (XP-D/CS) cells. *Proc Natl Acad Sci U S A*. 2013;110:E212-E220.
55. Epanchintsev A, Costanzo F, Rauschendorf M-A, et al. Cockayne's syndrome a and B proteins regulate transcription arrest after genotoxic stress by promoting ATF3 degradation. *Mol Cell*. 2017;68:1054-1066.e6.
56. Moriel-Carretero M, Herrera-Moyano E, Aguilera A. A unified model for the molecular basis of Xeroderma pigmentosum-Cockayne syndrome. *Rare Dis*. 2015;3:e1079362.
57. Oh K-S, Imoto K, Boyle J, Khan SG, Kraemer KH. Influence of XPB helicase on recruitment and redistribution of nucleotide excision repair proteins at sites of UV-induced DNA damage. *DNA Repair*. 2007;6:1359-1370.
58. Mocquet V, Lainé JP, Riedl T, Yajin Z, Lee MY, Egly JM. Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J*. 2007;27:155-167.
59. Godon C, Mourgues S, Nonnekens J, et al. Generation of DNA single-strand displacement by compromised nucleotide excision repair. *EMBO J*. 2012;31:3550-3563.

60. Sabatella M, Theil AF, Ribeiro-Silva C, et al. Repair protein persistence at DNA lesions characterizes XPF defect with Cockayne syndrome features. *Nucleic Acids Res*. 2018;46:9563-9577.
61. D'Errico M, Pascucci B, Iorio E, et al. The role of CSA and CSB protein in the oxidative stress response. *Mech Ageing Dev*. 2013;134:261-269.
62. Fisher LA, Samson L, Bessho T. Removal of reactive oxygen species-induced 3′-blocked ends by XPF-ERCC1. *Chem Res Toxicol*. 2011;24:1876-1881.
63. Hosseini M, Ezzedine K, Taieb A, Rezvani HR. Oxidative and energy metabolism as potential clues for clinical heterogeneity in nucleotide excision repair disorders. *J Invest Dermatol*. 2015;135:341-351.
64. Aamann MD, Sorensen MM, Hvitby C, et al. Cockayne syndrome group B protein promotes mitochondrial DNA stability by supporting the DNA repair association with the mitochondrial membrane. *FASEB J*. 2010;24:2334-2346.
65. Kamenisch Y, Fousteri M, Knoch J, et al. Proteins of nucleotide and base excision repair pathways interact in mitochondria to protect from loss of subcutaneous fat, a hallmark of aging. *J Exp Med*. 2010;207:379-390.
66. Pascucci B, D'Errico M, Romagnoli A, et al. Overexpression of parkin rescues the defective mitochondrial phenotype and the increased apoptosis of Cockayne syndrome a cells. *Oncotarget*. 2017;8:102852-102867.
67. Scheibye-Knudsen M, Tseng A, Borch Jensen M, et al. Cockayne syndrome group a and B proteins converge on transcription-linked resolution of non-B DNA. *Proc Natl Acad Sci U S A*. 2016;113:12502-12507.
68. Chatre L, Biard DSF, Sarasin A, Ricchetti M. Reversal of mitochondrial defects with CSB-dependent serine protease inhibitors in patient cells of the progeroid Cockayne syndrome. *Proc Natl Acad Sci U S A*. 2015;112:E2910-E2919.
69. Karikkineth AC, Scheibye-Knudsen M, Fivenson E, Croteau DL, Bohr VA. Cockayne syndrome: clinical features, model systems and pathways. *Ageing Res Rev*. 2017;33:3-17.
70. Jeppesen DK, Bohr VA, Stevnsner T. DNA repair deficiency in neurodegeneration. *Prog Neurobiol*. 2011;94:166-200.
71. van der Pluijm I, Garinis GA, Brandt RMC, et al. Impaired genome maintenance suppresses the growth hormone--insulin-like growth factor 1 axis in mice with Cockayne syndrome. *PLoS Biol*. 2007;5:e2.
72. Robinson AR, Yousefzadeh MJ, Rozgaja TA, et al. Spontaneous DNA damage to the nuclear genome promotes senescence, redox imbalance and aging. *Redox Biol*. 2018;17:259-273.
73. Compe E, Drané P, Laurent C, et al. Dysregulation of the peroxisome proliferator-activated receptor target genes by XPD mutations. *Mol Cell Biol*. 2005;25:6065-6076.
74. Compe E, Malerba M, Soler L, Marescaux J, Borrelli E, Egly JM. Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH. *Nat Neurosci*. 2007;10:1414-1422.
75. Arseni L, Lanzafame M, Compe E, et al. TFIIH-dependent MMP-1 overexpression in trichothiodystrophy leads to extracellular matrix alterations in patient skin. *Proc Natl Acad Sci U S A*. 2015;112:1499-1504.
76. Fassihi H, Sethi M, Fawcett H, et al. Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc Natl Acad Sci U S A*. 2016;113:E1236-E1245.
77. Horibata K, Kono S, Ishigami C, et al. Constructive rescue of TFIIH instability by an alternative isoform of XPD derived from a mutated XPD allele in mild but not severe XP-D/CS. *J Hum Genet*. 2015;60:259-265.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ferri D, Orioli D, Botta E. Heterogeneity and overlaps in nucleotide excision repair disorders. *Clin Genet*. 2019;1–13. https://doi.org/10.1111/cge.13545