UNIVERSITÉ DE STRASBOURG

Ecole Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE « des Sciences de la Vie et de la Santé »

IGBMC - CNRS UMR 7104 - Inserm U 964

THÈSE présentée par :

Anton ZADORIN

soutenue le : **28 Septembre 2012**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences du Vivant/ Aspects Moléculaires et Cellulaires de la Biologie

Le complexe TFIIH dans la transcription effectuée par l'ARN polymérase II et l'ARN polymérase III

THÈSE dirigée par :
Dr EGLY Jean-Marc — DR INSERM, IGBMC, Strasbourg

RAPPORTEURS :
Dr DAEGELEN Dominique — DR INSERM, Institut Cochin, Paris
Dr SCHUMACHER Björn — DR, université de Cologne
Pr LAUGEL Vincent — PUPH, CHU de Strasbourg

To Rexy

Contents

Résumé de la thèse de doctorat en français

Le complexe TFIIH dans la transcription effectuée par Pol II et Pol III

Dans l'organisme des mammifères, il existe trois ARN polymérases ADN-dépendantes principales designées comme Pol I, Pol II et Pol III. La première transcrit la plupart des gènes de l'ARNr, la deuxième est responsable de la transcription des gènes codant les protéines et de certains gènes de l'ARN non codant, et la troisième est nécessaire pour la synthèse de l'ARNr 5S, des ARNt et quelques autres petits ARN non codants. Les gènes eux-mêmes sont groupés dans un classe I, classe II ou classe III selon la polymérase qui les transcrit. Le TFIIH est un complexe protèique qui, au début, a été caractérisé comme un facteur général de transcription de Pol II. Plus tard, il a été montré participer au mécanisme de réparation par excision de nucléotides (NER), ainsi qu'être un facteur transcriptionnel de Pol I [44]. Il est constitué d'un cœur contenant les sous-unités XPB, p62, p52, p44, p34, p8, qui est relié par la sous-unité XPD au sous-complexe CAK composé de CDK7, cycline H et MAT1. Certaines mutations des protéines de TFIIH provoquent des maladies génétiques rares avec un risque 1000 fois plus élevé que la normale de développer un cancer (*xeroderma pigmentosum* ou XP) ou/et des anomalies sévéres du développement (le syndrome de Cockayne ou CS, la trichothiodystrophie, le syndrome COFS). Un petit nombre de mutations spécifiques dans les génes *XPB* et *XPD* cause le phénotype combiné XP/CS [4, 9, 26, 82]. La sévérité du XP/CS est déterminée par le lien inhérent entre la transcription et la NER. XPB et XPD sont deux hélicases du complexe TFIIH, et leur activité (à différents degrés) est nécessaire pour la fusion de l'ADN au cours de la NER et pour l'ouverture du promoteur dans la transcription de Pol I

et de Pol II. Bien que la contribution de la carence de la réparation de l'ADN au phénotype du XP/CS soit irréfutable, ces mutations causent la dérégulation de plusieurs voies transcriptionnelles. Au niveau cellulaire, les cellules XP/CS partagent le même arrêt global transcriptionnel soutenu après l'irradiation UV [9]. Pour les cellules CS, il a été démontré que cet arrêt provenait d'un probléme de remodélisation de la chromatine [18, 69, 27].

Partie I. L'analyse moléculaire des mutations dans le XPD liées au phénotype XP-D/CS

L'objectif de ces recherches fût de mieux comprendre comment les cellules réinitient la transcription après une irradiation UV, et d'étudier le mécanisme de l'arrêt global dans les cellules avec les mutations spécifiques dans le gène *XPD* provoquant le phénotype XP-D/CS. En utilisant la méthode de transcription inverse suivie d'une PCR quantitative, nous avons montré que les cellules XP-D/CS ne pouvaient pas ré-initient la transcription des génes constitutifs tels que *DHFR* ou *GAPDH*, après une irradiation UV. D'autre part, les cellules de type sauvage (WT) et les cellules NER-défectueuses possédant seulement les traits du XP étaient capables de le faire. Nous avons aussi observé que les cellules XP-D/CS étaient incapables de transactiver les gènes induits par les récepteurs nucléaires (NR) aprés une irradiation UV, alors que les autres cellules NER-défectueuses transactivaient les mêmes gènes dans les mêmes conditions. Par contre, la transcription du gène *GADD45α* (p53-dépendant), ainsi que d'autres gènes induits par le stress (*ATF3*, *p21*, et *MDM2*), n'est pas inhibée dans les cellules XP-D/CS. Cette conclusion était confirmée à partir de séquençage haut débit de transcriptome (RNA-seq). Tandis que les cellules WT, 24 heures après une irradiation UV, recouvraient presque entièrement le profil initial de transcription, et un mutant XPD possédant seulement les traits XP faisait cela d'une façon moins effective, la transcription dans les cellules XP-D/CS restait complètement déréglée avec une partie considérable de gènes dont l'expression était élevée. Nos résultats indiquent que l'arrêt transcriptionnel des gènes constitutifs dans le phénotype XP-D/CS peut être non seulement un sous-produit de l'absence de réparation de l'ADN mais aussi une conséquence du blocage de l'ARN pol II par les lésions de l'ADN comme cela était supposé auparavant.

A cet effet, nous étions intéressés à savoir comment les mutations dans XPD perturbaient la formation du complexe transcriptionnel. Pour

cela, nous avons étudié le recrutement des facteurs de transcription sur le promoteur *DHFR*, en utilisant la technique d'immunoprécipitation de la chromatine (ChIP) suivie d'une PCR quantitative. Dans les cellules WT, Pol II et tous les facteurs basals se dissocient rapidement du promoteur après irradiation et sont de nouveau recrutés 6 heures après l'irradiation UV. Dans les cellules XP-D/CS, la présence de Pol II sur le promoteur *DHFR* atteint 30% de celle initiale après 12 heures, et ne retrouve pas un niveau normal même 24 heures après irradiation. Cela nous a incité à examiner les changements dans les modifications de chromatine sur les promoteurs des gènes constitutifs dans les cellules XP-D/CS. L'analyse par ChIP des cellules WT nous a permis de découvrir que le promoteur *DHFR* accumulait les modifications H3K9ac, H4K16ac, H3K4me3 et H3K79me2 après irradiation. Ces marques positives de transcription sont en accord avec la présence de Pol II sur ces gènes. Au contraire, le promoteur *DHFR* dans les cellules XP-D/CS possède un niveau bas de certaines de ces marques chromatiniennes à mettre en parallèle avec l'absence Pol II et le niveau faible d'ARNm. De la même façon, nous avons détecté l'accumulation de marques d'hétérochromatine comme H3K9me2 et l'histone H1 sur le promoteur *DHFR* après l'UV, alors que dans les cellules WT, leur niveau ne soit pas élevé.

Puisque H3K9ac et H4K16ac sont des substrats pour la déacétylase SIRT1 [83], nous avons étudié sa présence sur le promoteur *DHFR* par ChIP. Nous avons détecté un niveau élevé de SIRT1 sur ce promoteur dans les cellules XP-D/CS aprs l'UV, alors que dans les cellules WT, l'enrichissement de la SIRT1 était insignifiant. L'inhibition de la déacétylase Sirt1 (à l'aide de différents inhibiteurs chimiques et de l'extinction du gène *SIRT1*) a permis de recouvrer l'expression de l'ARNm du gène *DHFR* après l'UV à un niveau comparable à celui des cellules WT. De plus, l'analyse ChIP a montré que l'inhibition de la SIRT1 dans les cellules XP-D/CS conduisait à l'augmentation du niveau des marques d'euchromatine H4K16ac et H3K79me2, par rapport aux cellules non inhibées. L'inhibition de la SIRT1 dans les cellules XP-D/CS conduisait aussi au rétablissement du niveau de Pol II sur ce promoteur après l'UV. L'analyse RNA-seq a démontré d'une façon indépendante le recouvrement de la transcription du *DHFR* 24 heures après irradiation dans les cellules XP-D/CS, ainsi que le recouvrement de l'expression de 400 autres gènes environ (à différents degrés).

L'ensemble de ces résultats nous a permis d'émettre l'hypothèse que SIRT1 était responsable de la génération d'hétérochromatine sur les promoteurs des gènes constitutifs dans les cellules XP-D/CS après irradia-

tion. Le lien entre SIRT1 et les mutations XP-D/CS a des implications importantes. Les homologues de SIRT1 dans les levures, les vers et les mouches régulent différents processus cellulaires ; sa surexpression est associée à une durée de vie prolongée chez les souris, alors que son insuffisance augmente l'instabilité génomique et réduit la durée de vie [24]. Par conséquent, la dérégulation de la fonction normale de SIRT1 peut contribuer à la progéria sévère, aux anomalies métaboliques et à d'autres manifestations cliniques des patients XP-D/CS, qui ne pourraient pas être expliquées par le défaut de la réparation de l'ADN.

Partie II. Le complexe TFIIH et la transcription par Pol III

La TBP caractérisée au début comme un composant du complexe transcriptionnel de Pol II, a été aussi montrée participer à la transcription par Pol I et Pol III [22]. Récemment, il a été démontré que le TFIIH jouait un rôle essentiel à la transcription par Pol I [44]. De tels exemples de facteurs communs entre les différents ARN polymérases nous ont incités à étudier la participation du TFIIH à la transcription dans tout le génome. De plus, nous voulions discriminer les gènes potentiellement existants dont la transcription n'exige pas l'activité de la kinase CDK7. Il est connu que cette kinase phosphoryle le domaine CTD de Pol II (qui n'est pas présent dans Pol I et Pol III) et certains NRs.

Pour cela, nous avons effectué un séquençage haut débit des échantillons de ChIP (ChIP-seq) contre le cœur et la partie CAK du TFIIH dans les fibroblastes normales humains. Il était inattendu pour nous de détecter que la partie de génome la plus occupée par le cœur était représentée par les gènes de classe III, incluant une fraction considérable des gènes ARNt, un pseudogène putatif, *RNU6*, *RN7SK*, *RNY1*, *RNY3*, ainsi que certains locus non annotés mais évolutionnellement conservés et analogues à ceux du profil ChIP-seq. En fait, de 512 gènes de l'ARNt listés dans la base de données GtRNAdb pour *Homo sapiens* [17] au moins 174 sont associés à au niveau élevé d'enrichissement par le cœur de TFIIH.

Par contre, les composants du CAK et la forme phosphorylée de Pol II n'ont pas été détectés sur ces sites. D'autres équipes [5, 70] ont réussi à identifier par ChIP-seq la présence de Pol II sur certains gènes de classe III, ainsi que certains autres facteurs transcriptionnels et des modifications d'histones associées aux sites de transcription active. Cependant, il n'était pas clair si TFIIH était lié d'une certaine maniére avec l'activité de Pol III, ou si tout la machinerie de Pol II participait à

la régulation de la transcription des génes de classe III.

Par l'analyse ChIP-seq contre Pol III, nous avons détecté que TFIIH et Pol III co-occupaient les mêmes gènes de classe III. Pour valider ces résultats, nous avons effectué un ChIP conventionnel suivie d'une PCR quantitative pour quelques génes d'ARNt sélectionnés. L'occupation de ces gènes a été confirmée pour les sous-unités du cœur (p44, p52, p62 et XPB), mais nous n'avons pas réussi à détecter CDK7. En même temps, Pol III a été clairement démontrés être présents sur ces sites. En ce qui concerne Pol II, sa présence sur les gènes de l'ARNt n'a pas été confirmée avec l'anticorps qui ne distinguait pas les formes phosphorylée et non phosphorylée, alors qu'elle a été détectée sur le promoteur du gène constitutifs *GAPDH*. Ces données montrent que la présence du cœur de TFIIH sur les gènes de l'ARNt est indépendante de Pol II et couplée à la transcription par Pol III.

Les recherches *in vitro* offrent plus de flexibilité que le ChIP qui d'habitude, ne montre que les associations. Pour étudier de manière plus profonde le lien entre TFIIH et la transcription par Pol III, nous avons réalisé une série des transcriptions *in vitro* avec un extrait nucléaire de HeLa. L'efficacité de la transcription a été évaluée par autoradiographie des produits ARN. En utilisant le gène RNA VAI de l'adénovirus humain, nous avons réussi à détecter une réduction du taux de sa transcription dans l'extrait nucléaire immuno-déplété par les anticorps contre p44 et p62 en comparaison de celui non déplété. L'addition du cœur purifié de TFIIH à l'extrait déplété a presque complètement recouvré le taux initial de la transcription. Nos données indiquent que TFIIH, quoiqu'il ne soit pas essentiel, participe à la transcription des gènes de classe III.

Puisque la présence de Pol II sur les gènes de l'ARNt était incertaine lors de la comparaison de nos résultats avec ceux précédemment obtenus [5, 70], sa vérification *in vitro* était trés souhaitable. L'explication possible de cette présence pouvait être la participation de Pol II à la transcription des gènes de classe III, alors que le TFIIH était recruté sur ces gènes comme une partie de la machinerie transcriptionnel de Pol II. Pour élucider ce point, nous avons d'abord effectué la transcription *in vitro* sur le gène *VAI* et sur un promoteur spécifique pour Pol II. L'immuno-déplétion de Pol II n'a pas influé sur la production de l'ARN VAI. Par conséquent, au moins dans ce système modèle, Pol II n'était ni nécessaire, ni participante à l'expression des gènes de classe III.

Au total, ces données nous conduisent à penser que le TFIIH influe sur la transcription des gènes de classe III indépendamment de la machinerie transcriptionnelle de Pol II, en démontrant la participa-

tion directe du TFIIH au processus transcriptionnel effectué par Pol III. La transcription des gènes de classe III est strictement controlée parce qu'elle influe directement sur la synthèse des protéines et la croissance cellulaire. Puisqu'il est connu que la perturbation de cette régulation est un trait principal de plusieurs transformations oncologiques, nous espérons que nos résultats permettront de mieux élucider le lien entre les différentes mutations dans les sous-unités du complexe TFIIH et un vaste spectre de leurs manifestations phénotypiques.

Conclusion

En conclusion générale, notre compréhension du rôle du complexe TFIIH en transcription a subi d'importants changements ces dernières années. D'abord identifié comme un des facteurs transcriptionnels essentiel de Pol II, ce complexe est reconnu à présent comme un acteur essentiel du système NER et comme un facteur transcriptionnel de Pol I. Nos recherches ont élargit notre connaissance, en montrant premièrement le lien entre TFIIH et SIRT1 pouvant conduire à une régulation différentielle de la transcription au niveau de la chromatine, et en deuxième lieu, en démontrant la participation du cœur de TFIIH dans la transcription des gènes de classe III, ce qui rend ce facteur universel parmi toutes les trois ARN polymérases principales de mammifères.

Introduction

In eukaryotic organisms, the process of DNA transcription is carried out almost exclusively by three primary multisubunit DNA-dependent RNA polymerases, *viz.* Pol I, Pol II and Pol III. Although they have similarities in structure and share some subunits, all three complexes transcribe different classes of genes, require different set of additional transcription factors to work and are regulated by a variety of polymerase and gene specific mechanisms. Respective to the polymerase that processes them, all genes are subdivided into three classes. Class I genes (templates for Pol I) and include most of rDNA. Class II genes (templates for Pol II) are represented by all protein coding genes and a growing number of genes for non-coding RNA. Finally, class III genes (templates for Pol III) encompass all tDNA, 5S rDNA and some other small non-coding RNA genes.

Pol II transcription is regulated in the most complex and the most studied way. For its function it requires not only a great number of essential transcription factors, but also the correct interplay with a vast array of DNA binding proteins, chromatin modifications, chromatin remodellers, DNA methylation, specialised genomic elements (flexible and variable promoter composition, enhancers, binding sites of nuclear receptors and other non-essential transcription factors). This versatility allows differential regulation of protein expression. The other two RNA polymerases mostly produce RNA for protein synthesis machinery. This task requires exclusively high yield but not so elaborate regulation. As a consequence, the structure of promoters, known transcriptional machinery composition and variety of additional transcriptional factors for Pol I and Pol III are less complex than for Pol II. At the same time the mechanism of their regulation is studied in lesser extend.

TFIIH is a multisubunit complex that was first characterised as an essential transcription factor of Pol II (hence the name). Later it was shown to participate in nucleotide excision pathway of DNA damage

repair. More recently it has been demonstrated to participate in Pol I transcription as well. A number of mutations in different subunits of TFIIH are known to be associated with specific genetic diseases with 1000-fold increased risk for skin cancer development comparing to normal state (*xeroderma pigmentosum*) or/and severe developmental defects (Cockayne syndrome, trichothiodystrophy, COFS syndrome). Previously these abnormalities were linked solely to the disruption of the function of TFIIH in DNA repair or in Pol II transcription.

In the course of this work specific mutations in one of the subunits of this complex, that led to the manifestation of a combined *xeroderma pigmentosum*/Cockayne syndrome phenotype, have been shown to cause inability to restart the transcription of housekeeping genes after genotoxic stress. Furthermore, that effect has been demonstrated to originate from faulty chromatin remodelling at the promoters of the genes, and not only from defects in DNA repair. At the same time the transcriptional arrest was not global, since there were a lot of up-regulated genes under the same conditions. This gene specific reaction on a mutated TFIIH provoked a whole genome wide survey of the localisation of the complex and its connection to the transcription. Surprisingly, that study ended up in a discovery of the fact, that TFIIH was strongly associated with class III genes.

The main goals of this work were to study the mechanism of gene specific response to genotoxic stress in cells with mutation in TFIIH mentioned above, and to investigate the involvement of TFIIH into the transcription of class III genes.

To achieve the first goal the following tasks were needed to be solved:

- to prove the case of differential transcriptional arrest;
- to investigate its connection to chromatin state of the affected genes;
- to identify a potential mechanism behind this process;
- to draw the link between this mechanism and the phenotypical manifestations of the responsible mutation in TFIIH.

To achieve the second goal the following questions were put forward:

- whether the observed association of TFIIH with class III genes were not an artifact;

- whether TFIIH resides on class III genes independent of the Pol II machinery;
- how does TFIIH interact with transcription process, mediated by Pol III?

The work is organised into two parts. The first part consists of the common literature review. The second part is dedicated to the stated goals with their own results and discussion in form of publishable articles. The obtained results and their implications are then summarised and conclusion is derived.

Part I

Literature Review

Chapter 1

Transcription by Pol II

1.1 DNA-directed RNA polymerase II

In eukaryotes, DNA-directed RNA polymerase II is the main enzyme responsible for transcription of all protein-coding genes. Although it is the smallest (in terms of the number of subunits) polymerase of the three main ones, its transcription machinery is the most intricate, consisting of more than 60 polypeptides, and has no match in Pol I or Pol III systems in complexity. Pol II is composed of 12 subunits that are designated RPB1 to RPB12 in order of the decrease of the mass of their yeast homologues (see Tab. 1.1).

Pol II subunits	Mass, kDa	Accession number	Orthologs in *E. coli*	Orthologs in *S. solfataricus*	Notes
RPB1	220	NP_000928	β'	A′ + A″	Contains CTD
RPB2	140	NP_000929	β	B	
RPB3	33	NP_116558	α	D	
RPB4	16.2	NP_004796		F	
RPB5	25	NP_004796		H	Common to all Pols
RPB6	14.5	NP_068809	ω	K	Common to all Pols
RPB7	19.2	NP_002687		E′	
RPB8	17.1	NP_006223			Common to all Pols
RPB9	14.4	NP_006224			
RPB10	7.6	NP_066951		N	Common to all Pols
RPB11	13.2	NP_006225	α	L	
RPB12	7.0	NP_005025		P	Common to all Pols

Table 1.1: Subunits of *Homo sapiens* Pol II and their relation to RNA polymerases of *Escherichia coli* and *Sulfolobus solfataricus*. Based on [89, 81, 46, 41].

Five subunits of Pol II (RPB5, RPB6, RPB8, RPB10, and RPB12) are commonly shared between all three main RNA polymerases. The other subunits have sequence and structural homology with subunits of Pol I and Pol III. All the subunits, except PRB8 and RPB9, have their counterparts in archaeal RNA polymerase, and Pol II itself is very close to it from the structural point of view. Moreover, RPB1, RPB2, RPB3, RPB6, and RPB11 are related respectively to β', β, α, ω, and α subunits of bacterial RNA polymerase (where RPB3 occupies the position of α_{I} and RPB11 occupies the position of α_{II} of the α_2 dimer). Although RPB1 is homologous to the largest subunits of all other mentioned RNA polymerases, it, nevertheless, contains a unique structural element, not found anywhere else: the so called **C**-**t**erminal **d**omain or CTD, which is very important in the regulation of Pol II and will be described later. RPB1 and RPB2 are the core of Pol II and are responsible for the phosphodiester bond formation [41, 46, 81].

CTD and its modifications

This part of Pol II is flexible and free. In *Homo sapiens*, it is composed of 52 tandem repeats of a peptidic sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS). CTD is a substrate for different post-translational modifications that contribute to the regulation of Pol II activity.

The main type of CTD post-translational modifications is the phosphorylation of its various residues. Depending on the level of phosphorylation, human Pol II can be separated into two forms: IIO and IIA. They alter between each other in course of transcriptional cycle. The IIA from has a hypo- or non-phosphorylated CTD and is normally involved in the assembly of the pre-initiation complex and transciption initiation. The IIO form is highly phosphorylated, primarily at Ser2 and Ser5 of the heptapeptide (numbering according the position of the residue in the YSPTSPS sequence). It is implicated in the promoter clearance and the elongation steps of the transcription. The IIB form of Pol II can also be separated from the nuclear extract. This form does not have CTD due to proteolytic degradation. Its importance and implication in transcription are not fully understood, however, it is transcriptionally active *in vitro* for the adenovirus major late promoter (AdMLP) and some nuclear genes [81].

In *H. sapiens*, there have been several protein kinases identified to which CTD is a substrate. The most important are **c**yclin-**d**ependent **k**inases (CDK): CDK7 (phosphorylates Ser5), CDK8 (phosphorylates

Ser5), and CDK9 (phosphorylates Ser2). All of them are parts of Pol II transcription machinery. The activity of these kinases is regulated by cyclins associated with them. CDK7 is controlled by cyclin H, CDK8 — by cyclin C, and CDK9 — by cyclin T. There are also other kinases that are able to phosphorylate CTD, but their impact and implication in the Pol II transcription is poorly understood [81].

Several protein phosphotases responsible for the dephosphorylation of Ser2 and Ser5 have been identified. *In vitro*, small CTD phosphotase 1 (SCP1) protein is able to remove the phosphate group from Ser5, while CTD phosphotase 1 (FCP1, a part of Pol II transcriptional machinery) mediates the removal of Ser2. The stimulation of the activity of the latter phosphotase increases the rate of transcription re-initiation [81].

In addition to the phosphorylation, CTD is also a subject to glycosylation *via* the covalent binding of N-acetylglucosamine to hydroxyl group of Ser of Thr residues. Interestingly, only the IIA form of Pol II may be glycosylated. The role of the glycosylation is probable to block the affected residues from the phosphorylation. Alternatively, it may have some impact on the structure of CTD. The exact role of CTD glycosylation is to be clarified in the future studies [81].

1.2 Promoters of class II genes

The promoters of Pol II transcribed genes have great variability. The crucial element of promoter that is absolutely needed for minimal non-regulated or basal transcription is the core promoter region (or simply the core promoter). It is the core promoter that serves as a base for the assembly of the main transcription proteins along with Pol II itself. It also specifies the **t**ranscription **s**tarting **s**ite (TSS) and spans about 30 base pairs (bp) up- and downstream it [37]. In addition to the core promoter the whole promoter complex usually contains different cis-regulatory elements. They include CpG-islands, that can span 0.5–2 kb (kilo bp) and are associated with roughly half of promoter of protein-coding genes [79] as well as stretches of binding sites of additional activating or repressing transcription factors that may be 0.2–3 kb long and contain clusters of up to 50 transcription factor recognition elements [64]. In mammals, most core promoters a compact and are able to initiate transcription only from a single TSS or from a closely spaced (usually within a short region of several nucleotides) clusters of alternative TSS. Such core promoters are generally referred to as focused. Recent genome-wide studies

have revealed a different type of core promoters with a number of TSS spread over a region of 50–100 bp. These disperse core promoters are not to be confused with alternative promoters. Although alternative promoters are separated by hundreds to thousands of bp, their architecture may conform with the focused type. The well characterised core promoter elements that are listed below (see Tab. 1.2) are typical only for focused core promoters, while the sequences and factors responsible for the initiation from disperse promoters are not well understood [47].

Element		Position (TSS = 0)	Consensus sequence $5' \to 3'$	Bound by
BRE^u		$-38 \ldots -32$	SSRCGCC	TFIIB
TATA		$-31 \ldots -24$	TATAWAWR	TBP
BRE^d		$-23 \ldots -17$	RTDKKKK	TFIIB
XCPE1		$-8 \ldots +1$	DSGYGGRASM	
Inr		$-2 \ldots +4$	YYANWYY	TAF1/TAF2
MTE		$+17 \ldots +28$	CSARCSSAACGS	
DPE		$+27 \ldots +33$	RGWCGTG	TAF6/TAF9
DCE	S_I	$+5 \ldots +10$	CTTC	TAF1
	S_{II}	$+15 \ldots +20$	CTGT	
	S_{III}	$+29 \ldots +33$	AGC	

Table 1.2: Consequence sequence and typical positions of a focused core promoter elements. The binding transcription factors are described in section 1.3. S ≡ C or G; W ≡ A or T; Y ≡ C or T; R ≡ A or G; M ≡ C or A; K ≡ T or G; D ≡ T or G or A; N ≡ A or C or G or T. Based on [47, 81].

TATA-box

The first historically recognised core promoter element was the TATA-box which is an A/T-rich region situated 25–30 bp upstream the TSS. It is bound by a special TATA-binding protein (TBP) [81]. Interestingly, TATA-box was first though to be a universal core promoter element for all Pol II dependant genes, however, recent genome-wide studies have demonstrated that only a small fraction of mammalian genes (10–20%; 22% in human genome) contain it. The majority of constitutively transcribed genes (so called housekeeping genes) posses TATA-less promoters, whereas TATA-containing genes are often activated in a

tissue-specific manner. However, it should be noted that the fraction of TATA-less promoters is identified *in silico* and the final result depends on an assumption of the consensus sequence. More direct methods are required to unambiguously determine whether or not these promoters have elements functionally equivalent to TATA-boxes [73].

Inr

The **in**itiato**r** element (Inr) is a seven nucleotide motif. If it is present, it encompasses the TSS. The first A nucleotide usually becomes the first transcribed nucleotide. In focused promoters, Inr is the most commonly found motif in focused core promoters. The genome-wide computational analysis of thousand of mammalian genes showed that Inr-less genes nevertheless have a consensus sequence YR at their TSS [47]. In *H. sapiens*, Inr belongs to 62% of TATA-containing promoters and to 45% of TATA-less promoters [81].

BRE^u and BRE^d

BRE (TFII**B** **r**ecognition **e**lement) is a disjoint binding element of TFIIB transcription factor. It consists of two sequences: BRE^u and BRE^d. The former is located upstream the TATA-box and the latter is situated immediately downstream it. BRE elements are also found on TATA-less promoters according to bioinformatical studies. In *H. sapiens*, 12% of TATA-containing promoters also contain BRE^u, whereas for TATA-less promoters this fraction is 28% [47, 81].

Downstream core promoter elements DPE, MTE, and DCE

Not all core promoter elements are located upstream the TSS. The **d**ownstream **p**romoter **e**lement (DPE) is an example of such regulating sequence. DPE cooperates with Inr and must be downstream of it at a specific distance to work properly [47]. Another downstream common downstream element is the **m**otif **t**en **e**lement (MTE). It is situated closer to the TSS that DPE ans as the latter must be strictly positioned relative to Inr. Both mentioned downstream elements may work in synergy with each other and with Inr, but can be found to be present independently. The binding protein for MTE is not known. Unlike MTE, the third known downstream element of the core promoter,

the downstream core element (DCE), is mutually exclusive with DPE. It consists of three short disjoint sequences designated S_I, S_{II}, and S_{III}. In human genome, DPE is found in 24% of TATA-containing promoters and in 25% of TATA-less promoters [81].

XCPE1

The **X** **c**ore **p**romoter **e**lement **1** (XCPE1) is a rare promoter element that encompasses the TSS. It is found only in 1% of human promoters that most often also contain the TATA-box. The factor that recognises this element has not been identified. XCPE1 do not act by itself. Instead, it works together with some sequence-specific activation factors, such as NF-1, NRF1, or Sp1.

1.3 General transcription factors

Several accessory factors needed for the basal transcription of class II genes were identified biochemically. They are called **g**eneral **t**ranscription **f**actors (or GTFs) and are named according to the historical names of protein fractions separated from the transcriptionally able Pol II machinery with TFII prefix (which stands for **t**ranscription **f**actor of Pol **II**). Six such factors were initially identified: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (see Tab. 1.3).

Not considered as GTFs but essential for the activated transcription are large multisubunit complexes Mediator and SAGA. They both facilitated **p**re**i**nitiation **c**omplex (PIC — the association of Pol II and its GTFs on the core promoter) formation, directly interacting with some GTFs during this process. In the case of an activation, when activators bind upstream control elements, SAGA and Mediator physically interact both with the activators and with PIC, mediating the activation signal [37, 57]. Mediator is a very large structure. The full Mediator is greater in geometrical size than the complete PIC. The composition of the Mediator complex is not unique. In *H. sapiens*, at least two different forms exist. They are though to interact with different activators. Its multiple subunits are thought to be targets of different regulatory factors [37, 81]. The link between activators and PIC is not the only role of SAGA and Mediator (see section 1.5).

Factor	Protein composition	Function in Pol II transcription
TFIIA	p35(α), p19(β), p12(γ)	Antirepression, stabilisation of TATA-TBP complex, coactivation
TFIIB	p33	Selection of TSS, stabilisation of TATA-TBP complex, recruitment of Pol II/TFIIF
TFIID	TBP, TAF1–TAF14	Binding core promoter, coactivation, protein kinase activity, ubiquitin-activating/conjugating activity, histone acetyltransferase activity
TFIIE	p56(α), p34(β)	Recruitment of TFIIH, initiation, promoter clearance
TFIIF	RAP30, RAP74	Recruitment of Pol II to the promoter, recruitment of TFIIE and TFIIH, selection of TSS, promoter escape, increase of elongation efficiency
TFIIH	XPB, XPD, p62, p52, p34, p8, CDK7, MAT1, cyclin H	ATPase activity for transcription initiation and promoter clearance, helicase activity for promoter opening, kinase activity for phosphorylation of Pol II CTD and nuclear receptors
Mediator	MED1–MED31, CDK8, cyclin C, variable composition	Cooperative binding with Pol II, kinase and acetylthransferase activity, stimulation of basal transcription, mediation of activated transcription
SAGA	5 TAFs, 2 STPs, 2 ADAs, 2 STAFs, GCN5, TRRAP, SGF29, USP22, 2 ATXNs, ENY2	Coactivation, histone acetyl transferase activity, ubiquitin protease activity

Table 1.3: General transcription factors and coactivators of *H. sapiens* Pol II. Based on [51, 74, 81].

1.4 Pol II transcription cycle

Promoter binding

The first step of Pol II transcription is the binding of gene-specific regulatory factors around the site of the transcription start. These factors either directly further interact with the transcription machinery, or prepare the future transcription site by recruiting chromatin modifying proteins. PIC is formed on the core promoter. At this stage Pol II with the general factors bind the promoter together but this complex is

not is not able to start transcription due to a different conformational state. PIC formation may happen in two different pathways: the sequential assembly pathway and the holoenzyme pathway. Both of them were identified *in vitro* and both are likely to exist *in vivo* as well. The sequential assembly starts from the promoter binding by TFIID. It is followed by TFIIA and TFIIB that stabilise TFIID on the promoter. Than the TFIIF-Pol II complex is recruited. After the stabilisation of this DNA-TFIID-TFIIA-TFIIB-TFIIF-Pol II complex the recruitment of TFIIE occurs. TFIIH is recruited at the very last step. The alternative pathway starts from the aggregation in solution of GTFs and other factor with Pol II to form a so called Pol II holoenzyme complex. Than it binds the promoter and forms PIC. Holoenzemes of different compositions have been reported [38, 81].

Initiation and promoter clearance

To progress further, strands of DNA must be separated around the TSS. This promoter melting occurs by the formation of a 11–15 bp long transcription bubble. The template strand is then placed in the active site of Pol II. From this point the transcription initiation starts. Usually, the transcription starts from several abortive runs, when short (shorter than 12 bp) RNA products are synthesised. At this stage Pol II stays bound to PIC. For successful initiation the phosphorylation of CTD is required at Ser5 residues of CTD. This releases Pol II from PIC. Upon release of Pol II, the scaffold complex is left on the core promoter. This complex consists at least of TFIIA, TFIID, TFIIE, TFIIH, and the Mediator. The scaffold complex is important for the transcription re-initiation [37]. In case of the induced transcription, promoter clearance is preceded by sequential binding of nuclear excision repair factors to the PIC in the order: XPC, XPA, RPA, XPG, and XPF-ERCC1 (see section 2.3) [21, 55].

Proximal pausing and elongation

Successful initiation does not guarantee productive elongation. Without further modification, travelling Pol II soon enters transcriptional arrest and, if nothing else happens, terminates the transcription. The length of the synthesised transcript before pausing is gene-specific. This arrest is mediated by factors DSIF and NELF and involves their interaction with the nascent RNA product. An action of the P-TEFb factor is needed to escape this pausing. One of modifications it performs is the change of

the phosphorylation status of Pol II. After the promoter clearance Pol II CTD is phosphorylated primarily at Ser5 as a consequence of the action of CDK7 during transcription initiation. CDK9, which is a subunit of P-TEFb, phosphorylates Ser2 residues of CTD [62, 68, 76]. Due to the constant action of phosphotases the phosphorylation at Ser5 is being progressively removed during elongation, while the phosphorylation at Ser2 lingers and even increases, since P-TEFb travels along with Pol II in a so called elongation complex. This is reflected in the fact that close to the promoter Pol II is phosphorylated mostly at Ser5 while approaching the terminator it is phosphorylated almost exclusively at Ser2 [68, 76]. Besides P-TEFb, Pol II is accompanied by elongation factors, the main of which are: TFIIS (it cleaves RNA transcript to release Pol II from occasional DSIF- and NEFL-independent pauses), TFIIF (the latter leaves PIC with Pol II), the ELL phosphotase, Elongin. This elongation complex also contains different chromatin interacting proteins, the FACT (contains chromatin remodellers) and the Elongator (contains a HAT) complexes being examples. They do not affect transcription on a naked DNA but are required for the Pol II progress through nucleosomes. In addition to this, the elongation complex serves as a platform for downstream RNA processing [76].

Termination and reinitiation

Unlike termination mechanisms of other RNA polymerases, termination of Pol II transcription is still poorly understood. If for Pol I and Pol III it is enough to encounter a simple termination sequence to release the transcript and to dissociate from the template, for Pol II such simple mechanism is not known. Instead, there are two known termination mechanisms that are linked to processing of the transcript $3'$ end: poly(A)-dependent and Senataxin-dependent [52].

Transcripts of most protein-coding genes (with exception of genes coding histones) and of many non-coding genes are polyadenylated. The gene itself contains the poly(A) signal AATAAA 10–50 bp donwstream followe by GT-rich region. During trascription the corresponding RNA sequence (AAUAAA) is recognised by Pol II associated RNA processing factor CPSF. This binding reduces the rate of Pol II progress and then causes its pausing. CstF, another factors associated with the elongation complex, binds to the downstream GU-rich signal. Interaction between CPSF and CstF leads to the RNA cleavage between poly(A) signal and GU-rich region. XRN2 exonuclease disintegrates the protruding $5'$ end

left after the cleavage. After that Pol II is released from the template. The Pol II stalling is crucial for termination but not sufficient. Polymerase pauses do not automatically lead to its dissociation from DNA. The physical interaction of XRN2 that is able to keep up with stopped or retarded Pol II is though to dislodge the polymerase. There is a body of evidence for a physical interaction between the scaffold complex on the promoter and Pol II during termination. This interaction involves chromosome looping and facilitates transcription reinitiation by the same Pol II complex [52].

For genes that do not contain poly(A) signal (most non-coding genes and histone genes) the termination mechanism is different. In this case the Senotaxin protein is responsible for the unwinding of the RNA-DNA hybrid inside the active site of Pol II, reminiscent of the action of ρ-factor in bacteria.

1.5 Chromatin and Pol II transcription

Early detailed knowledge on Pol II transcription was primarily derived from *in vitro* transcription of selected genes from naked DNA templates. In the nucleus, DNA is compactified in nucleosomes that are spaced in a more or less regular fashion with mean period of about 200 bp (146–147 bp are wrapped on a nucleosome, plus a short spacer DNA). The core of a nucleosome is composed of eight histone proteins (the so called histone octamer that consists of two histones H2A, two H2B, two H3 and two H4 or their variants). Histones bear a significant positive electrical charge in physiological conditions (more than +100 elementary charges), while DNA is negatively charged. The electrostatic interaction is the main reason for DNA wrapping around the octamer. Order of magnitude estimate for the binding energy of one nucleosome is from -10^{-19} to -10^{-18} J depending on the estimation method [33, 53, 87]. This energy is rather high. For comparison, the characteristic thermal energy at temperature $T = 300$ K equals $kT \approx 4 \cdot 10^{-21}$ J, while the Gibbs free energy released in ATP hydrolysis is about $7 \cdot 10^{-20}$ J per molecule under physiological conditions [2]. So even reading through a gene requires some additional activity from the transcription machinery, since nucleosomes positioned along the gene must be at the very least somehow overcome by Pol II. This cannot be done by mere sliding downstream due to the great length of most of class II genes. Histones may appear an obstacle on promoters as well hindering the initiation pro-

cess by screening binding sites of transcription factors. It implies that chromatin is one of key players in transcription *in vivo*.

Chromatin is not homogeneous. The most obvious difference in chromatin status is the distinction between euchromatin and heterochromatin. These two states were identified very early in the history of chromatin study. They reflect difference in structure. Euchromatin represents the so called beads-on-a-string structure when individual nucleosomes do not interact and are restricted only by DNA linkers between them. Transcription occurs only in euchromatic regions, though not all genes in euchromatin are necessarily expressed. Heterochromatin is silent in terms of transcription. Nucleosomes in it are though to form the so called 30 nm fibre. In this form they are much more closely spaced and bound to each other, so heterochromatin is more compact that euchromatin. The exact structure of the fibre is not known exactly. Different models are proposed for the winding of chromatin as well as for the forces that keep the fibre assembled, but the real ones are yet to be experimentally verified. 30 nm fibre is often associated with the linker histone H1 which supplies additional binding between nucleosomes, though it is not necessary for its formation *per se*. It should be noted that the existence of this 30 nm fibre *in vivo* was not unambiguously proven so far. Heterochromatin is expected to form higher orders of condensed structure than mere 30 nm fibre but the existence of such structures *in vivo* is even more elusive [35].

Nucleosomes are not identical throughout the whole genome. Histone proteins possess long loosely structured N-terminal domains sometimes referred to as tails. These histone tails are subjected to numerous porstranslational modifications: acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation, and SUMOylation. Multiple histone variants exist as well. Genome-wide studies have revealed that some these "chromatin marks" correlate positively or negatively with active transcription (with different degree of certainty) [57].

Physical basis for transcription regulation by chromatin

There are two approaches to understanding the meaning of chromatin modifications and their link to transcription. One of them looks for the physically (or physical chemically) sound and well understood basis for this connection. One of successful examples of this approach is the study of the effect of histone acetylation. Acetylation of histone tails is known to be associated with promoter regions of actively transcribed genes.

Transcription starts from the binding of some factors somewhere in this region. Every acetyl group added to a histone neutralises some of its excess positive charge, reducing thus histone-DNA binding energy [10]. This may dramatically change the equilibrium of wrapped-unwrapped nucleosome transition especially if the 147 bp nucleosome occupied part of DNA contains several binding sites for one o different factors. This mechanism is known as the cooperative nucleosome eviction model. It also explains why there is almost no order in the position of genomic elements for transcription activators and why these elements are often found in groups [77, 63, 64]. The acetylation of histones also prevents *in vitro* condensation of nucleosomes to the 30 nm fibre [58]. Other modification, with exception of methylation, change the net charge too.

Other chromatin marks, such as methylation of certain lysine residues of certain histones, which are known to be associated with the elongation regions of transcribed genes, may contribute in the chromatin opening (*i. e.* its accessibility to transcription) in a different way. During elongation, nucleosomes, obviously, do not constitute a serious obstacle for the elongation complex. Chromatin remodelling factors that accompany Pol II are apparently able to easily deal with them. Pol II is propelled forward in an active process using energy of hydrolysed nucleosides triphosphated. So the binding energy of DNA-octamer complex may not be of such importance here as in case of the initiation. The compactification of chromatin though may be a serious block. The idea is that theses special methylation of histones prevents nucleosome-nucleosome interaction, while the absence of acetylation, and hence high binding energy prevents non specific binding of different DNA-interacting proteins to some weak sites that happen to be in the region of elongation and, thus, inhibits cryptic transcription initiation.

Nucleosome positioning along DNA double strand may also depend on the local sequence. The rigidity of DNA double helix against mechanical deformations (twist, bend, stretch, slide, shift) is different for different nucleotide compositions. Because nucleosome wrapping induces substantial deformation in DNA structure, the net energy change in the transition form unbound to bound nucleosome differs for different genomic regions. Models, that take into account only the energetics of two adjacent stacked bases, are able to accurately explain and predict nucleosomal occupation preferences. For example, sequences rich in repeats of $3'-\mathrm{AT}-5'$ dinucleotides are the most rigid and often represent nucleosome-free regions. In contrast, $3'-\mathrm{AT}-5'$ and $3'-\mathrm{CA/TG}-5'$ dinucleotides rich regions are easily wrapped around the octamer and are

able to define strict positioning of a nucleosome. Such sequences can usually be found around promoters of genes with high basal transcription level (housekeeping genes). In this case the region upstream the core promoter is nucleosome-free, while a downstream nucleosome is strictly positioned. Interestingly, TATA-boxes differ in nucleosome binding capacity depending on their sequence. For instance, TATAAAAA TATA-box easily accommodate a nucleosome, whereas a nucleosome is easily displaced from TATATATA TATA-box only by competition with TBP binding. However, large portion of genomic sequence (95% in yeast) does not have any preferences in nucleosome positioning. Nucleosomes there are placed stochastically conserving only the internucleosomal distance which is determined by steric restriction or close-range attractive nucleosome-nucleosome interaction [65].

Unfortunately, for many modifications no physical mechanism of its correlation with transcription is known.

Histone code hypothesis

A different point of view focuses on the details and complexity of the chromatin marks themselves. The histone code hypothesis is proposed. It states that individual marks are recognised by specific factors which then recruit to chromatin other proteins and complexes that provide further regulation. In the scope of this paradigm associations of certain histone modifications with transcription level are investigated. Indeed, a number of proteins and specialised domains are found that are likely to specifically recognise some chromatin marks. A lot of correlation data is collected for different modifications and transcription activity (see Tab. 1.4).

This approach is very popular at present time. In this domain of research the terminology may slightly differ from the one used in structural or physical approaches to chromatin. For example, modifications that correlate with active transcription, such as acetylation (ac) of histones H3 and H4 or mono-, di-, or trimethylation of Lys4 residue of histone H3 (designated respectively H3K4me, H3K4me2, H3K4me3), are called euchromatin modifications. Modifications that are associated with non-transcribed genes, such as H3K9me or H3K27me, are commonly referred to as heterochromatin modifications. Chromatin immunoprecipitation analysis (ChIP) is usually applied to study these associations. The localisation of many modifications seems to follow its distinct pattern on

Modification	Position	Enzymes	Recognition domain (or protein)	Associated with
Methylation	H3K4	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromodomain, WD-40	Activation
	H3K9	Suv39h, G9a, Eu-HMTase I, ESET, SETBO1	Chromodomain (HP1)	Repression, activation
	H3K27	E(Z)	Ezh2, G9a	Repression
	H3K36	HYPB, Smyd2, NSD1	Chromodomain, JMJD	Repression of internal initiation
	H3K79	Dot1L	Tudor	Activation
	H4K20	PR-Set7, SET8	Tudor	Silencing
	H3R2	CARM1		Activation
	H3R17	CARM1		Activation
	H3R26	CARM1		Activation
	H4R3	PRMT1	(p300)	Activation
Phosphorylatoin	H3S10		(GCN5)	Activation
Ubiquitination	H2BK120	UbcH6	(COMPASS)	Activation
	H2BK123	UbcH6	(COMPASS)	Activation
	H2AK119	PRC1L		Repression
Acetylation	H3K56		(SWI/SNF)	Activation
	H4K16	MOF	Bromodomain	Activation
	Htz1K14			Activation

Table 1.4: Histone modifications associated with transcription in *Homo sapiens*. Based on [57].

upstream promoter region, the core promoter, and the open reading frame (ORF). Some of these patterns are shown in Tab. 1.5 [57].

Histone acetylation is catalysed by different histone acetyltransferase enzymes (HATs). In the context of histone code hypothesis, distinct functions are proposed for different patterns of lysine acetylation. However, acetylation is often carried out in a non-specific manner with one probable exception of H4K16ac. In contrast, for other types of covalent modifications specific enzymes are often known that take care of specific residues. Such modifications are suggested to provide unique functions in terms of regulation. In addition to HATs, histone methylases, and histone kinases, proteins maintaining the opposite reactions are known (histone deacetylases or HDACs, histone phosphotases and

Modification or histone variant	Localisation	Correlation with transcription rate
H2A.Z	Promoter	−/+
H3ac	Promoter	+
H4ac	Promoter	+
H3K4me1	3′ end of ORF	−/+
H3K4me2	ORF	+
H3K4me3	Core promoter and 5′ end of ORF	+
H3K36me2	ORF	−/+
H3K36me3	ORF	+
H3K79me	ORF	−/+
H3K9me[a]	ORF	+
H3K9me[a]	Promoter	−
H3K27me[a]	Core promoter	−
H2BUb1[a]	Entire gene	+
H2AUb1[a]	Distal elements and core promoter	−
Histone SUMO[a]	Entire gene	−

Table 1.5: Patterns of histone modifications in transcriptional context as deduced from genome-wide studies. [a]These data are based on a small number of genes. Based on [57].

demythelases) [57].

Not only the covalent modifications of histone tails are known to be associated with certain states of chromatin, but also nucleosomes may contain non-conventional variants of histones themselves. Several supposedly regulating variants of histones are known. Histone H3.3 and H2ABbd are associated with transcription activation. Histone variant H2A.X is associated with repressive chromatin. Histone macroH2A plays special role being involved in X chromosome inactivation. Histone variant H2A.Z is usually positioned at promoter-heterochromain boundary and its function is ambiguous since it is found to be associated both with active and inactive genes [57].

Though histone modifications are often portrayed as being involved in regulation and certain functions are asserted to them (such as activation or repression), one should not forget that the only information known so far is related almost exclusively to associations and correlations. Due to experimental challenge no cause-and-effect link has ever been demonstrated for any of chromatin mark. For instance, a chromatin mark, that is known to be associated with active transcription, may cause this transcription, it may facilitate further transcription, or it may be a consequence of the transcription itself not actually activating it.

Chapter 2

TFIIH — a multifunctional complex

TFIIH was first characterised as a general transcription factor of Pol II in 1989. At first it was purified form rat liver as transcription factor-δ. In HeLa, it was known as basic transcription factor 2. In yeast it was isolated as Pol II transcription factor b. Later, when homology of all these factors was revealed and the universal nomenclature for transcription factors names was proposed, it received designation TFIIH. It was shown to be absolutely required for transcription *in vitro*. TFIIH appeared to be a multisubunit complex of 10 polypeptides. At least four of them have catalytic activity. Surprisingly, TFIIH, as a transcription factor, was found to contain DNA repair proteins as its parts. ERCC2 (**e**xcision **r**epair **c**ross **c**omplementing **2**, also known as XPD) and ERCC3 (also known as XPB) were detected among its subunits. Later TFIIH has been recognised as a central player in **n**ucleotide **e**xcision **r**epair (NER). Furthermore, cyclin H is also a part of a subcomplex of TFIIH, that has been suggested to regulate cell cycle during the transition form G2 phase to M phase [21]. TFIIH is also known to be absolutely required for Pol I transcription [44]. There are several mutations in TFIIH subunits that induce rare genetic disease and are naturally found in *H. sapiens*. They are quite different in phenotypical manifestations, presumably reflecting the diversity of functions of this important complex.

2.1 The composition of TFIIH

Ten polypeptides are know to comprise the maximal TFIIH complex. They are organised in two relatively stable subcomplexes: the core TFIIH and the CAK. The core consists of six subunits: p8, p34, p44,

p52, p62, and XPB. Two of them are known to have catalytic activity (XPB is a DNA-dependent ATPase/ATP-dependent helicase, p44 is a ubiquitin ligase in yeast), whereas the other four proteins play structural role or mediate interaction with other factors. The core is connected *via* XPD subunit (the second DNA-dependent ATPase/ATP-dependent helicase of TFIIH) with the CAK subcomplex. In its turn, the CAK is composed of three units: CDK7 kinase, cyclin H, and MAT1. CDK7 is the fourth catalytic subunit of TFIIH. Tab. 2.1 summarises the composition of human TFIIH. Cryo-electron microscopy shows that in the full 10-subunit complex the core with XPD form a ring-like structure and the CAK is attached to its edge (through XPD) [21]. The components of TFIIH are not always found in the full complex form. In solution, the CAK subcomplex may be found independently [72]. The core may release the CAK when bound to DNA during NER [20]. p8 subunit of the core also exists in the cell in dimeric form separately from TFIIH [21].

XPB

XPB is the largest subunit of TFIIH. It is a part of the core. Its gene (also known as *ERCC3*) was first characterised as DNA-repair factor. The product of this genes was demonstrated to correct ultraviolet (UV) sensitivity in complementation group 3 DNA repair-deficient rodent mutants. *ERCC3*-dependent repair deficiency in rodents resembled human repair disorder *xeroderma pigmentosum* group B (or XP-B). The homologue of *ERCC3* in *H. sapiens* received the name *XPB*, since it too was able to correct the repair defect [84].

Structurally, human XPB protein contains putative nucleotide binding domain, chromatin binding domain, helix-turn-helix DNA binding domain and seven motifs found in DNA and RNA helicases [84]. It is a member of SF2 superfamily of monomeric helicases and is evolutionary conserved among eukaryotes. Homologues of human XPB exist also in bacteria [7]. XPB demonstrates two enzymatic activities: $3'$ to $5'$ helicase activity and DNA-dependent ATPase activity. Both helicase and ATPase activities of XPB are strictly required for the promoter opening and promoter escape by Pol II during transcription initiation. Interestingly, the helicase activities of XPB is dispensable in NER [21].

A number of mutations in XPB leading to genetic disorders are found in human population. Patients with mutated XPB exhibit surprisingly different clinical phenotypes: mild *xeroderma pigmentosum*, *xeroderma pigmentosum* combined with Cockayne syndrome as well as

Subcomplex	Subunit	Function	Known associated genetic disorders
Core	XPB	3′ to 5′ ATP-dependent helicase	Trichothiodystrophy, combined *xeroderma pigmentosum* and Cockayne syndrome
	p62	Structural subunit, interaction with transcription and NER factors	
	p52	Regulation of XPB ATPase activity	
	p44	E3 ubiquitin ligase (in yeast)	
	p34	Structural subunit, interacts with p44	
	p8	Regulation of XPB ATPase activity, interacts with p52	Trichothiodystrophy
	XPD	5′ to 3′ ATP-dependent helicase, bridge between the core and the CAK	Trichothiodystrophy, *xeroderma pigmentosum*, combined *xeroderma pigmentosum* and Cockayne syndrome, cerebro-oculo-facio-skeletal syndrome
CAK	CDK7	kinase	
	cyclin H	Modulation of CDK7 activity	
	MAT1	Stabilisation of the CAK	

Table 2.1: The composition of TFIIH complex in *H. sapiens*. Reproduced from [21].

trichothiodystrophy (their phenotypical features are described in section 2.4) [21].

p62

The core subunit p62 is a structural component of TFIIH that lack any known enzymatic activities. It is highly evolutionary conserved in metazoa [15]. It plays a central role in structure maintaining of TFIIH and has multiple contacts with other subunits of the complex. In Pol II transcription, p62 directly interacts with the α-subunit of TFIIE during recruitment of TFIIH to the core promoter. Moreover, *via* p62 TFIIH binds transcription activators and the tumour suppressor p53. The latter provide the way for p53-dependent transcription regulation [90].

No naturally occurring mutations in p62 that lead disorders are de-

scribed.

p52

Another non-enzymatic component of TFIIH was first characterised by co-purification and co-precipitation with the complex. Antibodies against p52 suppressed Pol II *in vitro* transcription and NER [61]. The primary role of p52 is to anchor XPB inside TFIIH. The depletion of p52 leads to DNA opening defect during transcription initiation [45].

No human genetic disorders are known to be associated with natural mutations in p52. Perhaps, due to the abolishing of the promoter opening that is with high probability lethal. However, mutants of *Drosophila melanogaster* has been described. The mutations in p52 in flies induce neurological defects, UV-sensitivity, cuticle defects, and the reduced level of TFIIH in cells. These mutations most likely impaired p52-p8 connection [1].

p44

The p44 subunit in the core TFIIH is a RING finger domain containing protein. Inside the TFIIH complex, p44 is in contact with XPD, p62, and p34. XPD contacts the RING finger which suggests that, upon DNA binding, p44 modulates enzymatic activity of XPD. Indeed, p44 was shown to stimulate XPD *in vitro* [19].

Interestingly, a yeast homologue of p44 (Ssl1) is a E3 ubiquitin ligase. It contains RING finger domain at its C-terminus which is critical for this activity. In human p44 this domain is highly conserved [80]. In *H. sapiens*, this region of the protein has been demonstrated to be important for the TFIIH complex integrity. However, no direct evidence of ubiquitin ligase activity is known for the human protein [21].

p44 has been suggested to be related to spinal muscular atrophies of type I. The gene resides in the spinal muscular atrophy region of chromosome 5. This region exists in two repeats: the telomeric and the centromeric ones. The telomeric copy derived protein differs by three amino acids. Both copies are expressed. In 15% of spinal muscular atrophy cases the telomeric copy is lost. However, this deletion does not affect either transcription or repair activities of TFIIH [11]. It is possible that the disorder originates from the deletion of the whole region but not only of p44 *per se*, since the region contains *SMN* gene, mutations in which are related to the disease. From the other hand, the duplication of

p44 may explain the absence of known repair of transcription disorders originating from its mutation despite its importance in the regulation of the XPD subunit.

p34

This core TFIIH subunit also contains zinc finger domain. However, no enzymatic or DNA binding activity is known for this protein. It is essential for both transcription and NER, though the exact role of p34 yet is to be revealed. It appears, that the primarily function of this subunit is to structurally stabilise the whole complex [80].

p8

For the long time the smallest subunit (with mass of only 8 kDa) of the entire TFIIH remained elusive. p8 (also known as TTD or TFB5) was the last identified subunit of the complex, detected as such only in 2004. This part of the core was demonstrated to be required for transcription both *in vitro* and *in vivo*. With no p8 TFIIH fails to bind the core promoter *in vitro*. Furthermore, the level of TFIIH complex is dramatically reduced in human cells deficient in p8 [71]. As it has already been mentioned, inside TFIIH p8 is conneted to p52. The structure of p8-p52 complex has been solved. It shows that p8 binds a hydrophobic surface of p52, presumably protecting it from solvent [48].

There are known mutations in p8 that cause human genetic disease trichothiodistrophy. Most likely, this happens because the loss of protection of p52 that leads to TFIIH instability [48].

XPD

XPD (or ERCC2, as it was first described in DNA repair) is the second largest subunit of TFIIH (after XPB). It is also the third polypeptide of the complex with proven enzymatic activity in *H. sapiens*. XPD is the second ATP-dependent helicase of TFIIH. Unlike XPB, XPD is a $5'$ to $3'$ helicase, contains iron-sulfur-cluster-binding domain, and is a member of superfamily 2 of DNA helicases. In addition to this domain, the protein contains two canonical helicase motor domains and the Arch domain. Homologues of XPD in eukaryotes and archaea are highly evolutionary conserved [59].

XPD is not a component of the core or CAK but rather a bounding structural bridge between these two subcomplexes of TFIIH. However, it may also be found in a separate XPD-CAK complex. In CAK it interacts with MAT1, while in the core it is attached to p44 and XPB. The former is known to stimulate the enzymatic activity of XPD [19, 21].

Interestingly, the helicase activity is dispensable in transcription and is absolutely required for NER, as demonstrated *in vitro* [21]. Nevertheless, several known mutations in human XPD provoke not only clearly repair-related disorders but also genetic diseases that are hard to explain only by deficiency in NER. The manifestations of these mutations are diverse and include *xerodemra pigmentosum*, trichothiodystrophy, combined *xeroderma pigmentosum* with Cockayne syndorm, and UV-sensitive Cerebro-oculo-facio-skeletal syndrome [21, 34].

CAK

This subcomplex of TFIIH is less evolutionary conserved than the core. In the beginning the whole subcomplex was isolated and characterised as the **C**DK-**a**ctivating **k**inase (hence the name CAK). CAK carries out phosphorylation of CDKs in CDK-cyclin pairs and this modification is required for the cell cycle progression [29]. Separated from the core TFIIH, CAK phosphorylates different CDKs, while in the large complex it changes the substrate specificity and phosphorylates different parts of transcription machinery.

CDK7

This kinase is the catalytic subunit of CAK. As the other CDKs, CDK7 is a member of the serine/threonine protein kinase family. It is unique in its dual role. While the other CDKs are involved either in cell cycle regulation (CDK1, CKD2, CDK4, and CDK6) or in transcription (CDK8 and CDK9), CDK7 is required for both [40]. In contrast to the conservation of the core TFIIH subunits, the CDK7 homologue in *S. cerevisiae* does not show the functional diversity of its human counterpart and is implicated only in transcription. Yeast CAK consists of a single polypeptide that is only distantly related to CDKs [88].

Cyclin H

Cyclin-depending kinases are called so, because they are regulated by cyclins. Cyclin-CDK association enables substrate phosphorylation. For CDK7 this regulation is carried out by cyclin H. This cyclin is also a subunit of CAK. Cyclin H provides another means of regulation of CAK activity. U1 snRNA binds this protein and affects TFIIH kinase activity [66].

MAT1

MAT1 is an acronym for french "**m**énage **à** **t**rois" named so as the third CAK subunits. This RING finger containing protein plays structural role in the complex, stabilising it. MAT1 is capable of individual binding either CDK7 or Cyclin H as well. This protein serves as an anchoring platform in the core TFIIH-CAK association, where XPD directly interacts with MAT1. The RING finger motif is not important for the formation of CAK or TFIIH but is essential for CTD phosphorylation and transcription [12].

2.2 TFIIH in transcription

TFIIH was initially characterised as a transcription factor of Pol II machinery. Later, it was shown to be involved in the production of rRNA by Pol I. However, most of studies on the role of TFIIH in transcription were focused on Pol II system. TFIIH is the last factor recruited in the sequential assembly of PIC at a core promoter. After the PIC formation XPB ATP-dependent helicase activity is responsible for promoter opening and promoter escape by Pol II. The activity of XPB is regulated by transcription factors. During the transcription of *MYC* gene, FBP factor has been demonstrated to stimulate XPB helicase activity, whereas FIR factor inhibits it in the same system [21].

As it has been already mentioned, promoter opening is not the only role of TFIIH during the initiation. CDK7 must phosphorylate CTD at Ser5 positions (and in lesser extent at Ser7 positions) for Pol II to enter productive elongation. Cyclin H is a target for the phosphorylation by Mediator subunit CDK8. Phosphorylated cyclin H in its turn represses CDK7. Some non-coding RNA, for example B2 RNA, can interfere with the phosphorylation of CTD [21].

In addition to these functions in basal transcription, TFIIH is involved in phosphorylation of other transcription factors that provide regulation of transcription. Among the targets of TFIIH are p53 tumour suppressor protein and FIR factor. Being a regulator of CDK7 activity, the latter closes the feedback loop of regulation. A vast group of TFIIH targets consists of different nuclear receptors (NR). DNA bound NRs are phosphorylated in their specific ways, so the substrates for CDK7 are either ligand bound or ligand free NRs. Phosphorylation without a ligand was reported for **r**etinoic **a**cid **r**eceptor-α1 (RARα1), RARγ, **p**eroxisome **p**roliferator **a**ctivated **r**eceptor-α (PPARα), PPARβ, PPARγ1, PPARγ2, androgen receptor, and thyroid hormone receptor-α1. Oestrogen receptor-α and thyroid hormone receptor-β are phosphorylated in response to their ligand binding. The role of this phosphorylation is not always known, but CDK7 is shown to be necessary for the normal activation of transcription. In case of RARγ the phosphorylation disrupts its interaction with vinexin-β which prevents normal function of this NR. TFIIH-mediated phosphorylation is also known to facilitate NR turnover by ubiquitin-proteasome machinery [21].

Different distinct roles of TFIIH in Pol II transcription and its interaction with numerous factors provides different ways for transcription regulation *via* this complex.

2.3 TFIIH in NER

The nucleotide excision repair pathway (or NER) is one of cellular mechanisms to deal with DNA lesions and to maintain genomic integrity. This particular process removes large DNA adducts that greatly disturb base pairing. UV-induced photochemical products of bases, such as cyclobutane pyrimidine dimers or 6-4 photoproducts (the most frequent covalent rearrangements in the cell), or *cis*-platin adducts are repaired though this pathway. The name "nucleotide excision" reflects the fact that in NER the lesion, along with 30 nucleotide around it, is excised from the damaged DNA strand. This gap is then filled by polymerisation with the undamaged strand as a template. TFIIH complex plays central role in this pathway [21].

NER is subdivide into two subpathways according to the initial damage recognition procedure: the **g**lobal **g**enome **r**epair (GGR) and the **t**ranscription **c**oupled **r**epair (TCR). In GGR lesions are detected throughout the whole genome regardless of the local transcription (though the

efficiency of detection may differ for eu- and heterochromatin). In TCR lesion detection is induced by Pol II stalling on the damaged cite during transcription, so TCR is limited only to ORFs of active genes [21].

GGR starts from lesion recognition and binding by XPC repair factor in complex with RAD23B. This may be accompanied with the binding of XPE-DDB1 complex to the same site. After the detection, the assembled platform of factors recruits the full 10 subunit TFIIH complex. At this step TFIIH unwinds DNA strands around the damaged site. Unlike in transcription, in NER DNA opening is carried out by helicase activity of XPD, while only ATPase activity of XPB is required. After DNA bubble formation, XPA is recruited to the site. XPA promotes dissociation of CAK from TFIIH and its departure from the lesion site. At the same time, XPA binding induces recruitment of RPA that binds single stranded DNA. CAK dissociation is essential for further DNA unwinding and the ability to accommodate the other repair factors. It also enables recruitment of two nuclease: XPF-ERCC1 complex and XPG. Upon their binding XPC-RAD23B complex is also released from DNA. XPG performs incision at the $3'$ end of the affected strand. XPF incises at the $5'$ end. This double incision cuts out DNA stretch of about 30 nucleotides including the damaged site. TFIIH dissociates from DNA following the release of this oligonucleotide fragment. From now on TFIIH does not take part in the downstream gap healing. The missing strand is rebuilt by DNA polymerase δ (polymerase ε and polymerase $\varkappa$ can be involved too) using the healthy strand as a template, and then ligation completes the repair [21].

The molecular mechanism and sequence of events in TCR is known with less confidence. Although this pathway is well characterised in bacteria, there is lack of full validation of TCR *in vivo* in eukaryotes. The main difference between TCR and GGR is in the lesion detection. While GGR requires XPC and other factors to detect alteration of DNA structure caused by its modification, TCR pathway utilises Pol II stalling on the obstruction as a signal for the repair. It is believed, that the main role in repair initiation is played by CSB protein. This protein travels along the ORF with the elongation complex of Pol II. CSB belongs to a family of SWI2/SNF2 chromatin remodellers and is capable of wrapping the double helix in ATP-dependent manner [6]. Though, its role in TCR is supposed to be the detection of Pol II stalling and the recruitment of downstream factors. When stalling happens, CSB tightly binds to Pol II. This recruits TFIIH, XPA and RPA to the damaged region. The repair mechanism then converges with the one in GGR. XPC, XPE and their

companions are not required. Such factors as CSA, TFIIS, XAB2 (XPA binding protein 2), HMGN1, and p300 have been shown to participate in TCR by ChIP analysis [39]. HMGN1 and p300 are though to clear the repair site removing surrounding nucleosomes. TFIIS is required to release Pol II by inducing RNA cleavage. The recent discovery of the presence of repair factors (XPA, XPF, XPG, and TFIIH itself) in the Pol II elongation complex may challenge this model [21].

2.4 Human disorders induced by mutations in XPD gene

XPD is one of the three subunits in TFIIH with known naturally occurring mutations that lead to human diseases. Different mutations in *XPD* gene induce a surprisingly broad variety of phenotypes. At least three distinct types of disorders are linked to XPD protein: *xeroderma pigmentosum* (XP), XP combined with Cockayne syndrome (XP/CS), and trichothiodystrophy (TTD) [56].

XP phenotype

XPD is not the only gene that induces XP phenotype. With different severeness it is also induced by some mutations in *XPB*, another TFIIH related gene, as well as in NER related genes *XPA*, *XPC*, *XPE*, *XPF*, and *XPG*. They are named after the disorder: first two letters in gene names are derived from initial letters of *xeroderma pigmentosum* [3].

Patients with XP are UV light sensitive. Usually, first signs of sunlight intolerance becomes evident at age of about 2 years with the appearance of intense freckling and heavy sunburn. Eye tissues are often affected. UV-induced skin cancer risk is very high in such conditions. It is more than 1000-fold higher that in the wild type. However, tumours rarely form metastases. UV-sensitivity is evident at the cellular level too. The complex of phenotypical features along with the NER specificity of majority of mutated proteins led to conclusion that XP is a manifestation of a simple repair defect. In case of *XPD* mutatins, it is in accordance with the crucial role of the enzymatic activity of XPD in NER and its inessentiality in Pol II transcription [3].

TTD phenotype

This phenotype is induced by distinct mutations in genes coding XPB, XPD, and p8 subunits of TFIIH. The clinical manifestations include progeroid features, such as neurological and skeletal degeneration, brittle hair and nails, ichthyosis, cachexia [75]. In this regard the mutation effect is close to Cockayne syndrome (CS) than to XP [54]. Unlike CS, however, these patients demonstrate additional kerotene related problems. The characteristic features of TTD, in addition to mentioned brittle nails and heir, is hyperkeratosis. Molecular basis for TTD development is poorly understood. It cannot be explained by GGR deficiency and must involve either TCR defect or/and transcription regulation problems [3].

XP/CS phenotype

This rare phenotype may be induced by specific mutations in *XPB*, *XPD*, and *XPG*. Only two XP/CS related mutations are known in *XPD*. Both of them are single amino acid change: Glycine 602 to Aspartic acid (G602D) and Glycine 675 to Arginine (G675R). G602 is located in the helicase motif V, whereas G675 is situated near helicase motif VI [56]. Both phenotypes express XP features like UV-sensitivity, skin pigmentation, multiple skin tumours (not observed in G675R, probably due to early death). In addition, typical CS features are also observed: developmental delay, mental retardation, microcephaly, cachectic dwarfism, retinal degradation, demyelinating neuropathy [3, 9, 54].

Although XP features indicate on GGR deficiency, it cannot explain the phenotype completely. Furthermore, NER is not entirely abolished in *XPD* derived XP/CS (XP-D/CS) cells. The repair synthesis in these cells is about 30% of normal [9]. The genotype-phenotype relation for XP-D/CS mutations is still poorly understood. One of strange observations is that mutations in very close positions may lead to drastically different phenotypes. For instance, mutation of an amino acid R601W just one position upstream of G602D is associated with pure XP, and mutation of close A594P gives rise to TTD. All of them are in the helicase motif V [56].

Chapter 3

SIRT1 histone deacetylase

SIRT1 is a **h**istone **d**e**ac**etylase (HDAC). It is a human homologue of yeast Sir2 (**s**ilent **i**nformation **r**egulator **2**), initially isolated as a silencing factor. All HDACs are classified into three groups: group I, group II, and group III. Group I and II HDAC enzymatic activity is inhibited by trichostatin, whereas group III HDACs are not sensitive to it. HDACs from group III are unique in the usage of NAD^+ as a cofactor in their deacetylase activity. As a consequence, enzymatic reaction by SIRT1 also turns NAD^+ into nicotinamide (NAM) [23].

SIRT1 specifically deacetylases H4K16ac and H3K9ac, but some acetylated non-histone proteins are also in the list of its substrates. These include transcription factors p53, MyoD, FOXO3, PPARγ, NF$\varkappa$B, E2F1, and transcription coactivators PGC-1α and p300, as well as coenzyme A synthetase AceCS [23, 28, 36].

NAD^+-dependence of SIRT1 suggests coupling of its activity to the cellular energy metabolism. However, the direct regulation has not been yet demonstrated *in vivo*. Conversely, the interaction of SIRT1 with PGC-1α and PPARγ suggests its important role in the regulation of glucose homeostasis at the organism level and mitochondria generation at the cellular level. The other SIRT1 targets imply its potential involvement in adipogenesis, vascular tissue development, neuronal development, and cell fate determination [28].

Interestingly, SIRT1 homologues in lower organisms are know to regulate lifespan. In yeast, calorie restriction increases lifespan, but this effect requires Sir2 activity. Sir2 is probably stimulated by the shift in NAD^+/NADH ratio. Surprisingly, SIRT1 homologue in *Caenorhabditis elegans* has also been shown to determine the lifespan of the worm. Anal-

ogous observation has been made on *Drosophila melanogaster*, where calorie restriction also increases lifespan, and SIRT1 homologue is involved in the process. A similar connection between SIRT1 and ageing is expected in mammals [8, 24].

Chapter 4

Transcription by Pol III

4.1 DNA-directed RNA polymerase III

Pol III is one of three main nuclear DNA-directed RNA polymerases in eukaryotes. It is the largest RNA polymerase and, like the other two, is composed of multiple polypeptides. Complete characterisation of the human Pol III complex identified 17 distinct subunits. Five subunits are unique to Pol III and do not have any parologues (subunits RPC3–RPC7), five subunits are unique to Pol III but are parologues of subunits of Pol II (RPC1, RPC2, RPC8–RPC10), two subunits are shared between Pol I and Pol III (RPAC1/RPA5, RPAC2/RPA9), and the rest five subunits are shared between all three polymerases (RPABC1/RPB5, RPABC2/RPB6, RPABC3/RPB8, RPABC4/RPB12, RPABC5/RPB10). In addition to this, RPC1, RPC2, RPAC1, RPAC2 and RPABC2 are orthologues to subunits of bacterial and archaeal RNA polymerases (see Tab. 4.1) [30, 43].

4.2 Genes transcribed by Pol III

Pol III transcribed genes are represented by short non-coding sequences. In mammalian genomes the most abundant Pol III targets are situated in numerous Alu repeats. The second most abundant are tRNA genes which comprise about 80% of non-Alu associated class III genes in human genome. Other Pol III dependent genes are responsible for the production of different short RNA of various function. These include 5S rRNA, 7SL RNA (a precursor of Alu associated genes), 7SK RNA, U6 snRNA, vault RNA, Y RNAs, H1 RNA (RNA part of RNase P),

Pol III subunits	Mass, kDa	Accession number	Paralogs in Pol II	Identical Pol I & Pol II subunits	Orthologs in *E. coli*	Orthologs in *S. solfataricus*
RPC1	156	AAB86536	RPB1		β'	A
RPC2	128	AY092084	RPB2		β	B
RPC3	62	NP_006459				
RPC4	53	AY092086				
RPC5	80	AY092085				
RPC6	39	NP_006457				
RPC7	32	NP_006458				
RPC8	23	NP_612211	RPB7			E′
RPC9	16.7	NP_055293	RPB4			F
RPC10	12.3	NP_057394	RPB9			
RPAC1	40	NP_976035	RPB3	RPA5	α	D
RPAC2	15.2	NP_057056	RPB11	RPA9	α	L
RPABC1	25	NP_004796		RPB5		H
RPABC2	14.5	NP_068809		RPB6	ω	K
RPABC3	17.1	NP_006223		RPB8		
RPABC4	7.0	NP_005025		RPB12		P
RPABC5	7.6	NP_066951		RPB10		N

Table 4.1: Subunits of *Homo sapiens* Pol III and their relation to subunits of *H. sapeins* Pol II and Pol I and RNA polymerases of *Escherichia coli* and *Sulfolobus solfataricus*. Based on [41, 43, 46].

MRP RNA (RNA part of RNase MRP). Pol III is also responsible for the transcription of human adenoviral VAI and VAII RNAs as well as Epstein-Barr virus small RNAs EBER I and EBER II. There has been recent reports on the involvement of Pol III in transcription of a small number of miRNA. However this requires additional study [31, 67, 86].

4.3 Promoter elements of class III genes

A special feature or Pol III transcribed genes is that some of them contain internal promoters which completely reside in the transcribed region. All promoters of known class III genes are subdivided into three types according to the composition of promoter elements and mechanism of their recognition. Type 1 and type 2 promoters are internal (also called internal control regions or ICRs, sometimes name ICR is reserved only for type 1 promoters) while type 3 promoters are situated up-stream the gene itself like in usual genes of other polymerases. 5S rRNA genes are

the only known genes with type 1 internal promoters. Typical genes with type 2 internal promoters are tRNA genes. U6 snRNA genes are the most studied examples of type 3 promoter containing genes. Our knowledge on the structure and the function of Pol III promoters comes primarily form biochemical study of three model genes: human 5S rRNA gene (type 1), adenoviral VAI RNA gene (type 2) and human U6 snRNA gene (type 3) [30].

Type 1 promoters

Mutation analysis of a *Xenopus* 5S rRNA gene type 1 promoter couples with *in vitro* transcription showed that its ICR spans from bp 50 to 97 counted from TSS. It is not a continuous structure but consists of several nucleotide blocks or boxes, *viz.* A-box, IE (**i**ntermediate **e**lement) and C-box. A-box stretches from bp 50 to bp 60 from TSS, the IE occupies region from bp 67 to bp 72 from TSS and C-box constitutes an interval bp 80–97. These elements are in fact binding sites of basal transcription factors of Pol III. During the initiation they are bound by TFIIIA (see below in 4.5). Changing of the sequence in the spacers between these elements but preserving their length does not affect the transcription efficiency. In contrast, changes of the spacing distance between promoter elements attenuates or completely blocks transcriptional activity. In budding yeast, the deletion of A-box does not completely suppress the transcription [31].

Type 2 promoters

Like type 1 promoters, type 2 promoters contain multiple ICRs separated by nucleotide sequence that does not affect transcription. In this case there are only two elements: A-box and B-box. During transcription initiation they are bound by TFIIIC factor (see below in 4.5). In this case the spacing distance is not of great importance. In naturally occurring promoters it varies greatly. A-boxes of type 1 and type 2 promoters are structurally similar. However, this similarity apparently comes from structural constrains on RNA itself rather than from a conserved function of promoters. As it has been already mentioned, they are recognised by different transcription factors. A- and B-boxes are very well conserved in tRNA genes, but this is probably due to structural constrains on viable tRNA. A-box corresponds to D-loop and B-box encodes T-loop of a mature tRNA. A-box of VAI RNA gene differs in more

degree from its tRNA counterparts than they differ from each other. In tRNA, A-box spans from bp 20 to bp 30 from TSS and B-box is usually situated from bp 80 to bp 90, though as it has already been stated, the distance from A-box may be different. It is A-box that determines TSS itself. Some genes contain a pseudo-A-box which can start transcription in case of mutationally inactivated primary A-box. At the same time, truncated tRNA genes with no B-box are able to initiate transcription *in vivo*, though its efficiency is attenuated [31, 74].

Type 3 promoters

Type 3 promoters are completely different and do not incorporate any ICRs. Instead they are composed of a **p**roximal **s**equence **e**lement (PSE) followed by a TATA-box. The former is also a core promoter of Pol II transcribed snRNA genes. The latter is the same as in Pol II promoters (not snRNA genes). Interestingly, deletion of TATA-box switches transcription specificity from Pol III to Pol II. Human U6 RNA genes also have an upstream **d**istal **s**equence **e**lement (DSE) which is important for transcription activation [31, 74].

4.4 Termination of class III gene transcription

For Pol III to terminate transcription it is enough to encounter a cluster of four or more T residues in CG-rich surroundings. This is the case of 5S rRNA genes and of most eukaryotic tRNA genes. However, some natural tRNA genes contain more Ts at the end. Deletions of extra Ts results in significant increase of read-through rate. *In vitro* Pol III dependent transcription also terminates within a cluster of 23 A residues of a mouse 5S RNA gene. In Alu repeats termination happens at a stretch of As preceded by an imperfect hairpin. The two last cases resemble prokaryotic ρ-dependent transcription termination sites. No subunit of Pol III has been identified to be responsible for its termination activity. Presumably, the termination process relies on a transcriptional arrest that is followed by separation of newly synthesised RNA from enzyme-DNA complex and returning Pol III machinery to a state that permit re-initiation. The two latter step may depend on additional factors. At least four factors have been found by biochemical studies to be implicated in efficient termination of Pol III: La protein, NF1, DNA topoisomerase I and PC4. La was found to bind poly(U) end of transcripts, and its de-

pletion from nuclear extract greatly reduced efficiency of multiple round *in vitro* transcription of class III genes. NF1 was shown, to be a part of TFIIIC transcription factor, and to bind DNA a consensus sequence 5′-YTGGCANNNTGCCAR-3′. This sequence is found downstream VAI RNA gene poly(T), however, it is not found anywhere near many other class III gene terminators. For DNA topoisomerase I and PC4 no direct mechanism has been identified. It should be noted that absolute necessity of any of these factors for the termination *in vivo* has not been demonstrated [31, 74].

4.5 Transcription factors

Early works on fractionation of a HeLa extract aimed for Pol III transcription factors characterisation identified three fractions called fraction A, B and C. It was found that transcription from type 1 promoters required all three fractions, whereas type 2 promoters could initiate with only B and C fractions. Later these fractions were identified with multisubunit complexes designated respectively TFIIIA, TFIIIB, and TFIIIC. An overview of their composition and function is outlined in Tab. 4.2. TFIIIB is absolutely required for Pol III transcription from any gene, albeit in different composition. The other two are initiation factors and are required to recruit TFIIIB in a gene specific manner. TFIIIA recognises type 1 promoters while TFIIIC is needed to start transcription from both type 1 and type 2 promoters. Type 3 promoters are bound directly by TFIIIB but their transcription relies on some additional factors mentioned below [30, 74].

Recruitment factor TFIIIA

TFIIIA is a founding member of the family of Cis_2-His_2 zinc finger DNA-binding proteins and contains 9 zinc fingers. Its only known role is to recognise ICR of 5S rRNA genes and to recruit TFIIIC factor to the promoter through protein-protein interaction. In *Saccharomyces cerevisiae*, strains lacking TFIIIA are viable if 5S rRNA gene is designed to start its transcription form a type 2 promoter [13]. In addition to its role in transcription, TFIIIA also binds mature 5S rRNA. The TFIIIA-RNA complex constitutes 7S storage ribonucleoprotein particle. This interaction is believed to serve an additional role of a negative feedback in 5S rRNA genes regulation, allowing the RNA-product to sequestrate the

Factor	Component	Alternative names	Notes
TFIIIA	TFIIIA		9 zinc fingers, binds ICR of type 1 genes
TFIIIB	TBP		Binds TATA-box of type 3 promoters, required for all promoters
	Brf1	TFIIIB90	Required for transcription of type 1 and 2 promoters, mutually exclusive with Brf2
	Brf2	BrfU, TFIIIB50	Required for transcription of type 3 promoters, mutually exclusive with Brf1
	Bdp1	B″, TFIIIB150	Contains SANT domain
TFIIIC	GTF3C1	TFIIIC220, TFIIICα	Binds B-box, histone acetyltransferase (HAT)
	GTF3C2	TFIIIC110, TFIIICβ	HAT
	GTF3C3	TFIIIC102, TFIIICγ	Contains TPR repeats
	GTF3C4	TFIIIC90, TFIIICδ	HAT
	GTF3C5	TFIIIC63, TFIIICε	Binds A-box
	TFIIIC1		4 uncharacterised polypeptides with masses of 70, 50, 45 and 40 kDa
	TFIIICU		Enhances U6 RNA genes transcription
	NF1		Polypeptides that regulate termination

Table 4.2: Components of the human Pol III core transcription factors.

factor needed for its production. At least *in vitro*, the transcription of 5S rRNA gene is inhibited by 5S rRNA through a competitive binding. It also has been suggested that competition of TFIIIA and ribosomal proteins for 5S rRNA may couple 5S rRNA transcription to ribosomal synthesis. In *S. cerevisiae*, 5S rRNA production rapidly responses to amino acid deprivation [31, 74].

Recruitment factor TFIIIC

TFIIIC is capable to bind class III genes promoters in two ways: recognising TFIIIA-ICR complex on type 1 promoters or directly binding A- and B-boxes of type 2 promoters. The means by which TFIIIC is recruited by TFIIIA are not well characterised so far. Structural analysis and electron microscopy studies on type 2 promoters suggest that it consists of two DNA-binding modules (A- and B-box specific) joined by a flexible linker capable of fitting to various spacing distances between the two promoter elements.

The human TFIIIC fraction separates into two distinct complexes: TFIIIC1 and TFIIIC2. It appears that, unlike TFIIIC2, TFIIIC1 complex is required for transcription of all three types of promoters. It also strengthens DNA-dinging of TFIIIC2. This fraction contains of four poorly characterised proteins and the TFIIICU factor that enhances human U6 snRNA gene transcription. In addition to TFIIIC1/TFIIIC2 complex, TFIIIC fraction also contains NF1 polypeptides. Their role in Pol III transcription is thought to be the regulation of termination, as described in section 4.4. Substoichiometric levels of topoisomerse I and Pol II co-activator PC4 can be purified from immunoprecipitated TFIIIC assemblies from human cells producing a tagged TFIIIC subunit. They are also suggested to participate in the termination process. However, as it has been already mentioned, no direct mechanism is found for their action [74].

TFIIIC2 is better characterised and contains polypeptides GTF3C1–GTF3C5 (see Tab. 4.2). The largest subunit GTF3C1 is responsible for the binding to the B-box. Relatively small GTF3C5 subunit effectuates A-box binding. GTF3C3 contains TPR domains. They mediate binding to Brf1 subunits of TFIIIB. Interestingly, human TFIIIC2 subunits, that are the most diverged from their yeast counterparts (GTF3C1, GTF3C2, GTF3C4), are histone acetyltransferases. This is in striking difference with fungi where no HAT activity has been detected for TFIIIC. Human TFIIIC2 is an uncommon case of a complex with three separate HATs. The role of this chromatin modifying activity is to be revealed. There is no doubt that the state of chromatin is important for the transcription by Pol III, and its transcription sites are marked by specific set of histone modifications, but no direct involvement of TFIIIC has yet been demonstrated in this process [30, 86]. The interplay between chromatin state and Pol III transcription will be reviewed in chapter 5. GTF3C4 is a key holding scaffold for the whole TFIIIC2: through its interaction with

GTF3C2 and GTF3C5 it binds together A- and B-box specific aggregates (GTF3C1-GTF3C2 and GTF3C3-GTF3C5). The way by which TFIIIC1 and TFIIIC2 are connected is not yet known [30].

Transcriptional factor TFIIIB

The composition of this factor varies depending on the promoter type. It is always composed of three subunits. On type 1 and 2 promoters TFIIIB is represented by a complex TBP-Brf1-Bdp1. Contrary, to type 3 promoters it binds as TBP-Brf2-Bdp1. In yeast, there is only one Brf. In human, the two forms are closely related. They show some homology to the Pol II transcription factor TFIIB, hence the name (Brf stands for TFII**B** **r**elated **f**actor) [30]. TBP is the TATA-binding protein, initially described as part of Pol II transcription initiation machinery as a subunit of TFIID transcription factor. It was shown to participate in Pol I and Pol III mediated transcription [25, 60]. It is through TBP that TFIIIB directly binds TATA-box of type 3 promoters. For the recruitment to other promoters it requires the interaction with TFIIIC. This is achieved *via* connection of Bfr1 and Bdp1 with GTF3C3. Brf1 and 2 subunits contain zinc ribbons. Bdp1 contains the specific SANT domain (identified in **S**WI/SNF and **A**DA complexes, the transcriptional corepressor **N**-Cor, and yeast **T**FIIIB Bdp1). The SANT domain is absolutely required for Pol III transcription which is TFIIIC dependent [30, 74].

One of the expected roles of TFIIIB is the binding to the TATA-box of type 3 promoters. It seems that the factor does not exist in its full composition in solution. *H. sapiens*, Brf1 was shown to be tightly associated with TBP in solution. On the other hand, Bdp1 association with TBP-Brf1 is weak and can be detected only with GST pull-down assays. Interestingly, Brf2 is also only weakly associated with TBP in HeLa cells. It suggests that TFIIIB complex assembles to its final forms only on DNA. Because of this, type 3 promoter binding stats from the direct association of the free TBP with the TATA-box. Bfr2 then recognises the TATA-box-TPB complex, like TFIIB. It has been shown *in vitro* that Bdp1 assembles on a complete TATA-box-TBP-Brf2 comples, although not very efficiently [74].

Another role of TFIIIB is to recruit Pol III onto bound genes. There is a number of protein-protein contacts known to exist between the factor and the polymerase. Firstly, TBP directly binds RPC6 subunit of Pol III. Secondly, Brf1 binds both RPC6 and RPC9. It is not known how Brf2

contacts Pol III. No data exist on contacts between Bdp1 and Pol III either.

Unlike Pol II machinery, where helicase activity of TFIIH factor is responsible for promoter opening in ATP-dependent manner, no ATPase-helicase subunit in any of Pol III factors or in Pol III itself has been described. However, it has been shown that DNA bubble forms in ATP-independent manner if a promoter is bound by TFIIIB and Pol III is added. It was demonstrated in yeast on *SUP4* $tRNA^{Tyr}$ gene and on a yeast specific B-box containing U6 snRNA gene. In case of *SUP4*, the promoter opening was shown to be non-coordinated and temperature dependent. The sequential promoter binding by TFIIIC, TFIIIB and Pol III happened normally at temperatures from 0 to 40°C. In contrast, the DNA strand separation did not happen at low temperatures. Very restricted or no bubble formation was observed at temperatures below 10°C, while its degree steadily grew throughout the wide range (up to the maximal temperature used in the experiment). It was demonstrated that some mutations in Brf1 homologue cease the full promoter opening at 20°C and thus deny transcription initiation. Such mutations can be circumvented by an artificial bubble formation at the appropriate site. This indicates that TFIIIB indeed has a role that extends beyond the recruitment of Pol III and is linked to the promoter opening [49, 50]. If this thermodynamical promoter opening is the only and the primary way of transcription initiation for Pol III is no very clear though. Such point of view seems not to agree well with the ability of *S. cerevisiae* to sustain cell growth at temperatures below 10°C [42, 78].

Additional recruitment factors

As it has been written, type 3 promoters do not normally contain any ICR, but rely on upstream elements. In addition to TATA-box, they include core element PSE and more distant DSE (see section 4.3). These two special elements are found not only in U6 snRNA genes and other Pol III dependent genes with type 3 promoters, but also in snRNA genes transcribed by Pol II. This specific promoter structure indicates on the involvement of additional factors that are common to Pol II and Pol III. Indeed, such factors have been identified. PSE is bound by the so called **sn**RNA **a**ctivator **p**rotein **c**omplex ($SNAP_c$). This multisubunit complex contains five polypeptides (SNAP190, SNAP50, SNAP45, SNAP43, and SNAP19 named according to their molecular mass in kDa). $SNAP_c$ specifically binds PSE primarily *via* four and a half repeats of an unusual

Myb domain in SNAP190. Immunodepletion of $SNAP_c$ from transcription extracts destroys U1 and U6 snRNA genes transcription while the addition of highly purified recombinant $SNAP_c$ restores it. The human DSE is a binding site of Oct-1 and STAF proteins. Oct-1 is a member of POU-homeodomain protein family, whereas STAF is a Cis_2-His_2 zinc finger protein. Oct-1 and $SNAP_c$ bind to DSE and PSE cooperatively. The interaction between Oct-1 and DSE is relatively weak, but it is stabilised by interaction with DNA-bound $SNAP_c$. From the other hand, the conformational changes cased by protein-protein interactions of $SNAP_c$ with Oct-1 open DNA-binding part of SNAP190 for the contact with DNA. The interaction of $SNAP_c$ with STAF has not yet been reported. Interestingly, protein-protein interaction between $SNAP_c$ and Oct-1 requires a formation of DNA loop, which is presumably achieved through the wrapping of the spacer between PSE and DSE around a properly positioned histone octamer [30, 74].

4.6 Regulation of Pol III activity

In the cell, the activity of Pol III is tightly regulated. Normally it strictly follows the cell cycle, since the main products of class III genes are involved in protein biosynthesis (a part of splicosome, a part of ribosome, a part of signal peptide recognition particle, tRNAs) and, hence, their turnover must response adequately to cell growth. In many oncogenically transformed cell Pol III activity is considerably up-regulated. Recent chromatin immunoprecipitation studies coupled with high throughput sequencing (ChIP-seq) showed that in human genome, in contrast to yeast, only a fraction of tRNA genes are occupied by Pol III. Furthermore, this fraction is cell line specific [67, 86]. For primate genomes, it is also important to inhibit parasitic transcription from Alu repeats. Indeed, only a very small fraction of all Alu-associated class III promoters are bound by Pol III under normal conditions [67]. There are several known ways of the direct non-specific regulation of Pol III transcription. All of them focus on the interaction with TFIIIB — the central transcription factor for all promoters types. The direct regulation *via* TFIIIB has been demonstrated for four proteins. Two of them are tumour suppressors, *viz.* p53 and the retinoblastoma protein (RB). Both proteins bind TFIIIB, preventing its interaction with TFIIIC2 and Pol III. Being naturally ubiquitous in the cell, normally they sequestrate this factor in an inactive form. There is also a strictly Pol III specific factor Maf1 that

acts in the same way. By the nature of the influence on Pol III, all this regulation is assumed to be general. The proto-oncogene c-Myc has also been shown to interact with TFIIIB. In contrast to the previous cases, this stimulates the recruitment of Pol III. There has been some evidence that DNA sequence upstream the promoter may influence the response on c-Myc. Interestingly, c-Myc is able to recruit TRRAP protein which in turn recruits GCN5 HAT to class III genes promoters. TFIIIB is also phosphorylated by several kinases, though the relevance of this for the transcription regulation has not been demonstrated *in vivo* [85, 86].

With advent of ChIP-seq, the growing body of data indicating on the potential involvement of chromatin modifications in the regulation of class III genes transcription. These findings will be reviewed in chapter 5.

Chapter 5

Interplay between Pol II and Pol III transcription machinery

The presence of three distinct transcriptional enzymes in eukarytic nucleus (particularly in vertebrates) had been first demonstrated by biochemical studies using ion exchange column fractionating, polyacrylamide gel electrophoresis, immunological assays and by the difference in sensitivity to amanitin. At that early stage of eukaryotic transcription research some similarities in polymerase composition had already been noticed. Polyacrylamide gel electrophoresis showed a subset of putatively common peptides for all three polymerases. That was also supported by immunological precipitation and inhibition of the transcriptional activity, where antiserum crossreactivity between the three enzymes was observed. All that allowed to suggest some degree of structural similarity [16]. Indeed, Pol I, Pol II, and Pol III later appeared to be structurally related. Not only they are composed of evolutionary related polypeptides but some of subunits are universal among all three polymerases (see Tab. 4.1). Such similarity could imply other shared parts of transcriptional machinery. Meanwhile, the sets of discovered additional transcription factors stayed different and polymerase specific for some time. Nevertheless, the point of view that each polymerase has its unique factors and regulation patterns began to change.

5.1 TBP as a universal transcription factor

The first transcription factor to be identified as common to Pol II and Pol III was TPB: the TATA binding protein, initially characterised as a

subunit of Pol II TFIID factor. It was separated by fractioning from a partially purified TFIIIB capable of transcription initiation on *S. cerevisiae* U6 rRNA gene. TBP was demonstrated to be absolutely required for Pol III transcription *in vitro* [60]. Latter it was shown to be tightly bound by Brf1, a part of TFIIIB that is a homologue of Pol II factor TFIIB [74]. Interestingly, earlier TBP had been shown to be an essential transcription factor for Pol I as a part of TIF-IB general transcription factor. This makes TPB the first discovered truly universal essential transcription protein for all eukaryotic polymerases [25].

5.2 Pol II, its factors and chromatin marks near class III genes

In recent years, several ChIP-seq based studies of class III genes revealed unexpected data. Four such works have been published so far. Unfortunately, they are not completely mutually corroborative and contain some partially contradicting details. Because of this, the articles will be reviewed separately one by one.

Article by Raha et al. [70]

In this work two immortalised cancerous human cell lines were used: K562 erythroleukaemic cells and GM12878 lymphoblastomic cells. The authors found close association of the hyperphosphorylated form of Pol II with Pol III bound class III genes. In total, from 60% (K562) to 25% (GM12878) of Pol III bound genes contained a nearby Pol II enrichment peaks. These peaks were generally shifted to 5' end of a gene with a high fraction of peaks centring at about –200 bp relative to TSS. However, this distance was not totally universal. Likewise, the association of Pol II factors c-Myc, c-Fos, and c-Jun was observed with near 70% to 80% of Pol III occupied class III genes in K562. The authors hypothesise that Pol II might help to open chromatin and to allow Pol III and its factors to access their targets.

Article by Oler et al. [67]

This work was performed on two other cancerous human cells: HeLa and lymphomic Jurkat T cell lines. The authors report that Pol III occupied tRNA genes were often (about 20% for HeLa cells) situated

2 kb (most often 300–900 bp) upstream to Pol II transcribed genes TSS. They also analysed ChIP-seq data for HeLa cells from public databases. And though 80% of occupied Pol III genes are situated outside annotated class II genes, they were associated with overlapping or closely spaced peaks of high levels of Pol II and marks of active chromatin, *viz.* histone H3 Lys4 monomethylation (H3K4me1), H3 Lys4 trimethylation (H3K4me3), H2A.Z histone variant, H3.3 histone variant were found close to top 50 the most highly Pol III occupied tRNA genes. The same analysis of published data for Jurkat cells revealed that top 50 the most highly Pol III enriched tRNA genes were associated with elevated levels of H3K9 acetylation (H3K9ac), H3K27ac, H3K18ac, H2BK5ac, H2BK20ac, H2BK120ac, H4K91ac, H3K36ac, H3K4me1 (usually with a rather large offset of –1 kb), H3K4me2, H3K4me3, H3K79me3, H3K9me1, and H2A.Z. In contrast to HeLa, Pol III occupied tRNA genes did not correlate well with Pol II peaks. In HeLa, H3K27me3 has been found to overlap with bottom 50 tRNA genes ranged by the level of Pol III enrichment. In Jurkat cells, this was the case for H3K27me3, H3K36me3, and H4K20me3. Overall, actively transcribed tRNA genes (detected by Pol III enrichment) seem to be generally closely associated (in a range of 2 kb) with pormoter of Pol II transcribed genes or with enchancer-like regions with marks of active chromatin.

Article by Barski et al. [5]

This work was done on human $CD4^+$ T and HeLa cells and was published jointly with the previous paper. The authors reported that Pol III enriched and actively transcribed (judged by RNA products) tRNA (302 analysed genes) and other class III genes are generally associated with enrichment by H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K23ac, H3K27ac, H3K36c, and H2A.Z, whereas H3K27me3 correlated with non-transcribed tRNA genes. These chromaitn marks are similar to those on the sites actively transcribed by Pol II. The difference was in the lack of H3K79me2 and H3K36me3, normally found associated with Pol II transcription. They also located Pol II near some active class III genes. Pol II enrichment was shown for a Pol II with unspecified phosphorylation status, for the unphosphorylated form of Pol II, for the Ser5 phosphorylated form, and for the Ser2 phosphorylated form. If Pol II was found to be associated with a class III gene, the maximum enrichment was generally observed at the 200 bp offset.

Article by Carrière at al. [14]

The work was performed on mouse embryonic stem cells. Active chromatin marks (H3K4me1, H3K4me2, H3K4me3, H3K9me3) and heterochromatic mark H3K27me3 were studied. Like in the previously reviewed studies, the association of Pol III bound class III genes with up- and downstream (400 bp) enrichment of H3K4me3 has been observed, while the TSS of that genes was free of those marks. On the contrary, the heterocromatic H3K27me3 was completely absent on these genes. H3K9me3 were slightly enriched around these genes, but not very significantly. Enrichment levels of H3K4me1, H3K4me2 were not significant. In contrast to the earlier works, no significant enrichment of phosphorylated Pol II (Ser2, Ser5, and Ser7) near tRNA genes was observed. Only a very low level of hypophosphorylated Pol II form was found upstream of these genes. Ser5 and Ser7 phosphorylated and hypophosphorylated Pol II enrichment was observed upstream type 3 promoters. Interestingly, Pol II transcription factor TFIIS was observed to be bound to about 60% of tRNA genes and 50% of type 3 promoter genes (from Pol III bound fraction). It should be noted, that the genome-wide colocalisation of TFIIS and Pol III had already been reported for budding yeast [32].

Part II

Results

Chapter 6

XPD/CS mutations and regulation of housekeeping genes

6.1 Publication 1

Submitted to PNAS on August 23, 2012.

Sirt1 regulates RNA synthesis after UV irradiation in TFIIH-mutated cells: new insight into the XP-D/CS cellular phenotype (122)

Classification: Biological Sciences, Genetics

Renier Vélez-Cruz[a,b], Anton Zadorin[b], Frédéric Coin, and Jean-Marc Egly*

Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/ULP, BP 163, 67404 Illkirch Cedex, C. U. Strasbourg, France.
[a]Current address: The University of Texas M.D. Anderson Cancer Center, Department of Molecular Carcinogenesis, Science Park Research Division, Smithville, TX, 78602, USA.
[b]These authors contributed equally to this work

Running title: Sirt1-induced repression of RNA synthesis after UV

*To whom correspondence should be addressed:

Jean-Marc Egly, Ph.D.
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)
1, rue Laurent Fries
67404 Illkirch, France
Tel: +33 (0)3 88 65 34 47
email: egly@igbmc.fr

Abstract (247)

Specific mutations in the *XPD* subunit of TFIIH result in combined xeroderma pigmentosum/Cockayne syndrome (XP-D/CS), a severe DNA repair disorder characterized at the cellular level by a transcriptional arrest following UV. This transcriptional arrest has always been thought to be the result of faulty transcription-coupled repair. In the current study we showed that following UV irradiation XP-D/CS cells displayed a gross transcriptional dysregulation when compared to a “pure” XP-D cells or WT cells. Furthermore, global RNA-seq analysis showed that XP-D/CS cells repressed the majority of genes after UV, while “pure” XP-D cells did not. Using housekeeping genes as a model we demonstrated that XP-D/CS cells were unable to re-assemble these gene promoters and thus to re-start transcription after UV. Furthermore, we found that the repression of these promoters in XP-D/CS cells was not a simple consequence of deficient repair but rather an active heterochromatinization process mediated by the histone deacetylase Sirt1. Indeed, RNA-seq analysis showed that inhibition of and/or silencing of Sirt1 changed the chromatin environment at these promoters and restored the transcription of a large portion of the repressed genes in XP-D/CS cells after UV. Our work demonstrates that a significant part of the transcriptional arrest displayed by XP-D/CS cells arises due to an active repression process and not simply due to a DNA repair deficiency. This dysregulation of Sirt1 function that results in transcriptional repression may be behind the various severe clinical features in XP/CS patients that cannot be explained by a DNA repair defect.

Introduction

The human genome is exposed to a variety of endogenous and exogenous insults that can alter the genetic information and physically interfere with critical cellular processes such as DNA replication and transcription (1, 2). The inability to remove these DNA alterations can lead to mutations or the halt of transcription and/or DNA replication. Mutations can ultimately cause cancers and the arrest of cellular processes can induce cell death, which can result in premature aging (3, 4).

The Nucleotide Excision Repair (NER) pathway is responsible for the removal of a variety of bulky DNA lesions, such as those induced by UV, and is subdivided into two sub-pathways. Global genome repair (GGR) is responsible for the removal of adducts from the whole genome and transcription-coupled repair (TCR) is responsible for the accelerated removal of lesions located on the transcribed strand of active genes (5, 6). While GGR is initiated by the damage-recognition proteins XPC-RAD23B, TCR is initiated by an RNA polymerase II (Pol II) stalled in front of a blocking lesion, and does not require XPC-RAD23B. Upon the stalled Pol II, the TCR-specific factors CSA and CSB are recruited to the site and recruit other chromatin remodeling factors (7, 8). The remaining steps of these mechanisms are thought to be identical for both sub-pathways (9, 10).

The absence of these repair mechanisms leads to severe genetic disorders such as: xeroderma pigmentosum (XP), trichothyodystrophy (TTD) and Cockayne syndrome (CS), which present an array of clinical symptoms including segmental progeria (Table 1) (3, 11-13). CS patients particularly, display a complex list of clinical features that are hard to reconcile with a sole defect in DNA repair, and argues for the involvement of CSA and CSB proteins in other cellular processes, such as transcription.

A limited number of specific mutations in NER genes (*XPB*, *XPD*, and *XPG*) have resulted in patients with a combined XP/CS phenotype (14-18). The clinical severity of combined XP/CS patients probably arises due to an inherent link between transcription and NER. For instance, XPB and XPD are both helicase subunits of the transcription/repair factor TFIIH (19). Furthermore, XPG was recently identified as a protein required for maintaining the integrity of the TFIIH complex and therefore also engaged in the transcription process (20). Even though the contribution of the DNA repair deficiency to the clinical features of XP, CS, and XP/CS patients is irrefutable, studies have shown a clear dysregulation of a variety of transcriptional pathways, which may also contribute to the clinical phenotype of these patients (21-27). Interestingly, at the cellular level, XP/CS cells share with CS cells a sustained global transcriptional arrest after UV, which has been always explained by the inability of these cells to perform TCR (17, 28). The fact that the so called global transcriptional arrest displayed by XP/CS and CS cells excludes genes that are activated upon DNA damage, such as p53-dependent genes, suggest that there must be an active transcriptional repression process, rather than a physical blocking of transcription. In this regard, CSB and other NER factors have been shown to affect chromatin remodeling for optimal transcription initiation (24, 29-31).

In the current study we showed that XP-D/CS cells display a gross transcriptional dysregulation upon UV, whereas WT and XP-D cells displayed no and a mild dysregulation, respectively. XP-D/CS cells (XPD-G675R and XPD-G602D) were never able to re-start transcription of housekeeping (HK) genes after UV, identical to the CS phenotype. We showed that the histone deacetylase (HDAC) Sirt1 was responsible for the establishment of a heterochromatin environment at these promoters after UV.

Treatment of XP-D/CS cells with a Sirt1-specific inhibitor or downregulation of Sirt1 by siRNA resulted in the restoration of the expression of a large proportion of the repressed genes in these cells. Many of the genes whose expression was restored play a role in multiple pathways including DNA repair and genomic stability and may help explain the severe phenotype of XP-D/CS cells.

Results

XP-D/CS cells display a global transcriptional dysregulation after UV irradiation

The inability to re-start transcription after UV (measured by [^{3}H] uridine incorporation) is the hallmark characteristic of CS cells (28). However, rRNA synthesis (which is rather high in growing cells) represents a large fraction of the incorporated [^{3}H] uridine in these studies and thus results may not be representative of global transcription of type II genes (32). Furthermore, a more recent analysis showed HK genes were repressed after UV in CS-B cells, while p53-inducible genes were not (24). XP-D/CS cells also displayed a global transcriptional arrest after UV, importantly, this transcriptional arrest is not observed in “pure” XP cells (17).

In order to study the global transcriptional response of XP-D/CS cells upon UV, we first performed RNA-Seq analysis in wild type (WT), XP-D/CS (XPD-G675R), and a “pure” XP-D (XPD-R683W) primary fibroblasts untreated and 24h after UV (10 J/m^2)(Figure 1A-C). The scatter plots and correlation analysis in WT cells showed that at 24h after UV this cells had re-established overall transcriptional equilibrium (*i.e.*, the expression of most genes had returned to basal levels, note the close distribution of the genes along the black diagonal line). On the other hand, XP-D/CS cells display a gross transcriptional dysregulation after UV (Figure 1B, note the spread of all the data points

away from the black diagonal). The UV treatment in XP-D/CS cells resulted in a total of ~2,000 genes whose expression changed more than 3-fold and was reproducible and statistically significant. The majority of the genes that changed in XP-D/CS cells were repressed (70%, dots below the black diagonal line Figure 1B), however, ~30% of the genes were overexpressed at 24h post-UV (~600 genes, dots above the black diagonal line, Figure 1B), thus challenging the global transcriptional arrest described for XP-D/CS. Interestingly, the same analysis on XPD-R683W fibroblasts resulted in a scatter plot that resembles more that of the WT cells (Figure 1C). Additionally, XP-D cells only repressed 26% of genes (compared to 70% repressed in XP-D/CS cells). XP-D/CS patients display a variety of severe clinical features associated with XP and CS phenotypes, while the "pure" XP patient displayed only UV hypersensitivity (Table 1 and 2). Interestingly, these differences in transcriptional dysregulation parallel the clinical severity of the XP-D/CS vs. XP-D patients thus underscoring their importance (Table 1) (13, 17, 33, 34).

XP-D/CS cells cannot re-start transcription of housekeeping genes after UV

In order to further dissect the mechanism by which XP-D/CS (XPD-G675R and XPD-G602D) cells repress transcription after UV irradiation we used HK genes as a model. WT cells displayed a slight decrease followed by the re-establishment of the mRNA levels of the HK gene *Dihydrofolate reductase* (*DHFR*, which was identified in our RNA-seq analysis) upon UV (Figure 2A). On the contrary, upon UV both XP-D/CS fibroblasts displayed a progressive decrease in the mRNA levels of *DHFR* (Figure 2B-C). Interestingly, UV irradiation of XPD-R683W cells displayed an initial decrease in the mRNA of this HK gene, followed by an increase in transcription (Figure 2D). These results are in agreement with our RNA-seq results where WT and XP-D recovered the

expression of many genes (97% for WT and 74% genes for XP-D) whereas XP-D/CS cells did not (see Figure 1A-C black dot for DHFR). Additionally, XP-C cells (devoid of GGR) displayed a decrease in the mRNA levels of *DHFR* upon UV irradiation (10J/m^2), which was recovered within 12h (Figure 2E). Finally, XP-A cells (devoid of NER) displayed also a decrease in the levels of DHFR mRNA after UV, which slowly recovered by 24h (Figure 2F). Similar results were observed for the *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*) HK gene (Figure S1A-F).

We next monitored the recruitment of the transcriptional machinery to the *DHFR* promoter using Chromatin Immuno-precipitation (ChIP) coupled to real time PCR. In WT cells, the transcriptional machinery re-assembled on the promoter of the *DHFR* gene at 6h as shown by the enrichment of Pol II and the transcription initiation factor IIB (TFIIB) (Figure 2G). Furthermore, we also observed at this time, the presence of the TFIIH transcription/repair factor, and the TCR factor CSB, which is recruited and required for the re-assembly of the transcriptional machinery at the promoters of activated genes (24). In WT cells, we observed agreement between the recruitment of Pol II, TFIIB, TFIIH, and CSB, *i.e.*, the re-assembly of the promoter 6h after UV irradiation (Figure 2G and M), which correlated with the restoration of the mRNA levels.

When we monitored the re-assembly of the *DHFR* promoter in XPD-G675R and XPD-G602D cells, neither of the two cell lines was able to re-assemble the transcriptional machinery at this promoter (Figure 2H-I, N-O). The amount of Pol II at the *DHFR* promoter decreased progressively to less than 30% of the initial amount at 12h for both XP-D/CS cells, and did not recover even 12h after UV. Furthermore, none of the transcription initiation factors, including TFIIB, or the repair factor CSB, were recruited to a significant extent or with a particular profile/pattern to these promoters (Figure 2H-I, N-

O). Importantly, the basal protein levels of these transcription and repair factors are similar between the different cell lines (Figure S1O). Interestingly, and in agreement with the mRNA expression data (Figure 2D) and RNA-seq data (Figure 1C), cells from a pure XP-D patient were able to re-assemble the promoters of HK after UV irradiation (Figure 2D, J, P). Thus, mutations in *XPD* that result in XP-D/CS do not allow the re-assembly of the transcriptional machinery on the DHFR promoter after UV, in agreement with the decreased mRNA levels of this gene after UV (Figure 2B-C). These results show that XP-D/CS cells, as CS-B cells, are unable to re-assemble the promoters of HK genes after UV (24). Furthermore, the re-assembly of the HK gene promoters after UV seems to differentiate an XP-D from an XP-D/CS phenotype.

We next asked whether proficient NER was required for the reassembly of the transcriptional machinery on the promoters of HK genes. XP-C cells recovered basal levels of Pol II at the *DHFR* promoter starting at 12h after UV. The recruitment of Pol II was concomitant with the recruitment of the CSB protein and TFIIH (Figure 2K and Q). Finally, in XP-A cells we also observed the recruitment of the CSB protein and Pol II on the promoter of this gene (Figure 2F, L, R). Additionally, very similar results were observed for the *GAPDH* HK gene (Figure S1E-F, K-L). It is important to note that in XP-C and XP-A cells we observed a statistically significant ($p < 0.05$) progressive increase in the levels of mRNA of these HK genes as well as the levels of Pol II recruited at these promoters, in complete contrast to what we observed with XP-D/CS cells, which was a progressive decrease of mRNA and Pol II at these promoters (compare panels in Figure 2 and also Figure S1A-F, M and N). While the removal of the DNA lesions *per se* may not be required for the re-assembly of the transcriptional machinery on these promoters, the presence of functional NER seems to aide the process.

XP-D/CS cells elicit a transcriptional stress-response upon UV irradiation

Since XP-D/CS cells were unable to re-assemble the promoters of HK genes after UV, we investigated whether these cells were able to launch a p53-induced transcriptional response upon DNA damage. WT, both XP-D/CS, and XP-D cells displayed an accumulation of the p53 protein as soon as 1h post-UV (Figure 3A-D). We analyzed the transcriptional response of the p53-inducible growth arrest and DNA damage inducible-alpha (*GADD45α*) gene upon UV. In agreement with the increasing p53 protein levels observed for all four cells lines, we observed an accumulation of the GADD45α mRNA immediately after UV (Figure 3E-H). ChIP analysis on the promoter of the *GADD45α* gene showed increased levels of the transcriptional machinery, Pol II, TFIIH, and p53 (Figure 3I-L). In addition to GADD45a, other DNA damage-inducible genes were also transcribed (Figure S1P-S). Taken together the above results demonstrate that both XP-D/CS cell lines are capable of launching a transcriptional response upon UV irradiation in the presence of these mutated TFIIH, and thus establishes different requirements for the re-assembly of promoters of these two gene families (HK and stress-induced genes). Finally, it is clear that the “global” transcriptional arrest does not include DNA damage-inducible genes.

XP-D/CS cells acquire heterochromatin marks on housekeeping genes

Euchromatin allows transcription and is characterized by acetylated (H3K9-Ac and H4K16-Ac), and methylated (H3K4me3, and H3K79me2) histone H3 and H4 (35-(35). Heterochromatin, on the other hand, inhibits RNA synthesis and is characterized by a different set of chromatin marks such as di- and tri-methylated H3K9 (H3K9me2-3)

and H3K27 (H3K27me2), the recruitment of histone H1, in addition to the loss of euchromatic acetylation and methylation marks (36-39). XP-D/CS cells were unable to re-start the transcription of HK genes after UV, we thus monitored these promoters at the chromatin level.

ChIP analysis of WT cells revealed that the *DHFR* promoter displayed increased levels of H3K9-Ac, H4K16-Ac, H3K4me3 and H3K79m2 upon UV (Figure 4A, E, I, and M). In striking difference, the *DHFR* promoter in both XPD-G675R and XPD-G602D cells displayed no significant increase in H3K9-Ac, H4K16-Ac, H3K4me3 or H3K79me2, but rather a decrease in some of these chromatin marks (Figure 4B-C, F-G, J-K, N-O). Furthermore, we observed a remarkable agreement between the lower levels of these euchromatic marks and the repression of these promoters (*i.e.*, the decreasing levels of mRNA, absence of Pol II, TFIIH, and CSB at this promoter, see Figure 2B-C, H-I, N-O) therefore suggesting the establishment of facultative heterochromatin. Interestingly, in XPD-R683W cells we observed the maintenance of euchromatic marks on the DHFR promoter (Figure 4D, H, L, P), in agreement with the presence of the transcription machinery, and the resumption of mRNA synthesis after UV shown by RNA-seq and qPCR analysis (see Figure 1C and 2D, respectively).

Since we observed a marked decrease in the amount of H3K9-Ac and H4K16-Ac, both of which are substrates for the type III family of HDACs known as sirtuins, we decided to determine whether Sirt1, a member of the sirtuin family and responsible for the formation of facultative heterochromatin, was recruited to these promoters (38, 40). While we only detected background levels of Sirt1 on this promoter in WT and XP-D cells, XP-D/CS cells displayed increased levels of Sirt1 recruited to these promoters (Figure 4Q-T). Furthermore, when we looked at other marks of facultative

heterochromatin such as histone H1 and H3K9me2 we found that, just like Sirt1, they were absent in WT and XP-D cells but they were present in XP-D/CS cells (Figure 4U-Z3) therefore confirming the existence of facultative heterochromatin on these promoters. Interestingly, the lack of active chromatin marks of transcription and the appearance of heterochromatin marks on HK genes was in agreement with the inability of the XP-D/CS cells to recruit the transcriptional machinery to these loci (see Figure 2H-I, N-O). Importantly, these differences in the recruitment of Sirt1 did not arise due to a different expression of this protein in different cell lines, since the basal levels of Sirt1 are very similar among the cell lines studied (Figure S1O). In addition to the *DHFR* gene, we observed a very similar pattern of heterochromatin formation and Sirt1 recruitment on the *GAPDH* HK gene (Figure S1G-J).

Altogether our results show that in XP-D/CS cells HK genes, such as *DHFR* and *GAPDH*, acquire specific heterochromatic marks, while in WT and XP-D cells these genes maintain euchromatic marks and are expressed. Moreover, this UV-induced transcriptional repression is not global, since p53-inducible genes, such as GADD45α, are expressed after UV in all cells studied (Figure 3).

Sirt1 mediates repression of housekeeping genes in XP-D/CS cells after UV

Since histone acetylation seems to be important for re-starting transcription after UV in XP-D/CS cells, we reasoned that inhibition of HDACs might relieve the transcriptional repression of these genes. Pre-treatment of XP-D/CS cells with *N*-butyrate (NaBut), which inhibits type I, II, and IV HDACs did not relieve the progressive decrease in DHFR mRNA upon UV of XP-D/CS cells (Figure S2A-C). To further confirm the involvement of Sirt1, a type III HDAC, in the transcriptional regulation of HK genes

after UV, we first used a pan-inhibitor for all sirtuins, nicotinamide (NAM). Pre-treatment of XP-D/CS cells with NAM resulted in the re-establishment of the mRNA levels of DHFR, while it had no effect in WT cells (Figure 5A-B). Furthermore, ChIP analysis showed that NAM pre-treatment in XP-D/CS cells resulted in the recovery of the recruitment of Pol II and H4K16-Ac levels on the *DHFR* promoter, while no significant changes were observed in WT cells (Figure 5G-H, M-N), further supporting the idea that a member of the type III HDAC family plays a role in the repression of HK genes after UV in XP-D/CS cells. Similar effects were observed with the GAPDH gene (Figure S3)

Since NAM inhibits all sirtuins and also other enzymes such as Poly(ADP-ribose) polymerases (PARPs), we used a Sirt1-specific inhibitor EX-527 (41, 42). Pre-treatment of XP-D/CS cells with EX-527 also resulted in the restoring of the transcription of DHFR in these cells (Figure 5C-D). Importantly, EX-527 pre-treatment also resulted in increased levels of Pol II and H4K16-Ac on the *DHFR* promoter, once again supporting the idea that Sirt1 mediates the repression of HK genes upon UV in XP-D/CS cells. No significant changes were observed in WT cells (Figure 5C-D, I-J, O-P).

Finally, in order to confirm that Sirt1 is responsible for the repression of DHFR upon UV in XP-D/CS cells we depleted cells of Sirt1 by transfecting them with siRNA targeting Sirt1 or a non-specific control. At 72h after siSIRT1 transfection Sirt1 levels were undetectable by Western blot (see WB panel Figure 5E-F). While XP-D/CS cells transfected with the non-specific control displayed the progressive decrease in DHFR mRNA upon UV, cells transfected with siSIRT1 restored the transcription of DHFR (Figure 5E-F). Importantly, XP-D/CS cells transfected with siSIRT1 also displayed re-established levels of Pol II and H4K16-Ac thus confirming that Sirt1 is responsible for the repression of DHFR in XP-D/CS cells after UV (Figure 5K-L, Q-U). It is important to note

that neither the inhibitors nor the siSIRT1 affected the basal expression of neither DHFR nor GAPDH in the absence of UV, thus suggesting that under normal conditions these genes are not under the regulation of Sirt1.

Most importantly, our RNA-seq analysis showed that pre-treatment of XP-D/CS cells with the Sirt1 inhibitor EX-527 significantly ameliorated the transcriptional dysregulation of these cells after UV. Out of the 1,400 genes downregulated more than 3-fold upon UV in XP-D/CS cells, 2h pre-treatment of cells with 50 μM of the Sirt1-specific inhibitor re-established the transcription of 484 genes in these cells (thus ~35% of the genes, see Figure 1B, D, red dots). This amelioration can be easily appreciated by comparing the slope of the red line (k values) for these genes (k=0.12 vs. k=0.49, non-treated vs. EX-527 treated XP-D/CS cells). This effect was not observed XP-D cells (see Figure 1C, E). Our results thus demonstrate that Sirt1 is responsible for the transcriptional repression of a significant number of the genes repressed upon UV in XP-D/CS cells.

Discussion

Discriminated transcription after UV irradiation in XP-D/CS

XP-D/CS cells cannot re-assemble the promoters or re-start transcription of HK genes, such as *DHFR* or *GAPDH*, after UV irradiation, (Figure 1 and 2) similar to CS-B cells (24). It is important to note that through all our studies we have not observed significant differences between the two XP-D/CS cells we studied (XPD-G675R and XPD-G602D). On the other hand, repair-deficient cells (XP-C, XP-A, and XP-D) were able to slowly re-assemble the transcriptional machinery and re-start transcription of these genes after UV (Figure 2). In the *unlikely* case that there were lesions in every HK

gene promoter, our results thus suggest that the re-initiation of transcription after UV does not depend (exclusively) on the removal of DNA lesion.

The reassembly of HK gene promoters did seem to occur concomitant with the recruitment of the CSB chromatin-remodeling factor (24). This SWI2/SNF2 ATPase is involved in transcription elongation, and chromatin remodeling after UV irradiation. The absence of CSB at the HK gene promoters in XP-D/CS cells (even though this protein was expressed Figure S1O), suggests that the XP-D/CS mutations do not support the recruitment and/or the function(s) of CSB. The potential modulation of the function(s) of CSB by TFIIH pinpoints the importance of CSB in transcription initiation, and may be the reason behind the inability of XP-D/CS cells to reassemble the promoters of HK genes after UV.

Interestingly, we also observed that XP-D/CS cells were unable to transactivate NR-inducible genes after UV irradiation (Figure S4), while other repair-deficient cells transactivated NR-genes under the same conditions. A deficiency in NR-transactivation could be explained by the weakened interaction between XPD and p44 that results from the XPD-G675R and XPD-R683W mutations, but not for the XPD-G602D, thus it is likely that the inability to re-start the transcription of HK and NR genes stems from another problem (25). In stark contrast with the transcription of HK and NR-inducible genes, the transcription of the p53-inducible GADD45α gene, as well as of other stress-inducible genes (such as *ATF3* and *p21*), was not impaired in XP-D/CS cells even though these genes may have also been damaged (Figure 3 and S1P-S). Previous work showed that the transcription of these genes does not require CSB (24). The difference between these gene families may lie in the fact that stress-response gene promoters are pre-

assembled awaiting for a stimulus to start elongation, while NR-inducible and HK genes must undergo cycles or assembly and disassembly (24, 29, 30, 43-45).

Finally, our study unveils a difference between the cellular XP, CS, and XP-D/CS phenotypes. It seems that the inability to re-start the transcription of HK genes after UV is strongly linked to the CS phenotype, while the exact mechanisms or reason for the repression mechanism for these genes may be different (Table 3).

Sirt1-mediated heterochromatinization of housekeeping genes

CS-B cells displayed impaired recruitment of the histone acetyl transferase (HAT) p300 and thus lower levels of H3K9-Ac, which caused impaired recruitment of the transcriptional machinery and transcription of the promoters of HK genes upon UV (24, 46, 47). We tested whether the mechanism behind the transcriptional repression in CS-B cells was the same as the one we report here for XP-D/CS cells, but it was not the case. While we did observe lower levels of euchromatin marks (H3K9-Ac, H4K16-Ac, H3K4me3) at the promoters of HK genes in CS-B cells after UV, we did not observe marks of facultative heterochromatin, and in agreement with this, pre-treatment of cells with HDACi NaBut or NAM (or EX-527) did not restore the transcriptional of HK genes in CS-B cells (see Figure S2D). Our results therefore suggest that *histone acetylation* at these promoters is very important for the re-start of transcription after UV and thus faulty histone acetylation (either impaired HAT recruitment for CS-B cells or uncontrolled HDAC recruitment for XP-D/CS cells, or a combination of both) can lead to transcriptional dysregulation and the inability of cells to re-start transcription after UV. It is possible that no one single chromatin modification may be responsible for the inability of CS and XP/CS cells to re-start transcription, but rather a combination of several

deficiencies may create a “perfect storm” for the subsequent heterochromatinization of these promoters. It is not surprising that XPD mutations result in a chromatin dysregulation since more recent studies place NER factors at the intersection between transcription and repair by regulating chromatin structure (24, 29-31).

The impact that the inhibition of Sirt1 had on the chromatin modifications of the promoters of HK genes strongly suggest that Sirt1 is directly changing the chromatin environment at these promoters and not repair indirectly through the deacetylation of another factor (48).

Why is Sirt1 mediating the heterochromatinization of HK genes? Interestingly, upon DNA damage and during normal aging, Sirt1 undergoes a re-distribution, thus abandoning (and thereby allowing the transcription of) typically repressed loci, and regulating another set of genes. This shift was named redistribution of chromatin modifiers (RCM) response and had been observed for oxidative damage and DNA double strand breaks (DSBs) (49-51). The type of DNA damage may be what finally sets apart XP-D/CS from other types of combined XP/CS cells and even CS cells, since XP-D/CS are the only type of XP/CS cells that have been shown to induce DSBs upon UV (18) (Table 3). We propose that the RCM response is also responsible for the repression of constitutively expressed loci, such as those of HK genes in XP-D/CS cells after UV. In support of this model, our RNA-seq analysis showed that pre-treatment with the Sirt1-specific inhibitor EX-527 corrected the expression of a large fraction (~35%, at least) of the genes downregulated in XP-D/CS cells upon UV (Figure 1B, D). Which factor(s) is responsible for the recruitment of Sirt1 to specific genes is unclear. The fact that the promoters that we studied became rather depleted from general transcription factors, including Pol II, TFIIH and CSB, made it difficult to identify a factor that recruits Sirt1 (we

did not detect any interactions between these factors and Sirt1 either) to these promoters and suggest that maybe the substrates for this enzyme (H3K9-Ac, H4K16-Ac, etc.) and the absence of chromatin modifications that would normally block Sirt1 recruitment such as the H3K79me2 (52-54) contribute to the recruitment of Sirt1 to these promoters.

XP-D/CS phenotype and Sirt1

XP-D/CS cells display genomic instability and higher cancer incidence, and a list of severe clinical features (Table 2) (14, 17, 34, 55). Although it would be unthinkable to try and explain all the different clinical features of these patients, studying the list of genes that are repressed upon UV we observed many genes involved in DNA repair (FANCA, FANCI, RAD51L1, RAD54L, POLQ), cell cycle control (RB1), neuronal development (OPTN, BDNF), among others (Table S1) that could potentially be responsible for these features. Importantly, the expression of these genes is restored when cells are treated with the Sirt1 inhibitor EX-527, thus opening the door for potential therapeutic avenues. Furthermore, in our study, inhibition of Sirt1 had a modest enhancement of the survival of XP-D/CS upon UV (Figure S3P-R), thus suggesting that the silencing of HK genes may thwart the ability of these cells to cope with DNA damage. Finally, the link between Sirt1 and XP-D/CS mutations has important implications, since this dysregulation of Sirt1 may contribute to the severe early onset progeria, metabolic problems, and other clinical features observed in these patients that cannot be explained by a DNA repair defect (Table 2).

Materials and Methods

Cell culture

Human primary fibroblasts WT, XP-D/CS (XP8BR, G675R and XPCS2, G602D), XP-D (XP34BE), CS-B (CS1PV), XP-C (GM11847), XP-A (XP39OS) were cultured under standard conditions. Experiments with sirtuin inhibitor nicotinamide (NAM, Sigma) cells were pre-treated with NAM (15mM) for 12h before the experiment (56, 57), UV irradiatied and incubated again with media containing 15mM NAM. Similarly, experiments with EX-527 (Sigma) were conducted by incubating cells with 50μM EX-527 for 2h, irradiating cells (UV-C, 254nm) and incubations cells again with media containing EX-527 (41).

mRNA expression

mRNA extracted using the GeneElute Kit (Sigma). The reverse transcription reaction was done with random primers and Superscript II (Invitrogen) followed by qPCR (Qiagen Syber Green) for the indicated genes. The gene expression was normalized to that of 18S. Primer sequences are available upon request. For the RNA-seq analysis Tag library preparation and high throughput sequencing were conducted on Illumina Genome Analyzer II sequencing system with sequencing depth of 72 nt. Image analysis and base calling were done with CASAVA 1.8.2 (Illumina). Tags alignment, transcript assembly, differential expression analysis and statistical significance calculation were performed in the Galaxy web-based environment using a pipeline TopHat -> Cufflinks -> Cuffdiff with hg19 human genome, FDR = 0.05 and minimal alignment count of 1000 (-c parameter for Cuffdiff).

Chromatin Immunoprecipitation (ChIP)

Experiments were carried out as previously described (29), (24). Briefly, cells were

crosslinked with 1% formaldehyde for 10 min at room temperature, rinsed with 200mM glycine, and cold PBS with protease inhibitors and harvested. Nuclear extracts were sonicated (Diagenode Bioruptor). Chromatin Immuno-precipitation (ChIP) experiments were performed with the indicated antibodies: RNA pol II, XPB, XPD, CSB (IGBMC antibody facility, 7C2, 1B3, 1B5, 1A11/3H8, respectively), TFIIB, p53, Sirt1 (Santa Cruz Biotechnology), H3K9-Ac, H3K9me2, H3K4me3, H3K79me2 (Cell Signaling), H4K16-Ac (Epigenetek), H1 (Millipore) followed by qPCR on the indicated gene promoters. Primer sequences are available upon request. All the results are presented as "fold recruitment" and represent the ratio of the % of input each time point relative to the non-irradiated cells (0h). Each point represents the average of three real time PCR reactions of three independent ChIP experiments. Statistical significance was determined by the student *t* test.

Immuno blots

Cells were UV irradiated (10 J/m^2), harvested at the indicated times in RIPA buffer (25mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% Sodium deoxycholate, 0.1% SDS, and proteases inhibitors cocktail), and whole cell lysates were sonicated. Lysate were separated by SDS-PAGE, transferred to nitrocellulose membrane, and probed with the indicated antibodies: p53, CSB (Santa Cruz Biotechnology), Tubulin a (Abcam), Sirt1 (Upstate). Tubulin was used as a loading control. A representative blot of three independent experiments is shown.

siRNA transfections

Cells were plated at 30% confluence 24h before transfection. Cells were transfected with 50nM non-targeting siCTRL or siSIRT1 using Lipofectamine 2000 (Invitrogen) and OPTI-

MEM media for 24h. After 24h regular media containing FCS was added to cells and incubated for additional 48h before exposing cells to UV irradiation.

Acknowledgements

We are grateful to E. Compe and N. Le May for their critical comments on the manuscript. This study has been supported by l'Agence National de la Recherche (ANR-08-GENOPAT-042), l'Association de la Recherche contre le Cancer, European Research Council Advanced Scientists grants (JME) and the Institut National du Cancer (INCA-2008-041 to FC). RVC was supported by the Chateaubriand and Association de la Recherche contre le Cancer post-doctoral fellowships. AZ was supported by advanced ERC.

References

1. Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461:1071–1078.

2. Tornaletti S (2009) DNA repair in mammalian cells: Transcription-coupled DNA repair: directing your effort where it's most needed. *Cell Mol Life Sci* 66:1010–1020.

3. Cleaver JE, Lam ET, Revet I (2009) Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet* 10:756–768.

4. Schumacher B, Hoeijmakers JH, Garinis GA (2009) Sealing the gap between nuclear DNA damage and longevity. *Mol Cell Endocrinol* 299:112–117.

5. Mellon I, Bohr VA, Smith CA, Hanawalt PC (1986) Preferential DNA repair of an active gene in human cells. *Proc Natl Acad Sci USA* 83:8878–8882.

6. Bohr VA, Smith CA, Okumoto DS, Hanawalt PC (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* 40:359–369.

7. Fousteri M, Vermeulen W, van Zeeland AA, Mullenders LHF (2006) Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol Cell* 23:471–482.

8. Lainé J-P, Egly J-M (2006) Initiation of DNA repair mediated by a stalled RNA polymerase IIO. *EMBO J* 25:387–397.

9. Mocquet V et al. (2008) Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J* 27:155–167.

10. Staresincic L et al. (2009) Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J* 28:1111–1120.

11. Lehmann AR (2001) The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev* 15:15–23.

12. Cleaver JE (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer* 5:564–573.

13. Kraemer KH et al. (2007) Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience* 145:1388–1396.

14. Ellison AR, Nouspikel T, Jaspers NG, Clarkson SG, Gruenert DC (1998) Complementation of transformed fibroblasts from patients with combined xeroderma pigmentosum-Cockayne syndrome. *Exp Cell Res* 243:22–28.

15. van Hoffen A et al. (1999) Cells from XP-D and XP-D-CS patients exhibit equally inefficient repair of UV-induced damage in transcribed genes but different capacity to recover UV-inhibited transcription. *Nucleic Acids Res* 27:2898–2904.

16. Nouspikel T, Lalle P, Leadon SA, Cooper PK, Clarkson SG (1997) A common mutational pattern in Cockayne syndrome patients from xeroderma pigmentosum group G: implications for a second XPG function. *Proc Natl Acad Sci USA* 94:3116–3121.

17. Broughton BC et al. (1995) Molecular and cellular analysis of the DNA repair defect in a patient in xeroderma pigmentosum complementation group D who has the clinical features of xeroderma pigmentosum and Cockayne syndrome. *Am J Hum Genet* 56:167–174.

18. Theron T et al. (2005) Transcription-associated breaks in xeroderma pigmentosum group D cells from patients with combined features of xeroderma pigmentosum and Cockayne syndrome. *Mol Cell Biol* 25:8368–8378.

19. Compe E, Egly J-M (2012) TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* 13:343–354.

20. Ito S et al. (2007) XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Mol Cell* 26:231–243.

21. Drané P, Compe E, Catez P, Chymkowitch P, Egly J-M (2004) Selective regulation of vitamin D receptor-responsive genes by TFIIH. *Mol Cell* 16:187–197.

22. Compe E et al. (2005) Dysregulation of the peroxisome proliferator-activated receptor target genes by XPD mutations. *Mol Cell Biol* 25:6065–6076.

23. Compe E et al. (2007) Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH. *Nat Neurosci* 10:1414–1422.

24. Proietti-De-Santis L, Drané P, Egly J-M (2006) Cockayne syndrome B protein regulates the transcriptional program after UV irradiation. *EMBO J* 25:1915–1923.

25. Dubaele S et al. (2003) Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol Cell* 11:1635–1646.

26. Cleaver JE, Hefner E, Laposa RR, Karentz D, Marti T (2007) Cockayne syndrome exhibits dysregulation of p21 and other gene products that may be independent of transcription-coupled repair. *Neuroscience* 145:1300–1308.

27. Laposa RR, Huang EJ, Cleaver JE (2007) Increased apoptosis, p53 up-regulation, and cerebellar neuronal degeneration in repair-deficient Cockayne syndrome mice. *Proc Natl Acad Sci USA* 104:1389–1394.

28. Mayne LV, Lehmann AR (1982) Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Research* 42:1473–1478.

29. Le May N et al. (2010) NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell* 38:54–66.

30. Le May N, Fradin D, Iltis I, Bougnères P, Egly J-M (2012) XPG and XPF Endonucleases Trigger Chromatin Looping and DNA Demethylation for Accurate Expression of Activated Genes. *Mol Cell*.

31. Fong YW et al. (2011) A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. *Cell* 147:120–131.

32. Hoppe S et al. (2009) AMP-activated protein kinase adapts rRNA synthesis to cellular energy supply. *Proceedings of the National Academy of Sciences* 106:17781–17786.

33. Cleaver JE, Revet I (2008) Clinical implications of the basic defects in Cockayne syndrome and xeroderma pigmentosum and the DNA lesions responsible for cancer, neurodegeneration and

aging. *Mech Ageing Dev* 129:492–497.

34. Takayama K et al. (1995) Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D. *Cancer Research* 55:5656–5663.

35. Steger DJ et al. (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* 28:2825–2839.

36. Kim J-H, Workman JL (2010) Histone acetylation in heterochromatin assembly. *Genes Dev* 24:738–740.

37. Stewart MD, Li J, Wong J (2005) Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol Cell Biol* 25:2525–2538.

38. Vaquero A et al. (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol Cell* 16:93–105.

39. Verzijlbergen KF, Faber AW, Stulemeijer IJ, van Leeuwen F (2009) Multiple histone modifications in euchromatin promote heterochromatin formation by redundant mechanisms in Saccharomyces cerevisiae. *BMC Mol Biol* 10:76.

40. Donmez G, Guarente L (2010) Aging and disease: connections to sirtuins. *Aging Cell* 9:285–290.

41. Lavu S, Boss O, Elliott PJ, Lambert PD (2008) Sirtuins--novel therapeutic targets to treat age-associated diseases. *Nat Rev Drug Discov* 7:841–853.

42. Haigis MC, Sinclair DA (2010) Mammalian sirtuins: biological insights and disease relevance. *Annu Rev Pathol* 5:253–295.

43. Muse GW et al. (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* 39:1507–1511.

44. Gilchrist DA et al. (2010) Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell* 143:540–551.

45. Gomes NP et al. (2006) Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev* 20:601–612.

46. Frontini M, Proietti-De-Santis L (2009) Cockayne syndrome B protein (CSB): linking p53, HIF-1 and p300 to robustness, lifespan, cancer and cell fate decisions. *Cell Cycle* 8:693–696.

47. Filippi S et al. (2008) CSB protein is (a direct target of HIF-1 and) a critical mediator of the hypoxic response. *EMBO J* 27:2545–2556.

48. Fan W, Luo J (2010) SIRT1 Regulates UV-Induced DNA Repair through Deacetylating XPA. *Mol Cell* 39:247–258.

49. O'Hagan HM, Mohammad HP, Baylin SB (2008) Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet* 4:e1000155.

50. O'Hagan HM et al. (2011) Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* 20:606–619.

51. Oberdoerffer P et al. (2008) SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. *Cell* 135:907–918.

52. Altaf M et al. (2007) Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin. *Mol Cell* 28:1002–1014.

53. Milner J (2009) Cellular regulation of SIRT1. *Curr Pharm Des* 15:39–44.

54. Kang M-R et al. (2010) Reciprocal roles of SIRT1 and SKIP in the regulation of RAR activity: implication in the retinoic acid-induced neuronal differentiation of P19 cells. *Nucleic Acids Res* 38:822–831.

55. Hasty P, Campisi J, Hoeijmakers J, Van Steeg H, Vijg J (2003) Aging and genome maintenance: lessons from the mouse? *Science* 299:1355–1359.

56. Grubisha O, Smith BC, Denu JM (2005) Small molecule regulation of Sir2 protein deacetylases. *FEBS J* 272:4607–4616.

57. Zhang T, Kraus W (2009) SIRT1-dependent regulation of chromatin and transcription: Linking NAD(+) metabolism and signaling to the control of cellular functions. *Biochim Biophys Acta*.

Figure 1

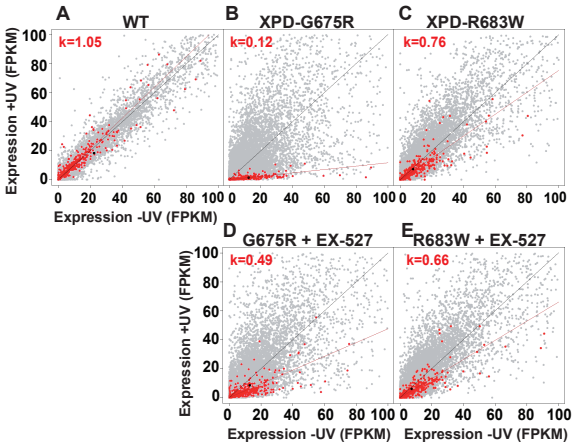


Figure 1: XP-D/CS cells display a gross gene dysregulation after UV. RNA-sequencing analysis scatter plots show the transcription of all genes read at 24h after 10J/m2 UV-C vs. untreated conditions for (A) WT, (B) XP-D/CS (XPD-G675R) and (C) XP-D (XPD-R683W). Each grey dot represents a gene. The black dot represents DHFR. (D) XP-D/CS and (E) XP-D cells pre-treated with the Sirt1-specific inhibitor EX-527 (50μM) for 2h before UV. The red dots represent the 484 genes that were repressed > 3-fold in XP-D/CS cells and whose expression was restored by the Sirt1-specific inhibitor EX-527. Red line represents the best linear fit for the genes in red. k is the slope for the red line. Axes presented as reads frame per kilobase per million (FPKM). Treatment of XP-D/CS cells resulted in the significant change in expression (statistically significant ± 3-fold change) of 2,024 genes (600 genes overexpressed and 1424 genes repressed).

Figure 2

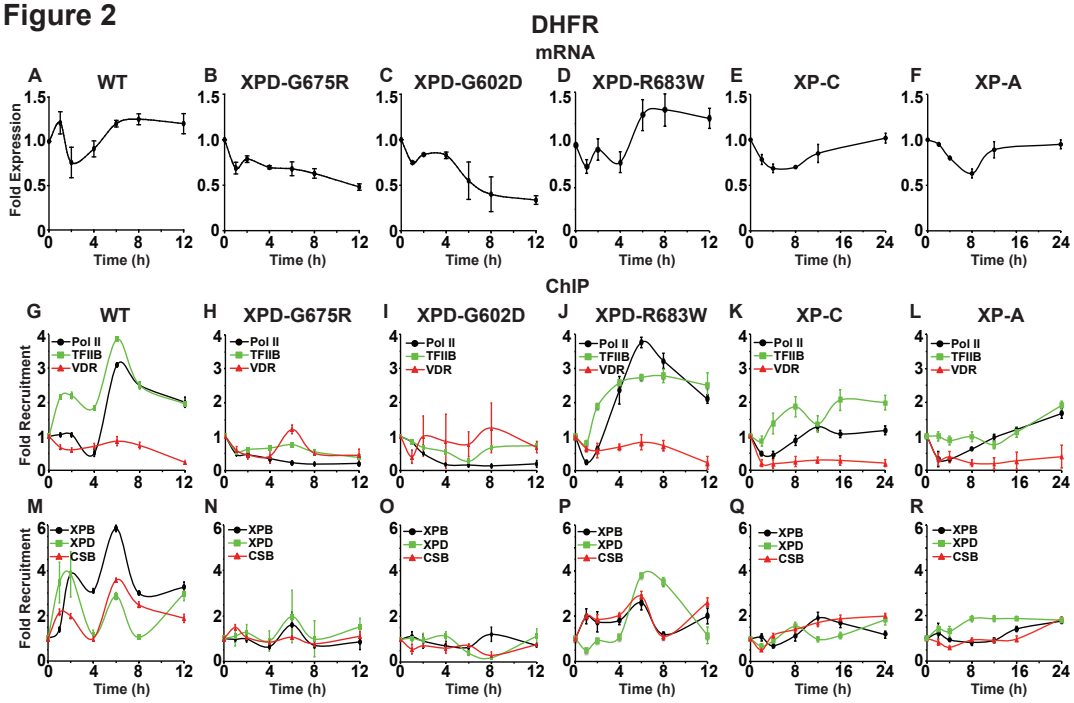


Figure 2: XP-D/CS cells cannot re-start transcription of housekeeping genes after UV. (A-F) Relative mRNA expression of DHFR after UV irradiation (10J/m2). DHFR mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments. (G-L) ChIP monitoring the occupancy of general transcription factor TFIIB, RNA pol II (Pol II), vitamin D receptor (VDR, as a negative control), (M-R) TFIIH subunits XPB and XPD, and the TCR factor CSB, at the promoter of the DHFR gene in WT and XP-D/CS (G675R and G602D), XP-D (R683W), XP-C, and XP-A cells. All the results are presented as "fold recruitment" which represents the ratio of the % input of each time point relative to that of the non-irradiated cells (t = 0h). Each point represents the average of three real time PCR reactions of three independent experiments and error bars represent SEM.

Figure 3

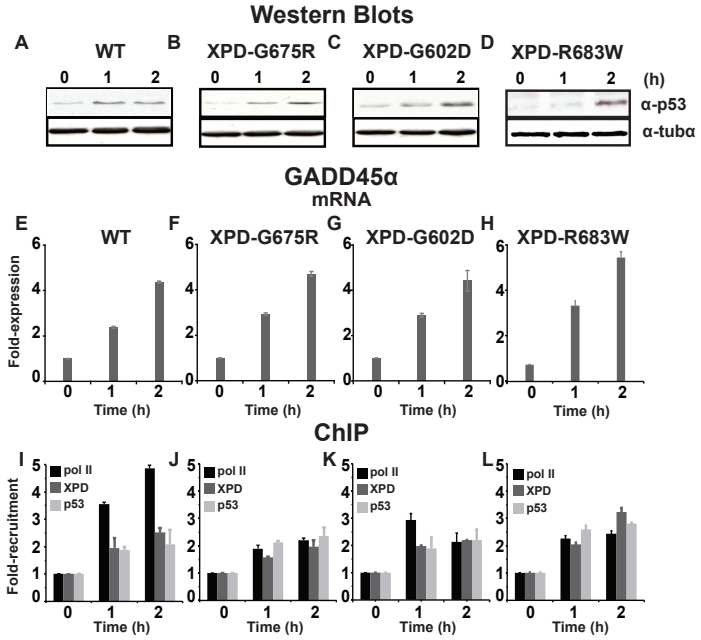


Figure 3: XP-D/CS cells transcribe GADD45α after UV irradiation. (A-D) p53 protein accumulates upon UV irradiation (10 J/m2) in WT, XP-D/CS (G675R and G602D), and XP-D (R683W) cells. Fifty micrograms of whole cell extracts were separated by SDS-PAGE, transferred to nitrocellulose membrane, and probed with the indicated antibodies. Tubulinα was used as a loading control. A representative blot of three independent experiments is shown. (E-H) GADD45α mRNA is expressed upon UV irradiation (10 J/m2). GADD45α mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" as previously described. Error bars represent the SEM of three independent experiments. (I-L) ChIP monitoring the occupancy of RNA pol II, XPD, and p53 at the promoter of the GADD45α gene of WT, XP-D/CS (G675R and G602D) and XP-D (R683W) cells. All the results are presented as "fold recruitment" as previously described. Each point represents the average of three real time PCR reactions of three independent experiments and error bars represent SEM.

Figure 4

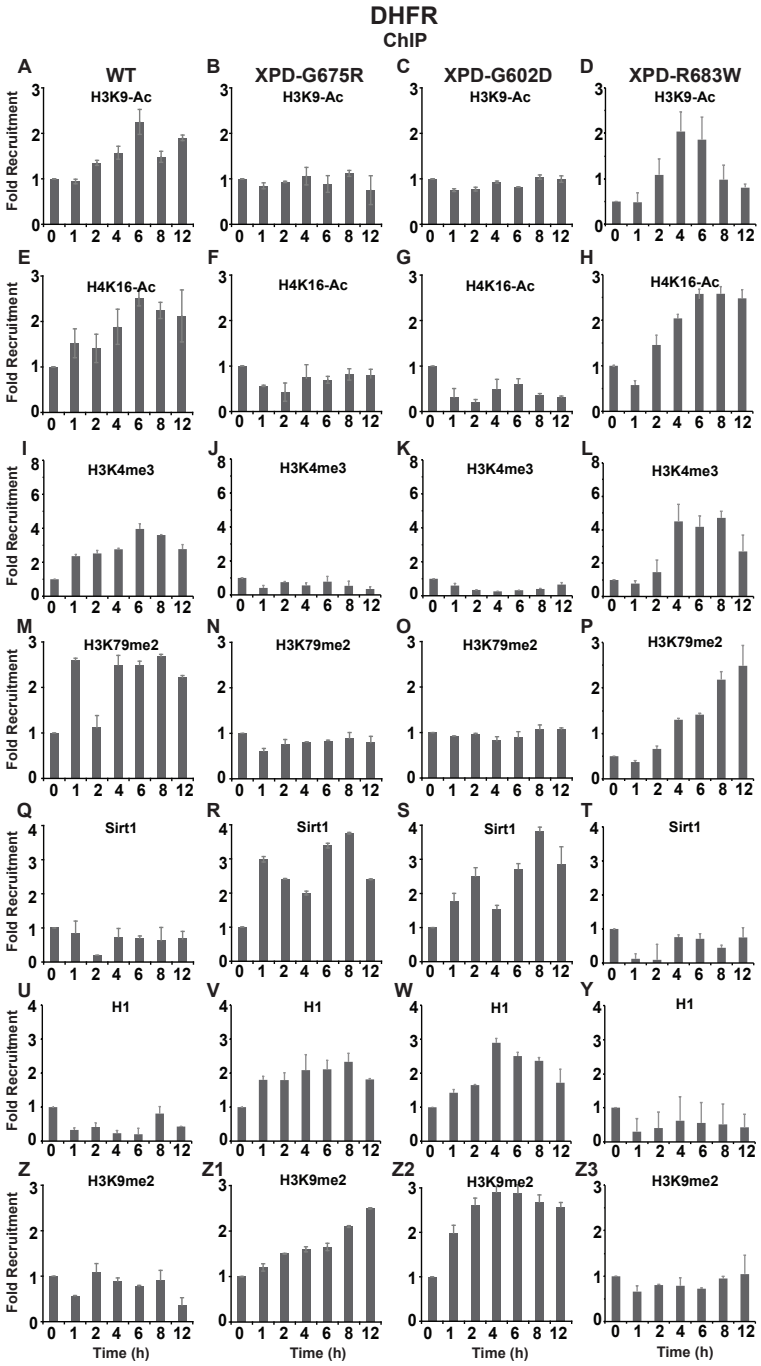


Figure 4: XP-D/CS cells loose euchromatin marks and acquire heterochromatin marks on the promoter of DHFR after UV. ChIP monitoring the occupancy of (A-D) H3K9-Ac, (E-H) H4K16-Ac, (I-L) H3K4me3, (M-P) H3K79me2, (Q-T) Sirt1, (U-Y) H1, (Z-Z3) H3K9me2 on the promoter of the DHFR gene in WT, XP-D/CS (G675R and G602D), and XP-D (R683W) cells. All the results are presented as "fold recruitment" as previously described. Each point represents the average of three real time PCR reactions of at three independent ChIP experiments and error bars represent SEM.

Figure 5

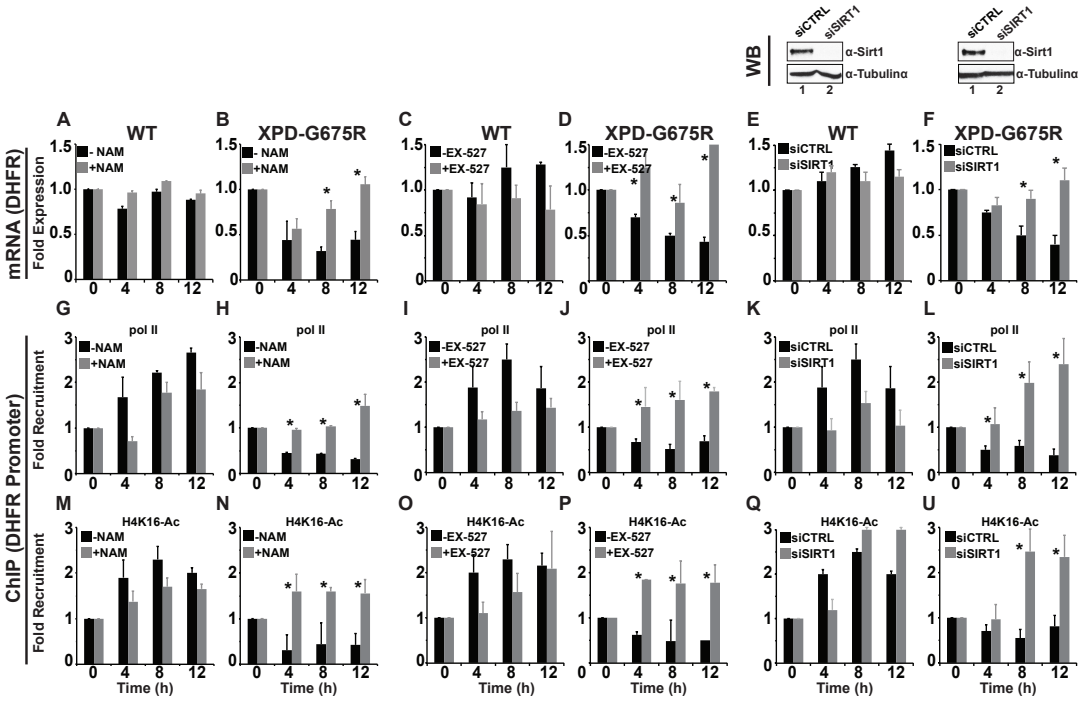


Figure 5: Inhibition of Sirt1 restores transcription of housekeeping genes in XP-D/CS cells. DHFR mRNA expression at the indicated times after UV irradiation (10J/m2) of WT of XP-D/CS (G675R) cells (A-B) pre-treated for 12h with 15 mM nicotinamide (NAM), (C-D) pre-treated for 2h with 50 μM EX-527, (E-F) previously transfected with siRNA targeting SIRT1 (siSIRT1) or a non-targeting control (siCTRL). The expression levels of the Sirt1 protein are shown on the immuno blot insert at the top of panels (E-F). ChIP monitoring the occupancy of (G-L) RNA pol II, (M-U) H4K16-Ac on the promoter of the DHFR gene in WT and XP-D/CS cells pre-treated with NAM, EX-527, or siSIRT1. All the results are presented as "fold recruitment" as previously described. Each point represents the average of three real time PCR reactions of three independent ChIP experiments and error bars represent SEM. The * denotes statistical significance (* = $p < 0.05$, ** = $p < 0.01$) calculated by student t test.

Table 1: Clinical Features of XP, TTD, CS, and XP/CS patients.

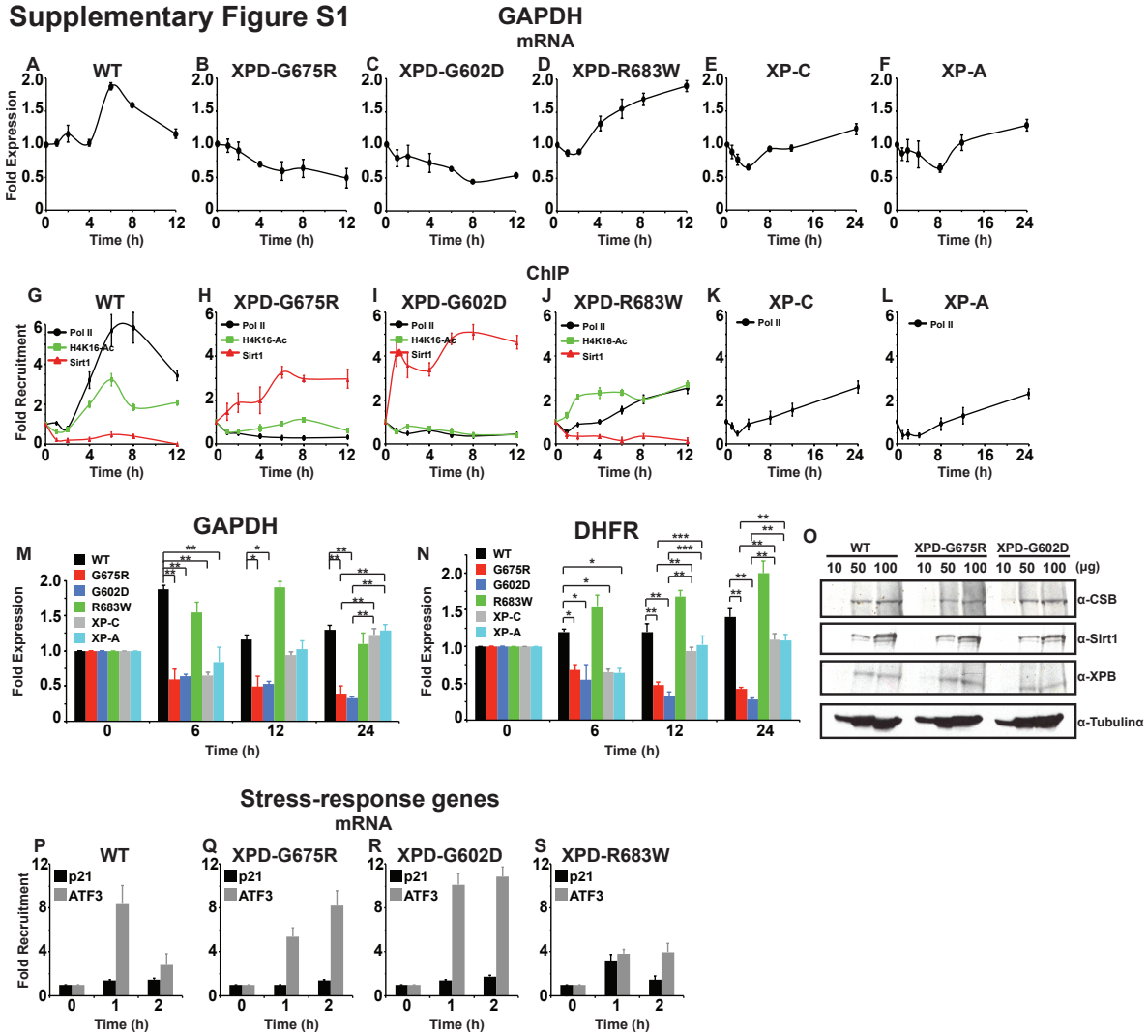
Genetic Syndrome	*Clinical Features*	*Mutated Genes*
XP	Sun hypersensitivity > 1,000-fold cancer susceptibility Abnormal skin pigmentation	XPA-G and XP-V
TTD	Brittle hair Mental retardation Ataxia Tremors	TTDA, XPD
CS	Sun hypersensitivity Cachectic dwarfism Severe mental retardation Skeletal and retinal abnormalities Segmental progeria	CSA, CSB
XP/CS	Extreme sun hypersensitivity Hair thinning and freckling Neurological development abnormalities Severe mental retardation Abnormal skin pigmentation High cancer proneness Segmental progeria	XPB, XPD, XPG

Table 2: Clinical and Molecular features of XP-D/CS patients and proteins

Mutation	Patient clinical symptoms	Molecular defects of XPD
XP8BR -G675R -fs669 (likely to inactivate XPD)	Developmental retardation Neurological development delay Failure to thrive Skin pigmentation abnormalities Freckling Thin hair Sensitivity to sun light Died at age 2.5 y.o.	Weakened interaction between p44 and XPD Very low helicase activity
XPCS2 -G602D -second allele not expressed	Growth and mental retardation Neurological development delay Acute sun sensitivity Freckled skin Skin cancers Died at 13 y.o.	No helicase activity No NER activity

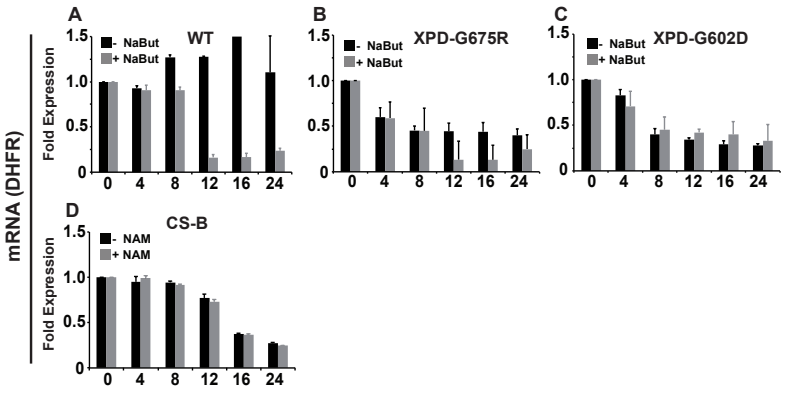
Table 3: Different characteristics of NER-deficient cells.

	WT	XP-A	CS-B	XP-D/CS	XP-D
Housekeeping genes (DHFR, GAPDH) after UV	+	+	-	-	+
p53-inducible genes (*GADD45a*, *p21*, *MDM2*)	+	+	+	+	+
Nuclear receptor genes (RARb2) after UV	+	+	**+/-**	-	-
Recruitment of CSB to HK gene promoters	+	+	-	-	+
Histone acetylation (H3K9-Ac, H4K16-Ac)	+	+	-	-	+
Heterochromatinization (Sirt1-mediated)	-	-	-	+	-
Transcriptional arrest after UV	-	-	+	+	-
Correction of the transcriptional arrest by NAM, EX-527, siSIRT1	N/A	N/A	-	+	N/A

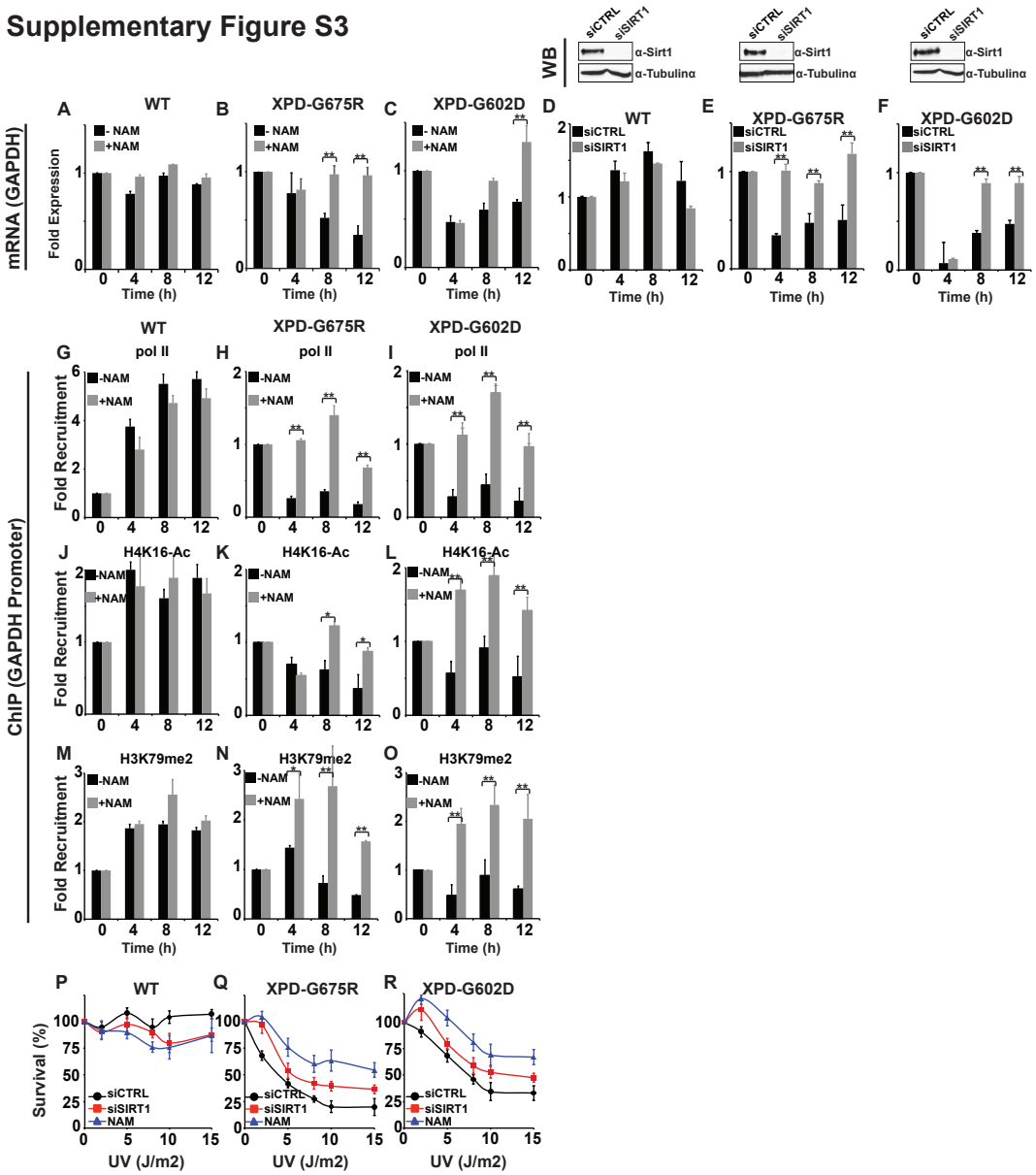


Supplementary Figure S1: XP-D/CS cells cannot re-start transcription of the GAPDH gene after UV. (A-F) GAPDH mRNA expression of cells after UV (10J/m2). GAPDH mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments. (G-L) ChIP monitoring the occupancy of RNA pol II (pol II), H4K16-Ac and Sirt1 on the promoter of the GAPDH gene in WT, XP-D/CS (G675R and G602D), XP-D (R683W), XP-C, and XP-A cells. ChIP results are presented as "fold recruitment" which represents the ratio of the % input of each time point relative to that of the non-irradiated cells (t = 0h). Each point represents the average of three real time PCR reactions of three independent ChIP experiments. (M-N) mRNA expression of DHFR and GAPDH genes at different time points for different cell lines showing the statistical significance. The * denotes statistical significance (* = $p < 0.05$, ** = $p < 0.01$) calculated by student t test. (O) Western blots showing protein levels of CSB, Sirt1, and XPB. (P-S) Expression of stress-response genes (p21 and ATF3) upon UV (10 J/m2). mRNA levels were normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments.

Supplementary Figure S2



Supplementary Figure S2: Inhibition of type I, II, and IV HDACs do not restore transcription of DHFR in XP-D/CS cells (A-C) WT and XP-D/CS cells were pre-treated with 15 mM Sodium butyrate (NaBut) for 12h and then irradiated with UV (10 J/m2) and mRNA was extracted at the indicated times. (D) CS-B cells were pre-treated with 15mM nicotinamide (NAM) for 12h and then irradiated with UV (10J/m2). DHFR mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments.



Supplementary Figure S3: Inhibition or downregulation of Sirt1 restores transcription of HK gene GAPDH after UV irradiation in XP-D/CS cells. (A-C) Relative GAPDH expression of cells pre-treated with nicotinamide (NAM, 0 or 15 mM) for 12h before UV irradiation (10J/m2). (D-F) Relative GAPDH mRNA expression after UV irradiation (10J/m2) of cells previously transfected with siRNA targeting SIRT1 (siSIRT1) or a non-targeting control (siCTRL). On both cases GAPDH mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments. (G-X) ChIP monitoring the occupancy of (G-I) RNA pol II, (J-L) H4K16-Ac, and (M-O) H3K79me2 at the promoter of the DHFR gene in WT and XP-D/CS (G675R and G602D) cells pre-treated with 0 (-NAM) or 15 mM (+NAM) nicotinamide (NAM). All the results are presented as "fold recruitment" which represents the ratio of the % input of each time point relative to that of the non-irradiated cells (t = 0h). Each point represents the average of three real time PCR reactions of at least two independent ChIP experiments. The * denotes statistical significance (* = $p < 0.05$, ** = $p < 0.01$) calculated by student t test. Inhibition of Sirt1 enhances survival of XP-D/CS cells upon UV. (A-C) WT, XPD-G675R, XPD-G602D cells were transfected with siCTRL, siSirt1 for 72h or pre-treated for 12h with 15 mM NAM prior to increasing doses of UV (254nm) and incubated with fresh media for 72h before measuring survival spectrophotometrically (595 nm) with crystal violet. Error bars represent the SEM of three independent experiments.

Supplementary Figure S4

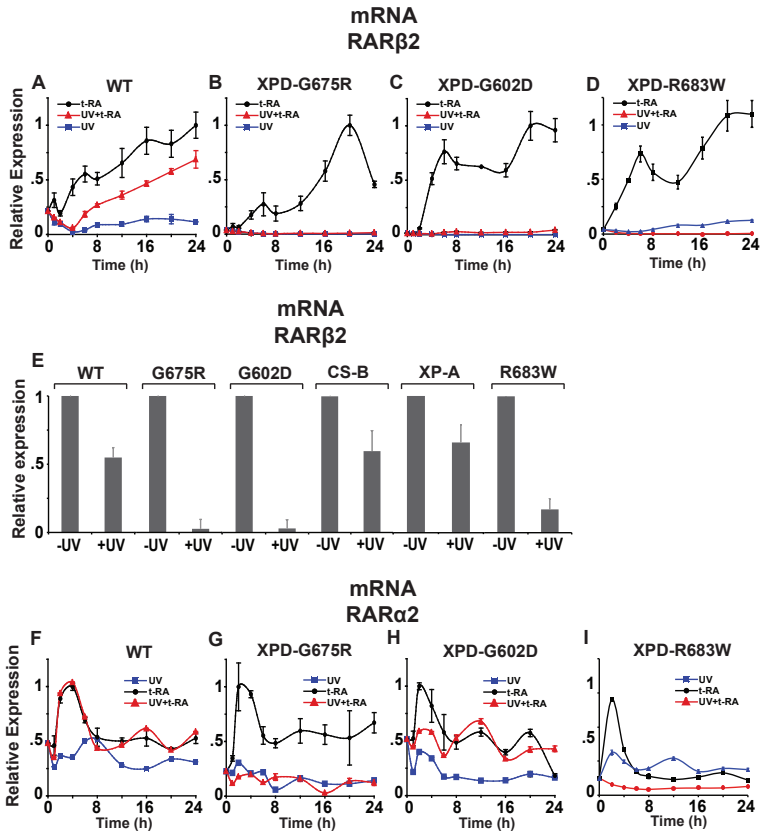


Figure S4: UV ablates NR-transactivation in XP-D/CS cells. (A-D) RARβ2 mRNA expression upon trans retinoic acid (tRA, 10 μM, black lines) treatment, UV irradiation (10J/m2, blue lines) or simultaneous t-RA and UV treatments (red lines). RARβ2 mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments. (E) Cells (WT, XP-D/CS, CS-B, XP-A, and XP-D (R683W) were treated with tRA (10μM) or tRA and UV (10J/m2) for 12h. (F-I) RARα2 mRNA expression upon trans retinoic acid (tRA, 10 μM) treatment. RARα2 mRNA was normalized to the amount of 18S rRNA and results are presented as "fold expression" which represents the ratio of each time point relative to the non-irradiated cells. Error bars represent the SEM of three independent experiments.

Table S1: Selected genes that were repressed upon UV in XP-D/CS and their expression was restored upon EX-527 treatment.

Gene	Gene function
MCC	candidate colorectal tumor suppressor gene that is thought to negatively regulate cell cycle progression.
MET	Proto-oncogene hepatocyte growth factor receptor and encodes tyrosine-kinase activity
FANCA	Fanconi anemia is a genetically heterogeneous recessive disorder characterized by cytogenetic instability, hypersensitivity to DNA crosslinking agents, increased chromosomal breakage, and defective DNA repair. Assembly into a common nuclear protein complex. Mutations in this gene are the most common cause of Fanconi anemia.
FANCI	Required for maintenance of chromosomal stability. Involved in the repair of DNA double-strand breaks by homologous recombination and in the repair of DNA cross-links. Participates in S phase and G2 phase checkpoint activation upon DNA damage. Promotes FANCD2 ubiquitination and recruitment to DNA repair sites
TRRAP	Adapter protein found in various multiprotein chromatin complexes with HAT activity. Component of the NuA4 HAT complex which is responsible for acetylation of nucleosomal histones H4 and H2A.
RAD51L1	Involved in the homologous recombination repair (HRR) pathway of double-stranded DNA breaks arising during DNA replication or induced by DNA-damaging agents.
RAD54L	Involved in DNA repair and mitotic recombination. Functions in the recombinational DNA repair (RAD52) pathway. Dissociates RAD51 from nucleoprotein filaments formed on dsDNA. Could be involved in the turnover of RAD51 protein-dsDNA filaments (By similarity). May play also an essential role in telomere length maintenance and telomere capping in mammalian cells
ATRX	belongs to the SWI/SNF family of chromatin remodeling proteins. The mutations of this gene are associated with an X-linked mental retardation (XLMR) syndrome most often accompanied by alpha-thalassemia (ATRX) syndrome.
POLQ	DNA polymerase on nicked double-stranded DNA and on a singly primed DNA template. The enzyme activity is resistant to aphidicolin, and inhibited by dideoxynucleotides. Exhibits a single-stranded DNA-dependent ATPase activity. Could be involved in the repair of interstrand cross-links
FOXO3	Transcriptional activator which triggers apoptosis in the absence of survival factors, including neuronal cell death upon oxidative stress.
OXR1	May be involved in protection from oxidative damage
USP9X	Deubiquitinase involved both in the processing of ubiquitin precursors and of ubiquitinated proteins. Regulates chromosome alignment and segregation in mitosis by regulating the localization of BIRC5/survivin to mitotic centromeres.
OPTN	Plays an important role in the maintenance of the Golgi complex, in membrane trafficking, in exocytosis, through its interaction with myosin VI and Rab8. Neuroprotective role in the eye and optic nerve.
UBE3A	E3 ubiquitin-protein ligase for RAD23A and RAD23B, MCM7 (which is involved in DNA replication), annexin A1, the PML tumor suppressor, and the cell cycle regulator CDKN1B.
DHFR	required for the de novo synthesis of purines, thymidylic acid, and certain amino acids. Its deficiency has been linked to megaloblastic anemia.
BDNF	The protein encoded by this gene is a member of the nerve growth factor family. It is induced by cortical neurons, and is necessary for survival of striatal neurons in the brain. Expression of this gene is reduced in both Alzheimer's and Huntington disease patients. This gene may play a role in the regulation of stress response and in the biology of mood disorders.
RSF1	Required for assembly of regular nucleosome arrays by the **RSF** chromatin-remodeling complex.
MYST3	Histone acetyltransferase which may be involved in transcriptional activation. May influence the function of ATM
RB1	Key regulator of cell division that acts as a tumor suppressor. Promotes G0-G1 transition when phosphorylated by CDK3/cyclin-C. Acts as a transcription repressor of E2F1 target genes. Directly involved in heterochromatin formation by maintaining overall chromatin structure and, in particular, that of constitutive heterochromatin by stabilizing histone methylation.

Chapter 7

TFIIH complex and transcription of class III genes

7.1 Publication 2

In preparation.

TFIIH complex is directly involved in the transcription by RNA polymerase III

Anton S. Zadorin, Frédéric Coin and Jean-Marc Egly

Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/ULP, BP 163, 67404 Illkirch Cedex, C. U. Strasbourg, France

TFIIH multisubunit complex was long recognized as an RNA polymerase II (Pol II) transcription factor and a an essential factor of nucleotide excision DNA repair. It also has been shown to participate in the transcription mediated by Pol I. The complex consists of two subcomplexes: the core and the CAK. In this study we investigated genome-wide localization of both subcomplexes of TFIIH, Pol II and Pol III in human primary fibroblasts. We found that the core TFIIH is strongly associated with Pol III transcribed genes, whereas they are free from the CAK and Pol II. *In vitro* transcription with a nuclear extract depleted of the core TFIIH and subsequent addition of the purified complex demonstrated the direct involvement of this factor in transcription mediated by Pol III. This unexpected finding reveals TFIIH as a truly universal factor for all three main eukaryotic RNA polymerases.

Introduction

In mammalians, there are three principal nuclear DNA-directed RNA polymerases (Pol), that take care of the RNA production. Pol I is responsible for transcription of most rRNA genes (class I genes). Pol II produces all mRNAs and a variety of non-coding RNAs such as majority of snR-

NAs and miRNAs (class II). Pol III is required for the transcription of a distinct group of short non-coding genes with specialised function (class III genes). Class III genes products include 5S rRNA, all tRNAs, U6 snRNA, 7SK RNA (a regulator of Pol II transcription), 7SL RNA, Y RNAs, RNA parts of RNase P and RNase MRP and some others[1]. All three polymerases have common evolutionary origin and share five core subunits[2]. Despite this, they require different sets of additional transcription factors. This difference is partially due to inability of a polymerase to recognize its promoter and to directly initiate its own transcription. Thus, the recruitment of transcription machinery to the transcription starting site is carried out by additional factors. They also determine the promoter specificity of a given polymerase and give means for regulation of genetic expression at transcription level.

Pol III promoters are subdivided into three types. Type 1 and type 2 promoters reside inside the transcribed portion of the genome. 5S RNA genes contain type 1 promoter, while transcription of almost all other class III genes (including tRNA genes) starts from type 2 promoters. In contrast, type 3 promoters, such as promoters of U6 RNA genes, are external and are situated upstream to the transcription starting site like promoters of other polymerases. Unlike for Pol II, there have been only few basal transcription factors identified for Pol III. The transcription regulation is still poorly understood, though it is clear that Pol III transcription is controlled. This follows from the fact that only a portion of tRNA genes are active in mammalian genome and the set of active tRNA genes is cell line specific. From the other hand, in many cancer cells Pol III transcription activity is significantly elevated[1].

TFIIH is a general transcription factor of Pol II that is also a key factor in nucleotide excision repair. Mammalian TFIIH is a ten subunit complex that contains two functional subcomplexes: the core TFIIH that consists of XPB helicase, p62, p52, p44, p34 and p8 subunits; and the CAK subcomplex consisting of CDK7 kinase, cyclin H and MAT1. These two parts are connected *via* XPD helicase. The role of TFIIH in Pol II transcription initiation is thought to be the promoter opening with XPB helicase activity and the phosphorylation of Pol II and various transcription factors with CDK7 kinase. Mutations in XPB, XPD and p8 subunits are know to cause unusually wide spectrum of disorders, including *xeroderma pigmentosum*, trichothiodystrophy, or the combined *xeroderma pigmetosum* with Cockayne syndrome[3].

Recently, there have been reports of close association of Pol II, some its auxiliary transcription factors and active chromatin marks with a fraction of Pol III bound class III genes in some cancerous and immortalized human cell lines[4, 5, 6]. A similar study on mouse embryonic stem cells, supporting peculiarities of chromatin, could not detect Pol II. Instead, the general transcription factor TFIIS was found to enrich Pol III transcribed genes[7].

Here we report that in normal human fibroblasts the core TFIIH is associated with active class III genes, as demonstrated by genome-wide study. This association is independent of Pol II. *In vitro* experiments show that it appears to be caused by the direct involvement of TFIIH in Pol III mediated transcription.

Results

Core TFIIH is enriched on Pol III occupied genes. We examined genome-wide co-occurrence of Pol II (RPB1 subunit), Pol III (RPC7 subunit), core TFIIH (p62 subunit) and the CAK subcomplex (CDK7 kinase) in normal lung foetal fibroblasts (MRC-5 cell line) by chromatin immunoprecipitation followed by high throughput parallel sequencing (ChIP-seq). To identify actively transcribed genes we used an antibody directed towards the phosphorylated form of the carboxy-terminal domain of Pol II. To our surprise, the visual examination of the obtained ChIP-seq profiles in UCSC genome browser[8] revealed high enrichment of the core TFIIH (visualized by p62) on some Pol III transcribed genes. ChIP-seq analysis showed a co-occupancy of class III genes by p62 and Pol III (Fig. 1a). We noticed that enrichment peaks were higher for p62 on class III genes, than for p62 on class II genes. For example, we observed the co-occupancy of p62 and Pol III on a cluster of tRNA genes on chromosome 6 and on *RNU6-1*, *RN7SK*, *RNY1*, three Pol III transcribed genes. Interestingly, neither Pol II nor Ckd7 were visibly present on those genes (Fig. 1a, see green marks). At the same time, on class II genes such as *HIST1H4H* and *RNU5A-1*, we observed an enrichment of CDK7, p62 and Pol II (Fig. 1a in blue for an overview of a cluster of histone coding genes and for these two genes).

To further have a quantitative evaluation of the visually observed co-localization of TFIIH and Pol III we used MACS software[9]. This tool is optimized for finding distinct more or less identical short peaks of enrichment in ChIP-seq which is particularly useful for peak detection on class III genes due to their short length. We selected only peaks with the false detection rate

(FDR) lower than 5%. Such strict selection disregarded the majority of TFIIH peaks (both p62 and CDK7) associated with Pol II due to their low height and varying shapes, but retained many peaks of Pol III on tDNA. Under those conditions, from overall 218 $(107 + 89 + 22)$ detected p62 peaks only 20 $(8+7+5)$ overlapped with the peaks for Pol II. Nevertheless, more than a half $(107+5)$ of p62 peaks still overlapped with Pol III enrichment peaks (Fig. 1b). A substantial number (143) of tRNA genes listed in GtRNAdb database[10] overlapped with these p62-Pol III peaks. With no FDR filtration this number increased to 166. Peaks on some other non tRNA class III genes were also detected (see Supplementary Table S1). The number of tRNA genes is greater than the number of peaks because a single identified region of enrichment often covered two or more closely spaced genes as it was the case for tRNA128 and tRNA129 on chromosome 6 (see Fig. 1a). It should be noted that under such conditions, no significant enrichment of Pol II on Pol III occupied genes was detected. In fact, high stringency peak calling found only 5 loci commonly bound by p62 and Pol III to be associated with Pol II, and 3 additional peaks, common to Pol II and Pol III only (Fig. 1b). The visual examination of enrichment profile showed very slight enrichment of Pol II on some class III genes along with p62, but the proportion of their peak heights was far from that on actively expressed class II genes (Fig. 1a). No CDK7 peaks were significantly associated with Pol III.

We next performed a conventional chromatin immunoprecipitation (ChIP) analysis with antibodies against p62, CDK7, Pol III, and Pol II; we chose a Pol II antibody that did not discriminate between phosphorylated and non-phosphorylated forms. ChIP followed by qPCR show that Pol II as well as p62 and CDK7 subunits of TFIIH were recruited at the promoter of glyceraldehyde-

3-phosphate dehydrogenase (*GAPDH*), a class II gene (Fig. 1c). Interestingly at the tRNA128 class III gene on chromosome 6, only p62 (but not CDK7) together with Pol III were recruited; this was in agreement with ChIP-seq data (Fig. 1a). As a control locus we used a geneless genomic region close to tRNA128 on chromosome 6 that was not enriched by any protein used in ChIP-seq. Further ChIP-qPCR assays demonstrated that XPB, p44, p62 (an alternative antibody as additional control) and p52 all of them being subunits of the core TFIIH were found on class III genes such as tRNA128 on chromosome 6, tRNA27 on chromosome 16, tRNA27 on crhomosome 17, tRNA5 on chromosome 2 and tRNA62 on chromosome 6 (Fig. 1d). As a negative control we used the same geneless locus as earlier. This analysis consistently showed strong association of different subunits of the core TFIIH with class III genes (Fig. 1d).

The core TFIIH is required for *VAI* transcription To first eliminate the possibility that Pol II might participate in some class III genes transcription we conducted *in vitro* assays using HeLa nuclear extracts (NE) and either the adenoviral major late promoter (AdMLP, run-off of 309 nt long) for class II transcription or the adenoviral *VAI* gene (VAI RNA, 130 nt long) that contains the same type of Pol III promoter as tRNA genes. Under those conditions, we observed transcription of both templates (Fig. 2a, lanes 1 and 7). To rule out the participation of Pol II in class III gene transcription we depleted the HeLa nuclear extract with an antibody against Pol II. That completely blocked transcription of the AdMLP and did not affect *VAI* (Fig. 2, lane 2). Immunodepletion of the HeLa NE with a non-specific antibody did not affect transcription of the class II and III genes (Fig. 2, lane 3). To next investigate if core TFIIH is directly involved in the transcription of class III genes, we immunodepleted HeLa NE with antibodies against either p44 or p62 subunits of the core

TFIIH. Such extracts were unable to allow class II and class III gene transcription when compared with extracts treated with a non-specific antibody as a control (Ab-Ctrl) (Fig. 2, compare lanes 5 and 8 with lanes 6 and 9 respectively). Addition of the recombinant core TFIIH to the Ab-p62 depleted HeLa NE restored transcription of both the AdMLP and *VAI* templates demonstrating that TFIIH is involved in both class II and class III RNA synthesis (Fig. 2, compare lane 11 and lane 12).

Previously uncharacterized candidates to class III genes Our ChIP-seq analysis revealed five novel genomic regions highly enriched with Pol III. They do not correspond to any previously characterized class III gene. These locations are described in Supplementary Table S2. One such region resides in a low complexity repeat region of $(TG)_n$ family 2 kb upstream the promoter of *FAIM3* gene. Another one is marked by high level of DNAse I hypersensitivity, according to UCSC Genomebrowser[8], and is adjacent to an Alu element situated 1.5 kb upstream *SLC7A2* gene promoter (this region has been already mentioned as Pol III enriched in another genome-wide study[5]). One of the Pol III enriched regions resides inside a cluster of immunoglobulin genes. The two other potential class III genes are linked to class II genes. One of them resides in an intronic part of *TBC1D16* gene. The last Pol III enrichment peak overlaps with the promoter region of *FEM1A* gene. Interestingly, no TFIIH was detected on any of these regions. This may imply that the potential genes are not actively transcribed.

In addition to this, we observed high enrichment of Pol III on four tRNA pseudo-genes two of which were also bound by the core TFIIH (see Supplementary Table S2 for more information).

Discussion

Despite the common evolutionary origin, structural similarity and shared subunits, biochemical studies on a small number of model genes for a long time supported the point of view that three eukaryotic RNA polymerases have completely different set of factors and recognize completely different classes of genes. The first transcription factor found to be common to Pol II and Pol III was TBP[11]. Interestingly, TPB appeared to be an essential factor also for Pol I[12], which made it a truly universal in transcription. Later, TFIIH was recognized as an essential transcription factor not only for Pol II but also for Pol I[13]. In this light, it is not entirely unexpected to find TFIIH to be involved in Pol III transcription as well.

Previous studies on cancerous or immortalized human cells[4, 5, 6] and on mouse ES cells[11] have found that only about a half of all tRNA genes were bound by Pol III. We observed the same pattern in normal human fibroblasts. In our study we observed that only 44% of tRNA genes found in human genome were enriched by Pol III. The core TFIIH was detected only on 28% of tRNA genes (omitting of FDR filtration increased this fraction to 33%) which constituted only about a half of Pol III bound genes. This might be a consequence of the fact that the presence of Pol III on a class III gene does not necessarily mean expression of the gene, and even such transcription factor as TFIIIB and TFIIIC that are absolutely required *in vitro* for Pol III transcription from type 2 promoters are not necessarily associated with all Pol III bound tRNA genes genome-wide[5, 7, 14].

The earlier reported strong association of Ser5 phosphorylated form of Pol II with actively transcribed class III genes[6] was not observed in our study. Pol II in other phosphorylation state

could still be there. However, that could not be the reason of the presence of TFIIH on that promoters. As our *in vitro* experiments demonstrated, TFIIH is required even in the context of type 2 promoter transcription from naked DNA, while the presence of Pol II is irrelevant.

The function of TFIIH in Pol III transcription has still to be clarified. Our experiments do not allow to identify it. The full TFIIH complex contains three ATP-dependent enzymes (XPD and XPB helicases and CDK7 kinase) that are involved in Pol II transcription and DNA repair. However, the fact that CDK7 is not found to be bound to class III genes may indicate that Pol III does not rely on phosphorylation by TFIIH. Indeed, RPC1 (the Pol III homologue of RPB1 subunit of Pol II) does not have a special C-terminal domain (CTD) which is usually phosphorylated in Pol II during transcription initiation and elongation. Nor any nuclear receptors, the other targets of CDK7, are known to regulate the transcription of class III genes. It is also known that TFIIH may not always operate as a complex with fixed composition during nucleotide excision repair[15], where the CAK subcomplex dissociates from the core. We observed that in our *in vitro* experiments where depletion of the nuclear extract with the antibody against p62 left CDK7 level unchanged (see Supplementary Fig. S1). From the other hand, the helicase activity of TFIIH is responsible for the promoter opening during Pol II transcription initiation. No subunit with helicase activity exist in Pol III or any of its general transcription factors. And though, in yeast, the promoter opening has shown to be possible *via* ATP-independent thermodynamically driven mechanism[16, 17] *in vitro*, its relevance and sufficiency *in vivo* is not known. Furthermore, this unaided promoter melting is temperature dependent, and the temperature of significant opening does not agree well with the ability of yeast to grow at even lower temperatures. This may imply that other mechanism may

facilitate DNA strand separation in energy dependent manner. TFIIH is a good candidate for this role.

Almost all class III genes products are important for protein biosynthesis, and hence, for cell growth. The involvement of TFIIH in their transcription may shed more light on the severity of some phenotypes induced by mutations in subunits of this complex. For example, the mixed *xeroderma pigmentosum*/Cockayne syndrome phenotype, induced by specific mutations in XPB or XPD, is characterized by developmental abnormalities, dwarfism, and premature aging. Such mutations compromise the function of TFIIH in DNA repair, but this alone cannot explain all the effects. If these mutations alter the expression profile of Pol III as well, this would add new dimension into the understanding of their manifestations.

Methods

Cells and Cell Culture For ChIP-seq and ChIP studies MRC-5 primary fibroblast were used (foetal lung non-transformed fibroblasts, obtained from Dr J. Hoeijmakers, Erasmus University, Rotterdam). The cells were cultured with the medium [DMEM(1 g/l glucose) + 10% foetal calf serum + Gentamicine (40 μg/ml)] at 37°C and 5% CO_2. Experiments were conducted with about 80% confluent cells in growing state.

ChIP-qPCR Plated cells were treated with 1% (mass fraction) formaldehyde directly in the culture medium for 10 minutes at room temperature, followed by quenching the excess of the cross-linker with 180 mM glycine for 2 minutes at room temperature. Cells then were rinsed three times with ice-cold PBS and collected by scraping with a rubber policeman in PBS supplemented with protease inhibitors cocktail (PIC,

Roche). Collected cells were swollen at 0°C for 15 minutes by 9 fold excess (compared to cell pack volume) of solution with composition: 25 mM HEPES pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5% v/v NP-40, 1 mM DTT, PIC. Then nuclei were pelleted and resuspended in approximately 1:10 proportion in the sonication solution (50 mM HEPES pH 7.6, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, PIC). The chromatin was sonicated with Bioruptor sonicator (Diagenode) to obtain fragments in 200–500 bp range. Chromatin amount containing 50 μg of DNA was incubated overnight with an antibody at -20°C. The following antibodies were used: CDK7, p69 (C-19 and Q-19 respectively, Santa Cruz Biotechnology), Pol II, XPB, p44, p62, p52 (7C2, 1B3, 1H5, 3C9, 5D6 respectively, IGBMC facility), and Pol III (RPC7/RPC32 subunit, H-136, Santa Cruz Biotechnology). Immunoprecipitation with 20 μl (packed volume) of sepharose beads coupled to protein G was performed during 3h at -20°C with subsequent washing 2 times with the sonication solution, 2 times with the high salt sonication solution (500 mM NaCl), and 2 time with Tris-EDTA (50 mM and 1 mM respectively). Elution was done by the same Tris-EDTA solution supplemented with 1% SDS 2 times at 65°C for 10 min. The precipitated chromatin was then decross-linked (overnight at 65°C with RNAse A and 200 mM NaCl) and treated with proteinase K for 2h at 42°C and released DNA was purified with QIAquick PCR purification kit (QIAgen). DNA was then subjected to qPCR analysis. The following primers were used:

GAPDH promoter

forward $5'-$GTCCCAGAGATGCCAGGAG$-3'$, reverse $5'-$CTCAGGCAAAGGCCTAGGAG$-3'$;

geneless region

forward $5'-$GGTCTCCTGACAGGCAACATA$-3'$, reverse $5'-$CCCAACATGGTGGATCTTAAA$-3'$;

tRNA128 chromosome 6

forward $5'-$TAAAACGTCATCGTTGCATTG$-3'$, reverse $5'-$CGACAAGGCTTCTTTTGTACG$-3'$;

tRNA27 chromosome 16

forward $5'-$ATCGAAATGACTGGAGCCTAAAT$-3'$, reverse $5'-$GACAGCGGTTCTATATTGAGCAT$-3'$;

tRNA27 chromosome 17

forward $5'-$GTGCGGGAACTCTTTAGAGGA$-3'$, reverse $5'-$AGGGAAAATGAAGACCACACC$-3'$;

tRNA5 chromosome 2

forward $5'-$ACAGCAGTACATGCAGAGCAAT$-3'$, reverse $5'-$AGACTCCTTACTTCTGGCAACG$-3'$;

tRNA62 chromosome 6

forward $5'-$GTTGTTGATTCTTTGTGTGACGA$-3'$, reverse $5'-$TTGCCAAATAAAGTGCTTACCAT$-3'$.

ChIP-seq Chromatin immunoprecipitation was performed with sepharose beads as described in **??**. At the step of immunoprecipitation the following antibodies were used: Ser-5 phosphorylated Pol II (H-14, Covance), CDK7 (C-19, Santa Cruz Biotechnology), p62 (Q-19, Santa Cruz Biotechnology), Pol III (RPC7/RPC32 subunit, H-136, Santa Cruz Biotechnology). DNA purification after protease A treatment step was performed using phenol-chloroform extraction and ethanol precipitation to increase the yield and to concentrate samples. An additional INPUT sample was used as a negative control. Magnetic beads coupled to protein G (Dynabeads, Invitrogen) were used instead of sepharose ones. High throughput sequencing was performed by IGBMC Microarray and Sequencing Platform using Illumina Genome Analyzer II accroding to manufacturer specifications. The sequencing depth was 36 cycles. Images analysis and base calling was performed using Illumina Pipeline software. The reads were aligned to GRCh37/h19 human genome assembly. Peak calling for the obtained files with tags in the interval format (.bed) was performed with Linux based MACS [9] software using the following additional parameters for CDK7, p62 and Pol II samples: `--tsize=36 --bw=300 --pvalue=1e-5 --nomodel --shiftsize=90`; and using the following additional parameters for Pol III: `--tsize=36 --bw=300 --pvalue=1e-5 --mfold=6`. High stringency peak calling implied selection of the peaks with false detection rate (FDR)

lower than 5%. No FDR filtration is referred to as low stringency peak calling. Peak intersection analysis was done with simple Linux shell scripts. WIG-files were generated using MACS with 50 nucleotide bins for all the samples except Pol III, where 25 nucleotide bins were used. These files were utilised for the visual inspection in the UCSC Genome browser [8].

***In vitro* transcription** *In vitro* transcription was carried out in the following mix: 10 mM Tris pH 8.0, 0.1 mM EDTA, 10 % (v/v) glycerol, 50 mM KCl, 1 mM DTT, 5 mM $MgCl_2$, 0.2 mM of each ATP, GTP and UTP, 10 μM CTP, 5–10 μCi of α^{32}P-CTP, 8 units of HeLa nuclear extract (depleted or native, HeLaScribe, Promega), $2 \cdot 10^{-13}$ mol of DNA template. The template for Pol II was 600 bp linear fragment with AdMLP and capable of producing run-off transcript of 309 nt. Pol III template was *VAI* gene incorporated into pVAI circular plasmid of 3,000 bp, no run-off was necessary due to easy termination by Pol III, and the length of the transcript was 130 nt. Reaction volume was 25 μl. First pre-incubation was carried out with no CTP in solution (at 25°C for 10 min.). Transcription was performed for 30 min. at 25°C and then stopped by stop solution (0.5% SDS, 50 mM CH_3COONa, 50 mg/l yeast tRNA). RNA was then extracted by phenol-chloroform and ethanol-precipitated. The transcription efficiency was evaluated by acrylamide (8%) gel electrophoresis and radioautography (4h for Pol III product and overnight for Pol II product). Immunodepletion of HeLa nuclear extract (NE) was performed using protein G magnetic beads (Dynabeads, Invitrogen). 10 μg of an antibody was incubated with agitation for 0.5h at 4°C with amount of beads corresponding to 100 μl of their initial suspension in 200 μl of the following solution: 10 mM Tris pH 8.0, 0.1 mM EDTA, 10 % (v/v) glycerol, 50 mM KCl, 1 mM DTT, 0.01% NP-40. Then the liquid was discarded and 100 μl of NE was added to antibody coupled beads with addition of NP-40 (0.01% final concentration). The suspension was incubated with agitation for 0.5h at 4°C. Then NE was separated and added to a new portion of antibody coupled beads, the incubation was repeated. At the last step NE was separated and

incubated with antibody free beads (to remove the excess of the antibody). The following antibodies were used: Pol II (7C2, IGBMC facility), p44 (1H5, IGBMC facility), p62 (Q-19, Santa Cruz Biotechnology). The efficiency of depletion was tested by immuno blot followed SDS-PAGE. Staining was performed by antibodies against XPB (1B3, IGBMC facility), GTF3C4 (TFIIIC component, abcam), RPC7/RPC3 (Pol III subunit, H-136, Santa Cruz Biotechnology), and CDK7 (2F8, IGBMC facility).

1. White, R. J. Transcription by RNA polymerase III: more complex than we thought. *Nature reviews genetics* **12**, 459–463 (2011).

2. Hu, P. *et al.* Characterization of human RNA polymerase III identifies orthologues for saccharomyces cerevisiae RNA polymerase III subunits. *Molecular and Cellular Biology* **22**, 8044–8055 (2002).

3. Compe, E. & Egly, J. M. TFIIH: when transcription met DNA repair. *Nature Reviews* **13**, 343–354 (2012).

4. Raha, D. *et al.* Close association of RNA polymerase II and many transcription factors with Pol III genes. *PNAS* **107**, 3639–3644 (2010).

5. Oler, A. J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Structural and Molecular Biology* **17**, 620–628 (2010).

6. Barski, A. *et al.* Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nature Structural and Molecular Biology* **17**, 629–634 (2010).

7. Carrière, L. *et al.* Genomic binding of Pol III transcription machinery and relationship with TFIIS transcription factor distribution in mouse embryonic stem cells. *Nucleic Acids Research* **40**, 270–283 (2012).

8. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Research* **12**, 996–1006 (2002).

9. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biology* **9**, R137 (2008).

10. Chan, P. & Lowe, T. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* **37**, D93–D97 (2009).

11. Margottin, F. *et al.* Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. *Science* **251**, 424–426 (1991).

12. Eberhard, D., Tora, L., Egly, J. M. & Grummt, I. A TPB-containing multiprotein complex (TIF-IB) mediates transcription specificity of murine RNA polymerase I. *Nucleic Acids Research* **21**, 4180–4186 (1993).

13. Iben, S. *et al.* TFIIH plays an essential role in RNA polymerase I transcription. *Cell* **109**, 297–306 (2002).

14. Canella, D., Praz, V., Reina, J. H., Cousin, P. & Hernandez, N. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Research* **1**, 710–721 (2010).

15. Coin, F. *et al.* Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Molecular Cell* **31**, 9–20 (2008).

16. Kassavetist, G., Blanco, J., Johnson, T. & Geiduschek, E. Formation of open and elongating transcription complexes by RNA polymerase III. *Journal of Molecular Biology* **226**, 47–58 (1992).

17. Kassavetist, G. A., Letts, G. A. & Geiduschek, E. P. The RNA polymerase III transcription initiation factor TFIIIB participates in two steps of promoter opening. *The EMBO Journal* **20**, 2823–2834 (2001).

Figure 1

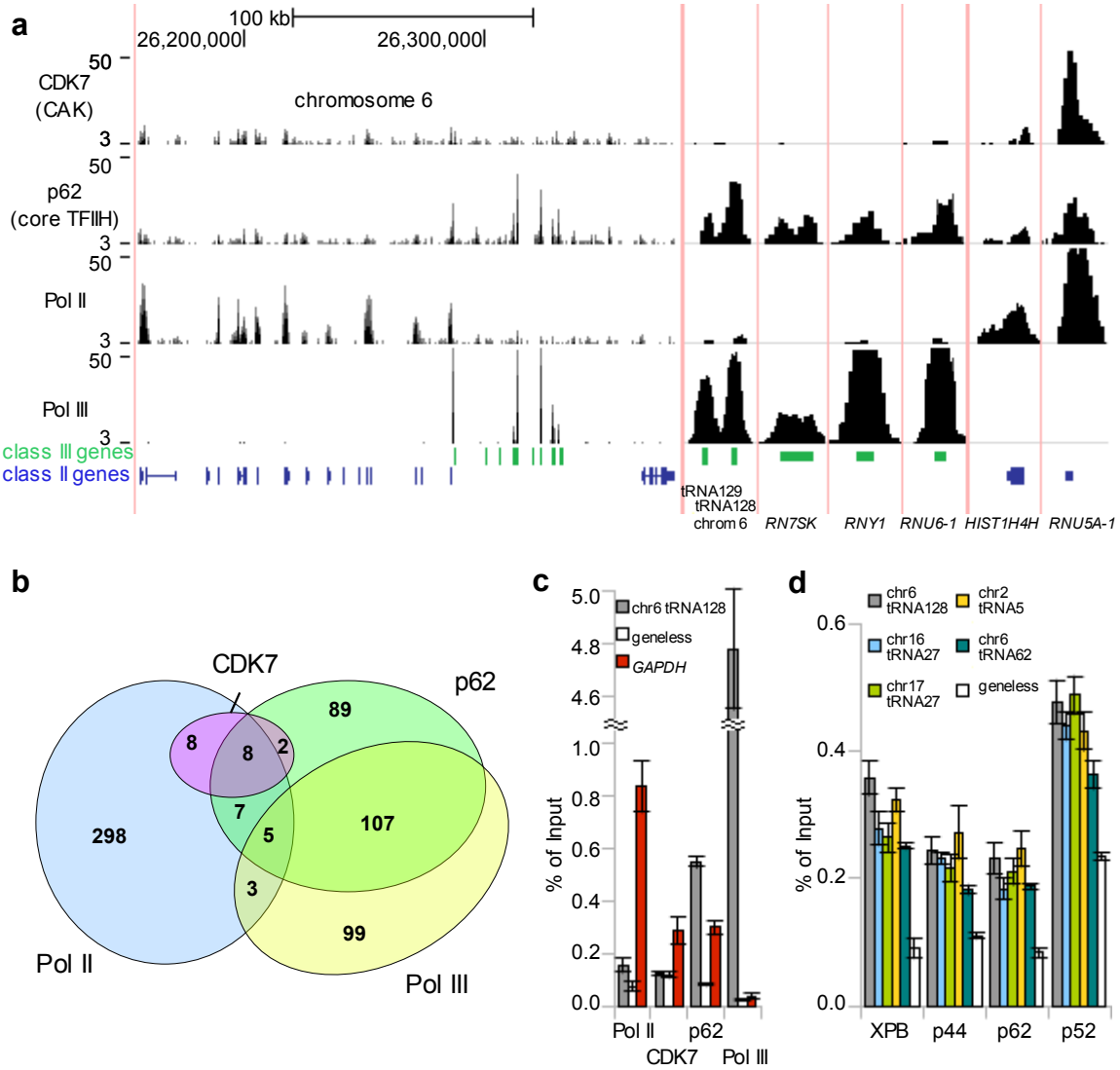


Figure 1: ChIP-seq analysis for CAK-subcomplex of TFIIH (CDK7), core TFIIH (p62), Pol II and Pol III. **a** — an overview of enrichment profile for a portion of chromosome 6, that contains a cluster of tRNA genes next to a cluster of actively transcribed protein coding genes (various histone genes) and enrichment profiles for some slected class III and class II genes (ChIP-seq data visualized using UCSC Genome browser). **b** — Euler diagram, showing intersections of the detected peaks of enrichment for all four proteins. **c** — ChIP-qPCR enrichment of a tRNA gene 128 on a chromosome 6, a geneless region and *GAPDH* housekeeping gene promoter by Pol II, p62 and CDK7; the same antibodies as for ChIP-seq except the antibody against Pol II, which does not differentiate between phosphorylated and non-phosphorylated forms. **d** — ChIP-qPCR enrichment of different tRNA genes by different subunits of core TFIIH; the antybody for p62 is different from the one, used in ChIP-seq. (In **c** and **d** numbers represent the percentage of the INPUT, error bars show the standard deviation of the mean.)

Figure 2

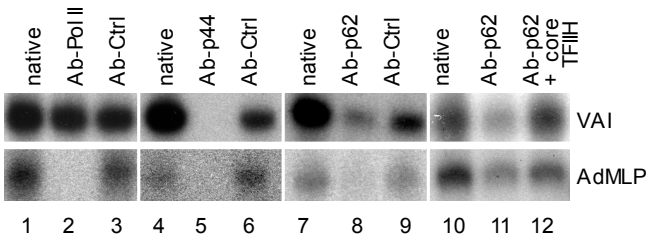


Figure 2: *In vitro* transcription experiments with class II and class III genes. Immunodepletion of HeLa nuclear extract with antibodies against Pol II (lanes 1–3) or against different subunits of TFIIH: 4–6 — anti-p44; 7–9 — anti-p62; 10–12 — anti-p62 with transcription recovery by purified core TFIIH.

Supplementary Data

Supplementary Table S1. Class III genes bound by Pol III and the core TFIIH.

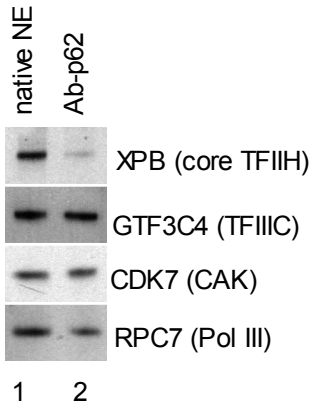
Promoter type	gene category	number of predicted genes in human genome	number of genes bound by Pol III	number of genes bound by Pol III and TFIIH with FDR≤5%	number of genes bound by Pol III and TFIIH with no filtration for FDR
Type 1	5S rRNA	17	1[a]	0	0
Type 2	tRNA	513	224	143	166
	pseudo tRNA	111	4	1	1
	7SL	2	1	0	1
	Vault	4	4	3	3
	Y	4	4	0	2
	Alu	11,081,124	5[a]	1	2
Type 3	7SK	1	1	1	1
	RNAse P	1	1	1	1
	RNAse MRP	1	1	1	1
	U6	9	4	3	3
	U6ATAC	1	1	0	0
	$tRNA^{SeC}$	2	1	0	0

[a]We used ChIP-seq tags with unique alignment.

Supplementary Table S2. Novel candidates for class III genes defined as uncharacterized genomic regions enriched by Pol III (FDR $\leq$ 5%).

chromosome	begining	end	p62 (FDR $\leq$ 5%)	notes
chromosome 1	207,097,120	207,098,104	−	$(TG)_n$ family repeat, 2 kb upstream *FAIM3* gene promoter
chromosome 8	17,352,403	17,353,040	−	DNAse I hypersensitive region, 1.5 kb upstream *SLC7A2* gene promoter[a]
chromosome 14	106,035,313	106,037,431	−	Resides between inmmunoglobulin genes
chromosome 17	77,998,007	77,998,808	−	Resides inside an intron of *TBC1D16* gene
chromosome 19	4,791,293	4,791,964	−	Overlaps with the promoter of *FEM1A* gene (class II)
chromosome 1	149,680,280	149,680,210	−	tRNA pseudogene[b]
chromosome 6	27,261,671	27,261,744	−	tRNA pseudogene[b]
chromosome 11	68,227,616	68,227,687	+	tRNA pseudogene[c]
chromosome 16	3,202,680	3,202,609	+	tRNA pseudogene[b]

[a]Previously identified as Pol III enriched region. [b]Listed in GtRNAdb. [c]Not listed in GtRNAdb, ENSEMBLE name AP000807.3-201.



Supplementary Figure S1. Immuno blot analysis for nuclear extract validation after depletion with antibody against p62. These extracts were used in recombinant TFIIH supplementation experiment. Lane 1 and lane 2 corresond to the native extract and immuno depleted extract respectively. The core TFIIH is represented by XPB, the CAK is represented by CDK7, Pol III machinery is represented by RPC7 subunit of Pol III and GTF3C4 subunit of TFIIIC factor. Note that although the core is depleted, CAK stays in the extract. Pol III machinery is not depleted.

Summary

The deeper we study TFIIH complex, the more intricate its story becomes. It entered our field of view as one of many transcription factors of Pol II which played a particular role in transcription initiation. Since then our understanding of its importance only grew wider. Not only it appeared to be a central participant in NER, a process very distant from transcription, but also to be involved in Pol I transcription and to take an active part in transcription regulation. Such broad functionality, of course, was expected from the manifold of phenotypical manifestations of mutations in subunits of TFIIH. However, our understanding of genotype-phenotype relations in these cases is still far from complete.

Mutations in XPD subunit and protein codding gene regulation

This study has broadened the known field of responsibility of TFIIH even wider. The extensive transcriptome analysis of two XP/CS mutations in XPD subunit of TFIIH has demonstrated their effect to be gene-specific. Cells with XP-D/CS mutations are clearly deficient in NER. Such cells are known to lower their global transcription intensity (reflected in overall RNA production) in response to UV-induced DNA damage. That observation allowed to suggest a global transcriptional arrest induced by unrepaired DNA blocking the way of RNA polymerases, directly hindering their elongation. Using RNA-seq technique, we were able to demonstrate that, despite the presence of a great fraction of genes, expression level of which was much lover even 24h post UV irradiation, than it was before the treatment, a substantial number of protein codding genes was significantly overexpressed in comparison with the initial state. In fact, with the statistical criteria that we had chosen, more

than 30% of statistically well represented genes experienced upregulation even 24h post-UV. Our finding, hence, challenged the hypothesis of the global transcription arrest by Pol II stalling. What is evident, is that XP-D/CS cells lose their ability to maintain homeostasis after UV-irradiation, since the overall transcription profile looks completely dysregulated especially when compared with the wild type cells. Even cell with mutation in the same gene but that displays only XP features, *i. e.* where exclusively repair function of TFIIH is affected, do not show transcriptional dysregulation of that extent. All this together with the low dosage used in the experiment (which means low probability of a lesion occurring in a given ORF) advocates for transcription regulation defect, where unrepaired damage sites only play a role of a trigger.

The investigation of the chromatin status on promoters of selected model housekeeping genes revealed the association of this downregulation with the activity of SIRT1 histone deacetylase. What is more important, both non-specific and specific inhibitions of enzymatic activity of SIRT1 and the silencing of *SIRT1* gene proved to be enough for the model genes to recover their transcription after UV. We also identified about 400 genes, expression response to UV of which is more or less restored with inhibition of SIRT1.

Why some genes are sustainably repressed while others are induced and well expressed after UV in XPD/CS mutants is yet to be understood. Our main hypothesis implies the relation of the silencing susceptibility to the nature of gene expression, *i. e.* if a gene is constitutively expressed (a housekeeping gene), or if its transcription must be activated or the initiation is stochastic. However, retinoic acid receptor regulated genes seem to complicate this simple picture. Indeed, their activation potential by retinoic acid is completely abolished by UV treatment in XP-D/CS mutants (Publication 1, Supplementary Figure S4). Yet, in normal conditions they are not expressed. Perhaps the structure of promoters or some other genomic structural information may provide a hint on why there is such striking difference between genes. All this is also to be investigated. This intriguing task is feasible now, when our research has yielded an extensive database of genes with relation to their reaction to UV damage in one of XP-D/CS mutants. The genes that are rescued by the inhibition of SIRT1 are of special interest, because SIRT1-dependent silencing is a unique feature of XP/CS related mutations in XPD and is not observed neither in the wild type, nor in pure XP mutants.

The crosstalk between TFIIH and SIRT1 mediated heterochromatisation may prove to be of great practical importance. XPD/CS mutations

leads to premature ageing phenotype. From the other hand, SIRT1 homologues are known to be implicated in the regulation of the processes that change longevity in some organisms, though it is not yet proven for mammals. In this regard, SIRT1 may represent a potential molecular target for therapy, that could, at least partially, soften the onset of malignant phenotypic manifestations of such mutations.

TFIIH and transcription by Pol III

The other unexpected finding related to TFIIH was the discovery of the genome-wide association of this factor with active Pol III transcribed genes. Because previously researches already reported putative association of Pol II machinery with active Pol III genes, the most obvious explanation to this fact (apart from mere coincidental antibody ill-specificity which was ruled out by the use of different antibodies for different components of the core TFIIH) would be involvement of Pol II in some class III related activity. In this scenario TFIIH would be brought to Pol III transcribed genes as one of Pol II factors, associated with it. In this work we clearly demonstrated that it is not the case.

First, we was unable to detect substantial enrichment of the elongation form of Pol II on class III genes. Conventional ChIP did not detect unphosphorylated Pol II either on a selected tRNA gene, previously identified as enriched by Pol III and the core TFIIH. Second, a nuclear extract depleted of Pol II retained its full capacity for the *in vitro* transcription of a class III *VAI* gene. The same experiment but with depletion of the core TFIIH significantly reduced the production of Pol III transcribed RNA. Of course, despite our precautions and test of depleted extracts on non-specific depletion of parts of Pol III machinery along with TFIIH, this still could be an explanation for the observed effect. We were physically unable to measure level of every singly Pol III subunit or of its transcription factors. The case of such mistake was ruled out when we added back recombinant TFIIH factor. That addition restored normal transcription by Pol III. We, hence, demonstrated the direct involvement of the core TFIIH in transcription of class III genes by Pol III. Furthermore, the role of TFIIH appeared to be stimulatory.

Even if Pol II is required for class III gene transcription *in vivo* for chromatin state modulation, TFIIH interacts with Pol III independent of it. Interestingly, we did not detect the CAK subcomplex on class III genes. Nevertheless, it is expected, since the role of the CAK is naturally

limited to Pol II transcription. Not only the other polymerases do not have CTD, but Pol III transcription is also not known to be regulated by transcription factors that may become a substrate for CKD7. Of course, the role of the core TFIIH is to be clarified in the future research.

Interesting observation is that the core TFIIH is not found on 5S RNA genes (with type 1 promoters). However, we were able to find only one 5S RNA gene enriched by Pol III itself. The problem with these genes is in their highly repetitious nature, since they originated initially by multiple duplications. Such loci are not suitable for ChIP-seq analysis. The length of DNA fragments, generated by ChIP, is not enough to uniquely map them. We opted for disregarding such regions and that meant the majority of 5S RNA genes. The only one, that could be unequivocally mapped with DNA fragments from ChIP against Pol III, was not enriched with TFIIH.

The interaction between TFIIH and class III transcription is very intriguing. It renders TFIIH a universal factor among three main DNA-directed RNA polymerases. Although in our case this discovery was accidental, it is not entirely unexpected. Indeed, all three polymerases are evolutionary and structurally related. Many factors and subunits are functional homologues between three of them, and some proteins, like five core subunits of polymerases or TBP, are the same. TFIIH is already known to be a transcription factor of Pol II and Pol I, but before now no counterpart has been found for it in Pol III system.

The involvement of TFIIH in Pol III transcription may contribute to the severity of disorders caused by some mutations in subunits of TFIIH. Almost all class III genes (not counting Alu associated ones) are implicated in translation (tRNAs, 7SL RNA, 5S RNA) or gene expression (RNA part of RNase P, 7SK RNA, U6 and U6ATAC snRNAs). Disruption of TFIIH-Pol III interaction and dysregulation of their expression may have very serious consequences. Furthermore, TFIIH may serve as additional means of regulation of Pol III transcription, which is now poorly understood.

Conclusion

In this work, two TFIIH-related phenomena were investigated: 1) the influence of specific mutations in TFIIH XPD subunits on the transcriptional response of housekeeping genes on UV irradiation and 2) the interaction between TFIIH and transcription of class III genes.

- For the first time the detailed investigation of transcriptome dynamics was carried out for the response of XP-D/CS mutant human cells to UV-irradiation.
- The transcription regulation nature of the observed selective gene expression dysregulation was clearly observed. Its relation to failure of transcription re-initiation and consequent heterochromatisation was demonstrated.
- SIRT1 histone deacetylase was identified as the main driver of the repressive chromatin establishment on the housekeeping genes upon UV. Inhibition of SIRT1 was found to recover normal expression of substantial number of affected genes.
- SIRT1 mediated mechanism was shown to be XP-D/CS specific. A potential link between this longevity related protein and progeria features of XP-D/CS mutants was hypothesised.
- Genome-wide study of the involvement of the core TFIIH in transcription revealed its association with active class III genes, not described previously.
- This association was demonstrated to be Pol II-independent.
- The core TFIIH was shown to be directly involved in Pol III mediated transcription *in vitro*.

TFIIH is a marvellous and fundamentally important protein complex. Its multifunctionality does not cease to astonish researches throughout decades. And yet, new unexpected details of its cellular life are discovered that may put it on even more important position among cellular molecular machines.

Acknowledgements

Foremost, I want to express my gratitude to the referees: Björn Schumacher, Vincent Laugel, Dominique Daegelen, for their careful reading and critical analysis of this work.

I am also grateful to my scientific advisers: Jean-Marc Egly and Frédéric Coin, who provided me with an opportunity of PhD research in one of the best international scientific centres.

My sincere thanks goes to Renier Véléz-Cruz, my colleague and mentor, in close cooperation with whom the first part of this work was carried out, and who taught me different useful molecular biological techniques.

I want to thank my fellow labmates from Egly group: Alexander Zhovmer, Satoru Hashimoto, Alexey Epantchintsev, Marc-Alexander Rauschendorf, Cathy Braun, Annabel Larnicol, and former members of the group: Valentin Oksenich, Hans Ulrik Kristensen, Yakov Vitrenko, for their invaluable help in research, professional advises, conversations on scientific topics, and many worm friendly moment of our daily life.

I want also to give thank to my parents, especially to my father, who greatly contributed to my interest in science and influenced my choice of a preocupation to be a researcher.

Last but not the least I want to thank my wife Elena and my cat d'Amour. Without their love and care this work hardly could appear.

Bibliography

[1] J. Aguilar-Fuentes, M. Fregoso, M. Herrera, E. Reynaud, C. Braun, J. M. Egly, and M. Zurita. p8/TTDA overexpression enhances UV-irradiation resistance and suppresses TFIIH mutations in a Drosophila trichothiodystrophy model. *PLoS Genetics*, 4(10):e1000253, 2008.

[2] R. A. Alberty. Standard Gibbs free energy, enthalpy, and entropy changes as a function of pH and pMg for several reactions involving adenosine phosphates. *The Journal of Biological Chemistry*, 244(12):3290–3302, 1969.

[3] J. O. Andressoo and J. H. J. Hoeijmakers. Transcription-coupled repair and premature ageing. *Mutation Research*, 577:179–194, 2005.

[4] J. O. Andressoo, J. R. Mitchell, J. de Wit, D. Hoogstraten, M. Volker, W. Toussaint, E. Speksnijder, R. B. Beems, H. van Steeg, J. Jans, C. I. de Zeeuw, N. G. Jaspers, A. Raams, A. R. Lehmann, W. Vermeulen, J. H. Hoeijmakers, and G. T. J. van der Horst. An Xpd mousemodel for the combined xeroderma pigmentosum/Cockayne syndrome exhibiting both cancer predisposition and segmental progeria. *Cancer Cell*, 10(2):121–132, 2006.

[5] A. Barski, I. Chepelev, D. Liko, S. Cuddapah, A. B. Fleming, J. Birch, K. Cui, R. J. White, and K. Zhao. Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nature Structural and Molecular Biology*, 17(5):629–634, 2010.

[6] N. Beerens, J. H. J. Hoeijmakers, R. Kanaar, W. Vermeulen, and C. Wyman. The CSB protein actively wraps DNA. *The Journal of Biological Chemistry*, 280(6):4722–4729, 2005.

[7] T. Biswas, J. M. Pero, C. G. Joseph, and O. V. Tsodikov. DNA-dependent ATPase activity of bacterial XPB helicases. *Biochemistry*, 48(12):2839–2848, 2009.

[8] L. Bordone and L. Guarente. Calorie restriction, SIRT1 and metabolism: understanding longevity. *Nature Reviews Molecular Cell Biology*, 6:298–305, 2005.

[9] B. C. Broughton, A. F. Thompson, S. A. Harcourt, W. Vermeulen, J. H. J. Hoeijmakers, E. Botta, M. Stefanini, M. D. King, C. A. Weber, J. Cole, C. F. Arlett, and A. R. Lehmann. Molecular and cellular analysis of the DNA repair defect in

a patient in xeroderma pigmentosum complementation group D who has the clinical features of xeroderma pigmentosum and Cochayne syndrome. *The American Journal of Human Genetics*, 56:167–174, 1995.

[10] B. Brower-Toland, D. A. Wacker, R. M. Fulbright, J. T. Lis, W. L. Kraus, and M. D. Wang. Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes. *Journal of Molecular Biology*, 346:135–146, 2005.

[11] L. Bürglen, T. Seroz, P. Miniou, S. Lefebvre, P. Burlet, A. Munnich, E. V. Pequignot, J. M. Egly, and J. Melki. The gene encoding p44, a subunit of the transcription factor TFIIH, is involved in large-scale deletions associated with Werdnig-Hoffmann disease. *The American Journal of Human Genetics*, 60(1):72–79, 1997.

[12] D. Busso, A. Keriel, B. Sandrock, A. Poterszman, O. Gileadi, and J. M. Egly. Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. *The Journal of Biological Chemistry*, 275:22815–22823, 2000.

[13] S. Camier, A. Dechampesme, and A. Sentenac. The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. *PNAS*, 92:9338–9342, 1995.

[14] L. Carrière, S. Graziani, O. Alibert, Y. Ghavi-Helm, F. al Boussouar, H. Humbertclaude, S. Jounier, J.-C. Aude, C. Keime, J. Murvai, M. Foglio, M. Gut, I. Gut, M. Lathrop, J. Soutourina, M. Gérard, and M. Werner. Genomic binding of Pol III transcription machinery and relationship with TFIIS transcription factor distribution in mouse embryonic stem cells. *Nucleic Acids Research*, 40(1):270–283, 2012.

[15] J. Castro, C. Merino, and M. Zurita. Molecular characterization and developmental expression of the TFIIH factor p62 gene from drosophila melanogaster: effects on the UV light sensitivity of a p62 mutant fly. *DNA Repair*, 1(5):359–368, 2002.

[16] P. Chambon. Eukaryotic nuclear RNA polymerases. *Annual Reviews of Biochemistry*, 44:613–638, 1975.

[17] P. P. Chan and T. M. Lowe. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research*, 37:D93–D97, 2009.

[18] E. Citterio, V. Van Den Boom, G. Schnitzler, R. Kanaar, E. Bonte, R. E. Kingston, J. H. J. Hoeijmakers, and W. Vermeulen. ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Molecular and Cellular Biology*, 20(20):7643–7653, 2000.

[19] F. Coin, J. C. Marinoni, C. Rodolfo, S. Fribourg, A. M. Pedrini, and J. M. Egly. Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature Genetics*, 20:184–188, 1998.

[20] F. Coin, V. Oksenych, V. Mocquet, S. Groh, C. Blattner, and J. M. Egly. Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Molecular Cell*, 31(1):9–20, 2008.

[21] E. Compe and J. M. Egly. TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology*, 13:343–354, 2012.

[22] B. P. Cormack and K. Struhl. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell*, 69(4):685–696, 1992.

[23] N. Dali-Youcef, M. Lagouge, S. Froelich, C. Koehl, K. Schoonjans, and J. Auwerx. Sirtuins: the 'magnificent seven', function, metabolism and longevity. *Annals of Medicine*, 39:335–345, 2007.

[24] G. Donmez and L. Guarente. Aging and disease: connections to sirtuins. *Aging Cell*, 9(2):285–290, 2010.

[25] D. Eberhard, L. Tora, J. M. Egly, and I. Grummt. A TPB-containing multiprotein complex (TIF-IB) mediates transcription specificity of murine RNA polymerase I. *Nucleic Acids Research*, 21(18):4180–4186, 1993.

[26] A. R. Ellison, T. Nouspikel, N. G. Jaspers, S. G. Clarkson, and D. C. Gruenert. Complementation of transformed fibroblasts from patients with combined xeroderma pigmentosumCockayne syndrome. *Experimental Cell Research*, 243(1):22–28, 1998.

[27] S. Filippi, P. Latini, M. Frontini, F. Palitti, J. M. Egly, and L. Proietti-De-Santis. CSB protein is (a direct target of HIF-1 and) a critical mediator of the hypoxic response. *the EMBO Journal*, 27:2545–2556, 2008.

[28] T. Finkel, C. X. Deng, and R. Mostoslavsky. Recent progress in the biology and physiology of sirtuins. *Nature*, 460:587–591, 2009.

[29] R. P. Fisher and D. O. Morgan. A novel cyclin associates with M015/CDK7 to form the CDK-activating kinase. *Cell*, 78(4):713–724, 1994.

[30] E. P. Geiduschek and G. A. Kassavetis. The RNA polymerase III transcription apparatus. *Journal of Molecular Biology*, 310:1–26, 2001.

[31] E. P. Geiduschek and G. P. Tocchini-Valentini. Transcription by rna polymerase III. *Annual Reviews of Biochemistry*, 57:873–914, 1988.

[32] Y. Ghavi-Helm, M. Michaut, and J. Acker. Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes and Development*, 22:1934–1947, 2008.

[33] J. M. Gottesfeld and K. Luger. Energetics and affinity of the histone octamer for defined DNA sequences. *Biochemistry*, 40:10927–10933, 2001.

[34] J. M. Graham Jr, K. Anyane-Yeboa, A. Raams, E. Appeldoorn, W. J. Kleijer, V. H. Garritsen, D. Busch, T. G. Edersheim, and N. G. J. Jaspers. Cerebro-oculo-facio-skeletal syndrome with a nucleotide excisionrepair defect and a mutated XPD gene, with prenatal diagnosis in a triplet pregnancy. *The American Journal of Human Genetics*, 69(2):291–300, 2001.

[35] S. A. Grigoryev. Higher-order folding of heterochromatin: Protein bridges span the nucleosome arrays. *Biochemistry and Cell Biology*, 79(3):227–241, 2001.

[36] I. Grummt and A. G. Ladurner. A metabolic throttle regulates the epigenetic state of rDNA. *Cell*, 133:577–580, 2006.

[37] S. Hahn. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural and Molecular Biology*, 11(5):394–403, 2004.

[38] M. Hampsey. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and Molecular Biology Reviews*, 62(2):465–503, 1998.

[39] P. C. Hanawalt and G. Spivak. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews Molecular Cell Biology*, 9:958–970, 2008.

[40] J. W. Harper and S. J. Elledge. The role of Cdk7 in CAK function, a retro-retrospective. *Genes and Development*, 12:285–289, 1998.

[41] A. Hirata, B. J. Klein, and K. S. Murakami. The X-ray structure of RNA polymerase from Archaea. *Nature*, 451:851–854, 2008.

[42] T. Homma, H. Iwahashi, and Y. Komatsu. Yeast gene expression during growth at low temperature. *Cryobiology*, 46:230–237, 2003.

[43] P. Hu, S. Wu, Y. Sun, C.-C. Yuan, R. Kobayashi, M. P. Myers, and N. Hernandez. Characterization of human RNA polymerase III identifies orthologues for saccharomyces cerevisiae RNA polymerase III subunits. *Molecular and Cellular Biology*, 22(22):8044–8055, 2002.

[44] S. Iben, H. Tschochner, M. Bier, D. Hoogstraten, P. Hozák, J. M. Egly, and I. Grummt. TFIIH plays an essential role in RNA polymerase I transcription. *Cell*, 109:297–306, 2002.

[45] A. Jawhari, J. P. Lainé, S. Dubaele, V. Lamour, A. Poterszman, F. Coin, D. Moras, and J. M. Egly. p52 mediates XPB function within the transcription/repair factor TFIIH. *The Journal of Biological Chemistry*, 277:31761–31767, 2002.

[46] S. H. Jun, M. J. Reichlen, M. Tajiri, and K. S. Murakami. Archaeal RNA polymerase and transcription regulation. *Critical Reviews in Biochemistry and Molecular Biology*, 46(1):27–40, 2011.

[47] T. Juven-Gershon, J. Y. Hsu, J. W. M. Theisen, and J. T. Kadonaga. The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology*, 20:253–259, 2008.

[48] D. E. Kainov, M. Vitorino, J. Cavarelli, A. Poterszman, and J. M. Egly. Structural basis for group A trichothiodystrophy. *Nature Structural and Molecular Biology*, 15:980–984, 2008.

[49] G. Kassavetist, J. Blanco, T. Johnson, and E. Geiduschek. Formation of open and elongating transcription complexes by RNA polymerase III. *Journal of Molecular Biology*, 226:47–58, 1992.

[50] G. A. Kassavetist, G. A. Letts, and E. P. Geiduschek. The RNA polymerase III transcription initiation factor TFIIIB participates in two steps of promoter opening. *The EMBO Journal*, 20(11):2823–2834, 2001.

[51] E. Koutelou, C. L. Hirsch, and S. Y. R. Dent. Multiple faces of the SAGA complex. *Current Opinion in Cell Biology*, 22(3):374–382, 2010.

[52] J. N. Kuehner, E. L. Pearson, and C. Moore. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, 12:1–12, 2011.

[53] I. M. Kulić and H. Schiessel. DNA spools under tension. *Physical Review Letters*, 92(22):228101–1–228101–4, 2004.

[54] V. Laugel, C. Dalloz, A. Stary, V. Cormier-Daire, I. Desguerre, M. Renouil, A. Fourmaintraux, R. Velez-Cruz, J. M. Egly, A. Sarasin, and H. Dollfus. Deletion of 5' sequences of the CSB gene provides insight into the pathophysiology of Cockayne syndrome. *European Journal of Human Genetics*, 16:320–327, 2008.

[55] N. Le May, D. Mota-Fernandes, R. Vélez-Cruz, I. Iltis, D. Biard, and J. M. Egly. NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Molecular Cell*, 38:54–66, 2010.

[56] A. R. Lehmann. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes and Development*, 15:15–23, 2001.

[57] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128:707–719, 2007.

[58] G. Li, R. Margueron, G. Hu, D. Stokes, Y. H. Wang, and D. Reinberg. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. *Molecular Cell*, 38:41–53, 2010.

[59] H. Liu, J. Rudolf, K. A. Johnson, S. A. McMahon, M. Oke, L. Carter, A. M. McRobbie, S. E. Brown, J. H. Naismith, and M. F. White. Structure of the DNA repair helicase XPD. *Cell*, 133(5):801–812, 2008.

[60] F. Margottin, G. Dujardin, M. Gerard, J. M. Egly, J. Huet, and A. Sentenac. Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. *Science*, 251(4992):424–426, 1991.

[61] J. C. Marinoni, R. Roy, W. Vermeulen, P. Miniou, Y. Lutz, G. Weeda, T. Seroz, D. M. Gomez, J. H. J. Hoeijmakers, and J. M. Egly. Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH. *The EMBO Journal*, 16:1093–1102, 1997.

[62] N. F. Marshall and D. H. Price. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and Cellular Biology*, 12(5):2078–2090, 1992.

[63] J. A. Miller and J. Widom. Collaborative competition mechanism for gene activation in vivo. *Molecular and Cellular Biology*, 23(5):1623–1632, 2003.

[64] L. A. Mirny. Nucleosome-mediated cooperativity between transcription factors. *PNAS*, 107(52):22534–22539, 2010.

[65] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Research*, 37(14):4707–4722, 2009.

[66] W. O'Gorman, B. Thomas, K. Y. Kwek, A. Furger, and A. Akoulitchev. Analysis of U1 small nuclear RNA interaction with cyclin H. *The Journal of Biological Chemistry*, 280:36920–36925, 2005.

[67] A. J. Oler, R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst, K. J. Chandler, P. A. Cassiday, C. A. Nelson, C. H. Hagedorn, B. J. Graves, and B. R. Cairns. Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Structural and Molecular Biology*, 17(5):620–629, 2010.

[68] D. H. Price. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and Cellular Biology*, 20(8):2629–2634, 2000.

[69] L. Proietti-De-Santis, P. Drane, and J. M. Egly. Cockayne syndrome B protein regulates the transcriptional program after UV irradiation. *The EMBO Journal*, 25:1915–1923, 2006.

[70] D. Raha, Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong, M. Gerstein, K. Struhl, and M. Snyder. Close association of RNA polymerase II and many transcription factors with Pol III genes. *PNAS*, 107(8):3639–3644, 2010.

[71] J. A. Ranish, S. Hahn, Y. Lu, E. C. Yi, X. Li, J. Eng, and R. Aebersold. Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nature Genetics*, 36:707–713, 2004.

[72] M. Rossignol, I. Kolb-Cheynel, and J. M. Egly. Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. *The EMBO Journal*, 16(7):1628–1637, 1997.

[73] A. Sandelin, P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, and D. A. Hume. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Reviews Genetics*, 8:424–436, 2007.

[74] L. Scharamm and N. Hernandez. Recruitment of RNA polymerase III to its target promoters. *Genes and Development*, 16:2593–2620, 2002.

[75] B. Schumacher, G. A. Garinis, and J. H. J. Hoeijmakers. Age to survive: DNA damage and aging. *Trends in Genetics*, 24(2):77–85, 2008.

[76] A. Shilatifard, R. C. Conaway, and J. W. Conaway. The RNA polymerase II elongation complex. *Annual Reviews of Biochemistry*, 72:693–715, 2003.

[77] M. Shogren-Knaak, H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, 311:844–847, 2006.

[78] A. Singh and T. R. Manney. Genetic analysis of mutations affecting growth of saccharomyces cerevisiae at low temperature. *Genetics*, 77:651–659, 1974.

[79] S. T. Smale and J. T. Kadonaga. The RNA polymerase II core promoter. *Annual Reviews of Biochemistry*, 72:449–479, 2003.

[80] Y. Takagi, C. A. Masuda, W. H. Chang, H. Komori, D. Wang, T. Hunter, C. A. P. Joazeiro, and R. D. Kornberg. Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. *Molecular Cell*, 18(2):237–243, 2005.

[81] M. C. Thomas and C. M. Chiang. The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, 41:105–178, 2006.

[82] F. Thorel, A. Constantinou, I. Dunand-Sauthier, T. Nouspikel, P. Lalle, A. Raams, N. G. J. Jaspers, W. Vermeulen, M. K. K. Shivji, R. D. Wood, and S. G. Clarkson. Definition of a short region of XPG necessary for TFIIH interaction and stable recruitment to sites of UV damage. *Molecular and Cellular Biology*, 24(24):10670–10680, 2004.

[83] A. Vaquero, M. Scher, D. Lee, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Molecular Cell*, 16(1):93–105, 2004.

[84] G. Weeda, R. C. A. van Ham, W. Vermeulen, D. Bootsma, A. J. van der Eb, and J. H. J. Hoeijmakers. A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell*, 62(4):777–791, 1990.

[85] R. J. White. RNA polymerase III transcription and cancer. *Oncogene*, 23:3208–3216, 2004.

[86] R. J. White. Transcription by RNA polymerase III: more complex than we thought. *Nature Reviews Genetics*, 12:459–463, 2011.

[87] T. Wocjan, K. Klenin, and J. Langowski. Brownian dynamics simulation of DNA unrolling from the nucleosome. *Journal of Physical Chemistry B*, 113:2639–2646, 2009.

[88] L. Wohlbold, S. Larochelle, J. C. F. Liao, G. Livshits, J. Singer, K. M. Shokat, and R. P. Fisher. The cyclin-dependent kinase (CDK) family member PNQALRE/CCRK supports cell proliferation but has no intrinsic CDK-activating kinase (CAK) activity. *Cell Cycle*, 5(5):546–554, 2006.

[89] R. Young. RNA polymerase II. *Annual Review of Biochemistry*, 60:689–715, 1991.

[90] M. Zurita and C. Merino. The transcriptional complexity of the TFIIH complex. *Trends in Genetics*, 19(10):578–584, 2003.

List of Symbols and Abbreviations

Abbreviation	Description	Definition
AdMLP	Adenovirus major late promoter	page 6
bp	Base pare	page 7
BRE	TFIIB recognition element (of class II genes)	page 9
CDK	Cyclin-dependent kinase	page 6
ChIP	Chromatin immunoprecipitation	page 17
ChIP-seq	ChIP coupled to high throughput sequencing	page 43
CS	Cockayne syndrome	page 31
DCE	Downstream core element (of class II genes)	page 10
DPE	Downstream promoter element (of class II genes)	page 9
DSE	Distal sequence element (of type 3 promoters of class III genes)	page 37
CTD	C-terminal domain of the Pol II RPB1 subunit	page 6
ERCCn	Excision repair cross-complementing (group) n	page 21
GST	Glutathione S-transferase	page 41
GTF	General transcription factor	page 10
HxK$y$me	Methylated Lys residue number y of histone x	page 17
HxK$y$ac	Acetylated Lys residue number y of histone x	page 17
HAT	Histone acetyltransferase	page 18
HDAC	Histone deacetylase	page 18
ICR	Internal control region (of class III genes)	page 35
IE	Intermediate element (of type 1 promoters of class III genes)	page 36
Inr	Initiator promoter element (of class II genes)	page 9
kb	Kilobase = 1000 bp	page 7
MTE	Motif ten element (of class II genes)	page 9
NAM	Nicotinamide	page 32
NER	Nucleotide excision repair	page 21
NR	Nuclear receptor	page 28
ORF	Open reading frame (of a gene)	page 18
PIC	Preinitiation complex (of Pol II transcription machimenry)	page 10

Abbreviation	Description	Definition
Pol I	DNA-directed RNA polymerase I	page 1
Pol II	DNA-directed RNA polymerase II	page 1
Pol III	DNA-directed RNA polymerase III	page 1
PPAR	Peroxisome proliferator activated receptor (nuclear receptor)	page 28
PSE	Proximal sequence element (of type 3 promoters of class III genes)	page 37
RAR	Retinoic acid receptor (nuclear receptor)	page 28
RB	Retinoblastoma tumour suppressor	page 43
UV	Ultraviolet light	page 22
TBP	TATA-binding protein	page 8
TCR	Transcription coupled repair	page 28
TFIIx	Basal transcription factor x of Pol II	page 10
TFIIIx	Basal transcription factor x of Pol III	page 36
TSS	Transcription starting site	page 7
TTD	Trichothiodystrophy	page 30
XCPE1	X core promoter element 1 (of class II genes)	page 10
XP	*Xeroderma pigmentosum*	page 22
XP/CS	XP combined with CS	page 30
XP-D/CS	*XPD* derived XP/CS phenotype	page 31

Anton ZADORIN

Le complexe TFIIH dans la transcription eféctuée par l'ARN polymérase II et l'ARN polymérase III

Résumé

Deux phénomènes liés au TFIIH ont été étudiés : l'influence des mutations spécifiques dans la sous-unité XPD de TFIIH sur la réponse transcriptionnelle de certains gènes après l'UV, et l'interaction entre le TFIIH et la transcription des gènes de classe III.

Pour la réponse des cellules humaines XP-D/CS à l'UV, une analyse détaillée de la dynamique du transcriptome a été effectuée. La dysrégulation sélective observée de l'expression des gènes était liée à l'incapacité pour la ré-initiation transcriptionnelle et à l'hétérochromatinisation, où l'histone désacétylase SIRT1 a été identifiée comme le principal facteur. Son inhibition a permis de recouvrer l'expression normale d'un nombre substantiel des gènes affectés.

Une étude de la participation pangénomique de TFIIH dans la transcription a découvert son association avec les gènes de classe III. Elle a été démontrée être indépendante de Pol II. TFIIH a été montré participer directement à la transcription effectuée *in vitro* par Pol III.

Mots-clés : TFIIH ; transcription ; chromatine ; stress UV ; Xeroderma pigmentosum/ Syndrome de Cockayne (XP/CS) ; SIRT1 ; l'ARN polymérase III ; RNA-seq ; ChIP-seq.

Résumé en anglais

Two TFIIH-related phenomena were investigated: the influence of specific mutations in TFIIH XPD subunits on the transcriptional response of certain genes on UV irradiation and the interaction between TFIIH and transcription of class III genes.

For the first time, the detailed analysis of transcriptome dynamics was carried out for the response of XP-D/CS mutant human cells to UV. The observed selective gene expression dysregulation was demonstrated to be related to failure of transcription re-initiation and consequent heterochromatisation, where SIRT1 histone deacetylase was identified as the main driver. Its inhibition recovered normal expression of substantial number of affected genes.

Genome-wide study of the involvement of the core TFIIH in transcription revealed its association with active class III genes, not described previously. This association was demonstrated to be Pol II-independent. The core TFIIH was shown to be directly involved in Pol III mediated transcription *in vitro*.

Keywords: TFIIH; transcription; chromatin; UV-stress; Xeroderma pigmentosum/Cockayne syndrome (XP/CS); SIRT1; RNA polymerase III; RNA-seq; ChIP-seq.