



HAL
open science

Mécanismes de l'activation de la transcription in vivo par le Médiateur

Fanny Eyboulet

► **To cite this version:**

Fanny Eyboulet. Mécanismes de l'activation de la transcription in vivo par le Médiateur. Biologie moléculaire. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112177 . tel-01202343

HAL Id: tel-01202343

<https://theses.hal.science/tel-01202343>

Submitted on 20 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 426 :
GÈNES GÉNOMES CELLULES

Laboratoire : *Régulation de l'expression des gènes et épigénétique*

THÈSE DE DOCTORAT

SCIENCES DE LA VIE ET DE LA SANTÉ

par

Fanny EYBOULET

Mécanismes de l'activation de la transcription *in vivo*
par le Médiateur

Date de soutenance : 19/09/2014

Composition du jury :

Directrice de thèse :	Julie SOUTOURINA	Chercheur (CEA/CNRS/Université Paris-Sud)
Rapporteurs :	Henri-Marc BOURBON Patrick SCHULTZ	Directeur de recherche (CNRS/Université de Toulouse) Directeur de recherche (IGBMC/CNRS)
Examineurs :	Cécile FAIRHEAD Alexis VERGER	Enseignant-Chercheur (IGM/CNRS/Université Paris-Sud) Chercheur (IRI/CNRS/Université de Lille)

*A quoi bon soulever des montagnes
quand il est si simple de passer par-dessus ?*

Boris Vian

REMERCIEMENTS

J'étais sur le point de débiter une page de remerciements sérieuse, mais à la réflexion, ce ne serait pas vraiment moi. De plus, cette section représente la seule occasion qui m'est donnée de montrer aux lecteurs de cet ouvrage (qui se compteront par milliers, je n'ai aucun doute la dessus) celle que je suis vraiment alors autant ne pas me cacher !

Tout d'abord je tiens à remercier chaleureusement les membres de mon jury de thèse : Henri-Marc Bourbon, Patrick Schultz, Alexis Verger et Cécile Fairhead qui ont accepté sans hésitation d'évaluer mon travail et qui ont eu le courage d'aller au bout des 350 pages de ce manuscrit ! Une petite mention particulière à Cécile qui a eu l'occasion de suivre mon travail en tant que tutrice durant ces 4 années de thèse et qui s'est toujours montrée très positive et optimiste.

Je remercie tout particulièrement Julie et Michel de m'avoir accueillie dans leur laboratoire. Si j'avais commencé à sauter partout dès cette époque, m'auriez-vous tout de même acceptée ? En tout cas, merci à toi Michel, notre Grand chef, qui malgré ta réunionite chronique a toujours su trouver du temps pour discuter de l'avancée de mes travaux chaque fois que j'en avais besoin. Merci également d'avoir supporté mon côté « surréaliste » et mes commentaires pas toujours drôles ni de très bon goût. Un immense merci à Julie, dont la porte est toujours ouverte que ce soit pour discuter de sciences ou de sujets variés. Tu as été une directrice de thèse hors pair, à l'écoute, patiente et toujours pleine de bonnes idées. Tu resteras pour moi la reine des manips et l'encyclopédie vivante du Médiateur et de la transcription en général !

Merci à Helen et Camille C, qui ont amené gaité, joie de vivre, terra cota, birkenstock, tenues loufoques, phrases cultes et autres réjouissances dans le labo. Merci à toi Helen d'avoir joué ma petite maman, toujours de bon conseil autant dans le domaine scientifique que personnel. Merci également de m'avoir initiée à la bioinfo, et par la même occasion merci à Lucie, notre pionnière dans ce domaine.

Merci à Olivier A, notre super bioinformaticien ! Je compte sur toi pour trouver un nom un peu plus original à notre script à tout faire, parce que redo, new_redo et very_new_redo ce n'est pas très parlant...

Remerciements

En tout cas, je garderai de très bons souvenirs de mes visites à Evry, même en hiver quand il faisait -15°C dans ton bureau. Mais les meilleurs moments resteront ceux passés à la super cantine du Généthon en compagnie de Claude que je tiens également à remercier pour son extravagance qui m'ont fait me sentir moins seule dans ce monde absurde !

Merci à Nuno mon ancien voisin de paillasse et co-proprétaire de notre calculatrice et de notre pipetman, créateur de la déesse du labo et maître spirituel de la Oie ! Merci pour ton écoute, tes conseils et ton aide à tous les niveaux. Merci à Camille B d'avoir supporté et même parfois soutenu ma bizarrerie, merci pour toutes ces discussions scripts et analyses ChIP-seq productives ! Merci à Thomas E mon voisin d'en face, au-delà du champ d'orchidées, pour les partages de solutions, les blagues intellectuelles et nos échanges toujours très élogieux sur les gens. Merci également à Adrien à qui incombe maintenant la responsabilité de maintenir un peu de folie dans le labo Werner-Soutourina ! Ces quelques mois à mon contact t'auront certainement donné suffisamment d'exemples de comment procéder pour cela !

Merci à nos voisins squatteurs du CNRS. Plus particulièrement, merci à Thaiz, si petite mais si attachante ! Un grand merci également à Jérémie, à côté de qui j'arrive presque à me sentir normale ! Merci d'avoir été mon partenaire de rock pendant presque une année complète et surtout merci pour toutes ces soirées à thème (scientifique évidemment) mémorables ! Merci aussi à nos voisins de la grotte Chauvet, Jean-Marc, Pauline et Véronique.

Ce projet n'aurait peut être jamais vu le jour sans Claire et Sandra qui ont eu la lourde mais essentielle tâche de construire la collection de mutants de Med17, base de mon travail de thèse. Merci Claire pour ton aide dans mes manip, nos discussions dos à dos et merci d'avoir contribué, en grande partie grâce à ton mari, à maintenir mon IMC au dessus de la moyenne ! J'espère que tu me montreras, avant que je parte, ton vortex inter-dimensionnel ! Merci également à Sylvie pour son soutien au quotidien, je garde en tête le « requiem du Furet » et tacherai de l'inculquer à mes nouveaux collègues.

Je remercie bien évidemment les personnes incontournables du bâtiment 144. Catherine et Chantal, nos supers secrétaires aussi sympas qu'efficaces ! J'ai eu beaucoup de plaisir à venir vous rendre visite régulièrement dans vos bureaux respectifs. Merci Chantal d'avoir réussi à sauver quelques unes de mes orchidées, et ce n'était pas toujours gagné d'avance. Merci Catherine pour nos échanges d'anecdotes, ta bonne humeur et ta patience face aux complexités administratives en tout genre. La grève d'Air France

Remerciements

n'aura pas eu raison de nous, non !! Merci également à Manu qui a (à peu près) toujours commandé dans les temps ce dont j'avais besoin et qui m'a toujours fourni, sans ronchonner, stylos, cahiers, gommes et autres petits matériels de bureau nécessaires à mon épanouissement. Et merci aussi à Bénédicte pour ses petites visites impromptues mais toujours agréables et porteuses d'infos exclusives dans notre labo.

Je remercie également toutes les autres personnes dont j'ai croisé le chemin durant ces 4 années : Arlette, Marina, Linda, Marie-Claude, François, Olivier L, Joël, Christine, Marie, Cécile, Gwenaëlle, Anne, Agathe, Céline, Christelle... et tous ceux que j'oublie sûrement...

Je tiens aussi à remercier mes amis qui sont toujours là, malgré tout : Elodie, Aurélie, David, Robin, Solène et Thibault, Stéphanie et Guillaume, Stéphane et Agathe... et qui sont désormais dans l'obligation de m'appeler Dr. Eyboulet ! Et enfin je remercie mon Thomas, qui a eu à me supporter durant la période la plus difficile de ma thèse mais a tenu bon malgré mes sautes d'humeur. Merci mon Choup de ta patience et de ta présence tout simplement, sans toi cette dernière année aurait été bien plus difficile...

Et pour finir, je remercie chaleureusement mes parents, pour m'avoir si bien réussie déjà (!!) mais également pour leur soutien immuable même après 10 ans d'étude... Promis je m'arrête là ! De façon plus générale je tiens à remercier toute ma famille et particulièrement ma tante Chantal dont la porte m'est toujours restée ouverte en tout temps (avec le double des clés, ça aide !).

Enfin je tiens à dédicacer ce manuscrit à ma grand-mère Raymonde et à mon grand-père Marcel, partis trop tôt, mais dont je sais qu'ils auraient été fiers de moi.

TABLE DES MATIERES

Table des matières	1
Table des illustrations	8
Abréviations	11
Avant-propos	15
Introduction	17
CHAPITRE I : La transcription des gènes chez les eucaryotes	19
I. L'ARN polymérase II	20
1. Structure et fonction des différentes sous-unités de la Pol II	20
2. Conservation des sous-unités des ARN polymérases	24
3. Le domaine carboxy-terminal de la Pol II	26
II. La transcription par l'ARN polymérase II	29
1. L'initiation de la transcription par la pol II	30
2. L'élongation de la transcription par la Pol II	39
3. La terminaison de la transcription par la Pol II	42
4. La réinitiation de la transcription par la Pol II	42
5. La transcription « pervasive »	42
III. Contexte chromatinien	45
1. Modification des histones	46
2. Variants d'histones	47
3. Remodelage de la chromatine	47
IV. Régulation de la transcription par la Pol II	48
1. Les éléments de régulation	48
2. Les facteurs de transcription (TFs)	49
3. Les co-activateurs	50
CHAPITRE II : Le Médiateur de la transcription	56
I. Le Médiateur de <i>S. cerevisiae</i>	56
1. Découverte du Médiateur chez <i>S. cerevisiae</i>	56
2. Structure et composition du Médiateur de <i>S. cerevisiae</i>	57
II. Le Médiateur des mammifères	61
1. Découverte du Médiateur chez les mammifères	61
2. Composition du Médiateur des mammifères	61
III. Conservation évolutive du complexe Médiateur	63
IV. Rôle du Médiateur dans l'initiation de la transcription	65
1. Interaction du Médiateur avec les activateurs	65
2. Mise en place du PIC	69
3. Recrutement du Médiateur et rôle <i>in vivo</i>	76
4. Le Médiateur, facteur général de la transcription	79
V. Régulation négative de la transcription par le Médiateur	80

VI.	Autres rôles du Médiateur	81
1.	Rôle du Médiateur dans les étapes post-recrutement du PIC	81
2.	Rôle du Médiateur dans la réinitiation de la transcription	82
3.	Rôle du Médiateur dans le maintien des télomères	83
VII.	Rôles fonctionnels et pathologies	84
1.	Rôle du Médiateur dans le développement et la différenciation	84
2.	Pathologies impliquant le Médiateur	87
CHAPITRE III : La réparation de l'ADN chez les eucaryotes.		91
I.	Les lésions UV-induites	92
1.	Les dimères cyclobutaniques de pyrimidine (CPD)	92
2.	Les photo-produits pyrimidiques (6-4 PP)	93
II.	Mécanisme de réparation des lésions par le système NER	93
1.	Reconnaissance de la lésion	96
2.	Le complexe de pré-incision	99
3.	Re-synthèse et ligation	104
4.	Séquences d'assemblage des facteurs NER <i>in vitro</i>	104
III.	Contexte chromatinien	108
IV.	Couplage de la réparation avec la transcription	109
V.	Pathologies liées à un défaut de réparation par le système NER	112
1.	Xeroderma pigmentosum (XP)	113
2.	Syndrôme de Cockayne (CS)	114
3.	Trichothiodystrophie (TTD)	115
Résultats et Discussion		117
Article I : Rôle du Médiateur dans la mise en place du PIC <i>in vivo</i>		119
I.	Contexte du travail	119
II.	Article 1	124
Article II : Rôle du Médiateur dans la réparation de l'ADN		189
I.	Contexte du travail	189
II.	Article 2	192
Conclusion et Perspectives		233
I.	Obtention d'une collection de mutants thermosensibles	235
II.	Rôle du Médiateur dans la mise en place du PIC <i>in vivo</i>	236
III.	Rôle du Médiateur dans la réparation couplée à la transcription	241
IV.	Conservation des fonctions du Médiateur chez les mammifères	243
Méthodes		247
Chapitre I : CHIP-seq		249
I.	Principe	249
II.	Schéma expérimental	250
1.	Fragmentation de la chromatine	250
2.	Contrôles utilisés en CHIP-seq	251
3.	Quantification de l'ADN	252
4.	Séquençage	252

Chapitre II : Analyses bioinformatiques	254
I. Format des données.....	254
1. FastQ	254
2. SAM/BAM.....	255
3. BED/bigBED	256
4. WIG/bigWIG	257
5. BEDGraph	259
II. Traitement des données	260
1. Analyses fastQC.....	261
2. Alignement sur le génome de référence.....	265
3. Conversion au format BED	269
4. Création d'un fichier de densité.....	269
5. Détection des régions liées avec MACS	272
6. Annotation des pics.....	277
7. Comparaison qualitative des Pics et analyses GO.....	282
8. Extraction des données de densité et du nombre de lectures	282
III. Bases de données d'annotation.....	288
IV. Normalisation des données	289
1. Méthode.....	289
2. Application	290
V. Elimination des artefacts.....	290
1. Données issues du « peakcalling ».....	291
2. Fichiers de densité	292
Bibliographie.....	293
Annexes	329
Annexe I : Protocole de CHIP	331
Annexe II : Protocole de dosage au Picogreen	336
Annexe III : Protocole de co-IP.....	337
Annexe IV : Protocole d'extraction des ARN totaux de levure.....	342
Annexe V : Protocole de RT-PCR	344
Annexe VI : Protocole de transformation de levure classique par LiAc	345
Annexe VII : Paramètres utilisés pour le traitements des données de CHIP-seq	347

TABLE DES ILLUSTRATIONS

Figure 1 : Les étapes de la transcription (Orphanides and Reinberg, 2002)	20
Figure 2 : Structure de l'ARN polymérase II de <i>S. cerevisiae</i> obtenue par diffraction des rayons X (Armache et al., 2005).....	21
Figure 3 : Profil de phosphorylation du CTD au cours de la transcription.	28
Figure 4 : Schéma récapitulant l'ensemble des étapes de la transcription par l'ARN polymérase II (Koch et al., 2008).....	29
Figure 5 : Organisation du promoteur eucaryote (d'après Smale and Kadonaga, 2003).	30
Figure 6 : Mise en place du complexe de pré-initiation in vitro (d'après Buratowski et al., 1989).	33
Figure 7 : Organisation chromatinienne de l'ADN (Griffiths et al., 2004, http://www.mun.ca/)	45
Figure 8 : Organisation des sous-unités du Médiateur de <i>S. cerevisiae</i> (Guglielmi et al., 2004).	57
Figure 9 : Carte des interactions entre les sous-unités du Médiateur (Guglielmi et al., 2004).	58
Figure 10 : Structures cristallographiques des sous-complexes et des modules de tête et du milieu du Médiateur chez <i>S. cerevisiae</i> et <i>S. pombe</i>	60
Figure 11 : Structure modulaire du complexe Médiateur des mammifères (Yin and Wang, 2014)	62
Figure 12 : Comparaison de la structure du Médiateur de <i>S. cerevisiae</i> et de l'homme (Tsai et al., 2014).....	64
Figure 13 : Modèle de l'interaction entre la Pol II et le Médiateur, obtenue par microscopie électronique (Cai et al., 2009).	70
Figure 14 : Modèle structural du PIC humain obtenu par microscopie électronique (Bernecky et al., 2011).....	72
Figure 15 : Etapes de l'activation de la transcription par le Médiateur <i>in vivo</i> (Malik and Roeder, 2010).....	77
Figure 16 : Influence du Médiateur sur la mise en place du PIC et la phosphorylation du CTD (Esnault et al., 2008; Soutourina et al., 2011).	78
Figure 17 : Les différents types de lésions de l'ADN et les mécanismes de réparation associés (Hoeijmakers, 2001).	91
Figure 18 : Mécanisme de formation des CPD.	93
Figure 19 : Mécanisme de formation des 6-4PP.	93
Figure 20 : Mécanismes de réparation par les voies TCR et GGR du système NER (Hoeijmakers, 2001).....	94
Figure 21 : Structure primaire de l'endonucléase XPG (Fagbemi et al., 2011)	102

Figure 22 : Structure primaire de l'endonucléase XPF-ERCC1 (Fagbemi et al., 2011)	103
Figure 23 : Modèle probable d'assemblage des facteurs de réparation dans la GGR et la TCR chez la levure (Boiteux and Jinks-Robertson, 2013).....	106
Figure 24 : Modèle probable d'assemblage des facteurs de réparation dans la TCR et la GGR chez les mammifères (Hanawalt and Spivak, 2008).....	107
Figure 25 : Conséquences des dommages à l'ADN et phénotypes observés en cas de défaut dans les voies GGR ou TCR (Menck and Munford, 2014).	112
Figure 26 : Influence des mutations <i>med17ts</i> sur le recrutement du Médiateur, de la Pol II, de TFIIH et de la TBP.	122
Figure 27 : Modèle de l'assemblage du PIC <i>in vivo</i> (figure modifiée d'après Esnault et al., 2008, complétée d'après Soutourina et al., 2011 et l'étude des mutants <i>med17ts</i> (Article 1)).	237
Figure 28 : Localisation des mutations portées par les mutants de Med17 sur la structure du module de tête du Médiateur (Larivière et al., 2012a).....	239
Figure 29 : Localisation des mutations de Med17 sur l'alignement des séquences de l'homme et de la levure (modifié d'après Bourbon, 2008)	243
Figure 30 : Mutations de XPG chez les patients XP et XP/CS (Scharer, 2008).	245
Figure 31 : Méthode du CHIP-seq.....	250
Figure 32 : Taille des fragments obtenus après sonication.....	251
Figure 33 : « Workflow » de l'analyse CHIP-seq suivi pour ce projet.....	260
Figure 34 : Description de la création d'un fichier de densité	269
Figure 35: Workflow de MACS (Feng et al., 2012).	272
Figure 36 : Illustration de la méthodologie de MACS pour estimer la taille des fragments séquencés.....	273
Figure 37 : Exemple de visualisation d'un fichier de pics issu de MACS2, superposé aux fichiers de densité des échantillons CHIP et Input.....	276
Figure 38 : Comparaison des pics de la protéine Rad3 entre la souche MED17 sauvage et le mutant <i>med17-98</i>	282
Figure 39 : Exemple de profil moyen de liaison des protéines Med15, Med17, Rad2, Rad3 et Kin28 sur une région comprise entre -500 et +500 autour des pics du Médiateur.....	284
Figure 40 : Exemple de « heatmap » des protéines Rad3 et Kin28 sur une région comprise entre -1000 et +1000 autour du TSS.	285
Figure 41 : Exemple de comparaison des taux d'occupation de la Pol II entre une souche sauvage et une souche mutante, sur les régions liées par cette protéine.	287

Table des illustrations

Tableau 1 : Fonction des différentes sous-unités de la Pol II.....	22
Tableau 2 : Composition en sous-unités des ARN polymérases bactériennes, archées et eucaryotes (Cramer et al., 2008; Werner and Grohmann, 2011).....	25
Tableau 3 : Fonction des différents facteurs généraux de la transcription.	32
Tableau 4 : Modification des histones (Li et al., 2007).....	46
Tableau 5: Variants d'histones (Li et al., 2007)	47
Tableau 6: Nomenclature unifiée pour les TAFs (Tora, 2002).....	51
Tableau 7 : Classification des remodeleurs de la chromatine (Clapier and Cairns, 2009).	53
Tableau 8 : Conservation des sous-unités du Médiateur chez les eucaryotes (Bourbon, 2008).	63
Tableau 9 : Interaction des sous-unités du Médiateur avec les facteurs de transcription (d'après Ansari and Morse, 2013; Malik and Roeder, 2010).....	67
Tableau 10 : Implication des sous-unités du Médiateur dans des phénotypes ou des processus développementaux (d'après Yin and Wang, 2014).....	85
Tableau 11 : Implication des sous-unités du Médiateur dans des pathologies humaines (d'après Schiano et al., 2014; Spaeth et al., 2011).....	87
Tableau 12 : Facteurs impliqués dans la réparation par excision de nucléotide (d'après Prakash and Prakash, 2000; Rechkunova et al., 2008; Shuck et al., 2008).....	95
Tableau 13 : Facteurs impliqués dans les pathologies de réparation de l'ADN par la voie NER et symptômes associés (Menck and Munford, 2014).	113
Tableau 14 : Tables utilisées pour l'annotation des données de CHIP-seq.....	288
Tableau 15 : Exemple de calcul du facteur de normalisation utilisé pour comparer les données de CHIP-seq des souches sauvages et mutantes.	290

ABREVIATIONS

- 3C : Capture de conformation de chromosome
- ABD : « Activator binding domain »
- AD : Domaine d'activation de la transcription (« Activation domain »)
- ADN : Acide désoxyribonucléique
- ARC : « Activator-recruited cofactor »
- ARN : Acide ribonucléique
- ARNi : ARN interférence
- ARNm : ARN messenger
- ARNnc : ARN non-codant
- ARNr : ARN ribosomal
- ARNsn : Petit ARN nucléaire (« Small nuclear RNA »)
- ARNsno : Petit ARN nucléolaire (« Small nucleolar RNA »)
- ARNt : ARN de transfert
- BER : Réparation par excision de base (« Base excision repair »)
- BRE : Elément de reconnaissance de TFIIB (« TFIIB recognition element »)
- Cdk = « Cyclin-dependent kinase »
- ChIP : Immunoprécipitation de chromatine
- ChIP-seq : ChIP suivi d'un séquençage à haut débit (« ChIP-sequencing »)
- CPD : Dimères cyclobutaniques de pyrimidine
- CRSP : « Cofactor required for Sp1 activation »
- CS : Syndrome de Cockayne
- CSN : Signalosome Cop9
- CTD : Domaine carboxy-terminal (« Carboxy-terminal domain »)
- CUT : « Cryptic unstable transcripts »
- DBD : Domaine de liaison à l'ADN (« DNA binding domain »)
- DDB : « Damage DNA binding protein »

Abréviations

dNTP : Désoxy-ribonucléotide tri-phosphate

DPE : Élément promoteur aval (« Downstream promoter element »)

DRIP : « Vitamin D receptor-interacting proteins »

GGR ou GG-NER : « Global genome NER »

GTF : Facteur général de la transcription (« General transcription factor »)

HFD : « Histone fold domain »

HR : Réparation par recombinaison homologue (« Homologous recombination »)

HSP : Protéine de choc thermique (« Heat choc protein »)

Inr : Séquence initiatrice

IP : Immunoprécipitation

KO : Invalidation génétique (« Knock-out »)

MMR : Réparation des mésappariements (« Mismatch repair »)

MTE : Motif de 10 éléments (« Motif ten elements »)

MudPIT : « Multidimensional protein identification technology »

NAT : « Negative regulator of activated transcription »

NER : Réparation par excision de nucléotides (« Nucleotide excision repair »)

NFR = « Nucleosome free region »

NHEJ : Réparation par jonction d'extrémités non-homologues (« Non-homologous end joining »)

nt : Nucléotide

NTP : Ribonucléotide tri-phosphate

ORF : Phase ouverte de lecture (« Open reading frame »)

pb : Paire de bases

PC2 : « Positive cofactor 2 »

PCNA : « Proliferating cell nuclear antigen »

PCR : Réaction en chaîne par polymérase (« Polymerase chain reaction »)

PIC : Complexe de pré-initiation (« Pre-initiation complex »)

Pol II : ARN polymérase II

PP : Photo-produits pyrimidiques

PPi : Pyrophosphate

P-TEFb : « Positive transcription elongation factor »

RFC : « Replication factor C »

RPA : « Replication factor A »

RT : Transcription inverse (« Reverse transcription »)

SAGA : Spt-Ada-Gcn5 acétyltransférase

SMCC : « SRB Mediator containing cofactor »

SRB : « Suppressor of RNA polymerase B »

SUT : « Stable unannotated transcripts »

TA : Activateur de la transcription (« Transcription activator »)

TADs : « Transactivation domains »

TAF : « TBP associated factor »

TBP : « TATA-binding protein »

TCR ou TC-NER : « Transcription coupled NER »

TF : Facteur de transcription (« Transcription factor »)

TPE (« Telomere position effect »)

TR : Répresseur de la transcription (« Transcription repressor »)

TRAP : « Thyroid hormone receptor associated protein »

ts : Thermosensible

TSS : Site de démarrage de la transcription (« Transcription start site »)

UAS : « Upstream activating sequence »

UBA : « Ubiquitin-associated domain »

URS : « Upstream repressing sequence »

XP : Xeroderma pigmentosum

XUT : « Xrn1-sensitive unstable transcripts »

SAM : « Sequence alignment map »

BAM : « Binary alignment map »

BED : « Browser extensible data »

HTS : Séquençage à haut débit « High-throughput sequencing »

NGS : « Next generation sequencing »

WIG : « Wiggle »

AVANT-PROPOS

Plus de 20 ans se sont écoulés depuis la découverte du Médiateur de la transcription chez la levure *S. cerevisiae*. Ce complexe, conservé chez tous les eucaryotes, joue un rôle clé dans la régulation de l'expression génique en facilitant l'étape d'initiation de la transcription. Bien que son rôle dans ce processus soit de mieux en mieux décrit, de nombreuses questions subsistent quant à son mode de fonctionnement à l'échelle moléculaire. De plus, de nouveaux rôles au-delà de l'initiation de la transcription ne cessent de lui être attribués, faisant de ce complexe un acteur majeur des processus biologiques ayant l'ADN comme substrat et donc, potentiellement, une cible thérapeutique pour le traitement de nombreuses pathologies cancéreuses ou du développement.

Lors de mon arrivée au laboratoire, un crible double hybride avait déjà permis d'identifier un grand nombre d'interactions physiques entre des sous-unités du Médiateur et divers complexes nucléaires. Une de ces interactions avait d'ailleurs été analysée avec succès grâce à l'utilisation de mutants thermosensibles d'une sous-unité essentielle du Médiateur. Forts de ce succès, nous avons décidé d'appliquer la même méthodologie à l'étude des autres protéines essentielles du complexe.

Nous avons obtenu une collection de 30 mutants thermosensibles (ts) de Med17, une sous-unité essentielle du module de tête du Médiateur. De plus, notre laboratoire a établi une interaction directe entre cette sous-unité et l'ARN Polymérase II (Pol II), essentielle au recrutement de cette enzyme sur l'ensemble des gènes de classe II. Mes deux premières années de thèse ont été dévouées à la caractérisation de cette collection de mutants afin de sélectionner les plus intéressants pour la suite du projet. Ces analyses ont consisté en des études d'interactions entre les sous-unités du Médiateur par co-immunoprécipitation (co-IP), de quantification des ARNm par PCR après transcription inverse (RT-PCR) et de localisation des facteurs de transcription sur l'ADN par immunoprécipitation de chromatine (ChIP). Par la suite nous avons souhaité étendre notre étude à l'ensemble du génome de *S. cerevisiae* par ChIP suivi d'un séquençage à haut débit (ChIP-seq), une méthode d'analyse qui n'avait encore jamais été appliquée au modèle levure dans notre laboratoire.

Mes deux années de thèse suivantes ont donc été majoritairement consacrées à l'acquisition et à l'analyse bioinformatique des données de ChIP-seq obtenues à partir des mutants de Med17

sélectionnés. Cette étude a abouti à une meilleure caractérisation de la fonction du Médiateur dans la mise en place du complexe de pré-initiation (PIC).

En parallèle de cette étude, nous avons analysé un lien fonctionnel entre Med17 et l'endonucléase Rad2 qui nous a permis de définir un nouveau rôle pour le Médiateur dans la réparation de l'ADN.

La première partie de ce manuscrit présentera l'état des connaissances concernant : (1) la transcription par la Pol II, (2) le Médiateur et (3) la réparation de l'ADN par excision de nucléotide. Je décrirai et discuterai ensuite les résultats acquis au cours de ma thèse sous forme de deux publications scientifiques présentant deux fonctions différentes du Médiateur *in vivo*. Puis ces résultats seront mis en perspectives de manière à présenter les futures analyses possibles pour améliorer notre compréhension des mécanismes moléculaires mis en jeu par le Médiateur. Enfin, une dernière partie sera consacrée aux méthodes mises à profit pour parvenir aux résultats présentés dans ce manuscrit, et notamment à la description détaillée du flux de travail suivi pour l'analyse des données issues du séquençage à haut débit.

INTRODUCTION

CHAPITRE I: LA TRANSCRIPTION DES GENES CHEZ LES EUCARYOTES

La réponse d'une cellule eucaryote aux signaux physiologiques et environnementaux exige une régulation stricte de l'expression de ses gènes. La transcription est la première étape de l'expression génique. Ce processus fait intervenir de nombreux complexes protéiques et notamment les ARN polymérases, au nombre de cinq chez les eucaryotes, chacune étant spécifique d'une classe de gènes. Les ARN polymérases I et III sont responsables de la synthèse des ARNs impliqués dans l'épissage des messagers et dans la traduction, comme les ARNs de transferts (ARNt) et les ARNs ribosomiaux (ARNr) (Werner et al., 2009). Les ARN polymérase IV et V ont été découvertes plus récemment chez les plantes et ne semblent pas posséder d'équivalent chez les autres eucaryotes. Elles permettent la transcription des petits ARNs impliqués dans les phénomènes de « silencing » et ne sont pas essentielles à la viabilité des cellules (Haag and Pikaard, 2011). Enfin, l'ARN polymérase II (Pol II) est responsable de la transcription de petits ARNs non codants (ARNnc) et des ARNs messagers (ARNm) nécessaires à la synthèse des protéines.

Un cycle de transcription se réalise en trois étapes (Figure 1) : (1) l'initiation, durant laquelle, suite à la liaison d'un activateur sur une séquence régulatrice du génome, l'ARN polymérase et les facteurs généraux de la transcription (GTF) sont recrutés sur un fragment de l'ADN appelé promoteur pour former le complexe de pré-initiation (PIC) ; (2) l'élongation, pendant laquelle se fait la synthèse de la molécule d'ARN ; et enfin, (3) la terminaison, qui permet à l'ARN polymérase de se détacher de l'ADN et au fragment d'ARN néo-synthétisé d'être libéré. Cette séparation en différentes étapes est artificielle car il existe des interactions entre les acteurs impliqués dans chacune d'entre elles et les phénomènes ne sont en aucun cas indépendants, cependant elle est indispensable pour la clarté de la description.

La suite de ce chapitre sera consacrée à la présentation de la machinerie de la transcription par l'ARN polymérase II basée principalement sur les études réalisées chez la levure *Saccharomyces cerevisiae* qui constitue un très bon modèle biologique pour les organismes eucaryotes. *S. cerevisiae* est adaptée aux outils de biologie moléculaire entre autres par la petite taille de son génome (≈ 12 Mb), la facilité à le manipuler en laboratoire, alliés à un système génétique bien défini puisque l'intégralité de son génome a été séquencé (Goffeau et al., 1996). *S. cerevisiae* peut être facilement modifiée génétiquement du fait de ses hautes capacités de recombinaison. Ces techniques sont notamment rendues possibles par l'existence de marqueurs

génétiques tels que les marqueurs d'auxotrophie et les vecteurs navettes. Enfin, il existe de nombreuses homologues entre les composants de la machinerie de la transcription chez la levure et chez les mammifères ce qui fait de cet organisme un modèle de choix pour l'étude de ce processus.

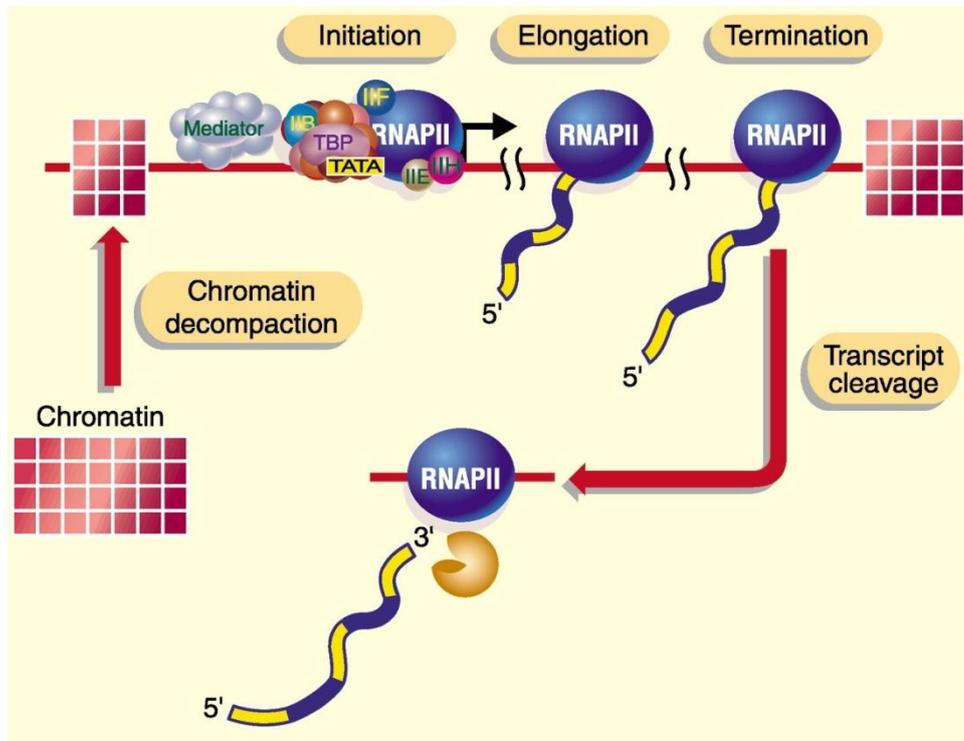


Figure 1 : Les étapes de la transcription (Orphanides and Reinberg, 2002)

Schéma simplifié représentant un cycle de transcription par la Pol II. L'initiation est l'étape correspondant à la mise en place du PIC. Au cours de l'élongation, la Pol II synthétise l'ARN puis le PIC se dissocie du promoteur. Enfin, la terminaison est l'étape durant laquelle la Pol II se détache et l'ARN néo-synthétisé est libéré.

I. L'ARN POLYMERASE II

1. Structure et fonction des différentes sous-unités de la Pol II

L'ARN polymérase II (Pol II) est une protéine hétéro-multimérique de 0,5 MDa composée de 12 sous-unités très bien conservées, codées par les gènes *RPB1-12* (Allison et al., 1985; Kolodziej and Young, 1989; Woychik and Young, 1990; Woychik et al., 1990). 10 de ces sous-unités sont essentielles chez *S. cerevisiae* (Archambault and Friesen, 1993).

La structure cristalline de la Pol II de *S. cerevisiae* (Figure 2) a été obtenue par l'équipe de R. Kornberg, prix Nobel de chimie pour ses travaux sur les mécanismes moléculaires de la transcription chez les eucaryotes, et sert désormais de référence pour toutes les formes de Pol II

(Armache et al., 2003, 2005; Bushnell and Kornberg, 2003; Cramer et al., 2000, 2001). L'obtention de cette structure a permis de mieux comprendre les mécanismes de fonctionnement de cette enzyme.

La masse centrale de la Pol II est formée par ses deux plus grandes sous-unités : Rpb1 et Rpb2 (Cramer et al., 2000, 2001) qui créent un sillon (« cleft ») chargé positivement par lequel s'introduit l'ADN pour accéder au site actif situé près du mur (« wall ») définissant le fond du sillon. Un « pore » part du site actif en s'élargissant vers l'extérieur, formant un entonnoir inversé (« funnel ») par lequel l'ARN néo-synthétisé est libéré (Cramer, 2004). Un des côtés du sillon comporte également une pince mobile (« stalk ») dont la structure est ouverte en l'absence d'ADN et se ferme lorsque la Pol II est complexée avec un hybride ADN/ARN. Cette pince permet donc par son changement de conformation le passage de l'étape d'initiation à l'étape d'élongation.

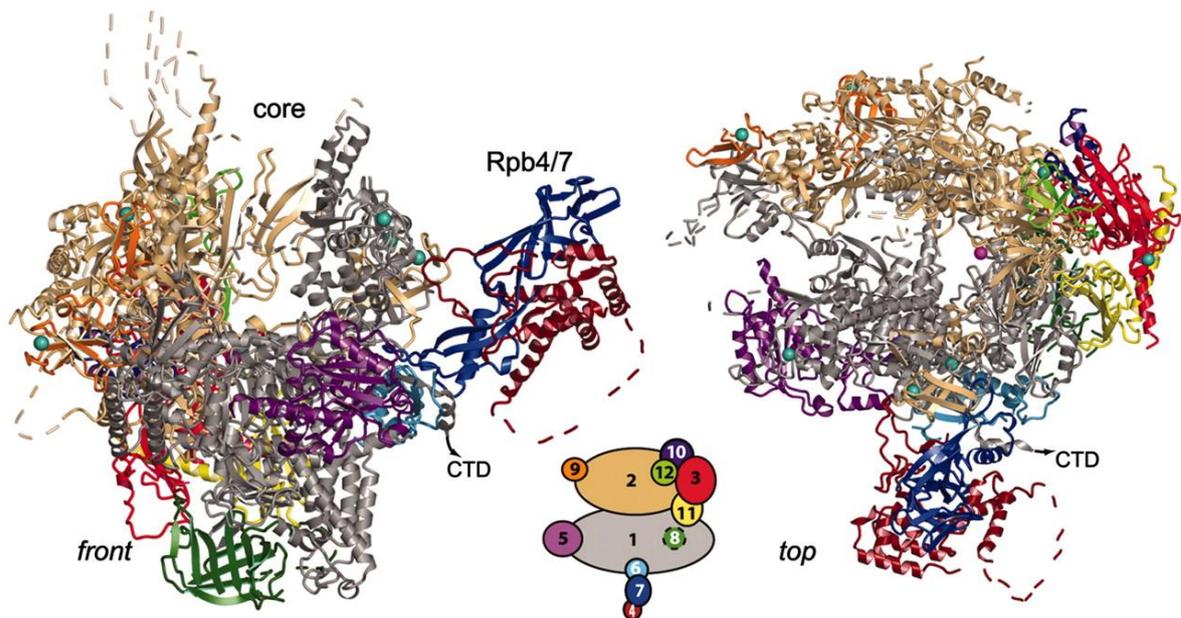


Figure 2 : Structure de l'ARN polymérase II de *S. cerevisiae* obtenue par diffraction des rayons X (Armache et al., 2005)

Deux vues différentes sont représentées : de face (à gauche) et de dessus (à droite). Les différentes sous-unités sont colorées suivant le schéma présenté au centre. La localisation du domaine carboxy-terminal (CTD) ainsi que celle des 8 ions zinc (sphères cyan), qui servent au maintien de la structure de la Pol II, et de l'ion magnésium du site actif (sphère rose) est également indiquée.

Des études génétiques et biochimiques ont permis d'éclairer la fonction de chacune des sous-unités de la Pol II (Tableau 1). Des mutations de Rpb1 et Rpb2 affectent la sélection du site de démarrage de la transcription (TSS) (Berroteran et al., 1994; Hekmatpanah and Young, 1991), ainsi que l'étape d'élongation (Archambault et al., 1992; Powell and Reines, 1996), entraînant des défauts de croissance chez la levure en présence de certaines drogues. Dans le cas de la sous-

unité Rpb1, ces défauts pourraient provenir d'une perte d'interaction avec le facteur d'élongation TFIIIS. De plus, cette sous-unité interagit avec TFIIB, un facteur général de la transcription (GTF) et cette interaction est importante pour la sélection du TSS (Berroteran et al., 1994).

Sous-unité	Masse	Propriétés
Rpb1	192 kDa	Impliquée dans la sélection du site d'initiation. Impliquée dans l'élongation de la transcription. Rôle dans l'assemblage de la Pol II. Porte le CTD et le site actif de l'enzyme. Interagit avec TFIIB et TFIIIS.
Rpb2	140 kDa	Impliquée dans la sélection du site d'initiation. Impliquée dans l'élongation de la transcription. Rôle dans l'assemblage de la Pol II.
Rpb3	45 kDa	Analogue de la sous-unité α chez <i>E. coli</i> . Rôle dans l'assemblage de la Pol II. Impliquée dans la réponse aux activateurs. Interagit avec le Médiateur
Rpb4	25 kDa	Impliquée dans l'initiation de la transcription. Rôle dans l'activation de la transcription. Rôle dans la réparation de l'ADN.
Rpb5	25 kDa	Rôle dans l'activation de la transcription. Interagit avec TFIIB.
Rpb6	18 kDa	Impliquée dans l'élongation de la transcription. Maintien du sous-complexe Rpb4-7.
Rpb7	19 kDa	Impliquée dans l'initiation de la transcription.
Rpb8	16 kDa	Rôle dans l'assemblage de la Pol II.
Rpb9	14 kDa	Impliquée dans la sélection du site d'initiation. Contrôle la fidélité de la transcription. Interagit avec TFIIE et TFIIIF.
Rpb10	8 kDa	Consolide l'architecture de la Pol II.
Rpb11	14 kDa	Rôle dans l'assemblage de la Pol II.
Rpb12	8 kDa	Consolide l'architecture de la Pol II.

Tableau 1 : Fonction des différentes sous-unités de la Pol II.

Les sous-unités essentielles chez *S.cerevisiae* sont indiquées en rouge. Les sous-unités de la levure pouvant être remplacées par leurs homologues humaines sont soulignées.

Chez les eucaryotes, la structure de la Pol II peut être divisée en deux modules : (1) un cœur catalytique de 10 sous-unités et (2) un hétérodimère constitué des sous-unités Rpb4 et Rpb7 (Edwards et al., 1991). *In vitro*, le module Rpb4-7 est indispensable pour reconnaître le promoteur et initier la transcription mais n'est pas requis pour l'étape d'élongation (Edwards et al., 1991; Orlicky et al., 2001; Sampath and Sadhale, 2005). Ces observations laissent suggérer que le complexe Rpb4-7 est présent au moment de l'initiation puis se dissocie de la Pol II. Cependant, les études *in vivo* viennent contredire cette hypothèse. En effet, Rpb7 est requis pour stabiliser la polymérase et reste associé à cette dernière durant toute l'étape d'élongation (Cojocar et al., 2008). Le complexe Rpb4-7 est présent sur la région transcrite de l'ensemble des gènes de *S. cerevisiae* et est indispensable aux étapes de maturation de l'ARN (Jasiak et al., 2008; Runner et al.,

2008). Ainsi, le modèle de la Pol II complète, avec ses 12 sous-unités est pertinent pour l'initiation et l'élongation puisque Rpb4-7 est nécessaire à l'initiation *in vitro* et que l'ensemble des sous-unités de la Pol II est associé au génome *in vivo*. Plusieurs données suggèrent également un rôle cytoplasmique pour les sous-unités Rpb4 et Rpb7 notamment dans l'export des ARNm et leur dégradation (Farago et al., 2003; Lotan et al., 2005, 2007). Cependant, cette hypothèse a récemment été contredite par une étude montrant que la diminution de la dégradation des ARNm observée dans les mutants *rpb4Δ* est due à un mécanisme général, capable de tamponner les quantités d'ARNm, mis en place par la cellule pour compenser le défaut de transcription causé par la délétion de Rpb4 (Schulz et al., 2014).

La position du sous-complexe Rpb4-7 au sein de la structure de la Pol II, à la base de la pince (Craighead et al., 2002), laisse suggérer un rôle dans la modulation des mouvements de cette dernière (Bushnell and Kornberg, 2003). Cette régulation pourrait se faire par une signalisation via la sous-unité Rpb6 qui sert de point de contact entre le sous-module Rpb4-7 et le cœur de la Pol II. Lorsque l'on modélise la pince dans un état ouvert, on observe une collision avec l'extrémité de Rpb7 (Armache et al., 2003). Ainsi, Rpb4-7 ne peut lier le cœur de la Pol II que lorsque la pince est fermée, et par conséquent, la pince ne peut s'ouvrir qu'après dissociation de Rpb4-7. Deux scénarios ont été proposés pour l'initiation de la transcription : (1) Le sous-complexe Rpb4-7 se dissocie du cœur de la Pol II, la pince peut alors s'ouvrir suffisamment pour laisser entrer l'ADN au sein de l'enzyme jusqu'au site actif. Rpb4-7 peut ensuite se lier de nouveau à la Pol II après que la pince se soit refermée. Cependant, bien que cette dissociation complète du sous-complexe Rpb4-7 de la Pol II ait pu être observée *in vitro* (Edwards et al., 1991), son existence *in vivo* n'a encore jamais été démontrée. (2) Une seconde possibilité est que le sous-complexe Rpb4-7 reste constamment lié à la Pol II, empêchant l'entrée de l'ADN qui ne peut alors se lier que bien plus loin du site actif. Cependant, après dénaturation de l'ADN, le brin matrice est suffisamment fin pour passer la pince et atteindre le site actif. Ce scénario est en accord avec ce qui a pu être observé précédemment avec la Pol II en solution, à savoir que, dans la majorité des cas, la position de la pince n'entrave pas l'accès au site actif de la Pol II et qu'aucune enzyme dans laquelle la position de la pince permettrait l'entrée d'un ADN double brin n'a été détectée (Craighead et al., 2002). Le film proposé par le laboratoire de P. Cramer décrivant les différentes étapes de la transcription par la Pol II, va également dans ce sens. Ce film débute par la fixation du complexe Rpb4-7 sur le cœur de 10 sous-unités afin de générer la forme complète de la Pol II à 12 sous-unités. Cette étape stabilise la pince dans une conformation fermée qui ne permet le passage vers le site actif que d'un ADN simple brin (Cheung and Cramer, 2012).

Deux types de transcription peuvent être décrits : la transcription basale, indépendante des activateurs, telle qu'elle peut être observée *in vitro* ; et la transcription activée, dépendante des

activateurs. Des mutations dans Rpb3 affectent la transcription activée, mais pas la transcription basale, indiquant un rôle pour cette sous-unité dans la réponse aux activateurs (Tan et al., 2000). Désormais, nous savons également que Rpb3 établit un contact direct avec la sous-unité Med17 du Médiateur, un co-activateur requis pour l'initiation de la transcription (cf. Chapitre II), et cette interaction est importante pour le recrutement de la Pol II sur l'ADN *in vivo* (Soutourina et al., 2011).

Les souches *rpb4Δ* et *rpb9Δ* ne possèdent pas de défaut de transcription dans les conditions standards de laboratoire, ces deux sous-unités ne sont donc pas essentielles à la viabilité cellulaire. Cependant, elles le deviennent dans les cellules soumises à un stress, tel qu'un défaut en nutriments ou un changement de température (Choder and Young, 1993; Miyao et al., 2001; Woychik and Young, 1989). Une étude plus approfondie a permis de montrer que ces sous-unités jouent un rôle dans les mécanismes de réparation de l'ADN par le système TC-NER (« transcription coupled nucleotide excision repair ») (cf. Chapitre III), qui répare les lésions bloquant la transcription (Li and Smerdon, 2002). Les souches *rpb9Δ* présentent une forte diminution de la fidélité de la transcription *in vitro*, avec une augmentation des mis-incorporations des ribonucléotides (NTP) (Walmacq et al., 2009). Ce défaut serait lié à une altération de l'interaction entre la Pol II, en l'absence de Rpb9, et TFIIF, un GTF important pour l'interaction de la Pol II et des autres composants du PIC avec l'ADN (Ziegler et al., 2003). Les modèles moléculaires de positionnement de TFIIF au sein du PIC, décrits par le laboratoire de S. Hahn sont venus conforter l'hypothèse d'une interaction directe entre Rpb9 et TFIIF (Eichner et al., 2010). Des tests double-hybride ont également montré que Rpb9 interagit avec TFIIE, un GTF responsable du recrutement de TFIIF, qui permet le passage en phase d'élongation (Van Mullem et al., 2002). Cette interaction serait importante pour le recrutement de TFIIF au sein du PIC.

2. Conservation des sous-unités des ARN polymérases

L'ARN est omniprésent chez tous les organismes vivants. Partant de cette observation, il n'est pas étonnant que les enzymes qui le synthétisent soient très conservées au cours de l'évolution. Toutes les ARN polymérases possèdent une base de structure commune et opèrent par des mécanismes moléculaires très proches. Les archées et les bactéries ne possèdent qu'un seul type d'ARN polymérase pour transcrire tout leur génome. Ceci les distingue des eucaryotes qui utilisent des ARN polymérases différentes selon les classes de gènes.

Toutes les ARN polymérases contiennent un cœur catalytique commun (Tableau 2). L'ARN polymérase bactérienne ne contient que ce cœur minimum, constitué des sous-unités : α , β , β' , et ω . Les polymérases archées et eucaryotes contiennent des sous-unités communes

supplémentaires : (1) un sous-complexe Rpb4-7 (Rpo4-7 chez les archées) capable de se dissocier du complexe principal (Grohmann and Werner, 2011); (2) une plateforme d'assemblage constituée des sous-unités Rpb10-12 (Rpo10-12) qui forment un complexe stable avec Rpb3-11 (Rpo3-11) (Werner and Grohmann, 2011); (3) deux sous-unités impliquées dans la dénaturation de l'ADN, Rpb5 (Rpo5) (Grünberg et al., 2010) et Rpb8 (Rpo8) (Kang et al., 2006).

Il existe également des sous-unités spécifiques à un seul domaine. La sous-unité Rpb9 de la Pol II, qui interagit avec le facteur TFIIF pour influencer la sélection du TSS (Ziegler et al., 2003), et ses paralogues dans la Pol I (A12) et dans la Pol III (C11) ne sont retrouvées que chez les eucaryotes. De la même façon, Rpo13, dont la fonction reste inconnue mais qui pourrait être impliquée dans les étapes d'initiation et d'élongation, n'existe que chez les archées (Korkhin et al., 2009). Cette sous-unité aurait des rôles similaires à ceux de certains GTFs chez les eucaryotes.

Bactéries	Archées	Eucaryotes				
		RNAPII	RNAPIII	RNAPI	RNAPIV	RNAPV
β'	Rpo1 (RpoA)	RPB1	C160	A190	NRPD1	NRPE1
β	Rpo2 (RpoB)	RPB2	C128	A135	NRPD/E2	NRPD/E2
αI	Rpo3 (RpoD)	RPB3	AC40	AC40	RPB3	RPB3
αII	Rpo11 (RpoL)	RPB11	AC19	AC19	RPB11	RPB11
ω	Rpo6 (RpoK)	RPB6	RPB6	RPB6	RPB6	RPB6
	Rpo5 (RpoH)	RPB5	RPB5	RPB5	RPB5	NRPE5
	Rpo8 (RpoG)	RPB8	RPB8	RPB8	RPB8	RPB8
	Rpo10 (RpoN)	RPB10	RPB10	RPB10	RPB10	RPB10
	Rpo12 (RpoP)	RPB12	RPB12	RPB12	RPB12	RPB12
	Rpo4 (RpoF)	RPB4	C17	A14	NRPD/E4	NRPD/E4
	Rpo7 (RpoE)	RPB7	C25	A43	NRPD7	NRPE7
		RPB9	C11	A12	NRPD9b	RPB9
	Rpo13		C37	A49		
			C53	A34.5		
			C82			
			C34			
			C31			

Tableau 2 : Composition en sous-unités des ARN polymérases bactériennes, archées et eucaryotes (Cramer et al., 2008; Werner and Grohmann, 2011).

Les sous-unités appartenant au cœur commun sont surlignées en rose. Rpb9, sous-unité spécifique aux eucaryotes, est indiquée en rouge. Les 10 sous-unités appartenant au cœur principal des polymérases eucaryotes, sont en gras et les 5 sous-unités communes aux 3 polymérases principales sont indiquées en bleu.

Les sous-unités de la Pol II sont pour la plupart bien conservées au sein des eucaryotes puisque chez la levure six peuvent être remplacées par leur homologue humain (McKune et al., 1995; Shpakovski et al., 1995). La Pol II possède une séquence d'acides aminés identique à 53% entre *S. cerevisiae* et l'homme, il existe donc une forte similarité de structure entre ces deux enzymes. De plus, la transcription dans *S. pombe* est très proche de ce qui est observée chez

l'homme notamment en ce qui concerne la structure du promoteur et la sélection du TSS (Spåhr et al., 2009). La structure de la Pol II humaine par cristallographie n'ayant à ce jour été déterminée que pour le sous-complexe Rpb4-7, l'étude de la Pol II des organismes modèles *S. cerevisiae* et *S. pombe* devient d'un intérêt grandissant.

La conservation de structure s'étend également aux différentes ARN polymérases eucaryotes. Ainsi, le cœur principal de 10 sous-unités est retrouvé dans les cinq ARN polymérases. Les ARN Pol I, II et III comprennent 5 sous-unités communes (Tableau 2) et leur site actif est très similaire. Toutes contiennent également un équivalent du sous-complexe Rpb4-7, mais la conservation des séquences correspondantes est très faible (Peyroche et al., 2002; Sadhale and Woychik, 1994; Siaux et al., 2003). La Pol I contient en plus 2 sous-unités spécifiques : A49 et A34.5, qui possèdent un équivalent dans la Pol III (C37 et C53) et partagent des similitudes avec le facteur général de la transcription par la Pol II TFIIF (Kuhn et al., 2007). La Pol III contient encore 3 sous-unités supplémentaires : C82, C34 et C31 impliquées dans l'initiation de la transcription (Wang and Roeder, 1997). Le sous-complexe C82/C34 possède des domaines homologues au facteur TFIIE (Carter and Drouin, 2010). Il est d'ailleurs intéressant de noter que TFIIE interagit avec la Pol II au niveau de la pince et que dans la Pol III, le sous-complexe C82/C34 est situé près de la pince également (Wu et al., 2012). Ces observations suggèrent qu'au cours de l'évolution, les sous-unités spécifiques à la Pol I et à la Pol III proviendraient du recrutement permanent des GTFs TFIIE et TFIIF.

3. Le domaine carboxy-terminal de la Pol II

Une structure spécifique remarquable, qui n'est retrouvée dans aucune autre ARN polymérase, est située à l'extrémité carboxy-terminale de la sous-unité Rpb1 et sert de plateforme de signalisation pour la Pol II. Ce domaine, appelé CTD (« carboxy-terminal domain »), est essentiel à la viabilité cellulaire (Corden, 1990). Sa localisation exacte n'a pas pu être déterminée, cependant, on peut le positionner approximativement en face amont de la structure de la Pol II, près du pore par lequel émerge l'ARN néo-synthétisé. Cette localisation lui permet d'interagir avec de nombreux partenaires. Ce domaine n'est pas essentiel à l'activité de la Pol II *in vitro* (Akoulitchev et al., 1995), mais il l'est *in vivo* (West and Corden, 1995).

Le CTD canonique est constitué d'une répétition de l'heptapeptide : Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Ce motif est très conservé au cours de l'évolution, cependant, le nombre de répétitions varie d'un organisme à l'autre. Ainsi, le CTD de *S. cerevisiae* en possède 24 et celui de l'homme 52 (Hampsey, 1998). Il existe également des eucaryotes possédant un CTD dit non-canonique, contenant un heptapeptide différent de celui décrit ci-dessus (Stiller and Hall, 2002). C'est le cas

des amibes *Mastigamoeba* et des parasites *Plasmodium*. Chez d'autres protistes parasites du genre *Tripanosoma* ou *Leishmania*, il n'existe pas de motif reconnaissable au niveau du CTD (Stiller and Hall, 2002).

Le CTD canonique est phosphorylable au niveau des sérines 2, 5 et 7 et ces phosphorylations entraînent des changements d'état de la Pol II (Hampsey, 1998). Au moins quatre kinases sont capables de phosphoryler le CTD. Elles sont toutes activées par des cyclines (Cdk « cyclin-depend kinase »). Deux d'entre elles sont associées aux complexes protéiques fixés à la Pol II, il s'agit de Cdk7 (associé à la cycline H) qui est une sous-unité du facteur de transcription TFIIH, et de Cdk9 (associée à la cycline T) qui est une sous-unité du facteur d'élongation P-TEFb. Des études d'immunoprécipitation de la chromatine (ChIP) ont permis de déterminer le profil de phosphorylation de la Pol II sur les gènes en cours de transcription (Figure 3). La Pol II lie le promoteur sous sa forme non-phosphorylée (forme IIA). Elle est ensuite phosphorylée (forme IIO) au niveau de la sérine 5 par Cdk7 (de TFIIH), entraînant la transition entre l'initiation et l'élongation (Hengartner et al., 1998), ainsi que sur la sérine 7 par Cdk7 également (Akhtar et al., 2009). On notera que la sous-unité Cdk8 du Médiateur est aussi capable, *in vitro*, de phosphoryler la sérine 5 de la Pol II avec, pour conséquence, une inhibition de la transcription (Hengartner et al., 1998) (cf. Chapitre II. V.). Au cours de l'élongation, la phosphatase Ssu72 (RPAP2 chez l'homme) déphosphoryle Ser5 tandis que Cdk9 (de P-TEFb) phosphoryle Ser2 (Ni et al., 2004). La phosphorylation sur Ser7, quant à elle, demeure à un niveau élevé jusqu'à l'étape de terminaison (Tietjen et al., 2010). A la fin du cycle de transcription, Ssu72 déphosphoryle Ser7, puis c'est au tour de Ser2 déphosphorylée par Fcp1 (Majello and Napolitano, 2001).

Les phosphorylations du CTD ne sont pas seulement essentielles aux différentes étapes de la transcription proprement dites mais également pour la coordination de la transcription avec d'autres processus tels que les modifications de la chromatine et la maturation des ARNs. Ainsi, les phosphorylation de Ser5 et Ser2 vont respectivement permettre le recrutement des histone méthyl-transférases Set1 et Set2 (Hampsey and Reinberg, 2003), deux enzymes qui catalysent le transfert de groupes méthyles sur les histones afin de modifier la structure de la chromatine. La forme IIO du CTD interagit avec l'ARN guanylyl-transférase et l'ARN méthyl-transférase, deux enzymes intervenant dans la synthèse de la coiffe en 3' de l'ARNm néo-synthétisé (McCracken et al., 1997a), mais également avec les protéines SR (« serine/arginine-rich ») qui jouent un rôle dans l'épissage (Mortillaro et al., 1996; Yuryev et al., 1996). Le CTD est également capable de lier les facteur CPSF (« cleavage and polyadenylation specificity factor ») et CstF (« Cleavage stimulatory factor ») qui interviennent dans le clivage du transcrit en 3' et l'ajout de la queue poly-adénylée (McCracken et al., 1997b).

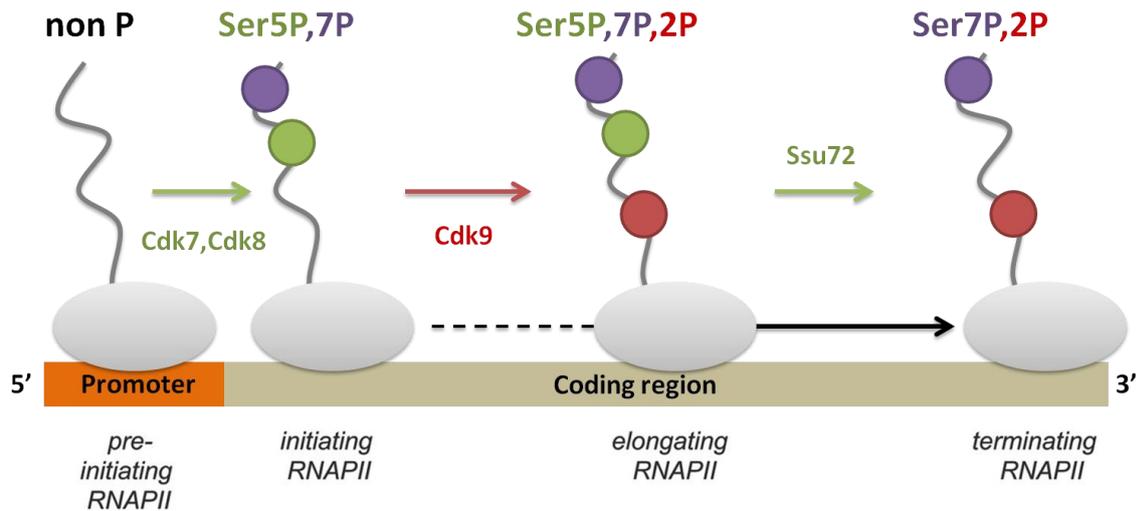


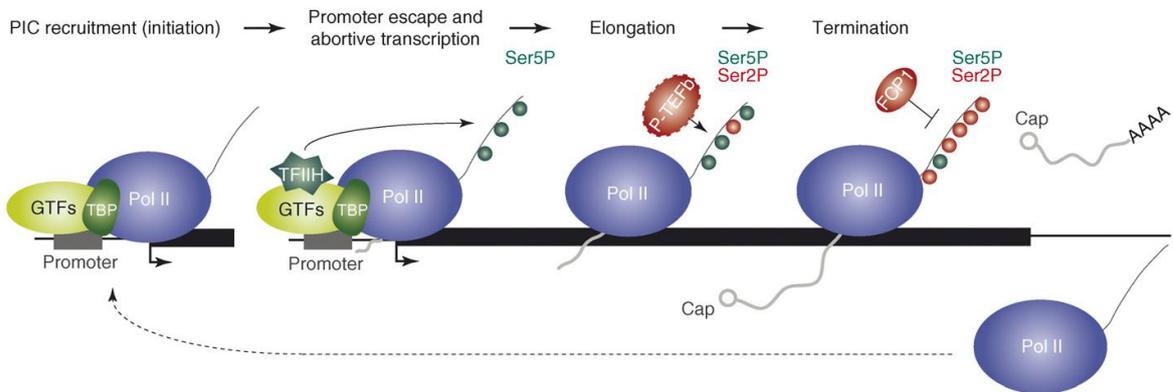
Figure 3 : Profil de phosphorylation du CTD au cours de la transcription.

En bas sont indiqués les divers éléments du gène. La Pol II est représentée en gris. Les phosphorylations sur la sérine 5 sont en vert, celles sur la sérine 7 en violet et celles sur la sérine 2 en rouge (d'après Phatnani and Greenleaf, 2006).

On sait désormais qu'il existe d'autres modifications du CTD que les phosphorylations des sérines décrites précédemment. On retrouve ainsi une forte phosphorylation de Thr4 dans le corps des gènes (Hsin et al., 2011) qui empêcherait la fixation sur le CTD des facteurs de terminaison de la transcription tels que Pcf11 (Meinhart and Cramer, 2004). Tyr1 est également phosphorylable et cette modification intervient en même temps que la phosphorylation de Ser2. Cependant, alors que Ser2 demeure phosphorylée jusqu'à la fin du gène, Tyr1 est déphosphorylée avant l'étape de terminaison. Cette modification aurait un rôle de prévention d'une terminaison prématurée de la transcription en empêchant la fixation des facteurs de terminaison sur le CTD (Mayer et al., 2012). Récemment, il a été montré que Ser2 et Ser5 peuvent également être glycosylées. Comme la glycosylation et la phosphorylation sont mutuellement exclusives, cette modification aurait pour rôle d'empêcher la phosphorylation du CTD avant le recrutement de la Pol II sur le promoteur (Ranuncolo et al., 2012). A l'extrémité du CTD, la position 7 de l'heptapeptide est majoritairement constituée de résidus arginine et lysine qui peuvent être méthylés (Sims et al., 2011). Cette modification inhiberait l'expression des petits ARN nucléaires (ARNsn) et des petits ARN nucléolaires (ARNsno). Chaque répétition du CTD contient 2 prolines en positions 3 et 6 qui peuvent subir une isomérisation. L'enzyme Ess1 (Pin1 chez l'homme) reconnaît le motif Ser-P/Thr – Pro et catalyse la transition entre les deux états isomériques de Pro3 et Pro6 (Zhang et al., 2012). Cette modification conformationnelle influence le schéma de phosphorylation du CTD et constitue un niveau de régulation supplémentaire pour le recrutement des facteurs liants le CTD.

II. LA TRANSCRIPTION PAR L'ARN POLYMERASE II

(a) 'Canonical' transcription



(b) Promoter proximal pausing and transition to elongation

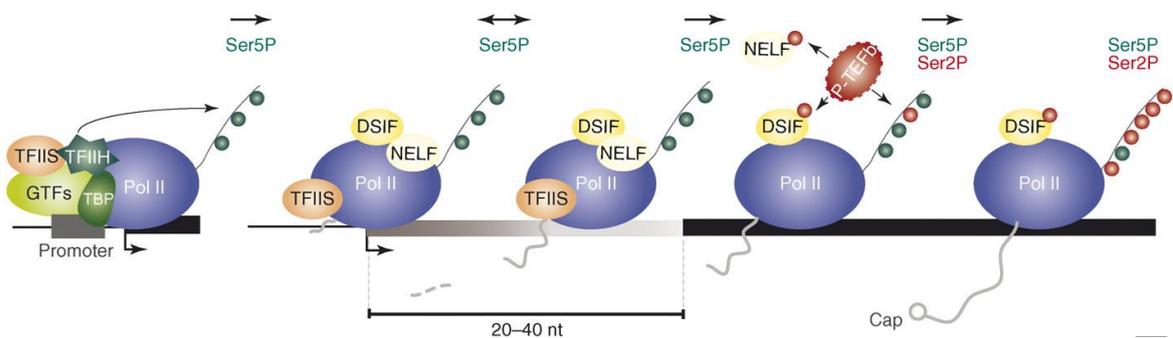


Figure 4 : Schéma récapitulatif l'ensemble des étapes de la transcription par l'ARN polymérase II (Koch et al., 2008).

Toutes les étapes de la transcription sont représentées de la gauche vers la droite.

(a) La transcription est initiée par le recrutement des facteurs de transcription constituant le PIC et de la Pol II au niveau du promoteur. La kinase Cdk7 de TFIID phosphoryle la sérine 5 du CTD, facilitant l'échappée du promoteur. L'élongation est initiée par la phosphorylation de la sérine 2 par la kinase Cdk9 du facteur d'élongation P-TEFb. La phosphorylation de la sérine 5 décroît, tandis que la phosphorylation de la sérine 2 augmente le long du gène. Une fois la Pol II arrivée au site de terminaison, la sérine 2 du CTD est déphosphorylée. L'ARN coiffé et polyadénylé est relargué, la Pol II peut être recyclée.

(b) La pause à proximité du promoteur est associée à une étape de transcription abortive. La Pol II est engagée en transcription et produit des transcrits de 20 à 40 nucléotides. Cependant, seule la sérine 5 du CTD est phosphorylée. Ce processus de pause est régulé par les protéines DSIF, NELF et TFIIS. La transition nécessite le recrutement de P-TEFb, qui phosphoryle la sérine 2 du CTD, et les facteurs NELF et DSIF. Lorsque NELF se dissocie, l'élongation productive peut être amorcée.

1. L'initiation de la transcription par la pol II

A. Le promoteur

Le promoteur désigne la séquence située immédiatement en amont du gène à transcrire et servant de plateforme à la mise en place du PIC. Chez les eucaryotes, il peut être divisé en éléments de base (« core elements »), qui constituent la portion minimale requise pour initier correctement la transcription *in vitro*, et en éléments régulateurs proximaux ou distaux (cf. IV.1. de ce chapitre).

Les éléments de base incluent (Figure 5) : la boîte TATA, la séquence initiatrice (Inr), l'élément de reconnaissance de TFIIB (BRE « TFIIB recognition element »), le motif de 10 éléments (MTE « motif ten elements ») et l'élément promoteur aval (DPE « downstream promoter element ») (Hampsey, 1998; Lim et al., 2004). Chez *S. cerevisiae*, les éléments Inr, BRE, MTE et DPE sont absents.

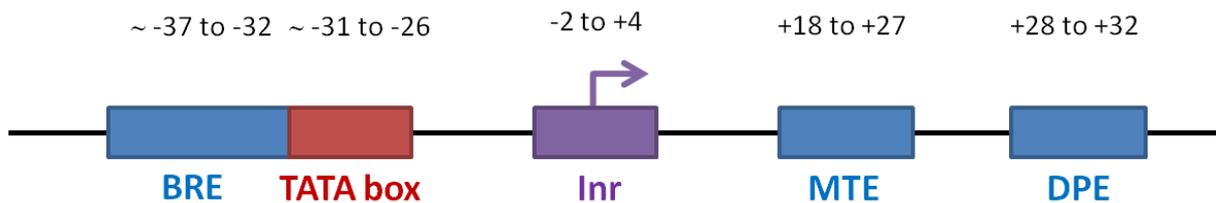


Figure 5 : Organisation du promoteur eucaryote (d'après Smale and Kadonaga, 2003).

a. La boîte TATA

La boîte TATA fut le premier élément du promoteur eucaryote à être mis en évidence. Elle contient une séquence consensus reconnue par la sous-unité TBP (« TATA-binding protein ») du facteur TFIID (Klein and Struhl, 1994). Cette association initie l'assemblage du PIC.

Chez la plupart des eucaryotes, notamment chez l'homme, la boîte TATA est localisée 25 à 30 pb en amont du site de démarrage de la transcription (Struhl, 1995). Dans ce cas, le PIC est assemblé au niveau de la boîte TATA et la sélection du TSS se fait automatiquement 25-30 bp en aval. Chez *S. cerevisiae*, la boîte TATA se situe à une distance beaucoup plus variable, pouvant aller de 40 à 120 pb (Hampsey, 1998). Chez cet organisme, le PIC est également assemblé au niveau de la boîte TATA, puis la Pol II va scanner l'ADN en aval à la recherche d'un TSS (Giardina and Lis, 1993; Kuehner and Brow, 2006).

Seuls 20% des gènes de *S. cerevisiae* (Basehoar et al., 2004) et 35% des gènes humains (Gershenson and Ioshikhes, 2005) semblent disposer d'une boîte TATA canonique, cependant la TBP peut également se lier au promoteur en son absence (Kuras and Struhl, 1999; Li et al., 1999).

La présence ou l'absence de boîte TATA au sein du promoteur a pu être reliée chez la levure (Basehoar et al., 2004, 2004; Mencía et al., 2002), ainsi que chez l'homme (Amir-Zilberstein et al., 2007), à deux voies différentes d'assemblage du PIC : (1) une dépendante de TFIID, en absence de boîte TATA, qui correspond surtout aux gènes exprimés de façon constitutive ; et (2) une dépendante de SAGA (Spt-Ada-Gcn5 acétyltransférase), en présence d'une boîte TATA, qui concerne plutôt les gènes de réponse au stress. Les travaux du laboratoire de F. Pugh ont récemment permis de préciser les mécanismes d'assemblage du PIC selon que le promoteur contient ou non une boîte TATA canonique (Rhee and Pugh, 2012). En présence d'une boîte TATA, SAGA permet le recrutement de la TBP sur le promoteur et entre en compétition avec le nucléosome +1 (premier nucléosome situé après le TSS). Ce mécanisme permet une plus grande souplesse d'expression des gènes en autorisant la perte du nucléosome +1. Avec cet obstacle en moins, la Pol II peut transcrire plus rapidement à un niveau plus élevé. De plus, le scan de la Pol II n'est alors plus limité et le TSS peut se situer plus loin en aval de la boîte TATA. En absence de boîte TATA canonique, la TBP est recrutée sur le promoteur sous une forme complexée avec TFIID. Dans ce second mécanisme, le PIC et les nucléosomes s'assemblent de manière coopérative et le nucléosome +1 joue alors un rôle indicatif quant à la position du TSS. Le scan de la Pol II est alors très limité, le TSS se situe donc beaucoup plus près, juste avant le nucléosome +1.

b. La séquence initiatrice (Inr)

La comparaison des séquences promotrices de différents gènes a permis de mettre en évidence l'existence d'une séquence consensus, incluant le TSS, suffisante à l'initiation de la transcription à la fois *in vitro* et *in vivo* (Corden et al., 1980; Smale and Baltimore, 1989). Cette séquence initiatrice notée Inr est retrouvée uniquement chez les eucaryotes pluricellulaires. De nombreux facteurs sont capables d'interagir avec l'Inr, le principal étant TFIID, qui s'y lie directement et spécifiquement via ses sous-unités TAF1 et TAF2 (Kaufmann and Smale, 1994; Verrijzer et al., 1995). La Pol II est également capable de reconnaître l'Inr et permet la mise en place d'un complexe stable constitué de la Pol II, la TBP, TFIIB et TFIIF sur les promoteurs sans boîte TATA (Carcamo et al., 1991; Weis and Reinberg, 1997).

c. L'élément de reconnaissance de TFIIB (BRE)

L'élément de reconnaissance de TFIIB (BRE) se situe en amont (BREu) ou en aval (BREd) de certaines boîtes TATA (Deng and Roberts, 2005). Il est reconnu par le complexe TFIIB grâce à une séquence consensus (Lagrange et al., 1998). Il peut avoir un effet positif ou négatif sur la transcription. A noter que cet élément ne fonctionne qu'en association avec une boîte TATA.

d. Le motif de 10 éléments (MTE) et l'élément promoteur aval (DPE)

Le motif de 10 éléments (MTE) se situe entre l'Inr et le BRE (Lim et al., 2004), tandis que l'élément promoteur aval (DPE) se situe en aval du site de démarrage de la transcription (Butler and Kadonaga, 2002). Tous deux sont retrouvés plus généralement dans les promoteurs dépourvus de boîte TATA. Ils sont fixés par les sous-unités TAF6 et TAF9 du complexe TFIID (Burke and Kadonaga, 1996; Theisen et al., 2010).

B. Les facteurs généraux de la transcription par la Pol II

L'initiation de la transcription constitue la première étape de la transcription. Durant cette étape, la Pol II ne se fixe pas directement sur l'ADN mais par l'intermédiaire de cofacteurs.

Les premiers à intervenir sont les facteurs généraux de la transcription (GTFs) (Reinberg et al., 1998) qui sont au nombre de six : TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH (Tableau 3). Ils interviennent notamment dans la reconnaissance du promoteur, le recrutement de la Pol II et l'ouverture de l'ADN. Leur assemblage va former le PIC qui correspond au complexe minimal requis pour initier la transcription.

Facteur	Sous-unités	Fonction
TFIID	14/15	Interagit avec la boîte TATA via sa sous-unité TBP Induit une torsion de l'ADN. Permet le recrutement de TFIIA et TFIIB.
TFIIA	2/3	Stabilise la TBP et permet la liaison TFIID/ADN. Rôle dans la réponse aux activateurs.
TFIIB	1	Stabilise l'interaction entre la TBP et le promoteur. Permet le recrutement du complexe Pol II/TFIIF. Rôle dans la sélection du TSS.
TFIIF	2/3	Interagit avec la Pol II et permet son recrutement au sein du PIC. Permet le recrutement de TFIIE et TFIIH. Rôle dans les étapes précoces de l'élongation.
TFIIE	2	Interagit avec le promoteur. Permet le recrutement de TFIIH et active son activité kinase.
TFIIH	10	Rôle dans la transcription et la réparation de l'ADN. Possède 3 activités : kinase qui permet de phosphoryler le CTD, ATPase qui permet l'échappée du promoteur, et hélicase qui permet l'ouverture de la bulle d'ADN.

Tableau 3 : Fonction des différents facteurs généraux de la transcription.

Des expériences d'addition *in vitro* de ces différents facteurs ont permis de déterminer l'ordre dans lequel ils interviennent (Buratowski et al., 1989) (Figure 6). TFIID est le premier facteur à être recruté sur le promoteur. Ce complexe qui contient la TBP, est suffisant pour initier la transcription basale *in vitro* sur les promoteurs pourvus d'une boîte TATA (Horikoshi et al., 1989). TFIIA et TFIIB sont ensuite recrutés et vont stabiliser la liaison entre la TBP et le promoteur. Puis la Pol II, liée à TFIIF, est également incluse au sein du PIC. Finalement, TFIIE recrute TFIIH qui va permettre l'ouverture de la bulle de transcription et le passage en élongation de la Pol II.

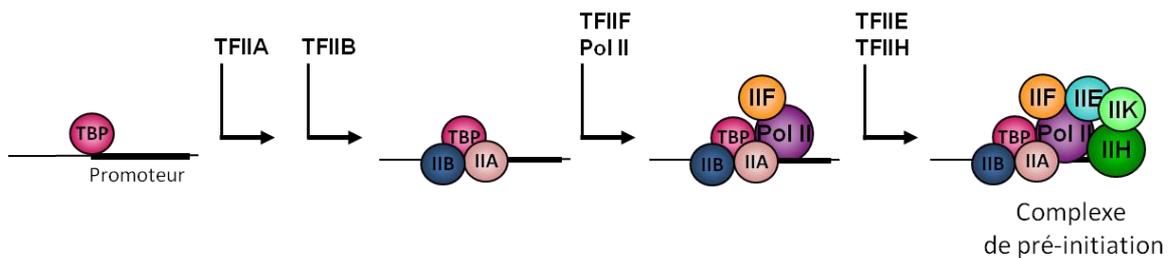


Figure 6 : Mise en place du complexe de pré-initiation *in vitro* (d'après Buratowski et al., 1989).

Récemment, des analyses par microscopie électronique des différents intermédiaires qui apparaissent durant l'assemblage du PIC ont permis de détailler le réseau d'interaction entre les différents facteurs. La première étude, réalisée par le laboratoire de E. Nogales, décrit la formation du PIC humain (He et al., 2013). Ce travail a permis d'intégrer les données biochimiques et structurales disponibles à ce jour en un schéma mécanistique unifié de l'initiation de la transcription. Il montre notamment une interaction directe entre le domaine de dimérisation de TFIIF et le domaine mobile de la Pol II lors de l'ouverture de la bulle de transcription. Cette observation permet de proposer un modèle mécanistique de la façon dont cet élément de TFIIF facilite et/ou maintient la séparation des brins d'ADN au moment de la fermeture de la pince de la Pol II. La structure obtenue montre aussi comment TFIIH peut à la fois interagir avec TFIIE à la base de la pince de la Pol II et placer sa sous-unité XPB sur l'ADN situé en aval. La visualisation des deux formes ouvertes et fermées de l'ADN permet également d'étudier les transitions structurales nécessaires à l'ouverture de la bulle de transcription. XPB fonctionne comme une ADN translocase et permet d'insérer environ 10 pb d'ADN dans le sillon de la Pol II. Cette action induit alors un surenroulement négatif de l'ADN près du TSS et TFIIF et TFIIB agissent alors ensemble pour maintenir l'ADN sous forme ouverte. Ce modèle est en accord avec le mécanisme proposé récemment par le laboratoire de S. Hahn pour la sous-unité Ssl2/XPB chez la levure qui agirait en tant qu'ADN translocase plutôt qu'hélicase comme il était suggéré jusqu'alors (Grünberg et al., 2012).

Cependant, les travaux réalisés au même moment chez *S. cerevisiae* par le laboratoire de D. Kornberg contredisent plusieurs aspects du modèle proposé par E. Nogales, notamment en ce qui concerne le positionnement de TFIIE, TFIIF et TFIIH et l'association de la Pol II avec l'ADN (Murakami et al., 2013). Cette étude associe la microscopie électronique au pontage et à la spectrométrie de masse. Les densités obtenues pour chacun des GTFs sont proches de 100% ce qui n'est pas le cas dans l'article de E. Nogales. Cette analyse décrit une division du PIC en deux parties, l'une associant les GTFs au promoteur, l'autre comprenant seulement la Pol II qui ne contacte pas le promoteur. Ce dernier est lié par 3 facteurs en amont de la pince de la Pol II : TFIIIB, TBP et TFIIE et par TFIIH en aval. Entre les deux, l'ADN est donc libre pour permettre l'action des hélicases. A l'inverse de ce qui est proposé dans le modèle de E. Nogales, ici ce n'est qu'après que l'ADN a été déroulé que le brin à transcrire peut entrer dans le sillon de la Pol II. Les GTFs permettent alors de stabiliser la bulle de transcription au niveau du promoteur. Bien que D. Kornberg justifie chaque aspect de son modèle en désaccord avec celui de E. Nogales, il n'est pas possible à ce stade, de définir lequel des deux schémas est le plus proches de la réalité et des analyses supplémentaires seront nécessaires afin de répondre à cette question.

In vivo, la situation est plus complexe puisqu'il faut tenir compte des co-activateurs. Ainsi, différentes séquences d'arrivée peuvent coexister avec différents intermédiaires plus ou moins abondants selon le promoteur (Esnault et al., 2008). Les analyses de H.S. Rhee et F. Pugh par ChIP-exo chez *S. cerevisiae* ont montré que dans tous les cas, la composition du PIC reste homogène, les différences consistant surtout en des variations de niveau d'occupation des divers GTFs selon le taux de transcription des gènes (Rhee and Pugh, 2012). La séquence de formation du PIC *in vivo* sera traitée plus en détail dans le chapitre consacré au Médiateur de la transcription, la description qui en a été faite partant de l'étude de ce complexe (cf. Chapitre II).

a. TFIID

In vitro, le promoteur est tout d'abord reconnu par le complexe TFIID, un facteur constitué de la sous-unité TBP, qui se lie à la boîte TATA, et de 14 polypeptides TAFs (« TBP associated factors ») qui interviennent en tant que co-activateurs (Cler et al., 2009; Matangkasombut et al., 2004; Sanders and Weil, 2000) (cf. IV. 3. de ce chapitre). 13 des TAFs de la levure sont conservés chez l'ensemble des eucaryotes, seule TAF14 est spécifique à la levure (Hahn, 2004) et TAF15 spécifique à l'homme. De plus, la TBP humaine est interchangeable avec celle de la levure (Buratowski et al., 1988) faisant de TFIID un facteur très conservé chez les eucaryotes.

Les structures atomiques de plusieurs modules TAFs ainsi que de la TBP ont été résolues par diffraction des rayons X mais aucune structure à haute résolution de TFIID dans son ensemble n'a

encore été obtenue. Cependant, le complexe a également été étudié par microscopie électronique et montre une architecture asymétrique trilobée en forme de pince (Andel et al., 1999; Schultz et al., 2000). Cette structure est très flexible autorisant d'importants changements conformationnels qui permettent à TFIID d'interagir à la fois avec les éléments régulateurs et le promoteur, même si la distance entre les deux est grande (Grob et al., 2006; Papai et al., 2009). Des analyses chez la levure, mais également chez l'homme, ont montré que 6 TAFs sont présents en deux copies formant le cœur symétrique de TFIID. Ce cœur est important pour l'assemblage de la TBP et des TAFs restants qui n'existent, quant à eux, qu'en une seule copie et forment une protrusion sur la structure générale du complexe (Bieniossek et al., 2013; Sanders et al., 2002). Toutes ces analyses montrent que les complexes TFIID de l'homme et de la levure partagent de nombreuses similarités structurales.

TFIID provoque une courbure d'environ 80° de la molécule d'ADN, ce qui a pour effet de dissocier légèrement les bases complémentaires de la double hélice au niveau de la boîte TATA pour permettre l'arrivée des autres facteurs généraux (Nikolov et al., 1996). *In vitro*, la TBP suffit à la transcription des gènes contenant une boîte TATA. Les protéines TAFs deviennent nécessaires pour les promoteurs qui en sont dépourvus ou pour répondre aux activateurs. La composition en TAFs va varier en fonction du type de promoteur sur lequel TFIID se fixe (Jacq et al., 1994). Chez la levure, TFIID interagit directement via ses TAFs avec l'activateur Rap1 et ce contact est indispensable à la transcription des gènes codant les protéines ribosomales *in vivo* (Garbett et al., 2007; Mencía et al., 2002; Papai et al., 2010). Cette observation indique que TFIID a également un rôle de co-activateur en formant un pont moléculaire direct entre les activateurs, liés sur les régions régulatrices, et la machinerie de transcription.

In vivo, des expériences de CHIP chez *S. cerevisiae* ont montré que TFIID est majoritairement retrouvé sur les promoteurs ne contenant pas de boîte TATA tandis que SAGA, qui partage 5 sous-unités avec TFIID, est majoritairement présent sur les promoteurs contenant une boîte TATA canonique (Basehoar et al., 2004). Cette observation révèle une bipolarité du génome, utilisant des voies de régulation différentes selon que le gène est exprimé de manière constitutive ou inductible comme il l'a été précédemment décrit dans la partie consacrée à la boîte TATA.

b. TFIIA

TFIIA se compose de 2 sous-unités chez *S. cerevisiae* et 3 chez l'homme (DeJong and Roeder, 1993; Ranish and Hahn, 1991). Ce facteur interagit à la fois avec la TBP et la boîte TATA et permet de stabiliser TFIID sur l'ADN (Imbalzano et al., 1994). Le complexe ainsi formé est actif, et l'ajout des GTFs restants et de la Pol II permet d'initier la transcription (Lieberman and Berk, 1994). La

comparaison des structures cristallographiques de l'assemblage TFIIA-TBP-ADN chez l'homme et la levure montre une forte conservation de l'architecture du complexe (Bleichenbacher et al., 2003).

TFIIA interagit également avec le co-activateur SAGA (Warfield et al., 2004) ainsi qu'avec de nombreux activateurs et répresseurs. Il permettrait ainsi la communication entre les activateurs et les composants de la machinerie de transcription en modulant l'activité des composants de SAGA et de TFIID et en inhibant l'action des répresseurs. Une analyse par microscopie électronique de la structure formée par TFIIA, TFIID et l'activateur Rap1 a permis de détailler le mécanisme de fonctionnement de TFIIA lors de la mise en place du PIC (Papai et al., 2010). Tout d'abord, Rap1 se lie sur la région régulatrice et recrute TFIID en interagissant avec TAF1. TFIID interagit à la fois avec le promoteur et la région régulatrice, créant une courbure de l'ADN qui permet de rapprocher les deux domaines. Puis TFIIA est recruté et forme un pont entre Rap1 et la TBP qui permet de stabiliser la boucle de l'ADN. Ces interactions induisent un changement de position pour TFIIA qui aurait plusieurs conséquences. Tout d'abord, il permettrait d'annuler l'inhibition de la reconnaissance de la boîte TATA par TFIID provoquée par l'interaction du domaine inhibiteur de TAF1 avec la TBP (Kokubo et al., 1998). Il stabiliserait également l'interaction de TFIID avec le promoteur et permettrait le recrutement ultérieur de TFIIB et de la Pol II. Ainsi, TFIIA conduit les signaux de régulation au sein de TFIID qui forme une plateforme d'assemblage pour le PIC.

c. TFIIB

TFIIB ne comprend qu'une seule sous-unité. Ce facteur interagit à la fois avec la TBP et l'ADN en se positionnant en aval de la boîte TATA. Plusieurs études ont montré que des mutations dans TFIIB entraînent un défaut de sélection du TSS (Pardee et al., 1998; Pinto et al., 1994; Zhang et al., 2002). Ce défaut serait notamment dû à une perte de l'interaction entre TFIIB et la sous-unité Rpb2 de la Pol II, fonctionnellement importante pour la sélection du TSS (Chen and Hampsey, 2004). TFIIB aurait également un rôle dans la terminaison de la transcription en se liant en 3' des gènes, permettant la formation de boucles entre le promoteur et le terminateur (Medler et al., 2011; Singh and Hampsey, 2007; Wang et al., 2010).

TFIIB participe au recrutement de la Pol II via une interaction directe avec TFIIF (Ha et al., 1993) ou la Pol II elle-même (Miller and Hahn, 2006) et va alors aider à son positionnement correct par rapport au complexe ADN-TBP pour que l'ADN atteigne le site actif de la Pol II. La structure de TFIIB associé à la Pol II a été étudiée par cristallographie et a permis de mettre en évidence le rôle de TFIIB dans l'ouverture de l'ADN au niveau du promoteur et la reconnaissance

de l'Inr (Kostrewa et al., 2009; Liu et al., 2010). La liaison de TFIIB à la Pol II induit également des changements dans la structure de cette dernière notamment au niveau de son site actif, importants pour stimuler la synthèse de l'ARN (Sainsbury et al., 2013).

d. TFIIF

TFIIF comprend 3 sous-unités chez *S. cerevisiae* et 2 chez l'homme (Burton et al., 1988; Henry et al., 1994). Ce facteur intègre le PIC sous une forme liée à la Pol II (Buratowski et al., 1989; Svejstrup, 2004). Cette fixation se fait via la face supérieure de l'enzyme, laissant TFIIF libre d'interagir avec d'autres facteurs (Chung et al., 2003). Ceci est important pour les étapes qui suivent puisque TFIIF va interagir avec les autres GTFs: TBP, TFIID, TFIIB et TFIIIE, pour stabiliser le PIC (Coulombe et al., 1994; Kim et al., 1997; Robert et al., 1996). TFIIF interagit notamment avec TFIIB et influence ainsi la sélection du TSS (Sun and Hampsey, 1995). Il empêche également les interactions non spécifiques avec l'ADN (Conaway and Conaway, 1993).

Des travaux récents suggèrent que TFIIF aurait également un rôle dans l'insertion et la stabilisation de l'ADN dans le sillon menant au site actif de la Pol II, ainsi que dans l'ouverture de la pince de la Pol II (He et al., 2013). TFIIF intervient aussi dans l'étape d'élongation en évitant les arrêts prématurés de la Pol II (Lei et al., 1999).

e. TFIIIE

TFIIIE interagit avec TFIIF et la Pol II et permet la stabilisation du PIC (Flores et al., 1989; Van Mullem et al., 2002). Des expériences de photo-pontage ont permis de positionner le site d'interaction de TFIIIE près de la pince de la Pol II (Chen et al., 2007), consistant avec un rôle pour ce facteur dans la stabilisation du brin non transcrit après l'ouverture de l'ADN (Forget et al., 2004). TFIIIE permet ensuite le recrutement de TFIIH tout en stimulant ses activités ATPase, hélicase et kinase (Ohkuma and Roeder, 1994; Ohkuma et al., 1995; Serizawa et al., 1994).

Le domaine C-terminal de TFIIIE pourrait jouer un rôle dans la transition entre l'initiation et l'élongation (Watanabe et al., 2003). Le domaine N-terminal est également impliqué dans l'échappée du promoteur (Goodrich and Tjian, 1994). Cependant, il n'est pas défini à ce jour si ces deux domaines agissent de façon indépendante ou coopérative pour aider la Pol II à passer en élongation.

f. TFIIH

TFIIH est un complexe multiprotéique constitué de 10 sous-unités réparties en 2 modules : un cœur de 7 sous-unités et un module kinase de 3 sous-unités (Gibbons et al., 2012). Sa kinase Cdk7 lui permet de phosphoryler le CTD de la Pol II sur sa sérine 5, modification qui marque le passage en phase d'élongation (Holstege et al., 1997; Ohkuma et al., 1995; Sakurai and Fukasawa, 1999). Le module cœur de TFIIH contient deux hélicases : XPB et XPD (Ssl2 ou Rad25 et Rad3 chez la levure). Seule l'activité hélicase ATP-dépendante de XPB serait requise pour permettre l'ouverture de la bulle de transcription au niveau du promoteur (Douziech et al., 2000; Holstege et al., 1996). Cependant, un article récent publié par le laboratoire de S. Hahn suggère un nouveau mécanisme d'ouverture de la bulle de transcription chez *S. cerevisiae* grâce à l'activité ADN translocase de Ssl2/XPB. Celle-ci permettrait d'insérer ~15 pb d'ADN dans le sillon de la Pol II tout en induisant l'ouverture de l'ADN (Grünberg et al., 2012). XPD n'intervient donc pas dans l'ouverture de la bulle de transcription mais permet de stabiliser le module kinase de TFIIH et de réguler l'activité kinase de Cdk7 (Keriel et al., 2002).

L'obtention de la structure des complexes TFIIH humain et de *S. cerevisiae* par microscopie électronique a permis de montrer la grande conservation de l'organisation de ce complexe chez les eucaryotes (Chang and Kornberg, 2000; Schultz et al., 2000). TFIIH apparaît sous la forme d'un anneau constituant le cœur de la protéine, surmonté par une protubérance qui correspond au module kinase. Le point d'ancrage entre les deux modules est constitué par Rad3/XPB.

TFIIH interagit avec de nombreux activateurs (GAL4-VP16, p53, Rb...). Ceux-ci pourraient faciliter son recrutement au sein du PIC ou activer ses activités enzymatiques (Zurita and Merino, 2003). A l'inverse, TFIIH pourrait également stimuler ces activateurs en phosphorylant certains de leurs résidus spécifiques. C'est le cas du domaine d'activation AF-1 des récepteurs nucléaires RAR, dont l'activité transcriptionnelle est stimulée après phosphorylation par TFIIH (Bastien et al., 2000; Rochette-Egly et al., 1997).

En plus de son rôle dans l'initiation de la transcription, TFIIH intervient également dans le processus de réparation de l'ADN par excision de nucléotide (Compe and Egly, 2012) qui sera décrit dans le dernier chapitre de cette introduction (cf. Chapitre III) et également dans le contrôle du cycle cellulaire (Larochelle et al., 1998).

C. La transition entre l'initiation et l'élongation de la transcription

La transition entre l'initiation et l'élongation est marquée par différentes modifications qui vont permettre à la Pol II de s'échapper du promoteur. Tout d'abord la sérine 5 du CTD est phosphorylée par TFIIF, avec pour conséquence une désolidarisation de TBP et TFIIE de la Pol II (Maxon et al., 1994; Usheva et al., 1992). D'autre part, la transcription ne peut pas débiter tant que les deux brins de la molécule d'ADN ne se sont pas séparés (Holstege et al., 1997). Cette étape est catalysée par la sous-unité Rad25 (homologue de XPB chez l'homme) de TFIIF et dépend de l'ATP (Guzder et al., 1994; Tirode et al., 1999; Yan and Gralla, 1997). Il est alors important de noter que durant les premières étapes de l'élongation, la position amont de la bulle de transcription va rester en place, près de la boîte TATA, tandis que la partie avale va se déplacer dans le sens de l'élongation, provoquant une extension de la bulle de transcription (Holstege et al., 1997; Pal et al., 2005).

La polymérisation qui débute ensuite est souvent abortive, c'est-à-dire qu'elle produit de petits ARNs de 5 à 10 nucléotides (nt) qui sont immédiatement libérés. Ce processus est surtout présent chez les procaryotes et peut conduire à plus d'une centaine de cycles avortés (Hsu et al., 2003). Chez les eucaryotes, il a été observé *in vitro* que la Pol II réalise beaucoup moins de cycles avortés. Ceci peut s'expliquer par le fait que la Pol II est assistée par TFIIF dans cette étape de transition, ainsi on constate qu'une inhibition de XPB empêche l'échappée du promoteur (Dvir et al., 1997; Lin et al., 2005; Moreland et al., 1999). TFIIF intervient également pour limiter la transcription abortive en augmentant la vitesse d'intégration des nucléotides au sein du transcrit naissant (Yan et al., 1999). Une fois que la longueur du transcrit atteint 10 nt, la bulle de transcription commence à se déplacer et on considère alors que la Pol II s'est échappée du promoteur.

2. L'élongation de la transcription par la Pol II

A. Pause à proximité du promoteur

Avant d'entrer en élongation productive, la Pol II est sujette à une pause à proximité du promoteur. Cette pause est induite par les complexes DSIF et NELF (Wada et al., 1998a; Yamaguchi et al., 1999). DSIF est constitué des facteurs d'élongation Spt4 et Spt5 qui sont conservés de la levure à l'homme (Hartzog et al., 1998). NELF est constitué de 4 sous-unités : NELF-A, B, C-D et E, et est conservé chez tous les mammifères, mais n'existe pas chez *S. cerevisiae* (Narita et al., 2003). L'induction de la pause par ces deux complexes pourrait se faire via une interaction avec l'ARN néo-synthétisé puisque NELF-E peut se lier à l'ARN (Fujinaga et al., 2004).

Cette pause existe dans environ 30% des gènes et peut intervenir à la fois sur des gènes inductibles et des gènes constitutifs, même si elle est plus fréquente sur les gènes impliqués dans les processus développementaux ou dans la réponse à divers stimuli (Core et al., 2008; Muse et al., 2007; Nechaev et al., 2010).

Chez la levure, la Pol II adopte une distribution relativement uniforme le long des gènes, la pause à proximité du promoteur étant absente dans la majorité des cas (Steinmetz et al., 2006). À l'inverse, chez la drosophile (Lee et al., 2008; Zeitlinger et al., 2007) et chez les mammifères (Guenther et al., 2007; Kim et al., 2005), la distribution de la Pol II est moins homogène sur le corps des gènes avec notamment un signal concentré à proximité du TSS. L'accumulation de Pol II s'accompagne de marques d'histones spécifiques aux gènes actifs (Guenther et al., 2007) et d'une bulle de transcription ouverte sur 20-60 pb en aval du TSS (Muse et al., 2007). De plus, l'inhibition de NELF induit la libération du promoteur de la Pol II pausée. Ces caractéristiques témoignent d'une Pol II engagée en élongation.

Depuis la découverte de l'existence de cette pause de la Pol II, de nombreux rôles ont pu lui être attribués. Tout d'abord, elle constitue un point de régulation en permettant l'intégration des signaux par la Pol II et également un point de contrôle afin de vérifier que la Pol II est prête à débiter l'élongation (Henriques et al., 2013). Elle permet également la déplétion en nucléosomes et le maintien de la chromatine dans cet état accessible aux différents facteurs de régulation (Shopland et al., 1995). Des expériences de localisation de la Pol II engagée dans la transcription ont révélé l'existence de deux types de pauses : une proximale en amont du nucléosome +1 et une plus distale au niveau du nucléosome +1, suggérant que les nucléosomes pourraient réguler ce mécanisme de pause en bloquant l'élongation à des sites spécifiques (Kwak and Lis, 2013). Enfin, cette pause de la Pol II pourrait également avoir un rôle dans l'ajout d'une coiffe en 3' de l'ARN néo-synthétisé, et cette modification pourrait être un pré-requis à la sortie de la pause (Pei et al., 2003).

La sortie de pause requiert l'activité kinase de P-TEFb qui va phosphoryler les complexes NELF et DSIF (Peterlin and Price, 2006), engendrant la dissociation de NELF tandis que DSIF reste associé à la Pol II mais passe d'un facteur inhibant à un facteur stimulant l'élongation (Wada et al., 1998b). Lors de son entrée en pause, la Pol II a un mouvement de recul sur l'ADN qui va déplacer du site catalytique l'extrémité 3' du transcrit. TFIIIS va alors jouer un rôle important en stimulant l'activité de clivage de la Pol II qui va permettre de réaligner l'extrémité 3' de l'ARN avec le site actif de la Pol II (Adelman et al., 2005). Sur la majorité des gènes, le recrutement de P-TEFb intervient immédiatement après l'initiation, la Pol II peut alors rapidement passer en élongation

productive. Cependant, à certains gènes, cet événement intervient plus lentement, avec pour conséquence une accumulation de Pol II pausée.

B. Progression de la bulle de transcription

Au cours de l'élongation, la bulle de transcription progresse le long de l'ADN qui s'ouvre au fur et à mesure sur 11-12 nucléotides. Durant cette étape, la molécule d'ARN est allongée par incorporations de NTPs et libération de pyrophosphate (PPi). L'ARN naissant est fusionné avec le brin d'ADN matrice sur 8-9 nt, formant un hybride ARN-ADN qui joue un rôle important dans la stabilité du complexe d'élongation (Kireeva et al., 2000).

In vitro, la vitesse de transcription par la Pol II n'est que de 100 à 300 nt par minute alors qu'elle est cinq fois plus élevée *in vivo*. Cette différence est due à la présence de nombreux facteurs d'élongation qui interviennent à plusieurs niveaux (Fish and Kane, 2002). Les facteurs TFIIF, ELL et les élonguines diminuent les pauses de la Pol II tandis que TFIIIS permet de réactiver la Pol II après une pause. HMG14 et FACT stimulent directement l'élongation par la Pol II. Enfin, d'autres facteurs tels que P-TEFb, DSIF, Elongator... ont également été impliqués dans la régulation de l'élongation mais leurs mécanismes d'action sont peu décrits.

Durant la phase d'élongation, la Pol II marque encore des pauses régulières. Le facteur d'élongation TFIIIS aide alors la Pol II à reprendre la transcription en stimulant son activité nucléase. Le transcrit naissant est clivé près de son extrémité 3' qui peut ensuite reprendre une position correcte au sein du site actif de la Pol II (Wind and Reines, 2000). TFIIF est quant à lui capable d'influencer la vitesse de progression de la Pol II en réduisant le temps que la Pol II passe sous forme pausée (Lei et al., 1999). Des analyses par NET-seq (« native elongating transcript sequencing ») réalisées chez *S. cerevisiae* ont montré que la Pol II passe autant de temps en pause qu'en élongation (Churchman and Weissman, 2011). De plus, les pauses interviennent près de régions riches en nucléosomes qui seraient la cause majeure des pauses observées. La délétion de *DST1*, le gène codant pour TFIIIS, induit un décalage du site de pause, qui correspond à la position de la Pol II avant son recul, et un changement global dans les séquences entourant le site de pause. Ceci confirme le rôle essentiel de TFIIIS dans la reprise de la transcription après une pause.

3. La terminaison de la transcription par la Pol II

La terminaison de la transcription est un processus complexe qui n'intervient pas à un site précis ou à une distance constante de l'extrémité 3' du nouveau transcrit. Chez les mammifères, la distance entre la région 3' et le site de terminaison de la transcription peut ainsi varier de quelques pb à plusieurs kb (Proudfoot, 1989).

La terminaison est importante afin d'éviter une interférence avec la transcription des gènes situés en aval du transcrit. Cette étape est dépendante de la poly-adénylation (poly(A)) de l'ARN néo-synthétisé. En effet, un défaut à ce niveau empêche la terminaison *in vivo* (Birse et al., 1998; Osheim et al., 1999). Suite au signal de poly(A), la Pol II décélère et l'hybride ARN/ADN est déstabilisé. La déphosphorylation du CTD par Fcp1 joue également un rôle déterminant dans la terminaison de la transcription en induisant la coupure de l'ARN néo-synthétisé, la libération de la Pol II et son recyclage (Kamada et al., 2003; Proudfoot et al., 2002). Par ailleurs, un mode de terminaison alternatif a été mis en évidence dans le cas des ARNnc faisant intervenir le complexe Ndr1, qui reconnaît des sites spécifiques à l'extrémité 3' de l'ARN. (Lykke-Andersen and Jensen, 2007)

4. La réinitiation de la transcription par la Pol II

Après initiation de la transcription par la Pol II, *in vitro*, TFIIA, TFIID, TFIIH, TFIIIE et le Médiateur restent associés sur la région promotrice formant une plateforme pour la réinitiation appelée « scaffold » (cf. Chapitre II.VI.2.). Ce complexe n'a pas encore été identifié *in vivo* mais pourrait exister au niveau des gènes les plus transcrits pour permettre une réinitiation plus rapide de la transcription en recrutant les facteurs manquants (Liu et al., 2004; Yudkovsky et al., 2000).

5. La transcription « pervasive »

Au cours des dernières années, il est devenu de plus en plus évident que l'expression du génome des eucaryotes est beaucoup plus complexe que ce que l'on pensait jusqu'alors. L'idée selon laquelle le transcriptome dériverait exclusivement des gènes codants pour des protéines ainsi que de quelques ARNnc (ARNsn, ARNsno, ARNt ou ARNr) a été remise en cause par de nombreuses études indiquant que la Pol II peut être retrouvée à n'importe quel endroit du génome. Le terme de transcription « pervasive » se réfère donc au fait que la transcription s'étend à tout le génome de façon omniprésente. Ce processus loin d'être futile semble jouer un rôle crucial dans le contrôle de l'expression génique et la plasticité du génome (Jensen et al., 2013).

Chez *S. cerevisiae*, on estime que 85% du génome est transcrit (David et al., 2006). Deux catégories de transcrits « pervasifs » peuvent être distingués : une première partie qui provient des régions intergéniques (Nagalakshmi et al., 2008) et une seconde qui se trouve au niveau d'ORFs connues, transcrite soit dans la même direction, soit dans le sens opposé du gène codant (Carninci et al., 2006; Katayama et al., 2005). Ce type de transcrit a pu être mis en évidence grâce à des mutations dans les facteurs d'élongation Spt6 et Spt16, protéines indispensables à la réorganisation correcte des nucléosomes après le passage de la Pol II. Leur mutation provoque une transcription intragénique parasite qui débute au sein de la région codante au niveau de promoteurs cryptiques (Kaplan et al., 2003). Peu après, une autre preuve de l'existence de cette transcription pervasive a été apportée grâce à l'utilisation de souches défectueuses dans certaines voies de dégradation de l'ARN. Ainsi, l'absence de Rrp6, une sous-unité catalytique de l'exosome de *S. cerevisiae*, induit l'accumulation d'un certain type de transcrits appelés CUTs (« cryptic unstable transcripts »), qui normalement ne s'accumulent pas jusqu'à un niveau détectable dans les cellules sauvages (Davis and Ares, 2006; Neil et al., 2009; Wyers et al., 2005). Un second type de transcrit a par la suite été découvert, moins sensible à l'activité de dégradation de Rrp6 et nommé de par ce fait SUTs (« stable unannotated transcripts »)(Xu et al., 2009). Enfin, une dernière classe de transcrit a pu récemment être mise en évidence, dont la dégradation dépend de l'exoribonuclease Xrn1 et donc nommés XUTs (« Xrn1-sensitive unstable transcripts ») (van Dijk et al., 2011).

CUTs, SUTs et XUTs sont tous transcrits par la Pol II. Les CUTs représentent 13% des transcrits trouvés chez la levure, les SUTs 12%. Ces deux types de transcrits proviennent quasiment exclusivement de régions pauvres en nucléosomes (« nucleosome-depleted regions » NDRs) situées dans les parties 5' ou 3' des gènes. La majeure partie de ces transcrits résulte d'une transcription divergente au niveau des promoteurs des gènes, laissant suggérer que la majorité des promoteurs sont divergents, leur polarité apparente résultant seulement du fait que l'un des deux transcrits produit est rapidement stoppé et dégradé (Core et al., 2008; He et al., 2008; Preker et al., 2008; Schulz et al., 2013; Seila et al., 2008; Xu et al., 2009).

La transcription pervasive existe également chez les mammifères. On estime ainsi que 75% du génome humain est transcrit. Dans ce type d'organisme, les ARNnc sont définis selon leur longueur. Il existe ainsi des ARNnc longs (LncRNA) qui ont une taille > 200 nucléotides et sont, le plus souvent, associés avec la transcription des ARNm ; et des ARNnc courts qui sont trouvés soit à la fin soit au début des gènes codants des protéines (Kapranov et al., 2007) et nommés pour cette raison PASRs (« promoter-associated small RNAs ») ou TASRs (« terminator-associated small RNAs »).

Une des hypothèses quant au rôle de la transcription pervasive serait que la Pol II fonctionne selon un système d'essais et tests d'erreurs et se lance parfois dans une transcription non fonctionnelle avant de transcrire les gènes codants des protéines (Berretta and Morillon, 2009). Cependant, plusieurs observations suggèrent que dans certains cas, la transcription pervasive pourrait avoir une fonction. Tout d'abord, tous ces transcrits ne sont pas instables, notamment les SUTs chez la levure et les PASRs chez l'homme. De plus, de nombreuses séquences qui encodent ces transcrits sont très conservées (Carninci et al., 2005). Enfin, il existe des transcrits qui peuvent être régulés selon les conditions de croissance (Xu et al., 2009) ou durant le développement (Stolc et al., 2004). Dans certains cas, la séquence du transcrit n'est pas conservée mais celle du promoteur l'est, et la transcription de l'ARNnc aurait alors un rôle d'interférence. C'est le cas de l'ARNnc *SRG1* qui interfère avec le promoteur du gène de réponse au stress *SER3* en empêchant la liaison des différents facteurs de la transcription (Martens et al., 2004).

Bien que cette transcription semble jouer un rôle physiologique, la majorité est simplement due à une transcription incontrôlée au niveau du site d'initiation. Des systèmes doivent donc intervenir afin de limiter ce « défaut » pour ne pas qu'il perturbe la transcription « classique ». De plus, un excès d'ARNnc pourrait être toxique pour la cellule. C'est pourquoi différents facteurs limitent ce processus. C'est le cas des remodeleurs de la chromatine Spt6 et FACT (Kaplan et al., 2003) ou de l'histone deacétylase Rpd3, qui assurent le maintien des nucléosomes sur l'ADN durant l'élongation pour éviter tout début de transcription non désiré. Dans le cas où cette première barrière ne suffit pas, des systèmes de terminaison de la transcription couplés directement à la dégradation du transcrit peuvent intervenir. Chez *S. cerevisiae*, ce processus dépend du système Nrd1 constitué des protéines de liaison à l'ADN Nrd1 et Nab3 et de l'hélicase Sen1 (Arigo et al., 2006; Thiebaut et al., 2006). Nrd1 et Nab3 permettent de reconnaître un signal de terminaison sur le transcrit naissant. Nrd1 interagit également avec le CTD de la Pol II et permet le recrutement de Sen1 qui va dissocier le transcrit de sa matrice ADN (Porrúa and Libri, 2013). Enfin, le système Nrd1 s'associe avec l'exosome et le complexe activateur TRAMP (Trf4-Air2-Mtr4 polyadenylation), tous deux impliqués dans la dégradation des CUTs.

III. CONTEXTE CHROMATINIEN

Au sein du noyau, l'ADN des eucaryotes est associé à des complexes protéiques qui forment la chromatine (Figure 7). Cette organisation permet de compacter l'ADN mais aussi de réguler les processus biologiques ayant l'ADN comme substrat tels que la transcription, la réplication, la recombinaison et la réparation, en modulant l'accès à l'ADN des différentes protéines intervenant dans ces processus.

La chromatine est principalement constituée d'histones composées de deux hétérodimères H2A-H2B et H3-H4, autour desquels s'enroule l'ADN sur 146 pb pour former un nucléosome. Sous cette forme compactée, l'ADN est moins accessible à la machinerie de transcription, cependant, il existe au sein du noyau des régions où la chromatine est sous forme relâchée formant l'euchromatine par opposition avec l'hétérochromatine qui représente les zones de structure plus dense. Les promoteurs des gènes sont situés dans ces régions où la densité de nucléosomes est moindre (Bernstein et al., 2004; Lee et al., 2004; Sekinger et al., 2005). Il existe ainsi des séquences d'environ 200pb autour des promoteurs des gènes qui sont dépourvues de nucléosomes (Yuan et al., 2005).

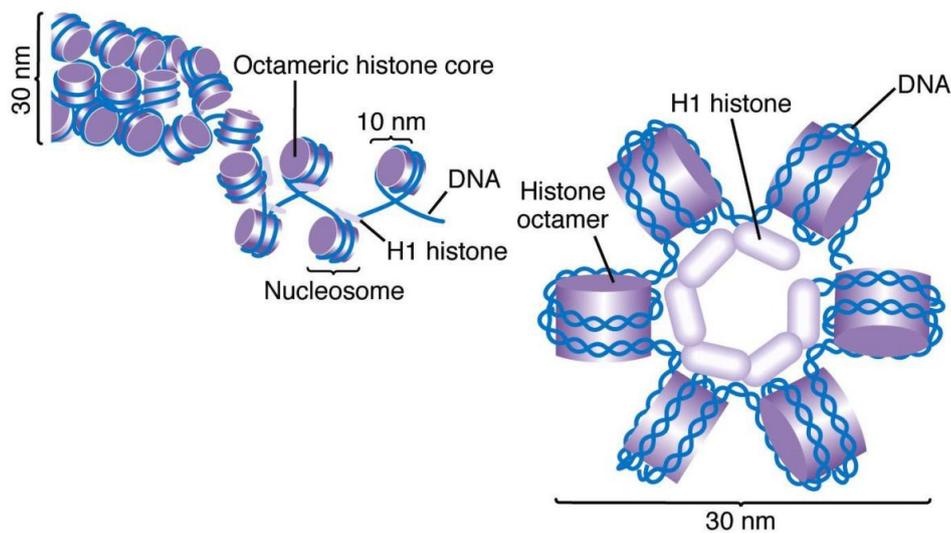


Figure 7 : Organisation chromatinienne de l'ADN (Griffiths et al., 2004, <http://www.mun.ca/>)

La cartographie des nucléosomes sur l'ensemble du génome de *S. cerevisiae* a pu être définie avec une très grande résolution en compilant les données obtenues par différents laboratoires (Jiang and Pugh, 2009). Tous ont utilisé la même méthode qui consiste à digérer les fragments obtenus par ChIP avec la MNase, une endonucléase qui digère l'ADN qui n'est pas protégé par les nucléosomes. Un séquençage à haut débit des fragments restants permet ensuite de localiser

précisément les nucléosomes (Wal and Pugh, 2012). Cette carte révèle que la position des nucléosomes varie très peu d'une cellule à l'autre et que la distance entre eux est globalement fixe. Elle montre également l'existence de motifs spécifiques d'organisation en 3' et en 5' des gènes. La localisation des nucléosomes dans les cellules humaines a également été étudiée en utilisant une méthode similaire (Gaffney et al., 2012; Schones et al., 2008; Valouev et al., 2011).

1. Modification des histones

Les histones subissent de nombreuses modifications post-traductionnelles qui vont changer leurs propriétés physico-chimiques (Kouzarides, 2007). Il s'agit de méthylations, acétylations, sumoylations, ubiquitinations, ADP-ribosylations, et phosphorylations des sérines, lysines et arginines (Tableau 4). Les différentes combinaisons de ces modifications peuvent être reconnues par des protéines spécifiques et accompagnent la réponse transcriptionnelle qui varie selon la position sur le génome.

Modifications	Position		Enzymes		Fonction dans la transcription
			<i>S. cerevisiae</i>	Mammifères	
Méthylation	H3	K4	Set1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	Activation
		K9	n/a	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Répression, activation
		K27		E(Z)	Répression
		K36	Set2	HYPB, Smyd2, NSD1	Recrutement de Rpd3S pour réprimer l'initiation interne
		K79	Dot1	Dot1L	Activation
	H4	K20		PR-Set7, SET8	« Silencing »
Méthylation Arg	H3	R2		CARM1	Activation
		R17		CARM1	Activation
		R26		CARM1	Activation
	H4	R3		PRMT1	Activation
Phosphorylation	H3	S10	Snf1		Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	UbcH6, RNF20/40	Activation
	H2A	K119		hPRC1L	Répression
Acétylation	H3	K56			Activation
	H4	K16	Sas2, NuA4	hMOF	Activation
	Htz1	K14	NuA4, SAGA		Activation

Tableau 4 : Modification des histones (Li et al., 2007)

Certaines modifications sont associées à une activation de la transcription comme c'est le cas pour l'acétylation des histones H3/H4 et les méthylations des lysines K4 et K36 de l'histone H3 (Pokholok et al., 2005). Ainsi, on retrouve au niveau des régions promotrices les marques H3K4me3, H3K9ac et H3K14ac, tandis que la marque H3K36me3 est présente sur l'ORF des gènes activement transcrits. A l'inverse, la désacétylation des histones et la méthylation des lysines K9 et K27 de l'histone H3 sont associés à une répression de la transcription (Li et al., 2007). Aussi, les marques H3K27me et H3K9me3 sont retrouvées exclusivement sur les gènes réprimés.

2. Variants d'histones

Il existe également des variants d'histones (Tableau 5) qui peuvent être incorporés à la place des histones canoniques et qui jouent un rôle dans la régulation de l'expression des gènes. Ils diffèrent des histones canoniques au niveau de la queue N-terminale, de l'HFD (« histone fold domain ») ou de plusieurs acides aminés dans la séquence protéique (Doyen et al., 2006; Henikoff and Ahmad, 2005). Leur incorporation a un impact sur la structure chromatinienne et peut être associée à une activation ou à une inhibition de la transcription.

Histone	Variante	Rôle dans la transcription	Localisation
H3	H3.3	Activation de la transcription	Région transcrite
H2A	macroH2A	Inactivation du chromosome X	Chromosome X inactivé
	H2AZ	Activation/ répression de la transcription	Promoteur, hétérochromatine
	H2ABbd	Activation de la transcription	Chromosome X actif et autosomes
	H2A.X	Répression	Canonique, distribué sur tout le génome

Tableau 5: Variants d'histones (Li et al., 2007)

3. Remodelage de la chromatine

Certains complexes protéiques utilisent l'hydrolyse de l'ATP pour altérer les contacts entre l'ADN et les histones avec pour conséquence un « remodelage de la chromatine » (Vignali et al., 2000). Le terme « remodelage » définit tous les événements qui altèrent la sensibilité aux nucléases de la chromatine (Aalfs and Kingston, 2000). Celui-ci peut consister en un déroulement de l'ADN, un déplacement des nucléosomes ou leur détachement complet et va permettre de modifier l'accessibilité de l'ADN aux divers facteurs de la transcription. Une description plus complète des complexes de remodelage de la chromatine sera faite dans la dernière partie de ce chapitre (cf. IV. 2. B. de ce chapitre).

IV. REGULATION DE LA TRANSCRIPTION PAR LA POL II

Les différents signaux de l'environnement cellulaire entraînent la fixation de facteurs de transcription (TFs), répresseurs (TRs) ou activateurs (TAs), sur des éléments régulateurs (UAS « upstream activating sequence » ou URS « upstream repressing sequence ») associés au promoteur des gènes afin d'en moduler l'expression. Des co-activateurs et co-répresseurs vont ensuite intégrer les signaux en provenance des activateurs et répresseurs spécifiques et les transmettre à la machinerie de base de la transcription dont l'activité pourra alors être augmentée ou diminuée.

1. Les éléments de régulation

Les éléments de régulation sont des séquences d'ADN pouvant être liées par divers facteurs de transcription impliqués dans le contrôle de l'expression génique. On en distingue deux types : les « enhancers » ou UAS et les « silencers » ou URS. Ces éléments sont majoritairement localisés dans des régions dépourvues de nucléosomes (Xue et al., 2004), facilitant ainsi l'accès aux TFs.

Les « enhancers » ou UAS sont des séquences d'ADN liées par des activateurs spécifiques. Ces séquences régulatrices semblent indispensables à la transcription de l'ensemble des gènes de classe II et leur nombre et position va influencer le niveau d'expression (Swamy et al., 2009). Les URS sont des sites de liaison pour des répresseurs spécifiques. Ils peuvent être situés en amont ou en aval du promoteur (Johnson, 1995).

Chez la levure, les UAS ne sont efficaces que si elles sont situées à une faible distance (quelques centaines de pb) en amont du promoteur de base (Dobi and Winston, 2007; Struhl, 1984). A l'inverse, chez les eucaryotes pluricellulaires, les « enhancers » sont fonctionnels à la fois en 3' et en 5' du gène (Webster et al., 1988) et peuvent se situer à une distance du promoteur supérieure à 1kb (Blackwood and Kadonaga, 1998). Or, les activateurs fixés sur l'« enhancer » recrutent des co-activateurs qui, à leur tour, vont se fixer au promoteur. Ce schéma d'interaction implique que l'activation de la transcription s'accompagne de la formation de boucles de l'ADN, afin de rapprocher le promoteur et l'« enhancer ». Des expériences de capture de conformation de chromosome (« chromosome conformation capture » 3C) (Dekker et al., 2002) ont permis de confirmer l'existence de ce phénomène sur différents gènes et de définir les protéines impliquées dans la formation des boucles de l'ADN (Jing et al., 2008; Vakoc et al., 2005). Par la suite, les méthodes 4C et 5C qui associent le 3C avec une analyse sur puce ou un séquençage (Dostie et al., 2006; Nora et al., 2012; Sexton et al., 2012; Simonis et al., 2006; Zhao et al., 2006), ont permis d'étudier la conformation de la chromatine à une plus grande échelle. Enfin, la méthode Hi-C est

apparue, qui associe le 3C avec un séquençage à haut débit, permettant ainsi l'analyse de la structure tridimensionnelle de la chromatine à l'échelle du génome (Dixon et al., 2012; Lieberman-Aiden et al., 2009). Les résultats obtenus montrent que les promoteurs et les éléments de régulation sont engagés dans de multiples interactions à longue portée et forment des réseaux complexes. Progressivement, ce type d'analyse permet de placer les gènes et les éléments de régulation dans un contexte tridimensionnel afin de révéler leurs relations fonctionnelles.

2. Les facteurs de transcription (TFs)

A ce jour, 180 protéines sont qualifiées de TFs dans la base de données de *S. cerevisiae* : Yeabstract. Ils se regroupent en différentes familles selon la séquence et la structure de leur domaine de liaison à l'ADN (Luscombe et al., 2000; Pabo and Sauer, 1992). On distingue, par exemple, les protéines à doigt de zinc, les protéines hélice-tour-hélice, les protéines à glissière à leucine, les protéines à feuillet β ...

Les méthodes de CHIP-chip et CHIP-seq (immunoprécipitation de chromatine suivie d'une hybridation sur puce à ADN ou d'un séquençage à haut débit) sont largement utilisées pour déterminer les sites de fixation des TFs, leurs gènes cibles et leur rôle dans la régulation (Guo et al., 2010; Zhu et al., 2009). Ces analyses permettent également de découvrir de nouvelles protéines interagissant avec l'ADN, notamment de nouveaux TFs, ainsi que les liaisons coopératives qui peuvent exister entre plusieurs TFs dans certaines conditions biologiques.

Les TFs se fixent préférentiellement dans les régions déplétées en nucléosomes (NFR) et contribuent à leur maintenance dans cet état (Badis et al., 2008). En surexprimant ou en délétant des TFs spécifiques, il est possible de déterminer leurs gènes cibles, sachant que dans la majorité des cas, plusieurs TFs peuvent réguler les mêmes gènes (Chua et al., 2006; Sopko et al., 2006).

A. Les activateurs (TAs)

Généralement constitués de deux domaines : un domaine de liaison à l'ADN (DBD « DNA binding domain ») et un domaine d'activation de la transcription (AD « activation domain ») (Triezenberg, 1995), les activateurs ciblent un site spécifique (UAS) et permettent le recrutement de co-activateurs (cf. IV.3. de ce chapitre). De ce fait ils favorisent le recrutement de la machinerie de transcription au promoteur. Par exemple, les activateurs Gal4 et Gcn4, très étudiés chez la levure, permettent le recrutement de SAGA, NuA4, Swi/Snf, TFIID et du Médiateur (Fishburn et al., 2005; Reeves and Hahn, 2005; Topalidou and Thireos, 2003).

La plupart possèdent plusieurs domaines d'activation, ce qui leur permet de stimuler l'expression de plusieurs gènes souvent impliqués dans des processus similaires. Ainsi, l'activateur

Hsf1 se lie sur ses gènes cibles suite à un stress thermique ou oxydatif et active, par exemple, les gènes codants les HSP (« heat shock protein ») (Hahn and Thiele, 2004). Leu2 agit à la fois en tant qu'activateur et répresseur dans la régulation de la biosynthèse des chaînes d'acides aminés branchées (Kohlhaw, 2003). La régulation des activateurs se fait au niveau post-traductionnel par phosphorylation ou ubiquitination qui favorise leur activité, leur transfert au noyau ou, au contraire, leur dégradation (Salghetti et al., 2001).

B. Les répresseurs (TRs)

Les répresseurs sont définis par le fait que leur délétion ou leur mutation entraîne une augmentation de la transcription des gènes cibles. Ils inhibent la transcription de trois manières différentes : (1) en renforçant la structure chromatinienne et bloquant ainsi la mise en place du PIC, par exemple le complexe Ssn6-Tup1 (Smith and Johnson, 2000), (2) en interagissant avec les GTFs ou les co-activateurs, par exemple, Mot1, NC2 (Auble et al., 1994; Goppelt et al., 1996) ou (3) en inhibant la fixation des activateurs, par exemple, Hsp90, Gal80 (Sil et al., 1999; Zou et al., 1998).

3. Les co-activateurs

Les co-activateurs sont requis pour l'activation de la transcription par les activateurs dont ils sont la principale cible (cf. IV.2.A. de ce chapitre). Bien qu'ils soient nécessaires aux mécanismes d'activation de la plupart des gènes *in vivo*, ils ne sont pas requis pour la transcription basale *in vitro*, et contrairement aux activateurs spécifiques ils ne lient pas l'ADN sur des sites spécifiques (Hampsey, 1998). Ils peuvent avoir deux types d'activités : (1) favoriser la formation du PIC (Médiateur, TAFs) en interagissant de manière directe avec la machinerie de transcription ; ou (2) modifier la structure de la chromatine (SWI/SNF, SAGA). Certains co-activateurs possèdent ces deux activités (Näär et al., 2001).

A. Les complexes agissant sur la mise en place du PIC

a. TAFs de TFIID

En plus de la TBP, TFIID contient 14 TAFs (« TBP associated factors ») dont 5 sont partagés avec le complexe SAGA (cf. IV.3.A.b. de ce chapitre). Ces protéines ont tout d'abord un rôle structural en permettant l'assemblage et la stabilisation du complexe TFIID. Par exemple, TAF1 est indispensable à l'association de la TBP avec le reste du complexe (Singh et al., 2004). Mais les TAFs jouent également un rôle d'adaptateur en reliant les activateurs spécifiques à la machinerie de transcription par la Pol II (Näär et al., 2001). Ils agissent donc en tant que co-activateurs, requis

pour la transcription activée mais dispensables pour la transcription basale *in vitro*. Une caractérisation plus approfondie de ces protéines a permis de montrer que les activateurs peuvent contacter certains TAFs de façon spécifique afin de recruter TFIID sur les différents promoteurs (Liu et al., 2009; Papai et al., 2010). Les TAFs auraient aussi pour rôle de stabiliser TFIID sur le promoteur en créant des contacts spécifiques avec l'ADN au niveau de l'Inr et du DPE (Smale et al., 1990; Verrijzer et al., 1995).

Les TAFs sont très conservés chez l'ensemble des eucaryotes, ce qui a permis la mise en place d'une nomenclature unifiée (Tableau 6). Chez la levure, la composition en TAF de TFIID est unique alors que chez les mammifères, la composition varie selon le type cellulaire ou la voie de régulation utilisée afin d'assurer une spécificité de liaison aux activateurs (Kolthur-Seetharam et al., 2008).

Nom	Homme	<i>S. cerevisiae</i>
TAF1	TAFII250	Taf145/130
TAF2	TAFII150	Taf150 ou TSM1
TAF3	TAFII140	Taf47
TAF4	TAFII130/135	Taf48 ou MPT1
TAF4b	TAFII105	
TAF5	TAFII100	Taf90
TAF5L	PAF65b	
TAF6	TAFII80	Taf60
TAF6L	PAF65	
TAF7	TAFII55	Taf67
TAF7L	TAF2Q	
TAF8	(BAB71460)	Taf65
TAF9	TAFII32/31	Taf17
TAF9L	TAFII31L	
TAF10	TAFII30	Taf25
TAF11	TAFII28	Taf40
TAF12	TAFII20/15	Taf61/68
TAF13	TAFII18	Taf19 ou FUN81
TAF14		Taf30
TAF15	TAFII68	

Tableau 6: Nomenclature unifiée pour les TAFs (Tora, 2002).

Les TAFs communs aux complexes TFIID et SAGA sont indiqués en rouge.

Des analyses d'expression des gènes à l'échelle du génome de *S. cerevisiae* après délétion de TAFs spécifiques à TFIID ou à SAGA ont permis de définir que 90% des gènes sont majoritairement TFIID-dépendants et 10% majoritairement SAGA-dépendants (Basehoar et al., 2004). Certains TAFs possèdent une ou plusieurs activités enzymatiques. Par exemple, TAF1 possède des activités kinase et histone acetyltransferase (HAT) (Mizzen et al., 1996). Ce TAF contient également deux bromodomains, des structures capables de reconnaître les résidus lysine acétylés des histones, que l'on retrouve dans les protéines impliquées dans les remaniements de la chromatine.

b. SAGA

Le complexe SAGA de *S. cerevisiae* régule l'activité d'environ 10% du génome. Il est plus particulièrement impliqué dans la transcription des gènes associés au stress, hautement régulés (Basehoar et al., 2004). On constate que ces gènes contiennent pour la plupart une boîte TATA canonique. Les autres gènes sont plutôt dépendant de TFIID, ce sont souvent des gènes exprimés de façon constitutive et qui, pour la majorité, ne contiennent pas de boîte TATA canonique.

SAGA possède de nombreuses fonctions incluant l'acétylation et la désubiquitination des histones et l'assemblage du PIC. En plus de ces rôles dans l'activation de la transcription, SAGA facilite l'étape d'élongation (Govind et al., 2007), l'export des ARNm néosynthétisés en dehors du noyau, la maintenance des télomères (Atanassov et al., 2009) et la réparation de l'ADN par le système NER (Guo et al., 2011). Ces activités, initialement décrites chez la levure, sont pour la plupart conservées jusqu'à l'homme.

SAGA est un complexe multiprotéique de 2 MDa qui peut contenir jusqu'à 20 sous-unités, dont les protéines Ada, Spt, TAF et les protéines Tra1 et Gcn5 (Timmers and Tora, 2005). Ces sous-unités sont organisées en 5 modules (Koutelou et al., 2010). Le module HAT contient l'histone acétyl transférase Gcn5 qui acétyle l'histone H3 et permet l'ouverture de la chromatine (Owen et al., 2000). Le module DUB contient l'enzyme Ubp8 capable de désubiquitiner l'histone H2B (Daniel et al., 2004), évènement indispensable au recrutement de la kinase Ctk1 (Cdk9 chez l'homme) qui phosphoryle le CTD. Le module SPT intervient dans l'assemblage du PIC en interagissant directement avec la TBP via les sous-unités Spt3 et Spt8 pour permettre son recrutement au promoteur (Laprade et al., 2007; Mohibullah and Hahn, 2008; Sermwittayawong and Tan, 2006). Ce module contient également la protéine Tra1 qui interagit avec les activateurs pour permettre le recrutement de SAGA au promoteur des gènes cibles (Brown et al., 2001). Enfin, le module TAF joue un rôle dans le maintien de l'architecture de SAGA.

Des preuves récentes ont permis de consolider l'hypothèse selon laquelle le complexe SAGA de la levure aurait divergé au cours de l'évolution en différents complexes protéiques « SAGA-like » chez les métazoaires, incluant des variants de SAGA bien conservés et des variants d'ATAC (« Ada two A containing ») plus distants (Guelman et al., 2006). La levure ne possède que le complexe SAGA, ainsi qu'une forme dérivée SLIK (« SAGA-like complex) ou SALSAs (« SAGA altered Spt8 absent ») qui ne contient pas la sous-unité Spt8 (Belotserkovskaya et al., 2000; Pray-Grant et al., 2002; Sterner et al., 2002). La drosophile possède 2 complexes « SAGA-like » : SAGA et ATAC, et l'homme 4 complexes : SAGA et ATAC contenant les sous-unités Gcn5 ou Pcaf.

c. Le Médiateur

Le Médiateur de la transcription sera plus amplement détaillé dans le chapitre II de cette introduction. Il s'agit d'un complexe multi-protéique conservé chez tous les eucaryotes qui crée un lien entre la Pol II et les activateurs spécifiques en transmettant les signaux de régulation (Kornberg, 2005).

B. Les remodeleurs de la chromatine

Afin de faciliter l'accès des différents facteurs de la transcription à l'ADN, un « remodelage » de la chromatine est nécessaire, pouvant nécessiter ou non une source d'énergie (ATP). Les différents remodeleurs de la chromatine partagent un domaine ATPase similaire, flanqué par des domaines variables qui permettent leur classification en 4 familles distinctes : SWI/SNF, ISWI, CHD et INO80 (Clapier and Cairns, 2009). Ces familles sont conservées de la levure à l'homme (Tableau 7).

Le rôle des remodeleurs de la chromatine ne semble pas restreint à la transcription, ils pourraient également intervenir dans d'autres mécanismes tels que la réparation et la réplication de l'ADN (Citterio et al., 2000; Guschin and Wolffe, 1999).

Familie	Organismes		
	Levure	Drosophile	Homme
SWI/SNF	SWI/SNF - RSC	BAP - PBAP	BAF - PBAF
ISWI	ISW1a - ISW1b - ISW2	NURF - CHRAC - ACF	NURF - CHRAC - ACF
CHD	CHD1	CHD1 - Mi-2/NuRD	CHD1 - NuRD
INO80	INO80 - SWR1	Pho-dINO80 - Tip60	INO80 - SRCAP - TRRAP/Tip60

Tableau 7 : Classification des remodeleurs de la chromatine (Clapier and Cairns, 2009).

a. SWI/SNF

Les membres de la famille SWI/SNF (« switch defecting/sucrose non-fermenting »), ont été initialement purifiés chez *S. cerevisiae*, en tant que gros complexes de 2 MDa et de 8 à 14 sous-unités (Cairns et al., 1994, 1996). Deux complexes ont été mis en évidence : SWI/SNF (BAF chez l'homme) et RSC (PBAF chez l'homme) qui ciblent des gènes différents. SWI/SNF est un complexe dispensable, qui régule environ 5% des gènes de *S. cerevisiae* (Holstege et al., 1998). A l'inverse, RSC est essentiel à la viabilité cellulaire et beaucoup plus abondant (Cairns et al., 1996) avec notamment un rôle dans la réponse au stress, l'activation de la transcription et la ségrégation des chromosomes.

Les remodeleurs de la familles SWI/SNF possèdent une activité ATPase intrinsèque stimulée par l'ADN (Laurent et al., 1993) qui leur permet de perturber l'arrangement des nucléosomes et de faciliter la liaison à l'ADN des facteurs de la transcription. Ils contiennent également plusieurs bromodomaines situés dans la région C-terminale de l'ATPase, capables de reconnaître les résidus lysine acétylés et essentiels à la fluidification de la chromatine.

Ces complexes sont recrutés par les activateurs au niveau de gènes spécifiques (Peterson, 2000). Ils peuvent également s'associer aux récepteurs nucléaires (Wallberg et al., 2000) ainsi qu'à des facteurs spécifiques tels que c-Myc ou C/EBP (Armstrong et al., 1998; Kowenz-Leutz and Leutz, 1999). En présence de chaperonnes des histones, ils sont capables d'engendrer le détachement complet du nucléosome (Boeger et al., 2003). Beaucoup de promoteurs ne dépendent pas de SWI/SNF pour leur activation. Une explication possible serait que ces complexes seraient uniquement requis quand les éléments critiques du promoteur sont positionnés sur des nucléosomes (Hirschhorn et al., 1992).

b. ISWI

D'autres complexes remodeleurs de la chromatine ont également été mis en évidence, qui présentent une grande homologie avec le domaine ATPase du complexe SWI/SNF et nommés pour cette raison ISWI (« imitation SWI »). Les premiers membres de cette famille : NURF (nucleosome remodeling factor), ACF (« ATP-dependent chromatin assembly and remodelling factor ») et CHRAC (« chromatin accessibility complex ») ont initialement été isolés chez la drosophile avant que leurs homologues humains et de levure ne soient découverts (Näär et al., 2001). Ces complexes possèdent, dans la région C-terminale de l'ATPase, les domaines SANT et SLIDE capables de se lier à la queue non modifiée des histones et à l'ADN, respectivement (Boyer et al., 2004).

Les complexes ISWI sont capables de déplacer les nucléosomes via un mécanisme de glissement d'octamères qui permet leur redistribution le long de l'ADN (Hamiche et al., 1999; Längst et al., 1999). Ils interviennent à la fois dans la répression de la transcription, en optimisant l'espacement entre les nucléosomes pour permettre l'assemblage de la chromatine (ACF, CHRAC), et dans l'activation de la transcription en redistribuant aléatoirement les histones (NURF).

c. CHD

Le premier membre de la famille CHD (« chromodomain, helicase, DNA binding ») : CHD1, a initialement été identifié chez la souris en tant que protéine nucléaire liant l'ADN (Delmas et al., 1993).

Les complexes CHD contiennent 2 chromodomaines (« chromatin organisation modifier ») en tandem dans la partie N-terminale de l'ATPase capables de se lier à l'histone H3 méthylée (Flanagan et al., 2005) ainsi qu'à l'ADN (Bouazoune et al., 2002) et à l'ARN (Akhtar et al., 2000). Ils possèdent également un domaine ATPase « SNF2-like » situé dans la région centrale de la protéine.

d. INO80

Les remodeleurs de la famille INO80 (« inositol requiring 80 ») ont initialement été purifiés chez *S. cerevisiae* par immunoprécipitation (Shen et al., 2000). Cette famille se caractérise par une insertion au milieu du domaine ATPase sur laquelle se fixent des protéines spécifiques.

Bien que très proche d'INO80, SWR1 possède une manière unique de restructurer les nucléosomes, en enlevant les dimères d'histones H2A-H2B canoniques et en les remplaçant par le variant H2A.Z-H2B (Kobor et al., 2004; Krogan et al., 2003; Mizuguchi et al., 2004).

CHAPITRE II : LE MEDIATEUR DE LA TRANSCRIPTION

La transcription basale, telle qu'observée dans un système reconstitué *in vitro*, s'effectue par la Pol II en présence des facteurs généraux de la transcription. Les activateurs sont capables de stimuler la transcription *in vivo*, mais pas dans ce système reconstitué. Cette constatation est à la base de la découverte du Médiateur, un co-facteur spécifique aux eucaryotes, qui intervient dans la régulation de la transcription par la Pol II en transmettant les signaux des activateurs vers la machinerie de transcription.

Ce chapitre sera consacré à la description de ce complexe multiprotéique chez *S. cerevisiae* et chez les mammifères, ainsi qu'à l'étude de son rôle dans la régulation de la transcription des gènes par la Pol II.

I. LE MEDIATEUR DE *S. CEREVISIAE*

1. Découverte du Médiateur chez *S. cerevisiae*

La notion de Médiateur est née dans les années 90 suite à la découverte du phénomène d'interférence entre activateurs par le laboratoire de R. Kornberg (Kelleher et al., 1990). Celui-ci stipule que l'ajout *in vitro* d'un activateur n'ayant pas la capacité de se lier au promoteur va induire une diminution de la transcription. Tout se passe comme si l'activateur non lié à l'ADN séquestre un facteur requis pour la transcription. Une fraction protéique capable d'annuler cette interférence a été mise en évidence chez *S. cerevisiae*. Cette découverte d'un facteur capable de stimuler, *in vitro*, la transcription en présence d'activateurs est à l'origine de la notion de Médiateur (Flanagan et al., 1991).

Une analyse génétique menée en parallèle est venue conforter cette notion. Un groupe de gènes capables de supprimer l'effet de mutations de troncature du CTD de la Pol II a été mis en évidence. Les protéines codées par ces gènes ont ensuite été purifiées au sein d'un complexe liant le CTD et favorisant l'initiation de la transcription *in vitro* (Hengartner et al., 1995; Nonet and Young, 1989; Thompson et al., 1993) et identifié depuis comme étant le Médiateur.

La caractérisation définitive du Médiateur s'est ensuite conclue grâce à sa purification sous la forme d'un complexe d'une vingtaine de sous-unités (Kim et al., 1994).

Depuis, le Médiateur a pu être mis en évidence dans de nombreux organismes tels que la drosophile (Park et al., 2001), le ver *C. elegans* (Kwon et al., 1999), la plante *A. thaliana* (Bäckström et al., 2007) et chez les mammifères (cf. II. de ce chapitre).

2. Structure et composition du Médiateur de *S. cerevisiae*

L'identification des protéines du Médiateur a permis de déterminer qu'il se compose chez la levure de 25 sous-unités (Guglielmi et al., 2004; Kornberg, 2005; Linder and Gustafsson, 2004), réparties en 4 modules : la tête, le milieu, la queue et le module CDK (Figure 8). Ce dernier module n'est pas toujours présent dans le complexe et lui conférerait, entre autres, un rôle répresseur de la transcription.

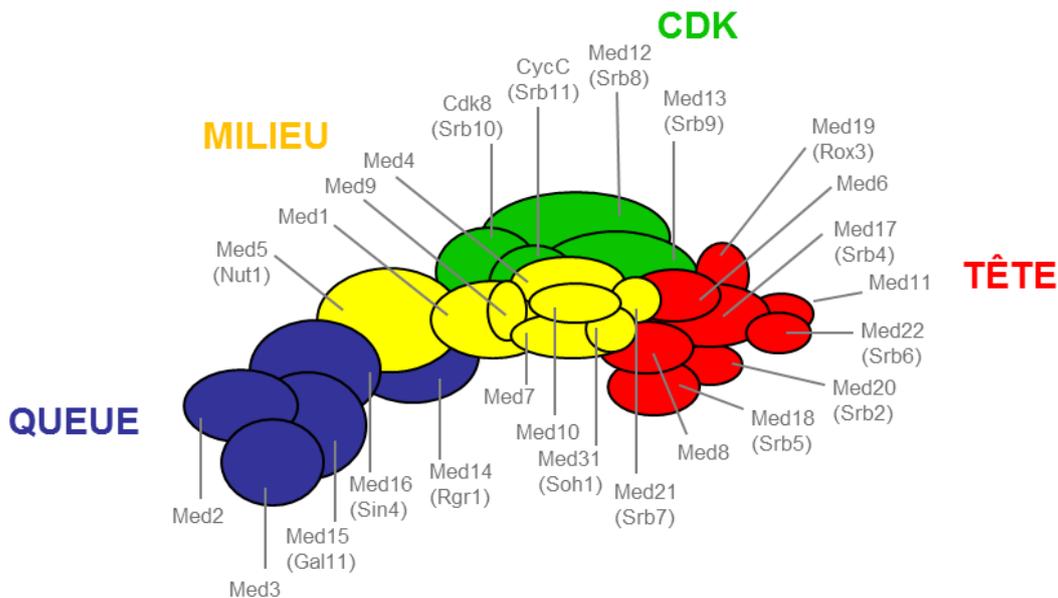


Figure 8 : Organisation des sous-unités du Médiateur de *S. cerevisiae* (Guglielmi et al., 2004).

Des expériences de co-IP et l'utilisation de cribles double-hybride ont permis de préciser l'organisation du Médiateur (Figure 9). Ainsi, le module de tête, composé de huit protéines (Med6, 8, 11, 17, 18, 19, 20 et 22), interagit avec le module du milieu (Med1, 4, 5, 7, 9, 10, 21 et 31) via une interaction entre Med17 et Med21. Le module de queue, composé de cinq protéines (Med2, 3, 14, 15 et 16), peut se replier sur celui du milieu grâce à une interaction entre Med3 et Med21. La sous-unité Med14, qui se trouve à l'interface entre le module du milieu et la queue, joue un rôle important dans l'organisation structurale du complexe (Lee et al., 1999). Enfin, le module Cdk8 vient s'ajouter au reste du complexe, il est composé de quatre sous-unités (Med12 et 13, Cdk8 et CycC) et pourrait s'associer grâce à une interaction entre Med13 et Med17.

Le Médiateur existe sous plusieurs conformations. Isolé, il forme une structure compacte, fermée, avec le module de queue replié sur celui du milieu, qui va s'étirer en présence de la Pol II pour donner une structure ouverte, propice aux interactions avec les différents facteurs de la transcription (cf. IV. 2. A. de ce chapitre). Ces changements de conformation sont importants pour l'activation de la transcription (Davis et al., 2002).

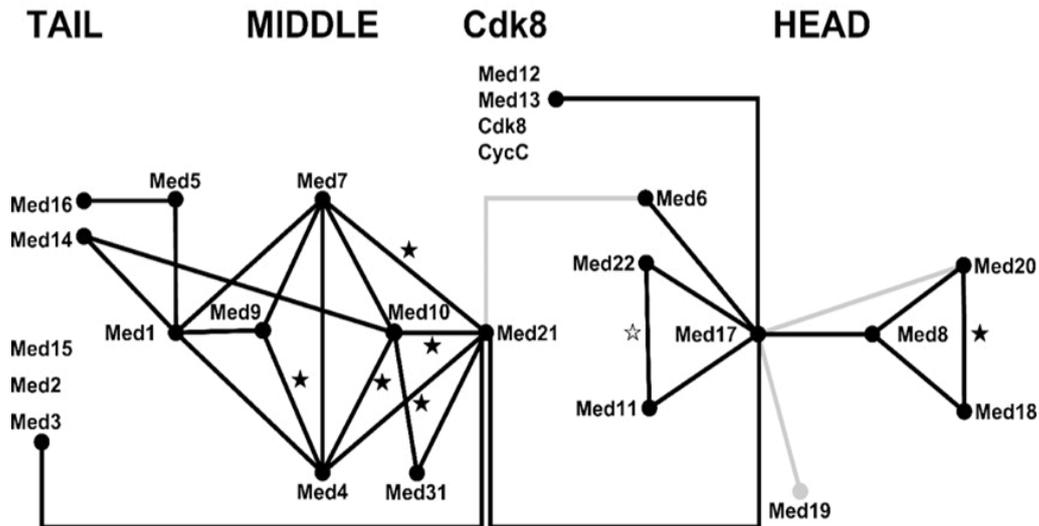


Figure 9 : Carte des interactions entre les sous-unités du Médiateur (Guglielmi et al., 2004).

Du fait de sa grande taille, de sa flexibilité intrinsèque importante et de sa faible abondance, aucune structure cristallographique du complexe Médiateur dans son ensemble n'a jusqu'à ce jour pu être obtenue. Cependant, le modèle structural de plusieurs sous-complexes ainsi que celle du module de tête dans son intégralité sont désormais disponibles (Figure 10). Plusieurs des sous-complexes caractérisés structuralement représentent des sous-modules fonctionnels, connectés au reste de la structure par des liaisons flexibles, et qui auraient des fonctions différentes, impliquées dans la régulation de différents groupes de gènes (Koschubs et al., 2009; Larivière et al., 2008). Ainsi, les sous-unités Med8-18-20 forment un sous-complexe important pour la formation du PIC, notamment via une interaction directe entre Med8 et la TBP (Larivière et al., 2006). De la même façon, le sous-complexe Med11-22 contient une zone très conservée essentielle à l'interaction avec TFIIH (Seizl et al., 2011). Une interaction directe entre la sous-unité Rad3 de TFIIH et Med11 a d'ailleurs été démontrée (Esnault et al., 2008). Le sous-complexe Med7-21 quant-à-lui joue un rôle architectural central au sein du module du milieu. Il possède plusieurs charnières flexibles capables d'engendrer des modifications conformationnelles importantes pour l'interaction avec la Pol II et d'autres facteurs (Baumli et al., 2005).

La structure du module de tête a récemment été obtenue à la résolution de 4,3 Å pour *S. cerevisiae* (Imasaki et al., 2011; Robinson et al., 2012) et 3.4 Å pour *S. pombe* (Larivière et al., 2012a) (Figure 10). Ce module est constitué de trois domaines distincts (Med6-8-11-22, Med17 et Med18-20), centrés sur un ensemble d'hélices appartenant à 5 sous-unités différentes. Cet ensemble hélicoïdal assure la stabilité des sous-unités de la tête et fournit des sites de liaison pour les GTFs et la Pol II. Le sous-module formé par Med18 et Med20 est mobile par rapport au reste de la structure et permettrait l'interaction avec les divers composants de la machinerie de transcription.

En associant la spectrométrie de masse au pontage protéique et en combinant les résultats obtenus avec les structures de sous-modules déjà connus, le laboratoire de P. Cramer a également résolu la structure du module du milieu de *S. cerevisiae* (Larivière et al., 2013). Ce module présente une forme très allongée (Figure 10), compatible avec l'idée que la surface d'interaction entre le Médiateur et la Pol II est très étendue. Il présente également plusieurs éléments très flexibles, sans doute importants pour l'interaction avec la Pol II. L'architecture de ce module est bien conservée entre la levure et l'homme.

Des analyses par microscopie électronique ont permis d'observer la structure globale du complexe Médiateur chez *S. cerevisiae* (Cai et al., 2009) et l'homme (Taatjes et al., 2002), cependant la faible résolution obtenue lors de ces études ne permettait pas de localiser les différentes sous-unités, ni même de positionner avec certitude les trois modules principaux. Dernièrement, une analyse par microscopie électronique réalisée par les laboratoires de F. Asturias et G. Cai ont permis de préciser la localisation des 25 sous-unités du Médiateur de *S. cerevisiae* et de comprendre comment les différents modules s'organisent entre eux (Tsai et al., 2014; Wang et al., 2014). Le module de tête est relativement compact, et sa liaison avec le reste du complexe est en grande partie réalisée par sa sous-unité Med17. Le module du milieu est plus étendu, avec Med1 et Med19 à chaque extrémité. Il interagit avec le module de tête via ses sous-unités Med7 et Med31. Enfin, le module de queue est constitué de plusieurs sous-complexe : Med5-Med16 à son extrémité distale, Med2-Med3-Med15 qui forme un coude qui relie l'hétérodimère Med5-Med16 au reste du complexe, et Med14 situé au centre du complexe et qui connecte les 3 modules du Médiateur. Cette étude fournit une vue relativement précise et complète de l'architecture du complexe Médiateur dans laquelle la position des différents modules est totalement redéfinie. Le module de tête préalablement positionné à la base de la structure, surmonté par les modules du milieu et de queue, se retrouve désormais adjacent au module du milieu, la base correspondant finalement au module de queue. Les auteurs ont ensuite comparé la structure obtenue chez *S. cerevisiae* avec celle du Médiateur humain (cf. III. de ce chapitre) et concluent en une très forte conservation de l'organisation des sous-unités du Médiateur chez les eucaryotes.

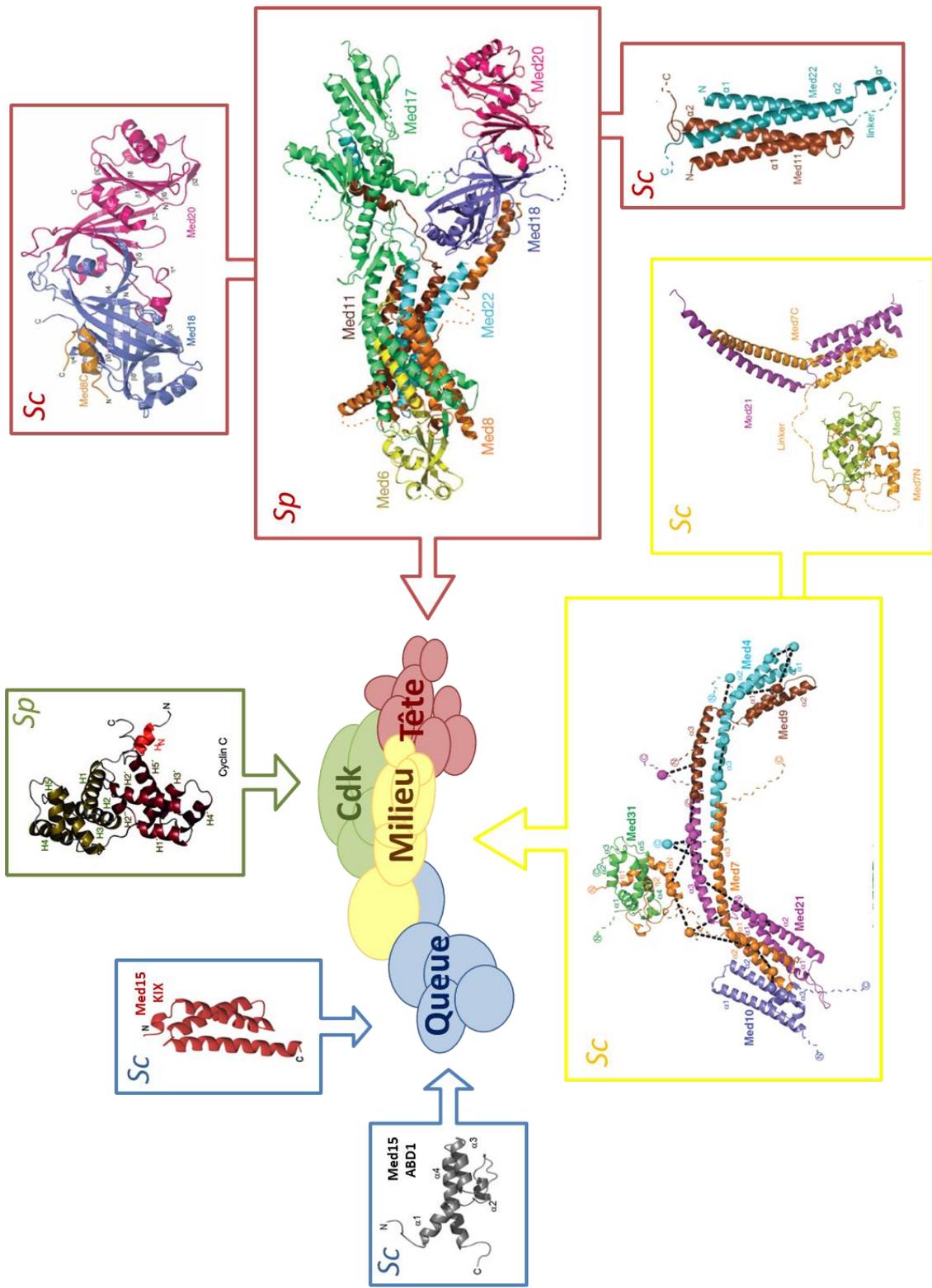


Figure 10 : Structures cristallographiques des sous-complexes et des modules de tête et du milieu du Médiateur chez *S. cerevisiae* et *S. pombe*.

Module de tête (Larivière et al., 2012), Module du milieu (Larivière et al., 2013), Sous-module Med8-18-20 (Larivière et al., 2006), Sous-module Med11-22 (Seizl et al., 2011), Sous-module Med7-21-31 (Koschubs et al., 2009), Med15 (Brzovic et al., 2011; Thakur et al., 2008), Cyclin C (Hoepfner et al., 2005).

Plusieurs études ont suggéré que les changements de conformation sont essentiels au fonctionnement du Médiateur. En superposant les structures cristallographiques des différents sous-complexes du Médiateur résolues à ce jour avec l'image obtenue par microscopie électronique, le laboratoire de F. Asturias a montré que la structure de chaque module est très stable (Tsai et al., 2014). Les changements de conformation du Médiateur ne seraient donc obtenus que par des changements de position des différents modules. Ces réarrangements sont importants pour l'interaction entre le Médiateur et la Pol II (cf. IV. 2. A. de ce chapitre).

II. LE MÉDIATEUR DES MAMMIFÈRES

1. Découverte du Médiateur chez les mammifères

Le laboratoire de R. Roeder est le premier à avoir identifié chez l'homme un complexe aux propriétés semblables à celle du Médiateur de *S. cerevisiae* (Fondell et al., 1996). Il fut nommé TRAP (« thyroid hormone receptor associated protein ») car isolé sous une forme liée au récepteur de l'hormone thyroïdienne. Ce co-activateur est capable de favoriser *in vitro* l'activation de la transcription en absence de TAFs (Oelgeschläger et al., 1998). Un complexe semblable a pu être purifié chez la souris (Jiang et al., 1998) et le rat (Brower et al., 2002).

Suite à cette découverte, plusieurs groupes se sont intéressés à la caractérisation de complexes Médiateurs favorisant la transcription *in vitro*. Les complexes CRSP (« cofactor required for Sp1 activation »), PC2 (« positive cofactor 2 »), DRIP (« vitamin D receptor-interacting proteins »), SMCC (« suppressor of RNA polymerase B (SRB) Mediator containing cofactor ») et NAT (« negative regulator of activated transcription ») ont pu être purifiés à partir de cellules HeLa (Gu et al., 1999; Malik et al., 2000; Rachez et al., 1998; Ryu et al., 1999; Sun et al., 1998). Des analyses par chromatographie d'affinité avec divers domaines d'activation de la transcription ont enfin permis d'identifier les complexes ARC (« activator-recruited cofactor ») et « Médiateur humain » (Boyer et al., 1999; Näär et al., 1999).

2. Composition du Médiateur des mammifères

L'analyse des protéines présentes dans les complexes Médiateurs des mammifères a permis de mettre en évidence une trentaine de sous-unités (Figure 11). Des études approfondies ont permis de montrer qu'il existe de nombreuses similitudes entre les sous-unités du Médiateur des mammifères et de la levure, faisant de ce dernier un complexe universel et bien conservé (cf. III. de ce chapitre).

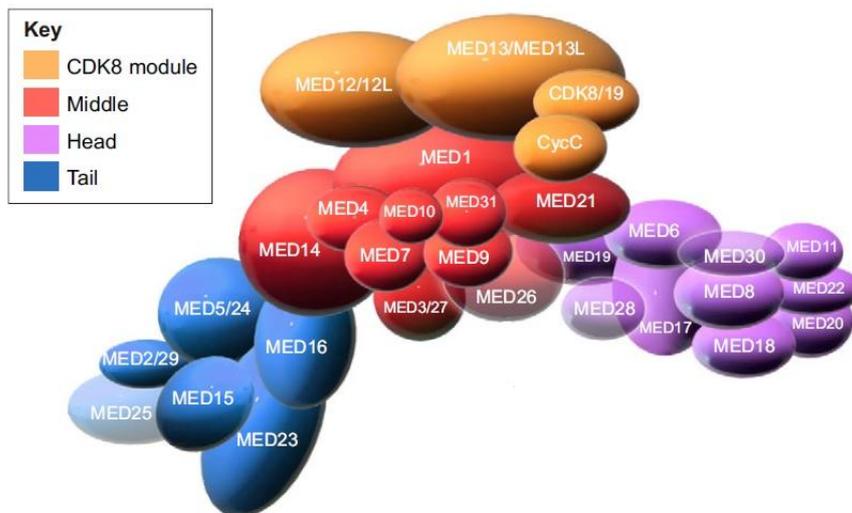


Figure 11 : Structure modulaire du complexe Médiateur des mammifères (Yin and Wang, 2014)

Ce schéma se base sur les données d'interactions et quelques structures cristallographiques publiées mais la localisation de chaque sous-unité n'est pas forcément exacte. Les positions de Med25, Med26, Med28 et Med30 sont hypothétiques (Larivière et al., 2012b), d'où leur représentation transparente.

Les complexes Médiateurs purifiés chez les mammifères par plusieurs laboratoires présentent des compositions peptidiques différentes. On distingue ainsi (1) des gros complexes de 1-2 MDa tels que TRAP/SMCC, DRIP et ARC, qui contiennent tout ou partie du module kinase ; (2) des plus petits complexes de 500-700 kDa tels que PC2, CRSP et NAT, dans lesquels le module kinase est absent (Conaway et al., 2005). Des différences en termes de composition en sous-unités apparaissent également entre ces deux groupes. Celles-ci pourraient être liées à la régulation tissu-spécifique de l'expression génique, et donc à des formes fonctionnelles du Médiateur, distinctes selon les cellules. Ainsi, les complexes Médiateurs contenant le module kinase ont peu d'effet, voire inhibent, la transcription *in vitro* des gènes activés par VP16, SREBP et Sp1 (Akoulitchev et al., 2000; Sun et al., 1998; Taatjes et al., 2002). A l'inverse, les complexes ne contenant pas le module kinase mais contenant la sous-unité Med26 entraînent une forte activation de la transcription par ces mêmes activateurs (Taatjes et al., 2002). Cependant, une étude de protéomique réalisée par le laboratoire de R.C. Conaway plaide pour l'existence d'un Médiateur de mammifère uniforme et suggère que les différences de composition observées entre les complexes seraient en majorité dues à l'utilisation de méthodes de purification trop peu sensibles et/ou à la perte de certaines sous-unités lors des étapes de purification. L'analyse MudPIT (« MUltiDimensional Protein Identification Technology ») utilisée pour cette étude, évite les ambiguïtés d'identification des protéines associées à la migration par électrophorèse. Au cours de cette analyse, toutes les sous-unités des complexes préalablement identifiés ont été détectées,

proposant ainsi un consensus où le Médiateur humain apparaît comme un complexe de 30 sous-unités (Sato et al., 2004).

Conservation évolutive du complexe Médiateur

Les premiers alignements de séquence n’ont permis d’identifier dans les complexes Médiateurs humains que 8 homologues des sous-unités du Médiateur de *S. cerevisiae* : Med14, Med4, Med6, Med7, Med10, Med17, Med21, Cdk8, CycC et Med31 (Rachez and Freedman, 2001). Cette constatation laissait suggérer que le Médiateur des mammifères était très éloigné en termes de structure et de fonction de son homologue chez la levure.

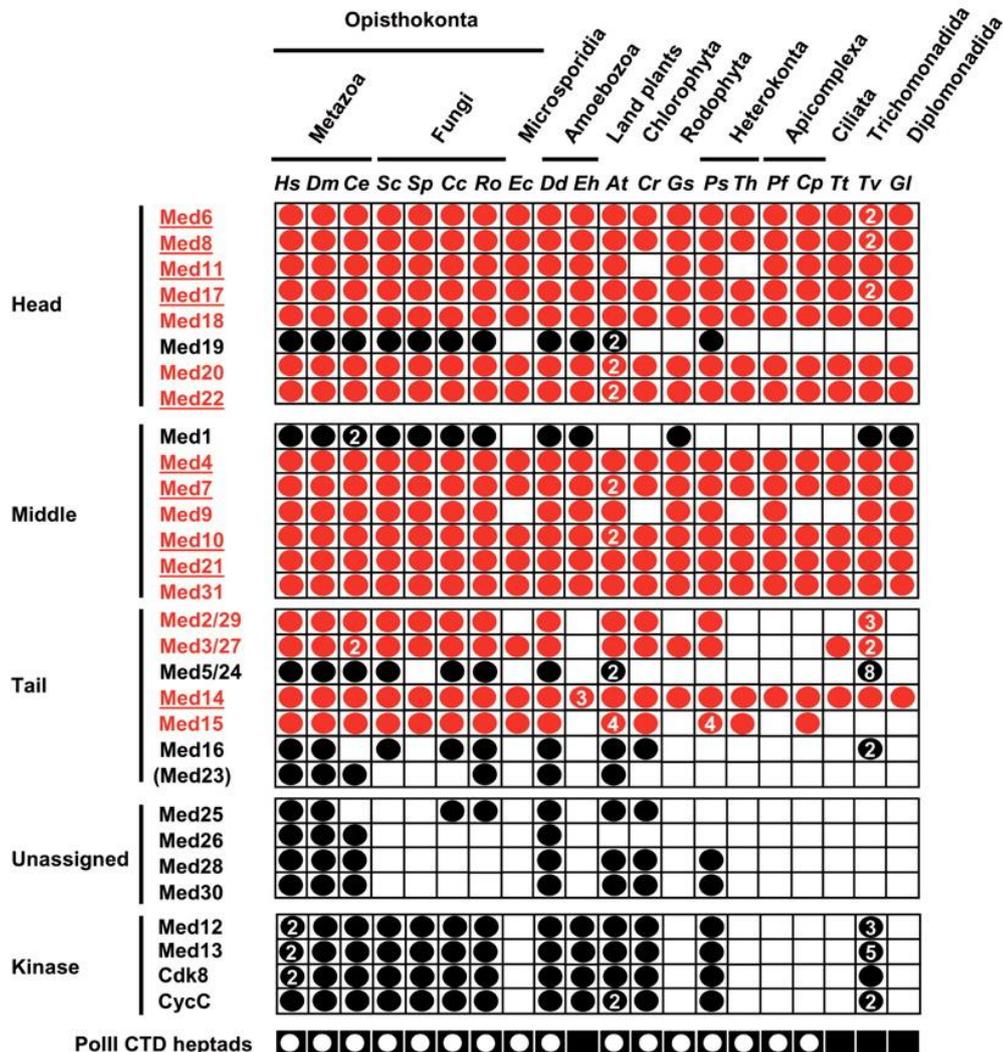


Tableau 8 : Conservation des sous-unités du Médiateur chez les eucaryotes (Bourbon, 2008).

Les sous-unités sont groupées selon le module auquel elles appartient conformément au modèle décrit chez *S. cerevisiae*. Celles qui sont essentielles à la viabilité cellulaire sont soulignées. Les points indiquent qu’au moins un homologue est présent dans le génome analysé, si plusieurs paralogues ont été détectés, leur nombre est indiqué. Les sous-unités les plus conservées sont indiquées en rouge. Celles-ci incluent les sous-unités essentielles pour la viabilité cellulaire chez la levure.

Plus tard, une approche de bioinformatique systématique exploitant des alignements de séquences multiples des sous-unités du Médiateur dans différentes espèces a permis d'identifier de nombreuses régions similaires entre les sous-unités des Médiateurs humains et de levure (Boube et al., 2002). Seules les sous-unités Med2, Med3 et Med5 du module de queue n'ont pu être détectées que chez la levure. Cependant, des analyses plus récentes se basant sur la présence de motifs similaires au sein des séquences ainsi que sur la taille des séquences, laissent suggérer que ces sous-unités seraient des homologues de Med29, Med27 et Med24, respectivement (Bourbon, 2008).

L'ensemble de ces études a permis de suggérer que la majorité des sous-unités de *S. cerevisiae* ont un homologue chez les métazoaires (Tableau 8). On constate que les sous-unités essentielles chez *S. cerevisiae* possèdent un homologue chez la majorité des eucaryotes étudiés. Une nomenclature commune à tous les complexes a ainsi pu être établie (Bourbon et al., 2004). Cependant, le degré d'homologie et la longueur des domaines restent très variables d'une sous-unité à l'autre.

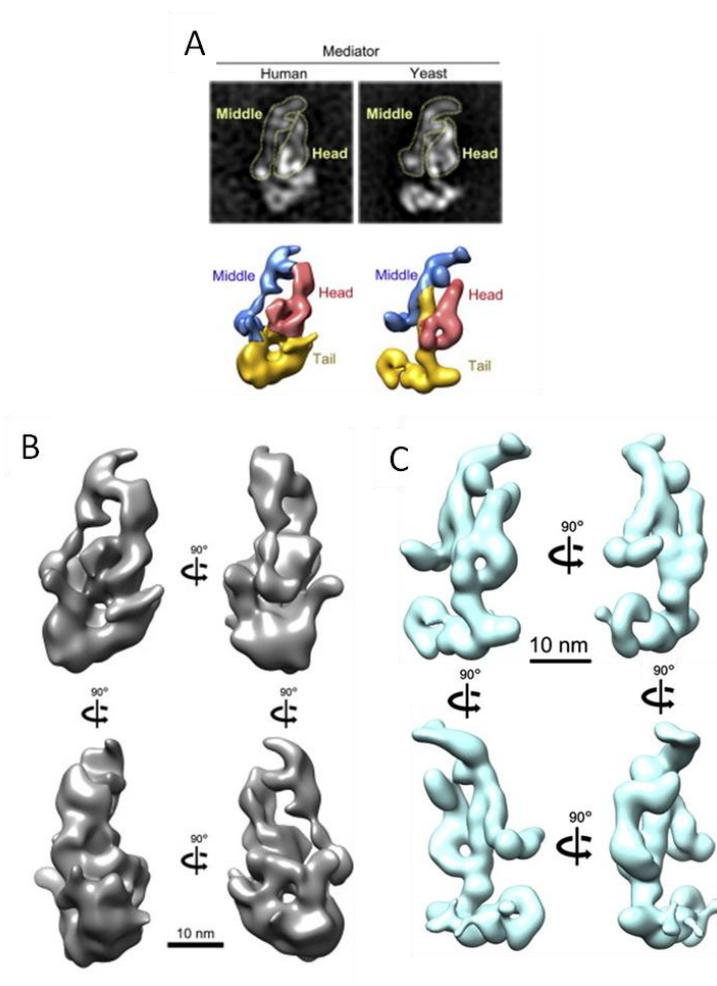


Figure 12 : Comparaison de la structure du Médiateur de *S. cerevisiae* et de l'homme (Tsai et al., 2014).

(A) Images des Médiateurs de levure et humain obtenus par microscopie électronique (EM) et comparaison des structures résultantes.

(B) Modèle 3D du Médiateur humain calculé à partir des images EM à la résolution estimée de 3 nm.

(C) Modèle 3D du Médiateur de levure calculé à partir des images EM à la résolution estimée de 1,8 nm.

Des visualisations en microscopie électronique ont également permis de comparer la structure du Médiateur de *S. cerevisiae* et de l'homme (Figure 12). Les images obtenues montrent une architecture globale similaire et une conservation de la position des différents modules. La forme des modules de tête et du milieu est très semblable chez les deux organismes suggérant que l'organisation du Médiateur est conservée de la levure à l'homme (Tsai et al., 2014). Ce maintien de la structure globale du complexe, malgré les divergences en termes de séquences des diverses sous-unités, pourrait s'expliquer par une homologie de séquences plus importante dans les régions impliquées dans les contacts entre les différentes sous-unités (Guglielmi et al., 2004).

III. ROLE DU MÉDIATEUR DANS L'INITIATION DE LA TRANSCRIPTION

Un grand nombre de fonctions dans l'initiation de la transcription ont pu être attribuées au Médiateur. Ce complexe est suffisamment large pour interagir à la fois avec les activateurs, les co-activateurs, les GTFs et la Pol II. Il sert alors d'adaptateur en transmettant les signaux de régulation des activateurs vers la machinerie de transcription et stimule la mise en place du PIC. Ce complexe est également capable de stimuler la transcription basale *in vitro* ainsi que la phosphorylation du CTD par TFIIF.

1. Interaction du Médiateur avec les activateurs

La fixation des activateurs sur les UAS induit un signal de régulation positive nécessaire à l'initiation de la transcription. Le Médiateur est capable d'interagir avec plusieurs de ces activateurs (Tableau 9) pour transmettre le signal à la machinerie de transcription. L'interaction se fait entre les TADs (« transactivation domains ») des activateurs et l'ABD (« activator binding domain ») du Médiateur.

Des expériences d'inactivation des gènes du module de queue chez *S. cerevisiae* ont abouti à une inhibition de la liaison avec les différents activateurs et une incapacité à stimuler la transcription activée *in vivo* tandis que la stimulation de la transcription basale demeure (Lee et al., 1999). À l'inverse, une inactivation d'une sous-unité n'appartenant pas au module de queue entraîne une diminution à la fois de la transcription activée et basale sans cependant inhiber l'interaction avec les activateurs. Le module de queue joue donc un rôle essentiel dans la liaison du Médiateur aux activateurs de la transcription. *In vitro*, les activateurs Gal4, VP16 et Gnc4 sont notamment capables d'interagir directement avec Med15 (Park et al., 2000) et ce même en l'absence des modules du milieu et de queue. Cependant, la sous-unité Med17 du module de tête

est également capable d'interagir avec l'activateur Gal4 (Koh et al., 1998), le module de queue n'est donc pas le seul impliqué dans ces interactions.

Chez les mammifères, la purification du premier complexe Médiateur associé au récepteur de l'hormone thyroïdienne (TR) a permis de suggérer que le Médiateur serait une interface entre les activateurs et la machinerie de transcription par la Pol II (Fondell et al., 1996). Depuis lors, l'étude des récepteurs nucléaires (NR) a permis de grandes avancées dans la compréhension du fonctionnement du Médiateur. La principale cible de cette famille d'activateurs est Med1, qui appartient au module du milieu, mais se trouve très proche de la jonction avec le module de queue (Guglielmi et al., 2004). Cette sous-unité contient deux boîtes NR (« nuclear receptor box »), caractérisées par la présence du motif très conservé LXXLL, qui lui permettent d'interagir avec le domaine AF2 (« activation function 2 ») des récepteurs nucléaires tels que VDR (« vitamin D receptor ») (Rachez et al., 1999), PPAR γ (« peroxisome proliferator-activated receptors ») (Ge et al., 2002), HNF4 (« hepatocyte nuclear factor 4 ») (Malik et al., 2002), GR (« glucocorticoid receptor ») (Hittelman et al., 1999) et ER (« estrogen receptor ») (Kang et al., 2002). Enfin, d'autres sous-unités du Médiateur sont également la cible de récepteurs nucléaires, ainsi Med14 interagit avec GR (Hittelman et al., 1999), HNF4 (Malik et al., 2002) et PPAR γ (Grøntved et al., 2010).

De nombreuses voies impliquées dans le développement sont sous le contrôle des sous-unités du module kinase : Med12, Med13, Cdk8 et CycC. Chez la drosophile, Med12 et Med13 interviennent dans la voie de signalisation Wnt importante pour le développement embryonnaire, via une interaction avec la β -caténine (Carrera et al., 2008). Chez Arabidopsis, Med25 intervient dans la floraison, la croissance des organes et la réponse au stress. Il est impliqué dans la régulation des gènes jasmonate-dépendants via son interaction avec les facteurs AP2/ERF59 (« Apetala2/Ethylene response factor 59 »), ERF1 (« Ethylene response factor 1 ») et MYC2 (Çevik et al., 2012). Chez l'homme, Med25 interagit avec les activateurs de la famille PEA3, impliqués dans la croissance et la différenciation cellulaire et également dans la tumorigenèse (Verger et al., 2013). Med15 interagit avec les activateurs SMAD2-4 et SMAD3-4 en réponse aux cytokines TGF β (« transforming growth factor »), impliquées dans le développement et la tumorigenèse (Kato et al., 2002).

Sous-unité	TF	Organisme	Fonction
Med1	TR	Homme	Récepteur de l'hormone thyroïdienne
	VDR	Homme	Récepteur de la vitamine D
	HNF4	Homme	Développement du foie
	ER	Homme	Récepteurs aux oestrogènes
	P53	Homme	Régulation du cycle cellulaire, suppresseur de tumeur
	GR	Homme	Récepteur aux glucocorticoïdes
	STAT2	Homme	Croissance, survie et différenciation en réponse à l'interféron
	GATA1	Souris	Régulateur majeur de l'érythropoïèse
	PPARγ	Souris	Régulateur majeur de l'adipogenèse
Med12	REST	Homme	Différenciation neuronale
	Sox9	Homme	Régulateur majeur de la chondrogenèse
	NANOG	Homme	Régulateur majeur de la pluripotence des cellules ES
	β-catenin	Drosophile	Développement embryonnaire dépendant de la voie Wnt
Med14	PPARγ	Homme	Régulateur majeur de l'adipogenèse
	HNF4	Homme	Développement du foie
	GR	Homme	Récepteur aux glucocorticoïdes
Med15	Gal4	<i>S. cerevisiae</i>	Réponse au galactose
	Gcn4	<i>S. cerevisiae</i>	Réponse à une carence en acides aminés
	Oaf1	<i>S. cerevisiae</i>	β -oxydation des acides gras
	Msn2	<i>S. cerevisiae</i>	Réponse au stress
	Pdr1	<i>S. cerevisiae</i>	Régulateur majeur pour les gènes de résistance aux drogues
	SREBP	Homme	Homéostasie des lipides
	SMAD2-3	Homme	Développement du mésenchyme en réponse à TGF β
	SBP-1	<i>C. elegans</i>	Homologue de SREBP1, impliqué dans le métabolisme des lipides
	NHR-49	<i>C. elegans</i>	Homologue de HNF4, impliqué dans le métabolisme des lipides et la survie cellulaire
Med17	VP16	HSV	TF dérivé du HSV
	P53	Homme	Régulation du cycle cellulaire, suppresseur de tumeur
	HSF	Drosophile	Réponse à un choc thermique
Med23	E1A	Homme	Gène exprimé durant la réplication des adénovirus
	ELK1	Homme	Adipogenèse insuline-dépendante
	C/EBP	Homme	Prolifération cellulaire, croissance et différenciation
Med25	VP16	HSV	TF dérivé du HSV
	IE62	Homme	TF associé au tégument des virions du VZV
	HNF4	Homme	Développement du foie
	Sox9	Homme	Régulateur majeur de la chondrogenèse
	PEA3	Homme	Croissance cellulaire et différenciation
	AP2/ERF59	<i>A. thaliana</i>	Défense contre les pathogènes en réponse à l'acide jasmonique
	ERF1	<i>A. thaliana</i>	Défense contre les pathogènes en réponse à l'éthylène
	MYC2	<i>A. thaliana</i>	Régulateur majeur de la résistance aux pathogènes
	ABI5	<i>A. thaliana</i>	Défense contre les pathogènes en réponse aux acides jasmonique et absidique
	DREB2A	<i>A. thaliana</i>	Réponse au stress

Tableau 9 : Interaction des sous-unités du Médiateur avec les facteurs de transcription (d'après Ansari and Morse, 2013; Malik and Roeder, 2010)

Le Médiateur intervient dans de nombreuses voies métaboliques. Chez l'homme, Med25 joue un rôle important dans le métabolisme des xénobiotiques et des lipides via son interaction avec HNF4 (Rana et al., 2011). Med15 interagit avec SREBP1 α (« sterol regulatory element-binding protein ») qui joue un rôle dans l'homéostasie du cholestérol (Yang et al., 2006). L'interaction se fait via le domaine KIX de Med15. Ce domaine est conservé chez *C. elegans* et lui permet d'interagir avec SBP-1, l'homologue de SREBP1. Chez *S. cerevisiae* aussi, Med15 régule le niveau d'acide gras dans la cellule via une interaction avec le facteur Oaf1 (« Oleate-activated transcription factor ») (Thakur et al., 2009). Le Médiateur est également impliqué dans la voie de signalisation MAPK (« Mitogen-activated protein kinases ») via une interaction entre Med23 et le facteur ELK1 qui régule le gène *Egr2* (« early growth response 2 ») impliqué dans l'adipogenèse insuline-dépendante (Wang et al., 2009).

Des analyses récentes ont mis en évidence l'existence de régulateurs majeurs (« master regulator »), déterminants des programmes de transcription spécifiques des lignées cellulaires. Beaucoup interagissent avec des sous-unités du Médiateur, par exemple Med1 qui interagit avec le régulateur majeur de l'érythropoïèse GATA1 (Stumpf et al., 2006), ou Med12 qui interagit avec Nanog, régulateur majeur de la pluripotence ou encore Med25 qui interagit avec Sox9, régulateur majeur de la chondrogenèse (Nakamura et al., 2011). Les travaux réalisés par le laboratoire de R. Young ont suggéré un lien entre l'organisation des régulateurs majeurs et le complexe Médiateur. Ce dernier serait capable de promouvoir la formation de domaines « super enhancers » qui correspondent à des groupes d'« enhancers » occupés à la fois par le Médiateur et les régulateurs (Whyte et al., 2013). Ces domaines contrôlent les gènes clés de l'identité cellulaire dans différents types cellulaires telles que les cellules ES, les lymphocytes T, les macrophages... On les retrouve également associés à des cellules cancéreuses (Lovén et al., 2013). Certains TFs, notamment Klf4 et Esrrb, importants pour l'expression des gènes des cellules ES et la reprogrammation des cellules somatiques en cellules souches pluripotentes, sont très enrichis au niveau de ces domaines. De plus, les gènes associés aux « super enhancers » sont très sensibles à une modification du taux de liaison des activateurs et co-activateurs. Ainsi, une diminution de la quantité de Médiateur ou des régulateurs majeurs au niveau des « super enhancers » induit une diminution de l'expression des gènes spécifiques à une lignée. Cette particularité serait à l'origine de la différenciation des cellules ES, en facilitant la transition vers un nouveau programme d'expression génique. En participant à la formation de ces « super enhancers », le Médiateur serait ainsi un coordinateur majeur (« master coordinator ») du développement et de la détermination du destin cellulaire.

La liaison aux activateurs va induire des changements conformationnels du Médiateur qui vont lui permettre d'interagir préférentiellement avec certains co-activateurs. Med15 possède 3

sites de liaisons pour l'activateur Gcn4. Les analyses structurales de l'interaction entre ces deux protéines, réalisées par le laboratoire de S. Hahn, montrent que celle-ci peut se faire selon différentes orientations (Brzovic et al., 2011). Ainsi, Gcn4 peut rapidement passer d'un domaine de liaison à un autre et ce mécanisme serait important pour permettre le recrutement du Médiateur mais également pour autoriser l'interaction de Gcn4 avec d'autres co-activateurs. Cette interaction modulable permet une adaptabilité et une réversibilité des interactions protéine-protéine, importante pour la fonction biologique de nombreux domaines activateurs. Des analyses MudPIT (« Multidimensional protein identification technology ») ont montrées que beaucoup de protéines associées au Médiateur lié à SREBP sont absentes de la forme liée à VP16 et inversement. De même, l'association entre le complexe TRAP et l'histone acétyltransférase Gcn5 ne peut se faire que lorsque le complexe TRAP est lié à un activateur (Ebmeier and Taatjes, 2010). Ainsi, les signaux de régulation des différentes voies de signalisation, impliquées par exemple dans la croissance cellulaire, le développement, la différenciation, le métabolisme... convergent vers le Médiateur à partir des activateurs. Le Médiateur intègre ces signaux et induit une réponse transcriptionnelle appropriée, notamment en recrutant les co-activateurs requis. On comprend dès lors que le Médiateur puisse être impliqué dans des pathologies aussi variées que des anomalies du développement, des cancers, des désordres métaboliques... (cf. VII. de ce chapitre).

2. Mise en place du PIC

A. Interaction du Médiateur avec la Pol II

Chez la levure, le Médiateur peut être purifié sous deux formes : libre ou associé à la Pol II (Kim et al., 1994). Chez les mammifères, cette seconde forme n'a encore jamais été purifiée, cependant elle a pu être reconstituée *in vitro* (Bernecky et al., 2011). La sous-unité Med26, spécifique du Médiateur des mammifères, est très enrichie dans la forme associée à la Pol II. Elle joue un rôle important dans l'activation de la transcription, à l'inverse du module kinase qui est absent lorsque le Médiateur est complexé à la Pol II (Malik et al., 2000; Taatjes et al., 2002). La liaison du Médiateur à la Pol II engendre un changement conformationnel important (Asturias et al., 1999) qui offre une structure ouverte dans laquelle l'interaction du Médiateur avec les autres partenaires est facilitée. Cet assemblage laisse libre 75% de la surface de la Pol II.

Plusieurs modèles de l'interaction Médiateur – Pol II (Med-Pol II) ont pu être obtenus par microscopie électronique (Figure 13). Chez la levure, la Pol II est placée avec les sous-unités Rpb3 et Rpb11 faces au Médiateur, mais avec deux orientations différentes possibles qui correspondent à une rotation de la Pol II d'environ 90° (Cai et al., 2009; Davis et al., 2002). Cette différence

pourrait être due à l'absence des GTFs et des activateurs qui en temps normal favoriseraient l'interaction dans une position donnée. Dans tous les cas, le modèle du complexe Med-Pol II montre que le Médiateur interagit via ses modules de tête et du milieu, sur une surface étendue de la Pol II, à l'opposé du canal d'entrée de l'ADN.

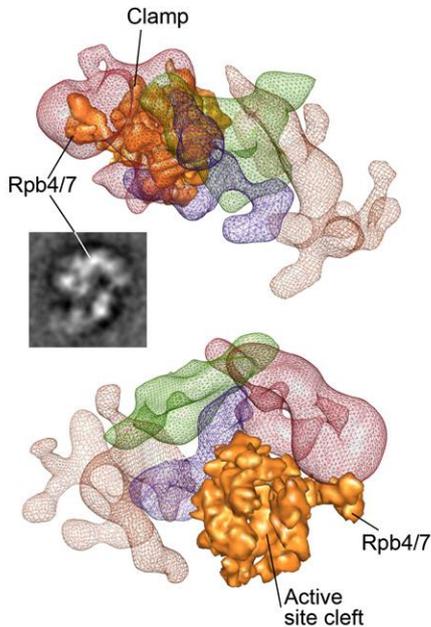


Figure 13 : Modèle de l'interaction entre la Pol II et le Médiateur, obtenue par microscopie électronique (Cai et al., 2009).

Les derniers travaux du laboratoire de FJ. Asturias par microscopie électronique proposent un modèle de l'interaction du Médiateur avec la Pol II (Tsai et al., 2014). La première étape consisterait en une liaison du CTD de la Pol II au niveau de la jointure entre les modules de tête et du milieu du Médiateur. Cette interaction faciliterait le changement de conformation de la tête requis pour la formation du complexe Med-Pol II. Cette structure pourrait ensuite être modifiée par les phosphorylations du CTD au moment de l'entrée en élongation. Cette observation est en accord avec les travaux de R. Kornberg sur le modèle de l'interaction entre le module de tête du Médiateur et le CTD obtenu par cristallographie qui montre au moins 4 points de contacts répartis le long du CTD (Robinson et al., 2012). De plus, des analyses *in vitro* à partir du Médiateur purifié de *S. cerevisiae* ont montré qu'en absence du CTD de la Pol II, le complexe Med-Pol II ne peut pas se former (Myers et al., 1998). Des analyses par chromatographie d'affinité dans des cellules HeLa ont également mis en évidence une interaction entre le Médiateur et le CTD importante pour le recrutement de la Pol II au promoteur (Näär et al., 2002). La liaison du Médiateur aux activateurs va également induire une rotation du module de queue qui facilite encore l'interaction avec la Pol II. La formation du complexe Med-Pol II se ferait donc en plusieurs étapes qui vont mener au modèle d'interaction final tel que décrit par les analyses EM. Cependant, ce modèle n'est pas forcément représentatif de la conformation du complexe Med-Pol II *in vivo*, et les contacts du Médiateur avec d'autres facteurs pourraient être importants pour stabiliser la Pol II.

Plusieurs contacts directs entre des sous-unités du Médiateur et de la Pol II ont été proposés au cours des dernières années. Tout d'abord, les premières études par microscopie électronique avaient permis de prédire les sous-unités de la Pol II impliquées dans cette interaction comme étant Rpb1, Rpb2, Rpb3, Rpb6 et Rpb11 (Davis et al., 2002). Un modèle plus récent, laisse quant à lui suggérer une interaction entre la tête du Médiateur, notamment les sous-unités Med18 et Med20, et les sous-unités Rpb4, Rpb7 ainsi que la pince de la Pol II (Cai et al., 2009, 2010). Cette interaction induit un changement de conformation de la Pol II rendant le site actif accessible à l'ADN et augmentant ainsi l'activité transcriptionnelle (Cai et al., 2012). Finalement, une analyse par photo-pontage *in vivo* chez *S. cerevisiae*, réalisée dans notre laboratoire, a permis de mettre en évidence une interaction directe entre la sous-unité Med17 du module de tête du Médiateur et Rpb3 requise pour la transcription de l'ensemble des gènes de classe II (Soutourina et al., 2011). La conservation de la sous-unité Rpb3 de la Pol II et de la sous-unité Med17 du Médiateur suggère que les contacts entre ces deux complexes seraient conservés chez tous les organismes. Cette étude a montré, grâce à l'utilisation de mutants conditionnels de Med17 et Rpb3, que l'interaction entre ces deux protéines est indispensable au recrutement de la Pol II sur le génome de la levure et à la mise en place du PIC *in vivo*. Ces résultats suggèrent que la majorité de l'information de régulation est transmise par l'interface entre le Médiateur et la Pol II.

Une étude *in vitro*, réalisée par le laboratoire de R.G. Roeder, a récemment mis en évidence deux formes de la Pol II chez les mammifères, l'une contenant le polypeptide Gdown1, l'autre non, la première forme étant celle retrouvée en abondance. Le polypeptide Gdown1 inhibe l'activation de la transcription en empêchant TFIIF de se lier à la Pol II (Jishage et al., 2012). La Pol II contenant ce polypeptide dépend du Médiateur pour son activation et répond à ce dernier. Ainsi, en l'absence du Médiateur, la réponse aux activateurs est moindre. L'ajout du Médiateur permet alors d'annuler l'effet inhibiteur de Gdown1 et de restaurer la transcription activateur-dépendante. A l'inverse, la Pol II ne contenant pas Gdown1 est capable de répondre très fortement aux activateurs même en l'absence du Médiateur et l'ajout de ce dernier n'a que très peu d'effet. Gdown1 aurait donc pour rôle d'empêcher une activation spontanée de la transcription (Hu et al., 2006).

B. Interaction du Médiateur avec les facteurs généraux de la transcription

Le Médiateur étant capable *in vitro* de stimuler à la fois la transcription basale et activée, il est logique de penser qu'il influence la mise en place du PIC en interagissant avec les GTFs. Des études impliquant des mutations dans les sous-unités du Médiateur montrent d'ailleurs une diminution de l'interaction avec les GTFs (Esnault et al., 2008; Takagi and Kornberg, 2006). Ces interactions se font principalement au niveau du module de tête (Takagi et al., 2006).

Récemment, un modèle structural de l'interaction entre le Médiateur et les GTFs a été obtenu chez l'homme (Figure 14). Ce modèle est en accord avec les études de l'interaction de la Pol II avec les GTFs puisque la surface laissée libre dans la structure Med-Pol II-TFIIF autorise l'interaction de la Pol II avec TFIIE, TFIIB mais également avec l'ADN. TFIIF joue un rôle important dans le positionnement de la Pol II humaine sur le Médiateur. Cette observation pourrait expliquer les différentes orientations de la Pol II dans les complexes Med-Pol II décrits chez la levure, TFIIF n'ayant pas été inclus dans ces études. Le modèle obtenu montre que le Médiateur constitue une plateforme d'assemblage pour les GTFs et la Pol II et coordonne l'assemblage du PIC (Bernecky et al., 2011).

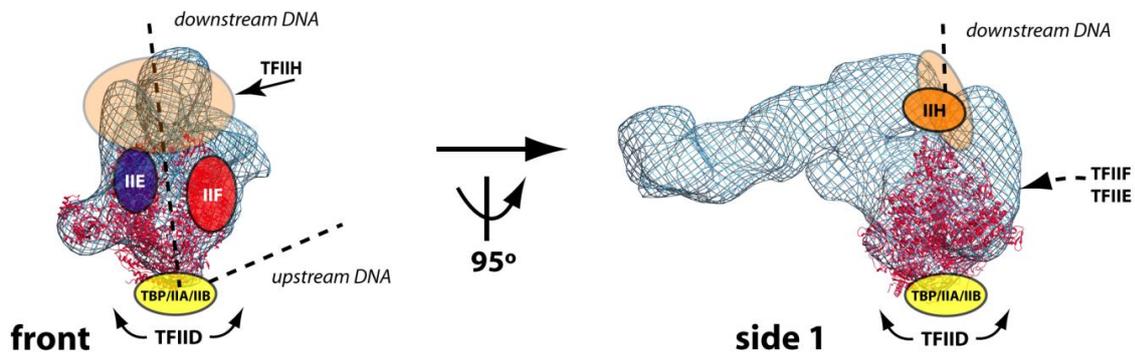


Figure 14 : Modèle structural du PIC humain obtenu par microscopie électronique (Bernecky et al., 2011).

Le modèle EM du complexe Médiateur – Pol II – TFIIF est indiqué en maillage bleu et la structure de la Pol II en rouge. La localisation des GTFs est en surimpression sur le modèle d'après les données publiées de cristallographie, microscopie électronique et pontage.

a. TFIID

Le facteur TFIID est le premier à intervenir dans la formation du PIC. Des études à partir de complexes humains ont démontré que le Médiateur participe à son recrutement et ont notamment révélé un rôle synergique de ces deux complexes, chacun favorisant la fixation de l'autre sur la chromatine (Baek et al., 2002; Johnson et al., 2002). Plus récemment, une analyse MudPIT a permis de montrer une interaction directe entre le Médiateur et TFIID via la sous-unité Med26 qui contient un site de liaison pour TFIID (Takahashi et al., 2011). Chez la levure, Med8 interagit avec la TBP *in vitro* et cette interaction est essentielle *in vivo* (Larivière et al., 2006). Plus généralement, le sous-module Med8-18-20 est important pour l'interaction avec la TBP et requiert la liaison du Médiateur avec un activateur. Ces observations ont pu être confirmées par la modélisation de la structure 3D du complexe Médiateur – TBP qui montre que la TBP se lie au niveau de l'interface entre les sous-modules Med11-17-22 et Med6-8-18. De plus, cette

interaction induirait un changement de conformation de la tête du Médiateur, importante pour son interaction avec la Pol II (Cai et al., 2010).

b. TFIIA

Aucune interaction directe entre TFIIA et le Médiateur n'a été démontrée, cependant le complexe TFIIID-TFIIA favorise *in vitro* le recrutement du Médiateur humain (Johnson et al., 2002). La structure TFIIID-TFIIA-Médiateur (DAMed) ainsi formée est suffisante pour constituer le PIC et servirait de plateforme à son assemblage. De plus, le pré-assemblage de ce complexe sur le promoteur permet de se passer de l'intervention d'un activateur (Johnson and Carey, 2003)

c. TFIIIB

Les modules de tête et du milieu de *S. cerevisiae*, plus précisément les sous-unités Med9, Med10 et Med17, sont connues pour interagir *in vitro* avec le facteur TFIIIB (Kang et al., 2001). De plus, la structure du complexe TFIIIB - Pol II a été résolue par cristallographie (Bushnell et al., 2004; Liu et al., 2010; Sainsbury et al., 2013). L'association de ce modèle avec celui du complexe Médiateur – Pol II suggère que TFIIIB se glisse dans l'espace laissé libre entre le Médiateur et la Pol II (Asturias, 2004). Enfin, des expériences *in vitro* ont montré que le Médiateur stimule la transcription basale en facilitant l'incorporation de TFIIIB au sein du PIC et qu'un excès de TFIIIB peut, en parti, contrebalancer l'absence de Médiateur dans ce système reconstitué (Baek et al., 2002, 2006).

d. TFIIIF

Lors de sa purification chez *S. cerevisiae*, la Pol II peut se trouver soit associée au Médiateur soit à TFIIIF mais jamais aux deux à la fois (Rani et al., 2004). Néanmoins, une étude du module de tête du Médiateur a révélé une interaction avec le complexe constitué par TFIIIF et la Pol II (Takagi et al., 2006). La modélisation de cet assemblage a permis de préciser le rôle de TFIIIF durant la mise en place du PIC. En absence de TFIIIF, l'orientation de la Pol II liée au Médiateur peut varier. A l'inverse, au sein du complexe Médiateur - Pol II - TFIIIF, la Pol II adopte une orientation stable, laissant suggérer un rôle pour TFIIIF dans la stabilisation de la liaison entre la Pol II et le Médiateur (Bernecky et al., 2011).

e. TFIIIE

La sous-unité Med15 du Médiateur interagit *in vitro* avec les deux sous-unités de TFIIIE (Sakurai et al., 1996). Il existerait également des contacts possibles entre le module du milieu et la grande sous-unité de TFIIIE via Med9 et Med10 (Kang et al., 2001). TFIIIE est important pour le

recrutement de TFIIH sur le promoteur. Cependant, des expériences *in vitro* ont montré que selon l'activateur, TFIIE peut ou non être requis. En son absence, TFIIH serait directement recruté par le Médiateur (Sakurai and Fukasawa, 2003).

f. TFIIH

TFIIH est le dernier facteur à intervenir au cours de l'initiation de la transcription *in vitro*, pour phosphoryler le CTD. Son interaction directe avec le Médiateur a pu être déterminée par double hybride et se fait par l'intermédiaire de la sous-unités Med11 du Médiateur et la sous-unité Rad3 du module cœur de TFIIH (Esnault et al., 2008). Une mutation de Med11 entraîne une diminution de l'occupation du promoteur par le module TFIK suggérant ainsi que le Médiateur stabiliserait l'association des sous-complexes de TFIIH en interagissant directement avec Rad3. Ce défaut de recrutement de TFIK empêche la stimulation de la phosphorylation du CTD par le Médiateur. La même étude a également défini qu'*in vivo*, l'ordre d'arrivée des GTFs n'est pas aussi séquentiel que ce qui a été décrit *in vitro*, et que le Médiateur est important pour coordonner la mise en place du PIC (cf. IV.2.E. de ce chapitre). Récemment, une étude *in vitro* a montré que la liaison du Médiateur à l'activateur p53 induit un changement conformationnel important pour l'activation de la Pol II et de TFIIH. Ceci confirme le rôle du Médiateur dans la stimulation de la phosphorylation du CTD (Meyer et al., 2010).

C. Interaction du Médiateur avec les co-activateurs

En plus des interactions avec les activateurs et les GTFs décrites ci-avant, le Médiateur est capable de se lier à divers co-activateurs. Ces derniers peuvent agir sur la structure de la chromatine, la mise en place du PIC ou les deux à la fois.

L'état de la chromatine peut être régulé de deux façons : (1) par le mouvement des nucléosomes sous l'action de remodeleurs dépendants de l'ATP ou (2) par des modifications des histones qui vont changer leurs propriétés physico-chimiques. Certaines de ces modifications, notamment l'acétylation des histones H3/H4 par p300, CBP et Gcn5, sont fortement corrélées avec l'activation des gènes. Le Médiateur interagit de manière directe avec p300 et stimule son recrutement sur l'ADN et donc l'acétylation des histones (Black et al., 2006). L'accès des autres facteurs est bloqué jusqu'à ce que la chromatine soit acétylée et que le facteur p300 se soit détaché. TFIID peut alors se lier au Médiateur et débiter la formation du PIC. Le Médiateur interagit également avec les complexes remodeleurs de la chromatine. Chez l'homme, le complexe STAGA recrute le Médiateur durant l'activation de la transcription par l'oncoprotéine Myc avec pour effet une stimulation de la prolifération cellulaire (Liu et al., 2008). Chez la souris, le Médiateur forme un complexe stable avec ATAC et régule la transcription d'un ensemble

d'ARNnc (Krebs et al., 2010). Chez la levure, le recrutement du Médiateur et de SAGA sur les gènes régulés par Gcn4 est interdépendant (Qiu et al., 2005) et permet le recrutement de Swi/Snf (Yoon et al., 2003). Tous les trois stimulent alors le recrutement de la TBP sur le promoteur et permettent d'initier la mise en place du PIC. Ces exemples montrent l'importance d'une coopération entre le Médiateur et les co-activateurs dans la régulation de la transcription.

Le Médiateur peut également interagir directement avec les histones et notamment la queue des histones H3 et H4 (Zhu et al., 2011a). Cette interaction est inhibée par l'acétylation H4K16 et une analyse par CHIP-on-chip montre que l'association du Médiateur et l'acétylation H4K16 sont mutuellement exclusives. Ceci laisse suggérer que la présence du Médiateur sur l'ADN pourrait être régulée par des modifications des histones. Une étude plus récente a permis de déterminer les sous-unités du Médiateur impliquées dans l'interaction avec l'histone H4, il s'agit de Med5 et Med17 (Liu and Myers, 2012).

Une étude récente a démontré l'existence d'une interaction entre la sous-unité Med1 du Médiateur murin et l'enzyme Chd1 (« chromodomain helicase DNA-binding protein 1 ») impliquée dans le remodelage de la chromatine (Lin et al., 2011). Chd1 se lie spécifiquement sur le promoteur des gènes actifs. Cette spécificité est en partie due à sa capacité à reconnaître la marque d'histone H3K4me3 qui est associée aux gènes activement transcrits. Cependant, chez la Drosophile, l'inactivation des chromodomaines de Chd1, indispensables à la reconnaissance des histones tri-méthylées, ne perturbe pas la localisation de l'enzyme sur la chromatine (Morettini et al., 2011). De la même façon, l'enzyme Chd1 de *S. cerevisiae*, dont les chromodomaines ne sont pas capables de reconnaître la tri-méthylation des histones, est recrutée spécifiquement au niveau des gènes activement transcrits (Ehrensberger and Kornberg, 2011). Par contre, en l'absence du Médiateur, le recrutement de Chd1 sur ces gènes est nettement diminué. Ainsi le Médiateur coordonne l'assemblage du PIC avec le recrutement de Chd1. L'action combinée du Médiateur et de la marque d'histone H3K4me3 permet de spécifiquement cibler Chd1 sur les gènes actifs.

D. Interaction du Médiateur avec d'autres facteurs importants pour l'initiation de la transcription

a. Interaction du Médiateur avec la cohésine

Pour permettre la communication entre les activateurs liés aux régions régulatrices, et la machinerie de transcription liée au promoteur, l'ADN doit former des boucles (« DNA looping »). La cohésine, un complexe protéique associé aux chromosomes et très conservé chez les eucaryotes, facilite la formation de ces boucles (Dorsett, 2011). Une analyse par CHIP réalisée

dans le laboratoire de R. Young à partir de cellules ES murines a montré que le Médiateur est présent à la fois sur l'« enhancer » et le promoteur des gènes activement transcrits (Kagey et al., 2010). Cette observation confirme que le Médiateur agit comme un pont entre les activateurs et la machinerie de transcription liée au promoteur. L'étude à l'échelle du génome par ChIP-seq a montré que la cohésine co-occupe les « enhancers » et les promoteurs liés par le Médiateur. Une inhibition par « knock down » de l'un de ces deux facteurs engendre la même diminution de l'expression des gènes Oct4-dépendants. Ainsi, la présence du Médiateur et de la cohésine est indispensable à l'expression de ces gènes. L'interaction du Médiateur et de la cohésine a par ailleurs été démontrée par co-IP. Enfin, une analyse 3C a permis de mettre en évidence la présence de boucles de l'ADN entre les promoteurs et « enhancers » liés par la cohésine et le Médiateur mais pas ailleurs. Tous ces résultats indiquent que le Médiateur, en association avec la cohésine, permet de connecter l'« enhancers » et le promoteur des gènes activement transcrits.

b. Interaction du Médiateur avec TFIIIS

Le rôle de TFIIIS a avant tout été décrit dans l'étape d'élongation durant laquelle il permet la reprise de la transcription après une pause de la Pol II en stimulant le clivage de l'ARN naissant pour qu'il puisse recouvrer une position correcte dans le site catalytique de la Pol II (Wind and Reines, 2000). Cependant, une activité dans l'initiation de la transcription a également été suggérée chez *S. cerevisiae*. Tout d'abord, la délétion du gène *DST1*, qui code pour TFIIIS chez la levure, induit une co-létalité lorsqu'il est associé à une délétion de *MED31*, supposant une activité redondante entre ces deux gènes (Malagon et al., 2004). TFIIIS est recruté sur le promoteur *GAL1*, et la délétion de *DST1* réduit le recrutement de la machinerie de transcription sur ce gène (Prather et al., 2005). Enfin, TFIIIS se fixe sur le promoteur des gènes *ADH1*, *VTC3* et *MET17* indépendamment de la Pol II et permet le recrutement de cette dernière en absence de Med31. Son domaine de liaison à la Pol II est requis pour cette activité tandis que son domaine de stimulation du clivage est dispensable (Guglielmi et al., 2007). Ces résultats sont en accord avec les observations *in vitro* qui montrent que TFIIIS participe à la formation et à la stabilisation du PIC indépendamment de son rôle dans l'élongation (Kim et al., 2007). Ainsi, TFIIIS et Med31 sont requis pour une occupation optimale de la Pol II sur les promoteurs *in vivo*.

3. Recrutement du Médiateur et rôle *in vivo*

L'ordre d'arrivée des complexes impliqués dans l'initiation de la transcription a été largement étudié *in vitro*. L'étude *in vivo* s'avère cependant plus complexe puisqu'il faut tenir compte, entre autres, des co-activateurs. La technique d'immunoprécipitation de chromatine (ChIP) a permis une nette avancée dans ce domaine et notamment de déterminer à quel moment de la mise en

place du PIC le Médiateur intervient. Des analyses cinétiques réalisées chez la levure, notamment sur les gènes *GAL* en présence de galactose, ont permis de mesurer l'arrivée des différents composants de la machinerie de transcription par la Pol II. Après ajout de galactose dans le milieu, l'activateur Gal4 se lie rapidement sur l'UAS et recrute d'abord le co-activateur SAGA puis le Médiateur qui, ensemble, permettent de recruter la Pol II et les GTFs (Bryant and Ptashne, 2003). Ainsi, le Médiateur fixe préférentiellement les régions régulatrices en amont du promoteur des gènes, avant le recrutement de la machinerie de transcription et donc indépendamment de la Pol II (Kuras et al., 2003; Park et al., 2001). Il permet alors la mise en place du PIC et sa stabilisation sur le promoteur (Cantin et al., 2003). Le schéma ci-après présente les étapes de l'activation de la transcription par le Médiateur *in vivo* (Figure 15).

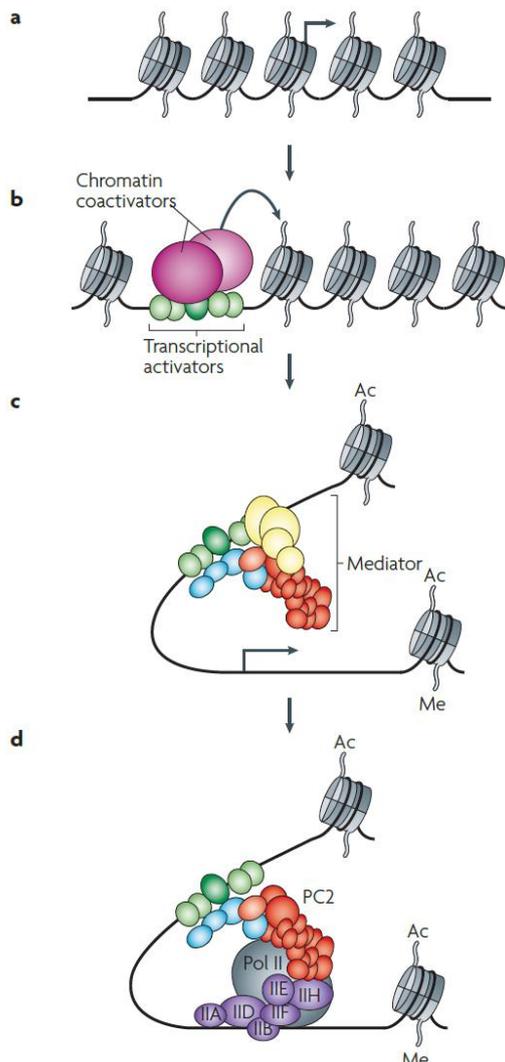


Figure 15 : Etapes de l'activation de la transcription par le Médiateur *in vivo* (Malik and Roeder, 2010).

(a) L'ADN est compacté sous forme de chromatine, enroulé autour des nucléosomes.

(b) L'activation de la transcription débute par la fixation d'un ou plusieurs activateurs sur une région régulatrice (UAS). Ceux-ci vont recruter des co-activateurs intervenant dans le remodelage de la chromatine (SAGA, Swi/Snf...)

(c) Médiateur est à son tour recruté par une interaction de son module de queue avec les activateurs. A cette étape, le module Cdk peut ou non être présent.

(d) Le Médiateur stimule l'assemblage du PIC en interagissant avec les GTFs et la Pol II via ses modules de tête et du milieu. A ce stade le module Cdk est absent car la liaison de la Pol II et du module Cdk sur le Médiateur sont mutuellement exclusives (cf. V. de ce chapitre).

Les travaux réalisés entre autres par notre équipe sur la sous-unité Med11 du Médiateur ont permis de démontrer une interaction directe entre cette sous-unité et Rad3, qui appartient au module cœur de TFIIF. Cette interaction est importante pour le recrutement de TFIIF. En effet, les mutations *med11ts* induisent un effet différentiel sur le recrutement des modules cœur et kinase

de TFIIH indépendamment de TFIIIE et de la Pol II. Les différentes données obtenues ont ainsi permis de démontrer *in vivo* le rôle du Médiateur dans le recrutement indépendant de TFIIIE, TFIIH et de la Pol II (Esnault et al., 2008). Différents intermédiaires de l'initiation de la transcription ont ainsi été mis à jour et ont permis de proposer un modèle où le Médiateur stimule la transcription en intervenant sur la mise en place du PIC (Figure 16). Il favorise notamment l'association de TFIIIE, TFIIH et de la Pol II au promoteur, ainsi que le recrutement de TFIIK sur le module cœur de TFIIH.

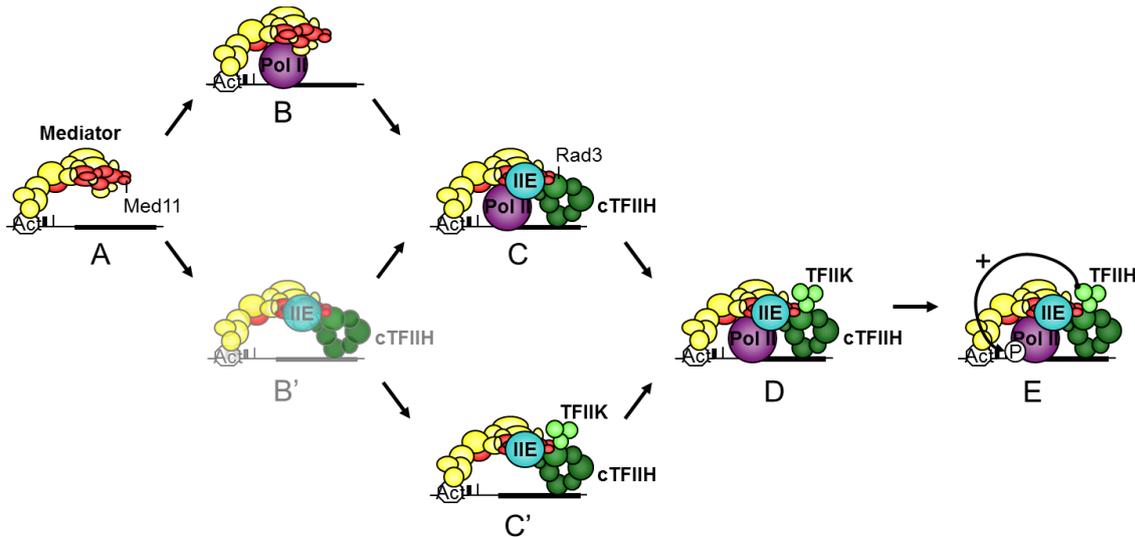


Figure 16 : Influence du Médiateur sur la mise en place du PIC et la phosphorylation du CTD (Esnault et al., 2008; Soutourina et al., 2011).

Les intermédiaires observés au cours de nos études sont en couleurs vives, ceux qui ont été déduits sont en couleurs transparentes.

La meilleure façon de comprendre le mécanisme d'action du Médiateur reste d'étudier l'effet des mutations qui affectent son interaction avec les différents facteurs de la transcription. Une carte d'interaction génétique réalisée chez la levure *S. cerevisiae* a ainsi pu voir le jour et a permis de démontrer que le module de queue est lié génétiquement aux activateurs, celui du milieu au facteur d'élongation TFIIS et celui de la tête à TFIIIF, TFIIH et à la Pol II (Collins et al., 2007).

Le Médiateur formant un complexe avec la Pol II et les GTFs, son interaction était attendue au niveau du promoteur des gènes. Cependant, une analyse par ChIP sur les gènes *GAL* de *S. cerevisiae* a montré sa localisation préférentielle sur les UAS, indépendamment de la Pol II et des GTFs (Kuras et al., 2003). Des études à l'échelle du génome, notamment par ChIP-on-chip chez *S. cerevisiae* (Andrau et al., 2006) et *S. pombe* (Zhu et al., 2006), ont par la suite confirmé que le Médiateur se fixe sur les régions UAS de la majorité des gènes de classe II, actifs ou inactifs. Les résultats obtenus suggèrent également que le Médiateur est recruté comme une seule entité avec son module Cdk, même si ce dernier n'est parfois présent que de manière transitoire. Les

différentes analyses réalisées à l'échelle du génome ont également soulevé la question de la localisation du Médiateur sur la chromatine. Bien que la majorité s'accorde à dire que le Médiateur se lie majoritairement sur les UAS, certains suggèrent aussi une présence sur le promoteur et même sur les ORFs. En ce qui concerne la présence sur les promoteurs, les travaux de C. Jeronimo et F. Robert ont apporté un premier élément de réponse en montrant que le Médiateur peut en effet s'associer aux promoteurs mais seulement de manière transitoire, tandis que sa liaison aux UAS est beaucoup plus stable (Jeronimo and Robert, 2014). Cette association serait régulée par l'état de phosphorylation du CTD de la Pol II par Kin28, la phosphorylation sur Ser5 provoquant la dissociation du Médiateur, et induisant ainsi l'échappée du promoteur (Wong et al., 2014). Pour ce qui est de la présence sur les ORFs, la majorité des signaux détectés ne seraient que des artefacts puisqu'ils sont également retrouvés dans les contrôles négatifs (ChIP à partir d'une souche non étiquetée ou avec un anticorps non spécifique) comme le confirment nos résultats non publiés. Plusieurs laboratoires ont d'ailleurs rapportés des artefacts similaires dans les ORFs des gènes fortement transcrits (Jeronimo and Robert, 2014; Park et al., 2013; Teytelman et al., 2013).

4. Le Médiateur, facteur général de la transcription

Les facteurs généraux de la transcription sont définis comme les facteurs indispensables à la transcription basale. Le Médiateur n'est pas requis dans ce contexte précis mais est capable de stimuler la transcription basale *in vitro*. Cette fonction a été montrée à la fois chez la levure (Kim et al., 1994) et chez l'homme (Mittler et al., 2001). De plus, il va permettre d'augmenter très fortement le niveau de la transcription suite à l'ajout d'un activateur (Baek et al., 2006; Kornberg, 2005; Mittler et al., 2001). La transcription basale a été largement étudiée *in vitro*, à la fois chez la levure et les mammifères, mais reste peu documentée *in vivo*. Une méthode mise au point récemment par le laboratoire de L. Kuras, utilisant le recrutement artificiel de la machinerie de transcription par la Pol II sur un promoteur test ne nécessitant pas d'activateur (« activator bypass »), a permis d'adresser cette question (Lacombe et al., 2013). Cette technique utilise un système dans lequel TFIIB est fusionné à RFX, un facteur de transcription humain, sans potentiel d'activation chez la levure. Le site de liaison de RFX est ensuite introduit en amont du promoteur de *MET17*. Les résultats obtenus montrent que le Médiateur peut être recruté *in vivo* indépendamment de la présence d'un activateur, mais via sa capacité à lier la Pol II et les GTFs. Ainsi le Médiateur appartiendrait à la machinerie de base de la transcription et pourrait influencer la transcription basale *in vivo*.

Une autre définition des GTFs stipule qu'ils sont requis pour la transcription de l'ensemble des gènes. Or chez la levure, des mutations de Med17 conduisent à une diminution de la

transcription des gènes, similaire à celle observée avec des mutations de la sous-unité Rpb1 de la Pol II (Holstege et al., 1998; Thompson and Young, 1995). Une étude récente montre également que le Médiateur est présent à la fois sur les gènes inductibles et constitutifs et est indispensable au recrutement de la Pol II sur leurs promoteurs (Ansari et al., 2009). Le Médiateur peut donc à ce titre être qualifié de GTF (Takagi and Kornberg, 2006).

Dans tous les cas, le Médiateur est un co-activateur requis de manière quasi-universelle pour initier la transcription, qui joue un rôle majeur et unique et contribue à activer la transcription par divers mécanismes.

IV. REGULATION NEGATIVE DE LA TRANSCRIPTION PAR LE MEDIATEUR

Bien que le Médiateur soit généralement décrit comme un co-activateur, de nombreuses sous-unités semblent impliquées dans la régulation négative de la transcription (Hampsey, 1998; Myers and Kornberg, 2000). Notamment, des analyses de transcriptome ont permis d'identifier de nombreux gènes régulés à la hausse en présence de mutations de diverses sous-unités du Médiateur (Holstege et al., 1998; Koschubs et al., 2009; van de Peppel et al., 2005).

Dans certains organismes, le Médiateur existe sous deux formes qui diffèrent principalement par la présence ou l'absence du module kinase (Liu et al., 2001; Sato et al., 2004; Spahr et al., 2003). Ce module comprend la kinase Cdk8 et la cycline CycC ainsi que deux sous-unités additionnelles : Med12 et Med13. Les premières études du module Cdk8 ont rapidement conclu en un rôle dans la régulation négative de la transcription. En effet la délétion de certaines sous-unités de ce module conduit à l'augmentation de l'expression de nombreux gènes chez *S. cerevisiae* (Hengartner et al., 1998; Holstege et al., 1998; van de Peppel et al., 2005). Le mécanisme de cette inhibition n'est pas encore totalement connu, cependant, il a été montré que la kinase Cdk8 est capable, *in vitro*, de phosphoryler les sérines 2 et 5 du CTD de la Pol II avant la mise en place du PIC ce qui empêcherait la formation du complexe Med-Pol II (Hengartner et al., 1998). De plus, la présence du module kinase au sein du Médiateur empêche son interaction avec la Pol II chez la levure et chez l'homme (Elmlund et al., 2006; Knuesel et al., 2009). Des analyses menées à l'échelle du génome montrent également que le module Cdk8 est capable de phosphoryler la sous-unité Med2, empêchant ainsi l'activation par le module de queue d'un ensemble de gènes impliqués dans la réponse à une carence en fer (van de Peppel et al., 2005). Chez les mammifères, Cdk8 peut également inhiber l'activité de TFIIF au cours de l'initiation de la transcription en phosphorylant la cycline H (Akoulitchev et al., 2000). Enfin, des travaux récents réalisés chez la levure montrent que le Médiateur est capable d'autoréguler sa propre fonction

(Gonzalez et al., 2014). Cdk8 peut en effet phosphoryler la sous-unité Med3, dont la fonction est requise pour l'activation de la transcription, ce qui engendre sa dégradation par l'ubiquitine-ligase Grr1 et inhibe ainsi la fonction co-activatrice du Médiateur.

Ces données montrent l'importance de Cdk8 pour réguler négativement la transcription. Ce module agit en inhibant trois acteurs essentiels pour l'initiation de la transcription : le Médiateur lui-même, TFIIH et la Pol II. Cependant, de plus en plus d'études montrent que Cdk8 peut également réguler positivement certains gènes notamment en facilitant le recrutement de la Pol II, du facteur P-TEFb ou de la TBP (Belakavadi and Fondell, 2010; Donner et al., 2010; Larschan and Winston, 2005).

V. AUTRES ROLES DU MEDIATEUR

Plusieurs études récentes suggèrent que le Médiateur pourrait intervenir au-delà de l'initiation, par exemple dans les étapes post-recrutement du PIC et dans la maturation des ARNm, mais également dans d'autres processus tels que le maintien des télomères.

1. Rôle du Médiateur dans les étapes post-recrutement du PIC

La réactivation de la Pol II après la pause à proximité du promoteur nécessite le recrutement de divers facteurs d'élongation tels que ELL/EAF et P-TEFb. La sous-unité Med26 du Médiateur intervient dans ce processus en liant tout d'abord le facteur TFIID lors de l'initiation puis en l'échangeant contre ELL/EAF et P-TEFb pour permettre l'échappée de la pause (Takahashi et al., 2011). Une coopération entre le Médiateur et TFIIS a également été mise en évidence, qui permet le transit de la Pol II au-delà du nucléosome +1 (Nock et al., 2012).

La sous-unité Cdk8 du Médiateur n'est pas requise pour le recrutement de la Pol II ni pour la sortie de pause. Cependant, en l'absence de Cdk8, l'élongation se fait plus lentement, en corrélation avec une phosphorylation de Ser2 et Ser5 diminuée (Donner et al., 2010). Cdk8 serait donc important pour l'étape d'élongation, notamment lors de l'assemblage du complexe d'élongation, en permettant le recrutement des kinases Cdk7 et Cdk9. Chez l'homme, Cdk8 stimule la sortie de pause de la Pol II sur les gènes H1F1A-dépendants en réponse à l'hypoxie en permettant le recrutement de P-TEFb et SEC (« super elongation complex »)(Galbraith et al., 2013). Dans les cellules ES murines *MED23* KO, le taux de transcription du gène *EGR1* (un gène modèle Med23-dépendant) est réduit de 4 à 5 fois et le recrutement de P-TEFb et de la Pol II sur l'ORF est significativement diminué. Par contre, l'occupation de la Pol II, du Médiateur ou des GTFs au niveau du promoteur reste inchangée (Wang et al., 2013). De plus, une interaction directe entre Med23 et la sous-unité Cdk9 de P-TEFb a été détectée. Ces résultats montrent que

Med23 et Cdk8 sont importants pour réguler l'étape d'élongation, notamment en recrutant le facteur P-TEFb.

Chez *S. cerevisiae*, la sous-unité Med18 du Médiateur est détectée sur la région terminatrice des gènes *CHA1* et *INO1*. De plus, dans les souches *med18Δ* les complexes de terminaison de la transcription CF1 (« cleavage factor 1 ») et CPF (« cleavage and poly-adenylation factor ») ne sont plus recrutés et la Pol II s'accumule en 3' du gène et dans la région intergénique (Mukundan and Ansari, 2011). Ces résultats suggèrent que le Médiateur est requis pour la terminaison efficace de certains gènes.

Une analyse par spectrométrie de masse a permis d'identifier plusieurs facteurs de maturation de l'ARN qui lient le Médiateur (Huang et al., 2012). HnRNP L qui intervient dans l'épissage alternatif et la polyadénylation interagit directement avec Med23 *in vitro* et *in vivo*. De plus, Med23 co-localise partiellement avec ce facteur sur l'ADN. Cette étude confère un nouveau rôle pour le Médiateur dans la régulation de l'épissage des ARNm.

2. Rôle du Médiateur dans la réinitiation de la transcription

Il est peu probable que le Médiateur reste lié à la Pol II après son échappée du promoteur. En effet, des analyses *in vitro* montrent que le Médiateur est retenu par les activateurs et forme une plateforme à partir de laquelle il peut faciliter la réinitiation (Yudkovsky et al., 2000). Pour comprendre ce processus, l'équipe de S. Hahn a développé un système de transcription sur matrice d'ADN immobilisée qui permet d'étudier la première initiation et la réinitiation de la transcription *in vitro*. Le PIC est formé à partir d'extraits nucléaires de *S. cerevisiae*. L'ajout de nucléotides au système pendant deux minutes permet de suivre un seul cycle de transcription. Dans ce système, la Pol II n'est pas retrouvée puisqu'elle est partie en élongation. Les facteurs TFIIB et TFIIF se désolidarisent du promoteur tandis que TFIIA, TFIID, TFIIF, TFIIE et le Médiateur restent associés sur la matrice. Pour observer un nouveau cycle de transcription, des extraits nucléaires sont de nouveau ajoutés. Si ces extraits ont été obtenus à partir de souches mutées dans des sous-unités essentielles des facteurs restés sur la matrice, la réinitiation peut tout de même avoir lieu. Les résultats obtenus suggèrent que ces facteurs forment un échafaudage (« scaffold ») pour la réinitiation de la transcription. Cette plateforme permettrait d'accélérer l'étape de formation du PIC et donc d'atteindre des niveaux de transcription plus importants. Cependant, ce phénomène de réinitiation a seulement été observé *in vitro*, aucune preuve de son existence *in vivo* n'a encore été apportée.

Deux mutants du Médiateur : *med3Δ* et *med16Δ*, affectent la réinitiation. Ils déstabilisent l'association du Médiateur avec la Pol II et n'entraînent pas de défaut pour le premier cycle de

transcription, mais diminuent le niveau de transcription au cours des cycles suivants (Reeves and Hahn, 2003). De plus, l'incorporation du module Cdk8 du Médiateur au sein de l'échafaudage permettrait d'« éteindre » le système de réinitiation après un certain nombre de cycles de transcription (Elmlund et al., 2006). Cette dernière observation est en accord avec un rôle pour ce module dans la régulation négative de la transcription (cf. V. de ce chapitre).

3. Rôle du Médiateur dans le maintien des télomères

Chez *S. cerevisiae*, les télomères consistent en une répétition d'une séquence dégénérée [(TG₁₋₃)_n] sur une longueur variant de 300 à 500 nucléotides (Walmsley et al., 1984) avec une extension 3' simple brin G-riche dont la longueur varie au cours du cycle cellulaire (Wellinger et al., 1993). Les gènes situés près des télomères subissent un « silencing » réversible (Ottaviani et al., 2008). Ce phénomène nommé TPE (« telomere position effect ») dont les mécanismes sont encore très peu connus, est très conservé chez les eucaryotes et est impliqué dans de nombreuses pathologies humaines, d'où l'importance de son étude. Les effets opposés des enzymes Sir2 et Sas2 induisent un gradient d'acétylation H4K16 qui marque la limite entre la chromatine active et inactive près des télomères. L'enzyme Sir2 permet le maintien du TPE en éliminant l'acétylation H4K16 tandis que l'HAT Sas2 possède un effet antagoniste à Sir2 en acétylant l'histone H4K16. Dans les cellules vieillissantes on constate une diminution de Sir2 s'accompagnant d'une augmentation de l'acétylation H4K16 et d'une diminution du TPE (Dang et al., 2009). La surexpression de cette enzyme augmente la longévité des cellules tandis que son inactivation la réduit. L'effet inverse est observé avec l'enzyme Sas2. Ainsi, la limite exacte entre la chromatine active et inactive, régulée par Sir2 et Sas2, peut influencer directement la durée de vie cellulaire. Une analyse par ChIP chez *S. cerevisiae* a démontré que le Médiateur s'associe avec l'hétérochromatine au niveau des télomères (Zhu et al., 2011b). La délétion de la sous-unité Med5 induit une diminution de l'occupation du Médiateur sur ces régions, augmente l'acétylation H4K16 et réduit le TPE avec pour conséquence un raccourcissement de la durée de vie des cellules. Le mécanisme proposé serait que le Médiateur influence la balance qui existe entre les enzymes Sir2 et Sas2. La délétion des sous-unités du module de queue du Médiateur induit une diminution de l'occupation de Sir2 et une augmentation de l'acétylation H4K16. Ce phénomène n'est pas observé après délétion des sous-unités des autres modules du Médiateur indiquant que seul le module de queue du Médiateur est requis et suffisant pour réguler le TPE (Peng and Zhou, 2012).

VI. ROLES FONCTIONNELS ET PATHOLOGIES

Diverses méthodes ont permis d'étudier les fonctions physiologiques des différentes sous-unités du Médiateur des animaux. Ces études ont notamment permis de confirmer le rôle crucial de ce complexe dans la transcription ainsi que son implication dans de nombreux processus biologiques et par conséquent dans diverses pathologies associées.

1. Rôle du Médiateur dans le développement et la différenciation

Le rôle essentiel du Médiateur dans le développement a été déterminé à partir d'études réalisées chez la drosophile, le poisson zèbre, la souris et l'homme (Tableau 10).

Le Médiateur est indispensable à un stade précoce du développement. Des expériences d'ARN interférence (ARNi) chez *C. elegans* ont en effet montré que les sous-unités Med6, Med7 et Med10 sont essentielles à l'embryogénèse (Kwon et al., 1999). De la même façon, des souris invalidées pour le gène *MED21* (KO « knock-out ») meurent au stade blastocyste sans qu'aucune cellule ES n'ait pu être obtenue (Tudor et al., 1999).

Des souris invalidées génétiquement pour différentes sous-unités du Médiateur ont été générées. Les souris *MED1* KO meurent à E11.5 (11,5 jours de développement de l'embryon) d'une insuffisance placentaire. Ces embryons montrent des malformations cardiaques, un défaut de développement neuronal, une nécrose hépatique et un défaut d'hématopoïèse (Ito et al., 2000). Les souris *MED12* KO ne terminent pas la gastrulation et meurent à E7.5 (Rocha et al., 2010). Les souris *MED23* KO ont un développement retardé et meurent entre E9 et E10.5 d'un défaut de circulation systémique (Balamotis et al., 2009). Les souris *MED24* KO présentent une sévère hypoplasie cardiaque, un défaut de développement du système nerveux central et ne dépassent pas E10.5 (Ito et al., 2002). Enfin, les souris *CDK8* KO meurent avant l'implantation et présentent des blastomères fragmentés (Westerling et al., 2007). Ainsi, toutes les souris invalidées pour les gènes cités ci-dessus meurent au stade embryonnaire, plus ou moins précocement, et présentent des phénotypes variés. Ceci suggère un rôle spécifique pour chaque sous-unité du Médiateur dans le développement embryonnaire.

Chez la drosophile, le module Cdk8 est impliqué dans la régulation de l'expression des gènes du développement de la patte, la différenciation cellulaire pendant le développement de l'œil, le développement des organes sensoriels externes et la définition des segments (Boube et al., 2000; Loncle et al., 2007). Med15 intervient dans le développement des ailes (Terriente-Félix et al., 2010) et Med31 dans la formation de l'axe antéropostérieur (Bosveld et al., 2008). De plus, Med17 est indispensable à la survie cellulaire (Boube et al., 2000).

Sous-unité	Phénotype / Processus	Organisme
Med1	KO léthal pour l'embryon à E11,5	Souris
	Adipogenèse, érythropoïèses	Souris
	Erythropoïèse et différenciation des glandes mammaires et du	Souris
	Différenciation des cellules luminales	Souris
Med12	KO léthal pour l'embryon à E7,5	Souris
	Fonction des cellules ES	Souris
	Différenciation neuronale	Souris
	Développement du cerveau, des oreilles, des reins, du cartilage et de l'endoderme	Poisson zèbre
	Chondrogenèse	Poisson zèbre
	Développement de la rétine et des ailes	Drosophile
	Développement de la vulve	<i>C. elegans</i>
Division des cellules T	<i>C. elegans</i>	
Med13	Développement de la rétine et des ailes	Drosophile
	Développement de la vulve	<i>C. elegans</i>
	Division des cellules T	<i>C. elegans</i>
	Développement du cœur et du cerveau	Homme
Med14	Adipogenèse	Souris
Med15	Développement des ailes	Drosophile
	Développement du mésendoderme	Xenope
	Métabolisme des lipides	Homme
	Développement du cœur	Homme
Med19	Différenciation neuronale	Souris
	Développement des oligodendrocytes	Homme
Med21	KO léthal pour l'embryon au stade blastocyte	Souris
	Différenciation des kératinocyte	Souris
Med23	KO léthal pour l'embryon à E9-10,5	Souris
	Adipogenèse	Souris
	Différenciation du muscle lisse	Souris
	Développement de la vulve	<i>C. elegans</i>
Med5/24	KO léthal pour l'embryon à E8-10,5	Souris
	Différenciation des cellules de la rétine	Poisson zèbre
	Adipogenèse	Homme
Med25	Chondrogenèse	Souris
	Sécrétion d'insuline	Homme
	Métabolisme des xénobiotiques et des lipides	Homme
	Développement du système nerveux périphérique	Homme
Med26	Différenciation neuronale	Souris
Med3/27	Différenciation des cellules de la rétine	Poisson zèbre
Med28	Différenciation du muscle lisse	Souris
Med31	KO léthal pour l'embryon à E16,5-18,5	Souris
	Chondrogenèse	Souris
	Formation de l'axe antéropostérieur	Drosophile
Cdk8	KO léthal pour l'embryon à E2,5-3	Souris
	Développement de l'oeil	Drosophile
	Réponse immunitaire et différenciation cellulaire	Homme
Cdk19	Réponse immunitaire et différenciation cellulaire	Homme

Tableau 10 : Implication des sous-unités du Médiateur dans des phénotypes ou des processus développementaux (d'après Yin and Wang, 2014)

Chez le poisson-zèbre, Med12 est indispensable au développement du cerveau, des cartilages, de l'œil, des reins et de l'endoderme (Hong et al., 2005; Rau et al., 2006; Shin et al., 2008; Wang et al., 2006b). Med23 intervient dans la différenciation des muscles lisses (Yin et al., 2012). Med12 et Med25 sont importants pour la chondrogenèse (Nakamura et al., 2011).

Toutes ces études montrent le rôle essentiel du Médiateur dans l'organogenèse à un stade précoce et expliquent que des mutations de ce complexe puissent conduire à des pathologies du développement.

Le Médiateur interagit avec des facteurs de transcription spécifiques et régule ainsi l'expression de différents groupes de gènes impliqués dans la différenciation et la prolifération cellulaire (Tableau 10). Des expériences ont montré que Med1 et Med21 stimulent la prolifération et la différenciation des kératinocytes tandis que Med10 possède un rôle inhibiteur (Oda et al., 2007, 2010). Med1 joue également un rôle crucial dans la différenciation des myélocytes au cours de l'hématopoïèse (Urahama et al., 2005). Med31 est requis pour la prolifération cellulaire intervenant lors de la croissance embryonnaire (Risley et al., 2010). L'inactivation de *MED23* abolit la transcription des gènes dépendant du facteur de croissance Egr1 qui joue un rôle important dans la différenciation et la transformation cellulaire ainsi que dans l'apoptose (Balamotis et al., 2009). La prolifération cellulaire pouvant être à l'origine du développement de nombreux cancers chez l'homme, il n'est donc pas étonnant que le Médiateur soit impliqué dans cette pathologie.

2. Pathologies impliquant le Médiateur

Sous-unité	Pathologie	Etat génétique/Mutation
Med1	Cancer de la prostate et du sein	Surexpression
	Mélanome	Surexpression
	Cancer du colon et du poumon	Sous-expression
Med4	Cancer cervical	Sous-expression
	Rétinoblastome	Sous-expression
Med12	Syndrome d'Opitz-Kaveggia (syndrome FG)	Mutation R971W
	Syndrome de Lujan	Mutation N1007S
	Léiomyome utérin	Mutation de l'intron 1 et exon 2
	Syndrome d'Ohdo	Mutation faux-sens
	Schizophrénie psychose	Polymorphisme
	Cancer du sein	-
	Cancer du colon	Mutation de l'exon 2
	Cancer du poumon	Sous-expression
Med13	Cancer de la prostate	Mutation
	Transposition des grosses artères	Translocation chromosomique
	Cancer du cerveau	Motif d'édition altéré
	Cancer du sein	Amplification de gène
Med14	Rétinoblastome	-
	Cancer du sein	-
Med15	Cancer de la prostate	Surexpression
	Syndrome de Di-George	Délétion chromosomique
Med17	Cancer du sein	-
	Cancer de la prostate	hyperméthylation
	Atrophie cérébrale infantile	Mutation L371P
Med19	Cancer de la prostate	Surexpression
	Cancer de la vessie, du sein, du colon, du poumon, de l'estomac de l'ovaire, de la prostate, du pancréas, de la langue	Surexpression
Med23	Ostéosarcome	Surexpression
	Déficit intellectuel non syndromique	Mutation R617Q
	Mélanome	Délétion chromosomique
	Cancer du sein	-
	Cancer de l'estomac	Mutation
	Phéochromocytome	Sous-expression
Med24	Cancer du poumon	Surexpression
	Cancer du sein et de la prostate	Surexpression
Med25	Maladie de Charcot-Marie-Tooth	Mutation A335V
Med28	Cancer du sein, du colon et de la prostate	Surexpression
Med29	Cancer du pancréas	Surexpression
Med30	Cancer du sein	Surexpression
Cdk8	Cancer du colon	Surexpression
	Mélanome	Surexpression
CycC	Cancer du colon	Surexpression
	Carcinome hépatocellulaire	Surexpression
	Leucémie	Délétion de région génomique
	Ostéosarcome	Sous-expression
Cdk19	Décollement de rétine congénital, microcéphalie et déficience intellectuelle	Inversion chromosomique péricentrique
	Myélome	Altérée
	Ostéosarcome	Surexpression

Tableau 11 : Implication des sous-unités du Médiateur dans des pathologies humaines (d'après Schiano et al., 2014; Spaeth et al., 2011).

A. Pathologies du développement cérébral

Chez l'homme aussi, le Médiateur joue un rôle important au cours de l'organogénèse (Tableau 10). Il intervient notamment dans le développement du système nerveux central et périphérique et peut ainsi être impliqué dans diverses pathologies cognitives (Tableau 11).

Des mutations de la sous-unité Med12 interviennent dans les syndromes d'Opitz-Kaveggia et de Lujan qui se traduisent par des malformations diverses et un retard mental (Risheg et al., 2007; Schwartz et al., 2007). L'interaction du Médiateur avec le répresseur de la différenciation neuronale REST est alors diminuée induisant une expression anormale des gènes neuronaux (Ding et al., 2008). Une mutation dans Med25 engendre un défaut d'interaction entre le Médiateur et les kinases Abl et Lck importantes pour le développement du système nerveux périphérique et conduit à la maladie de Charcot-Marie-Tooth, une neuropathie sensorielle et motrice (Leal et al., 2009). Une mutation dans Med17 conduit à un défaut de régulation des gènes importants pour le développement des oligodendrocytes et est responsable de l'atrophie cérébrale infantile (Kaufmann et al., 2010). La mutation correspondante chez *S. cerevisiae* conduit à une protéine inactive à 37°C. Le gène codant la sous-unité Med15 est délété dans le syndrome de DiGeorge (Berti et al., 2001). Dans les embryons de souris, Med15 est exprimée de façon ubiquitaire, cependant, son taux est plus important dans les régions fronto-nasales, les arcs pharyngiens et les bourgeons des membres, ce qui pourrait expliquer les défauts observés dans le syndrome de DiGeorge (insuffisance vélopharyngée, dysmorphie faciale...). Enfin, Med23 est impliquée dans la déficience intellectuelle non syndromique (Hashimoto et al., 2011). La mutation de cette sous-unité diminue l'activité des gènes de réponse précoce (IEG « immediate early genes ») Jun et Fos. Une dérégulation de ces gènes est également retrouvée chez les patients présentant des troubles neurologiques, indiquant que leur expression serait importante pour le développement cognitif.

Toutes ces études montrent un rôle important pour le Médiateur dans le développement cérébral. Plusieurs de ces pathologies sont associées à une cardiopathie, suggérant que le Médiateur intervient également dans le développement cardiaque. D'ailleurs, une forme tronquée de Med13 est retrouvée dans la transposition des gros vaisseaux (Muncke et al., 2003) qui est la malformation cardiaque cyanogène la plus fréquente du nouveau-né.

B. Médiateur et cancer

La carcinogenèse est un processus complexe caractérisé par la dérégulation de l'expression de nombreux gènes. Le Médiateur intervenant dans la régulation de la transcription, il n'est pas étonnant de le retrouver impliqué dans de nombreux cancers. Pourtant ce n'est que très récemment que les études de cancérologie ont permis de révéler que des altérations des composants de la machinerie de transcription, tels que le Médiateur, peuvent être oncogènes. Le rôle détaillé de ces modifications et l'effet qu'elles produisent reste largement inconnu, cependant il existe dans la littérature quelques exemples qui précisent le rôle fondamental du Médiateur dans la carcinogenèse. La suite de ce chapitre sera donc consacrée aux cas les mieux documentés d'implication des sous-unités du Médiateur dans des cancers.

Cependant le rôle du Médiateur dans le développement du cancer est bien plus important que ne le laissent suggérer ces quelques exemples, comme en témoigne le tableau récapitulatif précédent qui montre qu'un grand nombre de sous-unités est impliqué dans au moins un type de cancer (Tableau 11).

Exemple 1 : Implication de MED1 et MED17 dans le cancer de la prostate

Le complexe Médiateur est un important co-activateur pour de nombreux facteurs de transcription dont le récepteur aux androgènes (AR). *MED1/TRAP220* agit comme une cible directe du récepteur AR. Ce gène est surexprimé dans 50% des cancers de la prostate (Vijayvargia et al., 2007). L'utilisation d'ARNi ciblant *MED1* ou *MED17* diminue la prolifération androgène-dépendante et indépendante et augmente l'apoptose des cellules de cancer de la prostate. De plus, la perte de l'expression de *MED1* ou de *MED17* entraîne une diminution de l'expression AR-dépendante des gènes codant les protéines PSA (antigène prostatique spécifique) et KLK2 (« kallikrein-related peptidase ») dont des taux élevés sont associés au cancer de la prostate.

Le Médiateur est donc important pour la prolifération et la survie des cellules cancéreuses prostatiques et pourrait représenter une nouvelle cible thérapeutique pour traiter cette pathologie. De plus, du fait de la surexpression de *MED1* à un stade précoce, cette sous-unité pourrait servir de marqueur de diagnostic pour les investigations cliniques.

MED5/24, *MED14* (Wang et al., 2002) et *MED19* (Cui et al., 2011) sont également surexprimés dans les cellules issues de cancer de la prostate. Ces gènes jouent eux aussi un rôle important dans l'expression des gènes AR-dépendants.

Exemples 2 : Implication de *MED23* dans le cancer du sein

L'oncoprotéine Her2 est surexprimée dans 30% des cancers primitifs du sein et résulte souvent en une pathologie plus agressive, métastatique et résistante aux chimiothérapies. Le facteur de transcription principal activant *HER2* est ESX dont l'expression est augmentée dans les cancers du sein surexprimant Her2. Des analyses par SDS/PAGE, à partir des protéines isolées pour leur capacité à lier le domaine d'activation d'ESX, ont permis d'identifier la protéine Med23/DRIP130/CRSP130 (Asada et al., 2002). L'inhibition de l'interaction entre Med23 et ESX diminue l'expression de Her2 et entraîne une très nette diminution de la prolifération et de la viabilité des cellules cancéreuses en augmentant leur apoptose. Cette dernière observation est cohérente avec le fait qu'Her2 joue un rôle dans les voies anti-apoptotiques.

MED1 est également surexprimé dans les lignées cellulaires issues de cancer du sein. Son expression corrèle positivement avec la présence d'Her2 et contribue à la résistance au Tamoxifène, une molécule utilisée dans le traitement des cancers du sein hormonaux dépendants (Cui et al., 2012). Les sous-unités Med1 et Med23 pourraient donc être utilisées comme cibles thérapeutiques dans les cancers du sein surexprimant Her2.

Exemple 3 : Implication de *CDK8* dans le cancer colorectal

Une activation constitutive de la voie WNT/ β -caténine est retrouvée dans la majorité des cancers colorectaux et contribue à leur croissance, leur invasion et leur survie. Des expériences d'ARNi ont permis d'identifier les gènes régulant l'activité de la β -caténine et participant à la prolifération du cancer colorectal, parmi lesquels se trouve *CDK8* (Firestein et al., 2008). Ce gène est surexprimé dans la majorité des cancers du côlon et sa suppression induit une diminution de la prolifération cellulaire et un ralentissement du cycle cellulaire dans les lignées caractérisées par une activité élevée de la β -caténine. Ceci indique que Cdk8, via son activité kinase, est essentielle à la régulation de l'activité de la β -caténine dans les cellules cancéreuses du colon. Cette sous-unité pourrait donc servir de cible thérapeutique dans les cancers présentant une surexpression de la β -caténine.

CHAPITRE III: LA REPARATION DE L'ADN CHEZ LES EUCARYOTES.

Dans les conditions habituelles d'activité métabolique et d'exposition aux facteurs environnementaux, l'ADN des cellules subit chaque jour des dizaines de milliers de lésions. Celles-ci sont globalement néfastes pour la cellule qui doit être capable de les éliminer rapidement pour pouvoir survivre. Un ensemble très efficace de processus biochimiques est capable d'identifier, signaler et corriger les dommages de l'ADN et contribue ainsi à maintenir l'intégrité du génome et le bon fonctionnement cellulaire. Cette machinerie de réparation de l'ADN est en permanence active et finement contrôlée dans la cellule.

Les mécanismes de réparation diffèrent selon le type de lésion (Figure 17). Cinq grands systèmes de réparation existent au sein des cellules : (1) la réparation par excision de base (BER « base excision repair »); (2) la réparation par excision de nucléotides (NER « nucleotide excision repair »); (3) la réparation des mésappariements (MMR « mismatch repair »); (4) la réparation par jonction d'extrémités non-homologues (NHEJ « non-homologous end joining »); et (5) la réparation par recombinaison homologue (HR « homologous recombination »).

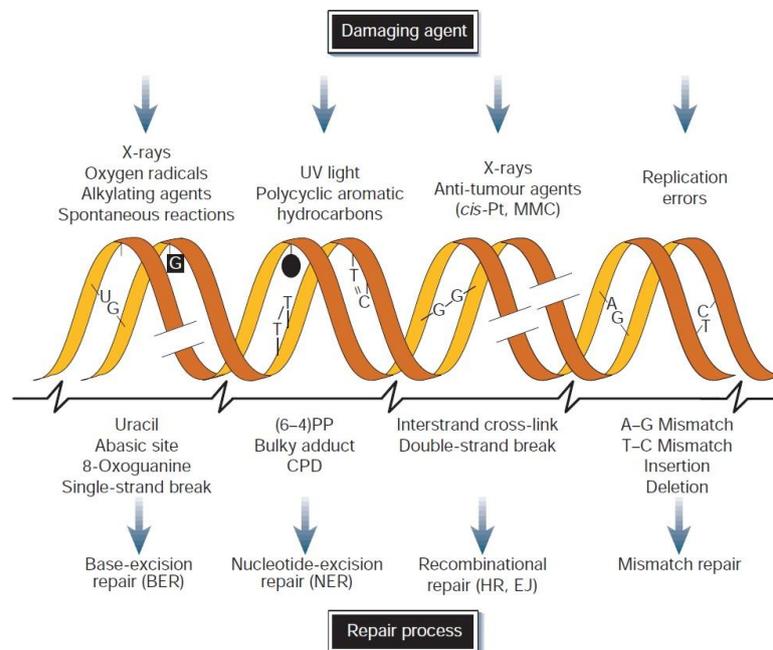


Figure 17 : Les différents types de lésions de l'ADN et les mécanismes de réparation associés (Hoeijmakers, 2001).

La suite de ce chapitre sera consacrée au système de réparation NER. Ce mécanisme ubiquitaire qui existe à la fois chez les procaryotes et les eucaryotes est de loin le plus versatile des systèmes de réparation en termes de reconnaissance des lésions. Il permet l'élimination des dommages à l'ADN qui génèrent une distorsion de la double hélice d'ADN, notamment causés par les rayonnements ultra-violet (UV).

I. LES LÉSIONS UV-INDUITES

Les lésions de l'ADN peuvent provenir à la fois de facteurs exogènes (environnementaux) et endogènes. Les facteurs exogènes comprennent notamment des agents chimiques carcinogènes et des agents physiques tels que les rayonnements UVs et les radiations ionisantes. Les sources endogènes sont plus nombreuses et proviennent surtout de l'hydrolyse spontanée des bases azotées ou de l'attaque par des radicaux libres. Cependant, des lésions peuvent également apparaître lors de processus tels que la réplication de l'ADN ou la ségrégation des chromosomes. La nature des lésions va donc varier selon l'agent qui l'aura causée. La majorité des carcinogènes environnementaux, tels que les rayonnements UVs ou les hydrocarbures contenus dans la fumée de cigarette, vont conduire à la formation de dommages de l'ADN qui peuvent bloquer les processus de réplication et de transcription (Hoeijmakers, 2001).

Les cycles aromatiques des bases constituant la molécule d'ADN absorbent l'énergie des photons. L'énergie absorbée au niveau de deux pyrimidines (cytosine ou thymine) adjacentes va permettre la formation d'une liaison covalente entre ces deux bases, au détriment des liaisons hydrogènes assurant la cohésion de la double hélice d'ADN. Les dimères de pyrimidine majoritairement induits par les UVs sont les dimères cyclobutaniques de pyrimidines (CPD) et les photo-produits pyrimidiques (6-4) pyrimidone (6-4 PP). Les UVs induisent également d'autres types de lésions sur les bases de l'ADN, comme l'hydrate de cytosine, les dimères d'adénine ou les adduits thymine-adénine, mais leur fréquence de formation est très faible.

1. Les dimères cyclobutaniques de pyrimidine (CPD)

Les CPD sont les lésions photo-induites les plus fréquentes et les plus cytotoxiques. Elles sont formées par la liaison covalente de deux pyrimidines adjacentes en une structure « cyclique » engageant les atomes de carbone C5 et C6 (Figure 18). Ce type de lésion crée une distorsion de la double hélice d'ADN d'environ 7 à 10° par rapport à sa conformation initiale (Kim et al., 1995). Cette distorsion bloque la progression des ADN polymérase lors de la réplication et de l'ARN polymérase lors de la transcription.

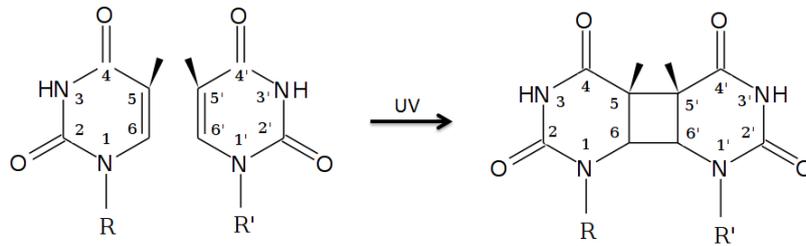


Figure 18 : Mécanisme de formation des CPD.

2. Les photo-produits pyrimidiques (6-4 PP)

Les 6-4 PP sont quantitativement les secondes lésions les plus fréquentes après une irradiation aux UVs. Le ratio entre le nombre de CPD et le nombre de 6-4 PP formés pour une dose UV donnée est estimé entre 8:1 et 3:1 (Douki et al., 2001; Perdiz et al., 2000). Les 6-4 PP résultent de la formation d'un pont stable entre les positions C6 et C4 de deux pyrimidines adjacentes (Figure 19). La déformation de l'ADN induite est beaucoup plus importante que dans le cas des CPD, elles créent une distorsion de 44° par rapport à sa conformation initiale (Kim et al., 1995). Cette distorsion bloque également la progression des ADN et ARN polymérase.

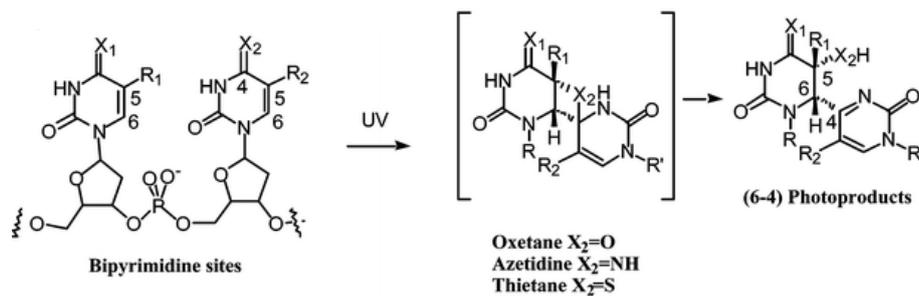


Figure 19 : Mécanisme de formation des 6-4PP.

II. MECANISME DE REPARATION DES LESIONS PAR LE SYSTEME NER

Le mécanisme de la réparation par excision de nucléotide a été largement étudié, notamment *in vitro*, à la fois chez la levure (Prakash and Prakash, 2000) et chez les mammifères (Park and Choi, 2006). La première étape consiste en la reconnaissance de la lésion par divers facteurs (Figure 20). Puis, des hélicases interviennent afin d'ouvrir la double hélice d'ADN et des endonucléases excisent le fragment d'ADN lésé. Enfin, l'ADN est resynthétisé à partir du brin non excisé qui sert de matrice aux ADN polymérase de la réparation et l'ADN ligase permet de lier l'extrémité 5' de l'ADN néosynthétisé au reste de la molécule.

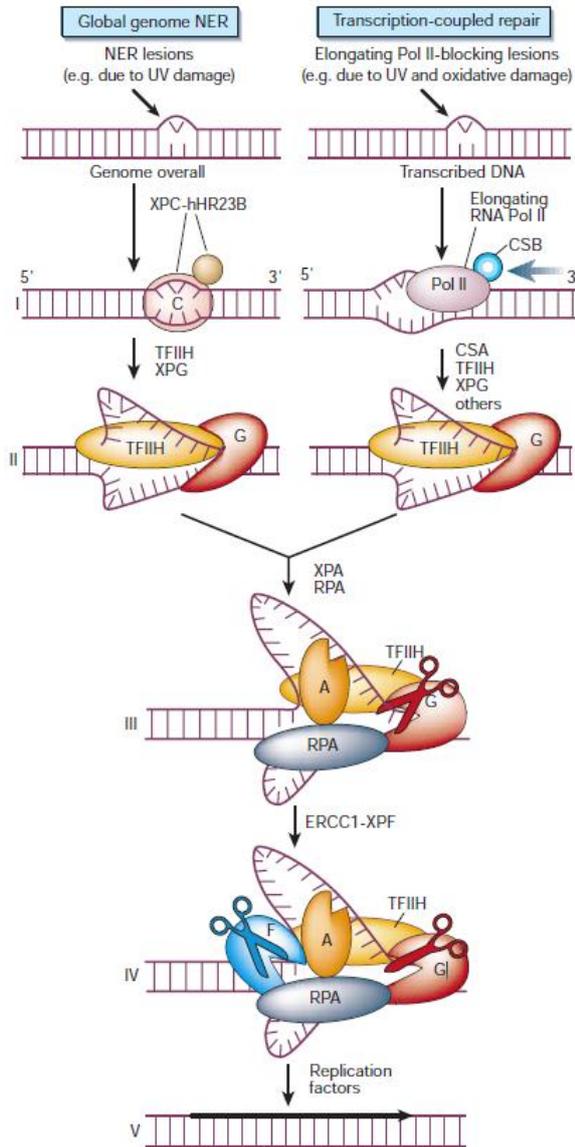


Figure 20 : Mécanismes de réparation par les voies TCR et GGR du système NER (Hoeijmakers, 2001)

En premier lieu, la lésion est reconnue soit par le complexe XPC-HR23B dans le cas de la GGR, soit par la Pol II dans le cas de la TCR. Dans ce second cas, la Pol II s'arrête et les facteurs CSB et CSA interviennent pour permettre la mise en place du complexe de réparation.

Les étapes suivantes sont les mêmes pour les deux systèmes. Les hélicases du complexe TFIIH permettent l'ouverture de l'ADN tandis que les complexes XPA et RPA confirment la lésion et stabilisent l'ADN sous forme ouverte. Puis les endonucléase XPG et XPF-ERCC1 clivent l'ADN respectivement en 3' et 5' de la lésion. Enfin, la machinerie de réplication de l'ADN intervient pour resynthétiser le brin d'ADN manquant.

Le schéma présenté ici est relativement simplifié, un schéma plus complet sera présenté plus loin dans ce chapitre.

Il existe deux types de mécanismes de réparation par excision de nucléotides (Laat et al., 1999) : (1) celui réparant les lésions sur l'ensemble du génome, appelé GG-NER ou GGR (« Global genome NER ») et (2) celui réparant les lésions sur le brin en cours de transcription appelé TC-NER ou TCR (« Transcription coupled NER »). Ces deux systèmes diffèrent seulement dans l'étape de reconnaissance de la lésion (Figure 20). Dans le système TCR, il est considéré que la Pol II bloquée au niveau de la lésion constitue un signal pour le recrutement de la machinerie de réparation alors que dans le système GGR, la distorsion induite par la lésion est reconnue par un complexe protéique spécifique.

Le tableau suivant recense les principaux facteurs impliqués dans le système NER chez l'homme et la levure (Tableau 12). La fonction de chacun sera décrite plus en détails dans les paragraphes qui suivent.

Facteur		Activité	Voie
Human	<i>S. cerevisiae</i>		
XPA	Rad14	Reconnaissance du dommage et assemblage du complexe	NER
XPB (ERCC3)	Rad25	Hélicase ATP-dépendante, Sous-unité de TFIIH	NER
XPD (ERCC2)	Rad3	Hélicase ATP-dépendante, Sous-unité de TFIIH	NER
XPC	Rad4	Reconnaissance du dommage	GGR / NER
HR23B	Rad23	Reconnaissance du dommage, forme un complexe avec XPC	GGR / NER
CEN2	Rad33	Reconnaissance du dommage, forme un complexe avec XPC	GGR / NER
XPE (DDB2)	Rad16	Forme un complexe avec E3-ub ligase, reconnaissance du dommage	GGR
DDB1	Rad7	Reconnaissance du dommage, forme un complexe avec XPE	GGR
XPF (ERCC4)	Rad1	5' Endonucléase	NER
ERCC1	Rad10	Forme un complexe avec XPF	NER
XPG	Rad2	3' Endonucléase	NER
RPA	RPA	Reconnaissance du dommage	NER
PCNA	PCNA	Réplication de l'ADN	NER
RFC1	RFC1	Recrute PCNA sur l'ADN	NER
CSA (ERCC8)	Rad28	Forme un complexe avec E3-ub ligase, interagit avec CSB	TCR
CSB (ERCC6)	Rad26	Remodelage de la chromatine et assemblage du complexe	TCR
XAB2	-	Facteur de transcription, lien entre XPA et Pol II	TCR
TFIIS	TFIIS	Stimulation du clivage du transcrit par la Pol II	TCR
LIG1	Cdc9	ADN ligase I	NER
LIG3	-	ADN ligase III	NER
XRCC1	-	ADN ligase, forme un complexe avec Lig3	NER
Pol κ	Pol κ	ADN polymérase κ	NER
Pol ϵ	Pol ϵ	ADN polymérase ϵ	NER
Pol δ	Pol δ	ADN polymérase δ	NER

Tableau 12 : Facteurs impliqués dans la réparation par excision de nucléotide (d'après Prakash and Prakash, 2000; Rechkunova et al., 2008; Shuck et al., 2008).

1. Reconnaissance de la lésion

A. Reconnaissance de la lésion par la voie GG-NER

Dans la voie GGR, les facteurs de réparation détectent la rigidité réduite de la double hélice d'ADN au niveau des lésions induites par les UVs. Plusieurs protéines interviennent dans cette étape. L'efficacité de la réparation varie le long du génome selon l'environnement chromatinien (Feng et al., 2003; Mullenders et al., 1991)

Le premier facteur à reconnaître la lésion est formé par l'association entre XPC (« Xeroderma pigmentosum complementation group C », Rad4 chez la levure) et HR23B (Rad23 chez la levure) en un complexe stable au niveau du dommage. XPC permet de lier l'ADN tandis que HR23B stimule sa fonction de liaison de l'ADN et permet de déplacer le complexe lors de l'étape suivante de la GGR (Sugasawa et al., 1998). Rad23 permettrait également de stabiliser la liaison de XPC sur l'ADN en empêchant sa dégradation par le protéasome (Lommel et al., 2002; Ng et al., 2003; Xie et al., 2004). *In vivo*, CEN2 (« Centrin 2 », Rad33 chez la levure) peut également s'associer à XPC et augmente encore l'affinité du complexe pour l'ADN (Nishi et al., 2005). Des analyses *in vitro* ont montré que XPC-HR23B se lie préférentiellement sur les molécules d'ADN présentant une distorsion (Hey et al., 2002; Trego and Turchi, 2006). Cette observation a été confirmée chez la levure par l'analyse cristallographique du complexe Rad4-Rad23 lié à une lésion CPD (Min and Pavletich, 2007). La structure obtenue montre que Rad4 lie l'ADN en 2 parties : les domaines TGD (« transglutaminase homology domain ») et BHD1 (« β -hairpin domain 1 ») se lient sur une partie de l'ADN non lésé, BHD2 et BHD3 sur le brin non lésé, face à la lésion CPD. Une autre portion du domaine TGD de Rad4 permet de lier Rad23. Cette structure montre également que Rad4 ne couvre que la partie 3' de la lésion, laissant la partie 5' libre pour les interactions ultérieures. Chez la levure, la délétion de Rad4 est létale, car TCR et GGR sont abolies, alors que chez l'homme, XPC n'est requis que dans la voie GGR (Venema et al., 1991; Verhage et al., 1994). A l'inverse, les souches *rad23Δ* ou *rad33Δ* sont viables, avec seulement un ralentissement de la réparation, mais la double délétion *rad23Δ rad33Δ* est létale, TCR et GGR étant également abolies (Dulk et al., 2006).

Lors de cette première étape de la GGR, si XPC n'est pas bien positionné il est rapidement dégradé, par un mécanisme indépendant du protéasome (Bernardes de Jesus et al., 2008). La liaison de XPC-HR23B au niveau de la lésion induit le premier changement de conformation de l'ADN (Evans et al., 1997; Tapias et al., 2004), indispensable au recrutement ultérieur de la machinerie de réparation, notamment de TFIIH (Mocquet et al., 2007; Sugasawa and Hanaoka, 2007).

Le facteur XPE ou DDB2 (« damage DNA binding protein 2 », Rad16 chez la levure) intervient également dans la reconnaissance de la lésion, indépendamment de XPC-HR23B (Luijsterburg et al., 2007). DDB2 interagit avec DDB1 (Rad7 chez la levure) et forme le complexe UV-DDB (« UV-damaged DNA binding protein ») requis pour la détection et la réparation des CPD *in vivo* (Tang et al., 2000). La structure du complexe UV-DDB humain lié à une lésion CPD a été résolue (Fischer et al., 2011). Celle-ci montre que DDB2 se lie à l'ADN contenant la lésion via une poche hydrophobe dans laquelle la lésion va venir s'insérer. Chez la levure, le complexe Rad7-Rad16 possède une activité ATPase qui est inhibée en présence d'un dommage lié aux UVs (Guzder et al., 1997, 1998). Une des hypothèses quant au rôle de ce complexe *in vivo* serait qu'il scanne constamment le génome, en utilisant l'hydrolyse de l'ATP pour se déplacer, et l'inhibition de son activité ATPase au niveau d'une lésion induirait sa stabilisation sur l'ADN. Lorsqu'il se lie, DDB2 induit une torsion de l'ADN qui serait indispensable à la reconnaissance de la lésion par les autres facteurs dans le cas d'une lésion ne modifiant pas la structure de l'ADN (Fitch et al., 2003). Chez la levure, le complexe Rad7-Rad16 est requis *in vivo* (Guzder et al., 1995), son association avec le complexe Rad4-Rad23 augmente l'efficacité de liaison à l'ADN lésé (Guzder et al., 1999).

Des analyses biochimiques suggèrent que UV-DDB fait partie du complexe E3 ubiquitine ligase (E3-ub ligase) qui contient, entre autres, l'ubiquitine ligase CUL4A (Shiyanov et al., 1999) et le signalosome Cop9 (CSN), un régulateur négatif de l'activité ubiquitine ligase (Groisman et al., 2003). L'activité ubiquitine ligase du complexe E3-ub ligase est augmentée suite à une irradiation UV et plusieurs de ses cibles ont été mises en évidence, notamment DDB2, XPC et les histones H2A, H3 et H4. L'ubiquitination des histones permet de conférer un environnement chromatinien dans lequel la liaison des facteurs de réparation est facilitée (cf. III. de ce chapitre). Elle permettrait également la liaison de XPC-HR23B sur l'ADN via son domaine UBA (« ubiquitin-associated domain ») (Bertolaet et al., 2001). Une fois cette première étape réalisée, DDB2 et XPC sont, à leur tour, ubiquitinylés. L'ubiquitination de DDB2 permet sa dégradation par le protéasome tandis que l'ubiquitination réversible de XPC accroît son affinité pour l'ADN (Sugasawa et al., 2005) et est indispensable pour le recrutement des autres facteurs (Gillette et al., 2006).

B. Reconnaissance de la lésion par la voie TC-NER

La voie TCR confère un système de réparation de l'ADN plus rapide et plus efficace sur le brin transcrit que les autres systèmes de réparation, afin de permettre une reprise rapide de la transcription des gènes (Bohr et al., 1985; Mellon and Hanawalt, 1989). Ce système reconnaît les lésions qui bloquent l'activité transcriptionnelle de la Pol II. Les dommages sont réparés plus

efficacement sur le brin transcrit que sur le brin non transcrit et proportionnellement au niveau transcriptionnel.

Le premier signal pour recruter la machinerie de réparation est conféré par la Pol II hyperphosphorylée, bloquée au niveau d'une lésion (Lainé and Egly, 2006; Sarker et al., 2005). Cependant, les facteurs CSA (« Cockayne syndrome A protein », Rad28 chez la levure) et CSB (« Cockayne syndrome B protein », Rad26 chez la levure) sont indispensables à l'initiation de la réparation, des mutations dans ces protéines engendrant chez l'homme une forte sensibilité aux UVs (Henning et al., 1995; Troelstra et al., 1992). Chez la levure, les mutations de Rad26 (van Gool et al., 1994), mais pas celles de Rad28 (Bhatia et al., 1996), affectent fortement la TCR. Cependant, la délétion de l'un ou l'autre de ces facteurs ne confère pas de sensibilité particulière aux UVs. Cette constatation a permis de mettre en évidence l'existence de deux voies de réparation TCR chez la levure, une dépendante de Rad26 et une totalement indépendante de Rad26 mais dépendante de Rpb9, une sous-unité non essentielle de la Pol II. Rpb4, une autre sous-unité non-essentielle de la Pol II, joue un rôle dans la régulation des deux voies en supprimant la voie Rpb9-dépendante afin de faciliter celle dépendant de Rad26. Les doubles mutants *rad26Δ rpb9Δ* sont totalement défectueux pour la TCR (Li and Smerdon, 2002).

CSB est une ATPase capable d'interagir avec l'ADN et qui partage de nombreuses homologies avec les membres de la famille des remodeleurs SWI2/SNF2 excepté l'activité hélicase (Guzder et al., 1996) (cf. III). Après irradiation UV, la Pol II interagit avec CSB (van den Boom et al., 2004). Le complexe stable ainsi formé va permettre le recrutement ultérieur de la machinerie de réparation (Fousteri et al., 2006; Lainé and Egly, 2006). Des analyses réalisées chez la levure suggèrent que Rad26 serait associée à la Pol II durant l'étape d'élongation de la transcription, elle pourrait ainsi être directement recrutée sur la lésion en cas d'arrêt de la Pol II (Malik et al., 2010). En accord avec cette hypothèse, des expériences de filtration sur gel, réalisées chez l'homme, ont montré que CSB et la Pol II co-purifient (van Gool et al., 1997). De plus, une mutation dans le domaine ATPase de CSB augmente la sensibilité aux UVs et l'apoptose et diminue la reprise de la transcription, confirmant le rôle essentiel de l'activité ATPase de CSB dans la réparation (Citterio et al., 1998; Muftuoglu et al., 2002; Selzer et al., 2002).

CSA, tout comme DDB2 dans la voie GGR, forme un complexe E3-ub ligase dans lequel, en partenariat avec CSN, il inhibe l'activité ubiquitine ligase. Cependant, la cinétique d'activation de ces deux complexes est bien différente. Dans la GGR, l'ubiquitine ligase est activée juste après irradiation UV et inhibée plus tard par l'association de CSN. A l'inverse, dans la TCR, l'ubiquitine ligase est inhibée au début de la réparation par CSN et devient active plus tard pour permettre la dégradation des facteurs de réparation à la fin du processus (Groisman et al., 2003). Une étude a

d'ailleurs montré qu'après quelques heures de culture suite à l'irradiation UV, CSB serait ubiquitinylé par le complexe E3-ub ligase puis dégradé par le protéasome et que cette dégradation serait importante pour la reprise de la transcription (Groisman et al., 2006). Cependant, des données plus récentes suggèrent que CSB est simplement libéré (Lake et al., 2010).

CSA et CSB interagissent *in vitro* ainsi que dans un système double hybride chez la levure (Henning et al., 1995). La liaison de CSA à la Pol II bloquée est dépendante des UVs et de CSB (Kamiuchi et al., 2002), alors que CSB peut également interagir avec la Pol II en absence de lésion (Tantin et al., 1997). CSA fonctionne comme une interface pour les interactions protéine-protéine (Bhatia et al., 1996; Henning et al., 1995) notamment via ses domaines WD-40. Il permet le recrutement de HMGN1 (« high mobility group N non-histone nucleosome binding protein ») (Lim et al., 2005); du facteur d'épissage XAB2 (« XPA-binding protein 2 »); et du facteur de transcription TFIIIS (Fousteri et al., 2006). A l'inverse de CSB, CSA n'est pas requis pour l'assemblage des facteurs de réparation.

XAB2 est une protéine impliquée dans l'épissage des ARNm, elle est capable de lier XPA et pourrait fonctionner comme un échafaudage pour l'assemblage des protéines du système TCR (Nakatsu et al., 2000). Au moment de l'arrêt de la transcription, la Pol II recule, avec pour conséquence un décalage de l'extrémité du transcrit en cours de synthèse par rapport au site actif. Le facteur TFIIIS va alors stimuler l'activité de clivage de la Pol II pour permettre le repositionnement correct du transcrit et la reprise de la transcription après élimination de la lésion bloquant la Pol II (Kalogeraki et al., 2005).

2. Le complexe de pré-incision

Les étapes suivantes sont les mêmes pour les deux voies GGR et TCR. Les analyses *in vitro* montrent que le facteur XPC-HR23B dans le premier cas et CSB dans le second, permet le recrutement des divers facteurs de réparations : XPA, RPA (« Replication factor A »), TFIIH, XPG et XPF-ERCC1 pour former le complexe de pré-incision (Lainé and Egly, 2006; Mocquet et al., 2008).

A. Ouverture de l'ADN

L'ouverture de l'ADN serait la première étape non réversible de la réparation par le système NER (Gillet and Schärer, 2006). Cette étape est réalisée par les deux hélicases XPB et XPD du complexe TFIIH. Elle nécessite également les facteurs RPA et XPA qui vont stabiliser et orienter le complexe de pré-incision et permettre le recrutement ultérieur des endonucléases.

a. TFIIH

TFIIH est un complexe constitué de 10 sous-unités réparties en 2 modules : un module cœur de 7 sous-unités et un module kinase CAK de 3 sous-unités, ce dernier étant dispensable dans la réparation (Mu et al., 1996; Sung et al., 1996). Il intervient à la fois dans la réparation, dans la transcription (cf. Chapitre I) et dans le cycle cellulaire.

Dans la voie GGR, quand XPC est correctement positionné au niveau du dommage, sa partie C-terminale peut adopter une structure tridimensionnelle qui permet le recrutement de TFIIH. XPC interagit avec plusieurs sous-unités de TFIIH : p62 (Tfb1 chez la levure) pour permettre le recrutement du complexe sur la lésion (Volker et al., 2001; Yokoi et al., 2000) ; et p8 (Tfb5 chez la levure), p52 (Tfb2 chez la levure) et XPB pour stimuler son activité ATPase (Bernardes de Jesus et al., 2008; Coin et al., 2006). La sous-unité p8 permet également de stabiliser TFIIH dans le complexe de pré-incision puisque la concentration en TFIIH au niveau de la lésion diminue fortement quand p8 est mutée (Coin et al., 2006; Giglia-Mari et al., 2004). Cependant, cette sous-unité est la seule non essentielle du complexe TFIIH, sa délétion chez *S. cerevisiae* a pour conséquence un ralentissement de la croissance et une sensibilité aux UVs (Ranish et al., 2004; Zhou et al., 2007). Le rôle des autres sous-unités est probablement de maintenir la structure de TFIIH via des interactions protéine-protéine. Certaines études suggèrent que TFIIH pourrait être le premier facteur à intervenir dans la TCR, en effet ce complexe est capable de se lier précocement à la Pol II bloquée par une lésion (Dvir et al., 1997).

Les sous-unités XPB (Ssl2 ou Rad25 chez la levure) et XPD (Rad3 chez la levure) de TFIIH sont des hélicases avec une activité ATPase dépendante de l'ADN. XPB déroule l'ADN de 3' en 5' et XPD de 5' en 3' (Schaeffer et al., 1993). Des mutations dans XPD conduisent à une diminution de la capacité de TFIIH à ouvrir l'ADN (Coin et al., 1998). Son activité hélicase est absolument requise dans la réparation (Dubaele et al., 2003; Tirode et al., 1999), des mutations de Rad3 conduisant à une sensibilité aux UVs. Par contre, elle n'est pas indispensable dans la transcription puisque des mutations affectant l'activité hélicase de XPD n'ont pas d'effet sur la viabilité cellulaire (Sung et al., 1988). A l'inverse, une mutation affectant l'activité hélicase de Rad25 est létale, montrant son rôle essentiel dans la transcription (Guzder et al., 1994; Qiu et al., 1993). Dans la réparation, XPB n'est cependant requise que pour son activité ATPase (Coin et al., 2007). Elle jouerait un rôle de « cale » (Fan et al., 2006) utilisant l'hydrolyse de l'ATP pour garder les deux brins d'ADN séparés pendant que XPD exerce son activité hélicase, déroulant l'ADN sur environ 30 nt autour de la lésion (Evans et al., 1997). XPB et p8 permettent également la réorganisation du complexe de pré-incision en repositionnant XPC-R23B en aval de la lésion pour permettre le recrutement ultérieur de XPA et RPA (Tapias et al., 2004). XPD jouerait également un rôle dans la vérification du

dommage, en association avec RPA et XPA (Mathieu et al., 2010), la fonction de XPC serait alors limitée à la reconnaissance d'une torsion de l'ADN, XPD permettant de vérifier la réelle présence d'une lésion (Reardon and Sancar, 2003). Sans cette étape de vérification, le processus de réparation est interrompu (Sugasawa et al., 2001)

b. RPA-XPA

XPC-HR23B permet le recrutement de RPA-XPA (RPA-Rad14 chez la levure) (You et al., 2003) qui va stabiliser le complexe déjà en place (Ikegami et al., 1998; Park and Sancar, 1994) et étendre la bulle d'ADN autour du dommage (Evans et al., 1997). Le complexe RPA-XPA possède une meilleure affinité pour le dommage que RPA ou XPA seul (He et al., 1995; Li et al., 1995; Matsuda et al., 1995). RPA et XPA vont alors orchestrer le recrutement des nucléases XPG et XPF-ERCC1. Chez la levure, le recrutement de RPA-Rad14 conduit à la libération de Rad4-Rad23-Rad33.

RPA possède une affinité pour l'ADN simple brin et permet de stabiliser l'ADN simple brin produit par TFIIH en interagissant avec les 2 bords opposés de la lésion (de Laat et al., 1998a). Plusieurs analyses suggèrent que RPA possède la capacité de séparer les deux brins complémentaires de l'ADN (Georgaki et al., 1992; Treuner et al., 1996). *In vitro*, RPA interagit directement avec TFIIH (Guzder et al., 1995). En absence de RPA, tous les facteurs du complexe de pré-incision sont tout de même recrutés au niveau de la lésion après irradiation UV. Cependant, RPA est requis pour stabiliser le complexe, l'orienter correctement et activer les endonucléases XPG et XPF-ERCC1 (Overmeer et al., 2011). En présence de XPA, RPA interagit préférentiellement avec le brin non lésé, le protégeant ainsi de l'activité des endonucléases (Lee et al., 2003). XPA quant à lui se lie préférentiellement au brin lésé (Guzder et al., 1993) et stabilise le complexe en inhibant la séparation des deux brins d'ADN induite par RPA (Patrick and Turchi, 2002). Son activité est requise *in vitro* et *in vivo* (Guzder et al., 1995). Sa liaison au complexe de pré-incision induit le détachement du module kinase (CAK) de TFIIH, qui aurait un rôle inhibiteur de la réparation, stimulant ainsi l'ouverture de l'ADN (Coin et al., 2008).

B. Excision de la lésion : XPG et XPF

L'ouverture l'ADN créée par TFIIH, XPA et RPA permet le recrutement des endonucléases XPG (Rad2 chez la levure) et XPF-ERCC1 (Rad1-Rad10 chez la levure) qui incisent l'ADN en 3' et en 5' de la lésion, respectivement (Habraken et al., 1993; Mu et al., 1995). L'excision n'a lieu que sur le brin lésé (Hess et al., 1997). XPG et XPF-ERCC1 sont des nucléases sensibles à la structure de l'ADN, capables de dégrader l'ADN simple brin, et qui vont agir préférentiellement près des jonctions ADN simple brin – ADN double brin (Sijbers et al., 1996). Aucune de ces deux nucléases

n'a cependant de spécificité pour l'ADN lésé où elles doivent donc être dirigées via leur interaction avec d'autres protéines.

XPG est une 3' endonucléase qui possède deux domaines nucléases N et I séparés par une région « spacer », ce qui la rattache à la même famille que Fen1 et Exo1 (Scherly et al., 1993) (Figure 21). Les analyses structurales suggèrent que les deux domaines N et I forment le cœur catalytique de l'enzyme (Hosfield et al., 2001). La région « spacer » est importante pour les interactions avec TFIIH et RPA (He et al., 1995; Thorel et al., 2004) et permet le recrutement de XPG sur la lésion (Dunand-Sauthier et al., 2005). Cette région forme une boucle, qui doit subir une modification de structure pour rendre l'enzyme catalytiquement active. XPG possède plusieurs sites d'interaction avec TFIIH (Araújo et al., 2001; Iyer et al., 1996), créant un lien très fort et très stable entre ces deux facteurs. Dans la GGR, c'est cette interaction qui va permettre le recrutement de XPG au sein du complexe de pré-incision (Zotter et al., 2006). Dans la TCR, XPG pourrait être recruté au tout début de la réparation, en même temps que CSB, en coopération avec ce dernier ou indépendamment, et interagirait directement avec la Pol II bloquée (Sarker et al., 2005). Avant d'exercer son rôle d'endonucléase, XPG possède un rôle structural. Arrivé sur la lésion, il forme un complexe très stable avec TFIIH et RPA-XPA. A ce stade son activité endonucléase est nulle mais son recrutement induit un changement de conformation requis pour le recrutement de XPF-ERCC1 (Dunand-Sauthier et al., 2005). Plusieurs facteurs ont été proposés comme régulateurs de l'activité nucléase de XPG. RPA est capable de stimuler l'incision par XPG *in vitro* (Overmeer et al., 2011) et serait donc un bon candidat pour l'activation de la fonction nucléase de XPG. PCNA interagit avec XPG (Gary et al., 1997) et permet l'incision de l'ADN par Fen1 durant la réplication (Chapados et al., 2004). Un rôle similaire peut donc être imaginé dans l'activation de XPG. Ainsi, les activités de liaison à l'ADN et de clivage de l'ADN sont totalement découplées chez XPG et XPF empêchant ainsi les incisions non spécifiques.



Figure 21 : Structure primaire de l'endonucléase XPG (Fagbemi et al., 2011)

Les domaines nucléases N et I sont représentés en bleu clair. La région « spacer » est indiquée en rose, elle contient un site de liaison pour l'ubiquitine (UBM) en bleu foncé. Les résidus des sites actifs qui ont été analysés sont indiqués en rouge. Enfin sont également présentés les sites d'interaction avec TFIIH, RPA et PCNA (PIP).

XPF-ERCC1 chez l'homme et Rad1-Rad10 chez la levure, forment un hétérodimère obligatoire, chacune des deux protéines étant instable en l'absence de l'autre (van Vuuren et al., 1993) (Figure 22). Leur dimérisation s'effectue grâce à deux motifs hélice-épingle-hélice (HhH « helix – hairpin – helix ») à l'extrémité C-terminale de chacune des deux protéines (de Laat et al., 1998b). Le motif HhH ainsi que la partie N-terminale de XPF sont également importants pour la liaison du complexe sur l'ADN (Tsodikov et al., 2005). Le domaine nucléase est porté par XPF (Enzlin and Schäfer, 2002). L'hétérodimère XPF-ERCC1 interagit spécifiquement avec XPA qui permet son recrutement (Li et al., 1994). Un modèle de l'interaction entre XPA et ERCC1 a été obtenu par cristallographie (Orelli et al., 2010; Tsodikov et al., 2005). Celui-ci montre que seulement une petite région de XPA est requise pour l'interaction avec ERCC1. *In vitro*, ce peptide de XPA en solution inhibe la réparation par le système NER. Suite à l'interaction avec ERCC1 il passe d'une structure désordonnée à une structure ordonnée et n'est alors plus inhibiteur.

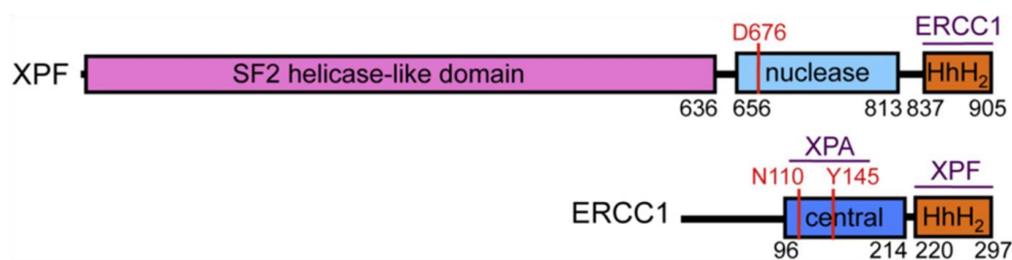


Figure 22 : Structure primaire de l'endonucléase XPF-ERCC1 (Fagbemi et al., 2011)

Les domaines hélice-épingle-hélice (HhH) par lesquels XPF et ERCC1 interagissent sont indiqués en orange. Le domaine nucléase de XPF est représenté en bleu avec le site actif en rouge. Le domaine assimilé aux hélicases de la famille SF2 est indiqué en rose. Enfin est également représenté le site d'interaction avec XPA, les résidus important pour cette interaction sont notés en rouge.

Chez la levure, Rad1-Rad10 forme un complexe avec Rad14 (Guzder et al., 2006), tandis que Rad2 interagit avec TFIIH via Tfb1 (LaFrance-Vanasse et al., 2012). Ces interactions permettraient leur recrutement au sein du complexe de pré-incision. Rad2 crée alors une incision en 3' de la lésion et Rad1-Rad10 en 5'. L'oligonucléotide généré, toujours lié par les facteurs de réparation, peut alors être libéré et les étapes de re-synthèse et ligation peuvent avoir lieu. Chez l'homme, un mécanisme différent a été proposé. L'incision en 5' par XPF-ERCC1 est nécessaire pour révéler l'activité endonucléase de XPG (Constantinou et al., 1999; Tapias et al., 2004; Wakasugi et al., 1997). Elle va également permettre la libération des facteurs XPC, XPA et TFIIH qui pourront être recyclés (Mocquet et al., 2008). Réciproquement, l'activité endonucléase de XPF-ERCC1 requiert XPB puisque des mutations de ce facteur inhibent l'incision en 5' (Coin et al., 2004). Enfin, l'initiation de la synthèse d'ADN par la polymérase pour réparer la lésion ne nécessite pas l'activité catalytique de XPG et peut donc démarrer juste après l'incision générée par XPF-ERCC1, avant l'incision par XPG (Staresincic et al., 2009).

Les endonucléases XPG et XPF-ERCC1 ne sont pas spécifiques au système de réparation NER. XPF-ERCC1 intervient également dans les systèmes BER et HR (Lyndaker and Alani, 2009) et XPG dans le système BER (Bessho, 1999; Klungland et al., 1999). De plus, le phénotype très grave observé chez les patients présentant des mutations dans XPG, avec notamment des retards de croissance, suggère que ce facteur pourrait également avoir un rôle en dehors de la réparation (van Gool et al., 1997). Chez la levure, un rôle transcriptionnel a été suggéré pour Rad2 (cf. IV. de ce chapitre). Cette protéine aurait également un rôle dans la régulation du cycle cellulaire après une irradiation UV (Kang et al., 2014).

3. Re-synthèse et ligation

L'excision du dommage est suivie par l'assemblage du complexe de post-incision qui va permettre la re-synthèse d'ADN et la ligation des extrémités du nouveau fragment synthétisé à la molécule d'ADN initiale.

Après libération de l'oligonucléotide contenant la lésion, la protéine PCNA (« proliferating cell nuclear antigen ») qui appartient à la famille des clamps glissants, est recrutée par RFC (« Replication factor C ») (Kelman, 1997). Elle forme alors une pince à laquelle différentes ADN polymérases peuvent s'associer. Chez l'homme, Pol ϵ , Pol δ et Pol κ sont capables de réparer la lésion en utilisant le brin non endommagé comme matrice (Aboussekhra et al., 1995; Araújo et al., 2000; Ogi and Lehmann, 2006). Le choix de l'une ou l'autre des polymérases pour effectuer la réparation dépendrait de l'étape du cycle cellulaire. Ainsi, Pol δ et Pol κ seraient les polymérases préférentielles des cellules quiescentes tandis que les cellules en division nécessiteraient également Pol ϵ (Ogi et al., 2010). Chez la levure, Pol ϵ et Pol δ pourraient réaliser cette étape (Budd and Campbell, 1995).

L'étape de ligation qui suit est effectuée par XRCC1-ligase III (XRCC1-LIG3) dans les cellules quiescentes et nécessite en plus la ligase I (LIG1) dans les cellules en division (Moser et al., 2007). Chez la levure, il semblerait que LIG1 (CDC19) réalise seule cette étape (Budd and Campbell, 1995), aucun homologue de LIG3 n'ayant à ce jour été révélé chez cet organisme.

4. Séquences d'assemblage des facteurs NER *in vitro*

Des systèmes reconstitués *in vitro* ont permis d'étudier l'ordre d'assemblage des différents facteurs de réparation notamment dans la voie GGR (Aboussekhra et al., 1995; Park and Sancar, 1994; Wood et al., 1988). Six facteurs sont essentiels et suffisants pour reconnaître la lésion et l'exciser à la fois chez l'homme et la levure : XPC-HR23B, TFIIH, RPA, XPA, XPG et XPF-ERCC1 (Guzder et al., 1995; Mu et al., 1995).

Plusieurs modèles ont été proposés quant au mode d'assemblage des facteurs de réparation *in vitro*. De nombreux complexes multiprotéiques pré-assemblés existent à la fois chez la levure et chez l'homme : XPF-ERCC1-XPA (Rad1-Rad10-Rad14=NEF1), XPC-HR23B (Rad4-Rad23=NEF2), XPG-TFIIH (Rad2-TFIIH=NEF3), DDB1-XPE (Rad7-Rad16=NEF4) et pourraient être recrutés tels quels en un assemblage séquentiel et ordonné. Chez la levure, la lésion serait reconnue par NEF2 et NEF4 combinés. NEF1 et NEF3 arriveraient en second lieu suivis par RPA (Guzder et al., 1999). Chez l'homme, aucun consensus n'a encore permis de déterminer de façon définitive le premier facteur à reconnaître la lésion. Les candidats pour cette fonction sont les facteurs ayant le plus d'affinité pour l'ADN lésé : XPC-HR23B, XPA, DDB et RPA. Cependant, aucun de ces facteurs ne possède seul l'affinité suffisante à la reconnaissance du dommage, une coopération est donc nécessaire. La majorité des modèles suggèrent que XPC-HR23B serait le premier à se lier (Sugasawa et al., 1998). Ainsi une lésion pré-incubée avec ce facteur est plus rapidement réparée qu'une lésion pré-incubée avec XPA-RPA (Hey et al., 2002). Dans ce cas, XPA-RPA n'interviendrait qu'en second lieu, pour aider TFIIH à ouvrir l'ADN. Cependant, sur certains types de lésions, XPA et RPA interviennent avant XPC (Volker et al., 2001; Wakasugi and Sancar, 1999). Les modèles plus récents montrent que XPC peut s'accumuler sur la lésion en absence de XPA, alors que le recrutement de XPA nécessite XPC. De plus, TFIIH, RPA et XPG ne nécessitent pas XPA pour leur recrutement (Rademakers et al., 2003; Reardon and Sancar, 2002).

Des modèles mathématiques ont été conçus afin de trouver un consensus dans le mode d'assemblage des facteurs de réparation. Les résultats obtenus sont disparates selon les laboratoires, même si tous s'accordent sur le fait qu'un processus ordonné résulte en une excision plus efficace (Politi et al., 2005). Finalement, il se pourrait que le recrutement des facteurs de reconnaissance de la lésion n'ait pas d'ordre précis. La lésion serait reconnue par XPA, RPA ou XPC. L'effet coopératif des trois permettrait de créer le complexe de réparation. Un modèle mathématique récent va dans ce sens d'un recrutement aléatoire. Il permettrait d'atteindre une spécificité de reconnaissance des lésion plus importante et rendrait le processus plus flexible et multifonctionnel (Kessler et al., 2007).

Les deux schémas ci-après présentent les modèles d'assemblage proposés chez la levure (Figure 23) et chez les mammifères (Figure 24).

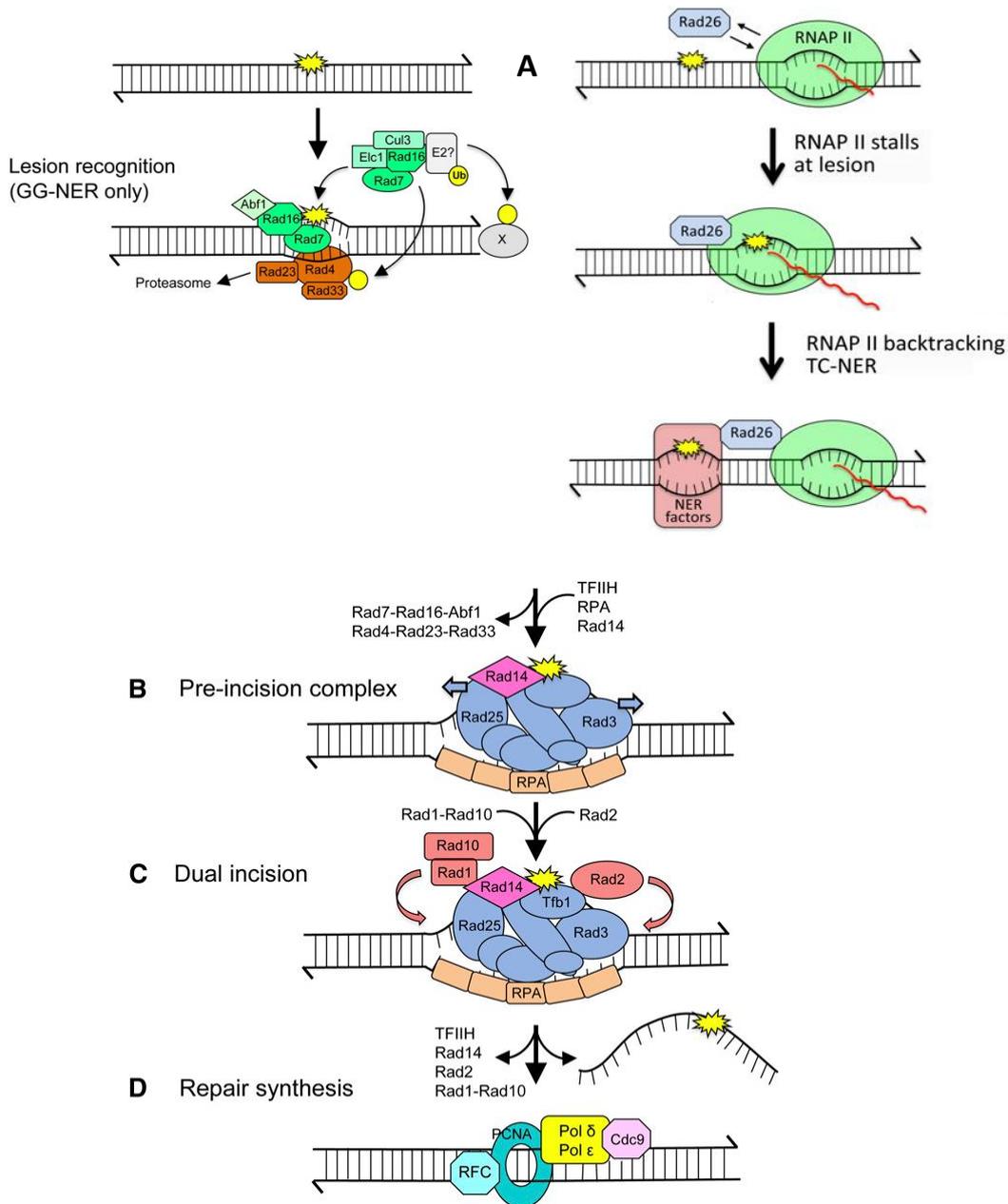


Figure 23 : Modèle probable d'assemblage des facteurs de réparation dans la GGR et la TCR chez la levure (Boiteux and Jinks-Robertson, 2013).

(A) Dans la GGR, une torsion de l'ADN est reconnue par les complexes Rad4-Rad23-Rad33 et Rad7-Rad16 (A gauche). Le complexe Rad7-Rad16-Abf1 a une activité de remodelage de la chromatine, tandis que Rad7-Rad16-Elc1-Cul3-E2 a une activité ubiquitine-ligase qui modifie Rad4 et va lui permettre de reconnaître la lésion et d'ouvrir l'ADN. Dans la TCR, la Pol II bloquée au niveau d'une lésion stabilise son interaction avec Rad26 (A droite). La Pol II recule ce qui rend la lésion accessible aux facteurs de réparation. (B) Puis TFIIH, Rad14 et RPA sont recrutés pour former le complexe de pré-incision qui vérifie la lésion et déroule l'ADN de part et d'autre de celle-ci. (C) Rad4-Rad23-Rad33 et Rad7-Rad16-Abf1 sont libérés et les endonucléases Rad1-Rad10 et Rad2 incisent la lésion. (D) L'oligonucléotide contenant la lésion est libéré, suivi par la resynthèse et la ligation.

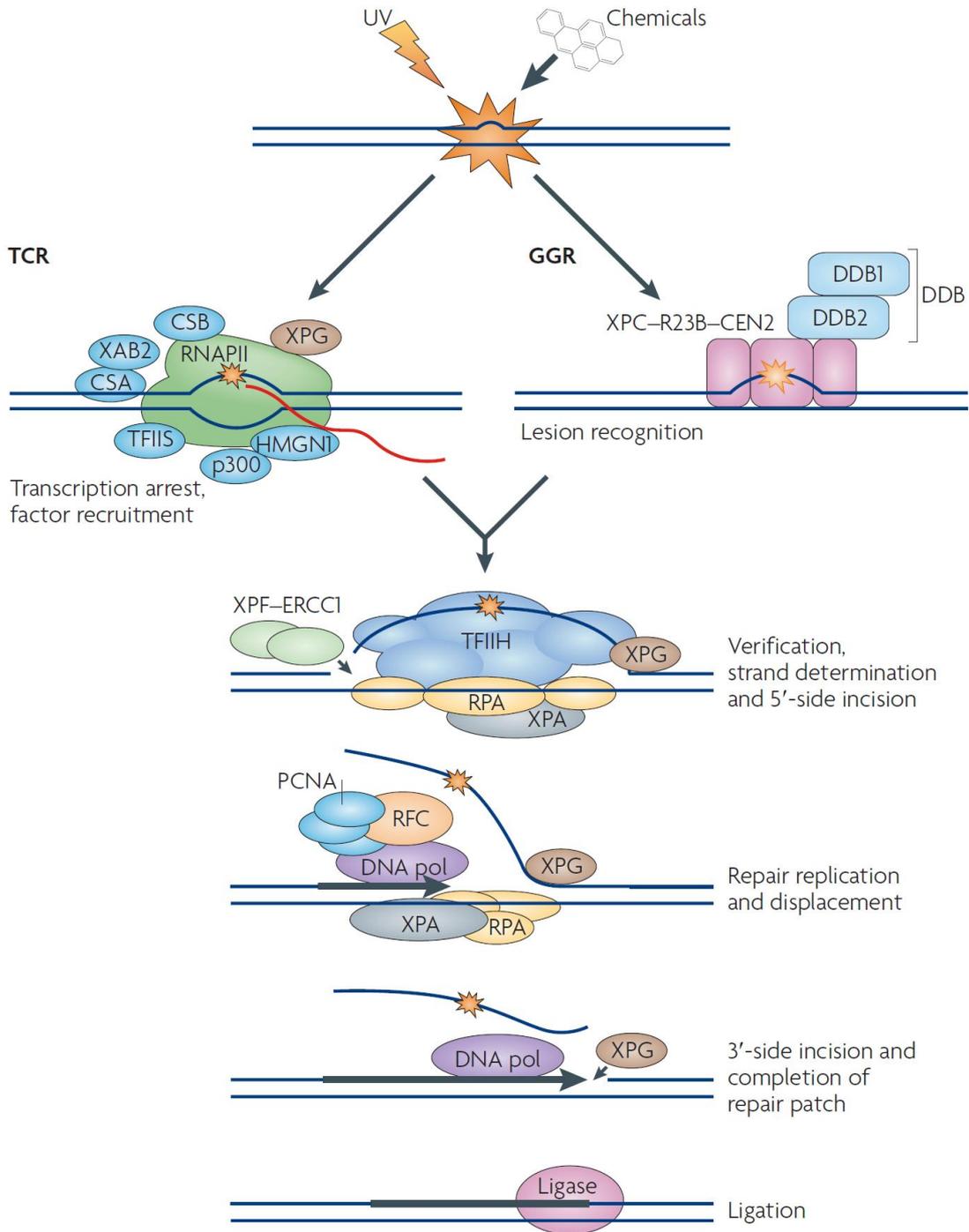


Figure 24 : Modèle probable d'assemblage des facteurs de réparation dans la TCR et la GGR chez les mammifères (Hanawalt and Spivak, 2008).

La lésion est reconnue soit par la Pol II bloquée (TCR à gauche) soit par un des complexes DDB ou XPC-R23B-CEN2 (GGR à droite). Puis les deux voies convergent. RPA et XPA interviennent afin de vérifier la lésion et stabiliser le complexe de pré-incision. Le facteur TFIIH est recruté en même temps que XPG. Les sous-unités XPD et XPB de TFIIH ouvrent l'ADN. L'endonucléase XPF-ERCC1 incise en 5' de la lésion et RFC recrute PCNA qui permet l'arrivée de l'ADN polymérase. La re-synthèse d'ADN débute avant que n'ait eu lieu l'incision en 3' par XPG. L'étape finale de ligation est réalisée par LIG1 ou LIG3-XRCC1.

III. CONTEXTE CHROMATINIEN

La réparation fonctionne dans un contexte chromatinien et nécessite donc des modifications de la chromatine pour avoir accès à la lésion. Il existe deux moyens de modifier la structure de la chromatine (cf. Chapitre I. III.) : (1) la modification des histones, afin de diminuer leur affinité pour l'ADN et/ou (2) l'intervention de remodeleurs qui vont modifier les contacts histones – ADN.

On sait depuis longtemps que la chromatine subit des réarrangements pendant la réparation. Ainsi en 1978, Smerdon et Lieberman décrivent le modèle ARR : accès, réparation, restauration (Smerdon and Lieberman, 1978) dans lequel la chromatine subit une première modification qui va donner accès aux facteurs de réparation, puis sa structure est restaurée à l'état initial en fin de processus. Ce modèle est soutenu par les analyses *in vitro* qui montrent que la vitesse de réparation est bien plus rapide sur un ADN nu que sur des nucléosomes reconstitués et que dans ce dernier cas, la vitesse peut être augmentée par ajout de remodeleurs tels que SWI2/SNF2 (Hara et al., 2000; Ura et al., 2001).

UV-DDB (DDB2-DDB1), qui intervient dans la voie GGR, possède un rôle clé en facilitant la réparation dans un contexte chromatinien. Ce complexe s'associe avec CUL4A et ROC1 et va ubiquitinyler les histones H2A et H3, réduisant leur interaction avec l'ADN, avec pour conséquence une déstabilisation des nucléosomes (Bergink et al., 2006; Guerrero-Santoro et al., 2008; Wang et al., 2006a; Kapetanaki et al., 2006). Cependant, DDB2 possède également une fonction indépendante de CUL4A et ROC1 en facilitant la polyribosylation par l'enzyme PARP1 (« poly-ADP ribose polymerase 1 ») qui permet le recrutement de ALC1, un remodeleur de la famille SWI/SNF (Pines et al., 2012). Dans les cellules humaines, une délétion d'ALC1 induit un défaut partiel de la réparation des lésions UV-induites et une sensibilité accrue aux UVs. Chez *S. cerevisiae*, Rad16 comprend un domaine SWI2/SNF2 et possède une activité d'ADN translocase (Yu et al., 2004, 2009). Ce facteur contrôle également l'acétylation de l'histone H3 sur les lysines 9 et 14 par Gcn5 (Teng et al., 2008). Ainsi, l'occupation de Gcn5, et par conséquent l'acétylation de H3, augmente en réponse aux UVs (Yu et al., 2011). L'incorporation du variant Htz1 de H2A facilite également la NER. Sa suppression diminue l'occupation de Gcn5 et donc l'acétylation de H3 avec pour conséquence une diminution du recrutement de Rad14. CSB, qui intervient dans la voie TCR, possède une activité de remodelage intrinsèque et se lie aux histones *in vitro* (Citterio et al., 2000). Ce facteur est également important pour le recrutement de l'HAT p300 tandis que CSA permet le recrutement du remodeleur HMGN1 (Fousteri et al., 2006).

De nombreux remodeleurs de la famille SWI/SNF interviennent dans la réparation par le système NER. C'est le cas des facteurs BRG1 (ou SMARCA4) et SNF5 chez l'homme (Ray et al.,

2009; Zhao et al., 2009), et SNF6 chez la levure (Gong et al., 2006). Ces remodeleurs sont recrutés par XPC/Rad4 et interviendraient dans les étapes finales de la réparation. La délétion de INO80, un autre membre de la famille SWI/SNF induit une légère sensibilité aux UVs et un ralentissement de la réparation des lésions CPD et 6-4PP (Jiang et al., 2010). Chez l'homme, ce remodeleur interagit avec DDB1 qui pourrait permettre son recrutement, et permet le recrutement ultérieur de XPA et XPC. Ces données suggèrent qu'INO80 intervient dans les étapes précoces de la réparation. A l'inverse, chez *S. cerevisiae*, INO80 interagit avec Rad4-Rad23 et participe à la restauration de la chromatine après la réparation (Sarkar et al., 2010). Ce remodeleur intervient donc à des étapes différentes de la réparation chez l'homme et la levure.

L'hétérodimère CAF1 (« chromatin assembly factor 1) et la chaperone d'histone ASF1 sont recrutés par PCNA à la fin du processus de réparation pour permettre le réassemblage de la chromatine en recyclant les histones (Gaillard et al., 1996; Green and Almouzni, 2003). CAF1 incorpore également le variant d'histone H3.1 sur l'ADN néosynthétisé ce qui permettrait la maintenance de l'information épigénétique mais également de conserver la trace d'une lésion passée (Polo et al., 2006).

IV. COUPLAGE DE LA REPARATION AVEC LA TRANSCRIPTION

Les arrêts de la transcription causés par une lésion de l'ADN sont très délétères pour la cellule en affectant l'expression des gènes et, par conséquent, la croissance et la survie cellulaire. C'est pourquoi la cellule a mis au point un système de réparation dépendant de la transcription (TCR) qui assure une réparation rapide des lésions sur le brin transcrit et empêche un blocage de la transcription. Bien que les acteurs principaux soient maintenant bien décrits *in vitro*, de nombreuses questions demeurent quant au couplage de ce processus avec la transcription, notamment *in vivo*.

La première interrogation concerne le devenir de la Pol II au moment de la mise en place du système de réparation. En effet, lorsque celle-ci se trouve bloquée au niveau de la lésion, elle bloque physiquement l'accès de la machinerie de réparation à l'ADN (Mei Kwei et al., 2004). Elle doit alors soit être déplacée, soit être dégradée. Chez la bactérie, l'ARN polymérase est systématiquement libérée (Selby and Sancar, 1994). Chez les eucaryotes, les premières analyses suggéraient une ubiquitination de la Pol II dépendante de CSA et CSB qui conduirait à sa dégradation (Beaudenon et al., 1999). Cependant, celle-ci ne serait requise que pour la reprise de la synthèse d'ARNm, mais pas pour la réparation elle-même (McKay et al., 2001). En accord avec cette hypothèse, la Pol II isolée à partir de la chromatine de cellules irradiées aux UVs n'est pas ubiquitinylée, alors que la Pol II isolée de la fraction soluble l'est (Groisman et al., 2003). Chez la

levure, Rad26 interagit avec DEF1 (« RNAPII degradation Factor 1 »), une protéine impliquée dans l'ubiquitination de la Pol II et sa dégradation par le protéasome (Reid and Svejstrup, 2004; Woudstra et al., 2002). Une autre hypothèse suggère que l'ubiquitination et la dégradation de la Pol II n'aurait lieu qu'en cas d'arrêt prolongé de la transcription, suite à une impossibilité à réparer la lésion (Harreman et al., 2009; Somesh et al., 2005; Wilson et al., 2013). Une étude *in vitro* dans des cellules HeLa a d'ailleurs montré que l'assemblage des facteurs de la TCR peut avoir lieu en présence de la Pol II (Tremeau-Bravard et al., 2004). Ainsi, dans le modèle désormais couramment admis, la Pol II n'est pas éliminée du site contenant la lésion mais seulement déplacée en amont pour laisser suffisamment de place à l'assemblage des facteurs de réparation. Si la réparation est efficace, la Pol II peut alors repartir en élongation, à l'inverse, en cas d'arrêt trop long, elle pourra être ubiquitinylée et dégradée.

La seconde question concerne la spécificité des facteurs de réparation et de la transcription. En effet, certains d'entre eux, initialement identifiés comme des composants de l'un des deux systèmes pourraient jouer un rôle dans les deux processus. Le premier facteur à s'être vu attribuer cette double fonction est TFIIH. Ce complexe a été identifié dans les années 90 comme un composant essentiel de la machinerie de transcription par la Pol II (Flores et al., 1992). Une étude par microséquençage a par la suite révélé qu'il contient les sous-unités XPB et XPD (Schaeffer et al., 1993, 1994), deux hélicases impliquées dans la réparation de l'ADN (Weber et al., 1990; Weeda et al., 1991). Depuis, le complexe TFIIH s'est également vu attribuer un rôle dans le contrôle du cycle cellulaire, plus précisément durant la transition entre les phase G2 et M (Larochelle et al., 1998). La démonstration selon laquelle TFIIH aurait plusieurs fonctions a été grandement facilitée par l'observation que des mutations dans ses sous-unités XPB ou XPD causent diverses pathologies telles que la trichothiodystrophie (TTS), le Xeroderma pigmentosum (XP) ou une combinaison de XP avec le syndrome de Cockayne (CS), dont certains symptômes ne peuvent s'expliquer uniquement par un défaut de réparation mais pourrait également provenir d'un défaut de transcription. Cependant, les fonctions de TFIIH dans les deux processus sont bien distinctes. TFIIH joue un rôle dans l'ouverture de l'ADN autour de la lésion (dans le cas de la réparation) ou au niveau du promoteur (dans le cas de la transcription). Dans le premier cas, XPB n'est requis que pour son activité ATPase, l'activité hélicase étant apportée par XPD (Oksenysh et al., 2009; Tirode et al., 1999). A l'inverse, dans la transcription, XPB est requis pour son activité hélicase (Holstege et al., 1996) alors que cette même activité pour XPD est dispensable. La transcription requiert le module CAK de TFIIH qui contient la kinase Cdk7, indispensable à la phosphorylation du CTD de la Pol II, tandis que dans la réparation, le module CAK est dispensable et rapidement éliminé pour permettre la liaison des autres facteurs de réparation (Coin et al., 2008; Svejstrup et al., 1995). La sous-unité p8, dispensable dans la transcription est requise dans

la réparation pour stimuler l'activité ATPase de XPB et promouvoir le recrutement de XPA. Ainsi, bien que le facteur TFIIH intervienne à la fois dans la réparation et la transcription, les sous-unités requises pour chacun des deux processus sont différentes.

Des analyses chez la levure ont montré que la délétion de Rad26 entraîne une sensibilité au 6AU indiquant un défaut d'élongation, de plus la croissance des souches *rad26Δ dst1Δ* est ralentie. Un rôle dans l'élongation de la transcription a ainsi été proposé pour Rad26 (Lee et al., 2001). Un rôle transcriptionnel a également été suggéré pour Rad2, la double délétion de Rad2 et Rad26 affectant plus fortement la croissance des souches que la simple délétion (Lee et al., 2002). Les auteurs ont donc proposé que Rad2 serait important pour promouvoir la transcription efficace par la Pol II. De plus, les travaux récents réalisés par l'équipe de JM. Egly montrent la présence des facteurs de réparation, notamment XPG et XPF, sur le promoteur et les régions distales des gènes dépendants des récepteurs nucléaires, en absence de stress génotoxique (Le May et al., 2010). Ceux-ci permettraient de promouvoir la modification des histones et la méthylation de l'ADN nécessaire à la transcription. Cette équipe a également suggéré l'implication des nucléases XPF et XPG dans la formation des boucles de chromatine entre le promoteur et le terminateur du gène humain *RARβ2* activé, importante pour la régulation de l'expression de ce gène (Le May et al., 2012). Enfin, XPG serait également important pour la transactivation des récepteurs nucléaires en stabilisant le module CAK de TFIIH qui permet leur phosphorylation (Ito et al., 2007). Cependant, aucune de ces données ne permet aujourd'hui de certifier l'implication de ces facteurs de la réparation dans la transcription.

Certains activateurs de la transcription, tels que GAL4-VP16 ou RAR, sont également capables de stimuler la réparation de l'ADN (Frit et al., 2002). A l'inverse, un complexe contenant XPC peut également fonctionner en tant que co-activateur des facteurs OCT4 et SOX2 durant la transcription dans les cellules ES (Fong et al., 2011). Enfin, des mutants de PAF et Ccr4-Not, deux complexes impliqués dans l'élongation de la transcription, sont déficients dans la réparation par le système NER, suggérant qu'ils possèdent également un rôle dans ce processus (Gaillard et al., 2009).

Ainsi, les processus de réparation et de transcription sont bien plus liés que ce que l'on imaginait initialement, et des pathologies que l'on définit comme dûes à un défaut dans l'un de ces deux systèmes, pourrait en fait résulter de défauts dans les deux mécanismes.

V. PATHOLOGIES LIEES A UN DEFAUT DE REPARATION PAR LE SYSTEME NER

La réparation est importante pour le maintien de l'intégrité du génome, une instabilité génétique pouvant mener au cancer, ou à la mort de la cellule et donc au vieillissement prématuré de l'organisme. Des défauts dans le mécanisme de réparation NER sont responsables de divers syndromes chez l'homme : Xeroderma pigmentosum (XP), syndrome de Cockayne (CS), trichothiodystrophie (TTD)... que l'on retrouve parfois associés chez un même patient. Ces pathologies ont pour symptôme commun une hypersensibilité aux agents qui endommagent l'ADN, notamment les UVs. Cependant, bien qu'affectant le même processus, la réparation par le système NER, ces trois syndromes ont des conséquences cliniques bien différentes selon que la protéine mutée appartient à la voie TCR, GGR ou les deux (Figure 25).

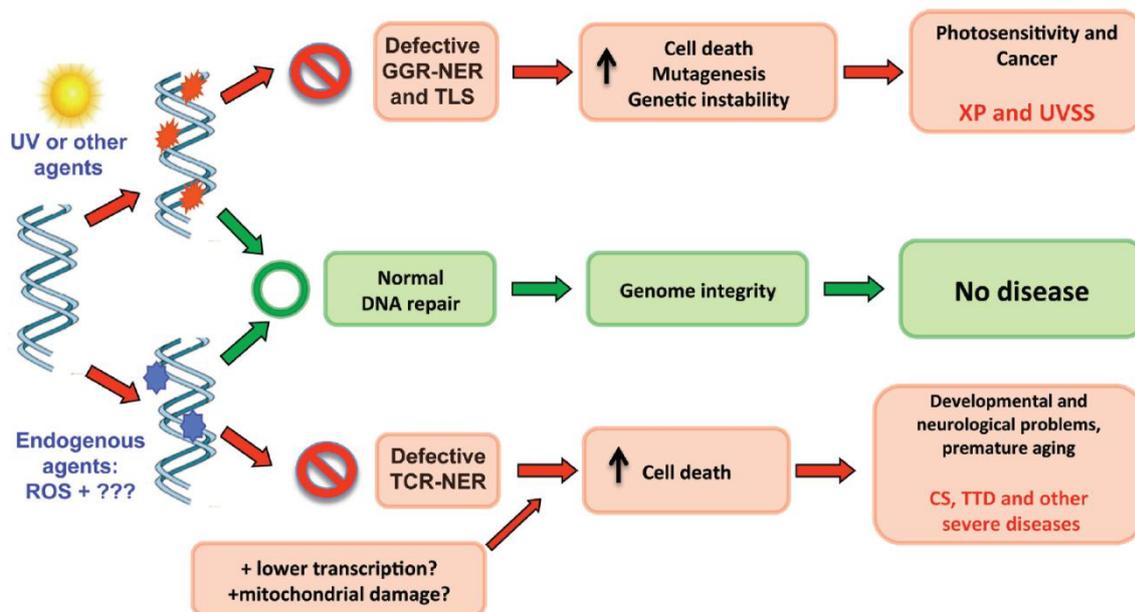


Figure 25 : Conséquences des dommages à l'ADN et phénotypes observés en cas de défaut dans les voies GGR ou TCR (Menck and Munford, 2014).

Gene	Pathologie	Voie défectueuse	Symptômes cliniques
<i>XPA</i>	XP	NER	Photosensibilité, cancer de la peau, défauts neurologiques de moyen à sévère
<i>XPB</i>	XP, XP/CS, TTD	NER	Photosensibilité, cancer de la peau, défauts neurologiques moyens
<i>XPC</i>	XP	GGR	Cancer de la peau
<i>XPD</i>	XP, XP/CS, TTD	NER	Photosensibilité, cancer de la peau
<i>XPE</i>	XP	GGR	Cancer de la peau
<i>XPF</i>	XP, CS	NER	Photosensibilité, cancer de la peau
<i>XPG</i>	XP, XP/CS	NER	Photosensibilité, cancer de la peau, parfois des défauts neurologiques
<i>XPV</i>	XP	NER	Photosensibilité
<i>CSA</i>	CS, UVSS	TCR	Photosensibilité, retard mental, microcéphalie, défauts de croissance
<i>CSB</i>	CS, UVSS	TCR	Photosensibilité, retard mental, microcéphalie, défauts de croissance

Tableau 13 : Facteurs impliqués dans les pathologies de réparation de l'ADN par la voie NER et symptômes associés (Menck and Munford, 2014).

1. Xeroderma pigmentosum (XP)

Découvert en 1870 par M. Kaposi, le Xeroderma pigmentosum (XP) est une maladie autosomique récessive rare qui se caractérise par une hypersensibilité de la peau aux UVs, une pigmentation anormale et une prédisposition au cancer (risque multiplié par 1000).

Les cellules des patients XP montrent une fréquence de mutation augmentée après exposition aux UVs (D'Errico et al., 2000), corrélée avec un défaut de réparation des dommages de l'ADN qui conduisent à une fréquence des tumeurs augmentée (Cleaver, 2000; van Steeg and Kraemer, 1999).

Des tests de complémentation (De Weerd-Kastelein et al., 1972) ont permis d'identifier 7 gènes responsable du XP : XPA, XPB, XPC, XPD, XPE, XPF et XPG (Tableau 13). Tous ces gènes sont impliqués dans la GGR et un défaut de ce processus serait donc la principale cause des symptômes observés. Plus tard, un nouveau cas de patient a été mis en évidence, chez lequel la NER est fonctionnelle mais qui présente un défaut de re-synthèse de l'ADN après réparation de la lésion (Fujiwara et al., 1981). Le gène responsable de ce type de XP a été identifié comme XPV (XP variant) et code pour l'ADN polymérase η (Masutani et al., 1999).

Dans 20 à 30% des cas, des problèmes neurologiques s'ajoutent aux symptômes XP (DiGiovanna and Kraemer, 2012). Cette caractéristique est due à une association de deux syndromes : Xeroderma pigmentosum et syndrome de Cockayne (XP/CS) dans lesquels les deux voies TCR et GGR sont affectées. Les troubles associés au syndrome de Cockayne seront décrits

dans la partie suivante de ce chapitre. La pathologie XP/CS résulte de mutations dans les protéines XPB, XPD et XPG qui interviennent dans les deux voies de réparation (Oh et al., 2006; Scharer, 2008; Theron et al., 2005). Ainsi, une mutation de XPG affectant seulement son activité nucléase va conduire au XP, dans lequel une réparation résiduelle peut encore avoir lieu. A l'inverse, une mutation résultant en la troncature de la partie C-terminale de la protéine, importante pour son interaction avec les autres facteurs de réparation, va conduire au double syndrome XP/CS dans lequel tout le système de réparation NER est aboli (Emmert et al., 2002). Le phénotype plus grave observé dans ce second cas serait notamment dû au fait que XPG intervient également dans le système BER pour réparer les dommages oxydants de l'ADN. L'étude des allèles portés par les patients XP-B, XP-D et XP-G présentant une association des complexes XP et CS permet d'en apprendre davantage sur les mécanismes moléculaires à l'origine des symptômes observés mais également sur le rôle de ces facteurs dans la réparation par le système NER ou d'autres processus.

2. Syndrôme de Cockayne (CS)

Découvert en 1936 par E.A. Cockayne, le syndrome de Cockayne (CS) est une maladie autosomale récessive rare caractérisée par un retard de développement causant un nanisme, un dysmorphisme facial, une microcéphalie, une neurodégénération responsable de retards mentaux et un vieillissement prématuré. Ces patients présentent également une hypersensibilité aux UVs sans prédisposition au cancer.

Les cellules des patients CS présentent une hypersensibilité aux UVs alors que la réparation a lieu (Schmickel et al., 1977). Cette constatation a permis de déterminer que la cause principale de cette pathologie serait un défaut de reprise de la transcription après la réparation du dommage par la voie TCR (Mayne and Lehmann, 1982). Les souris CS présentent une prédisposition au cancer qui n'est pas retrouvée chez l'homme (van der Horst et al., 1997). Or la TGR est plus efficace dans les cellules humaines et pourrait ainsi protéger du cancer les patients CS. Une autre hypothèse serait que les cellules ne pouvant pas reprendre la transcription envoient un signal d'apoptose, notamment par accumulation de p53 (McKay et al., 2001), empêchant le développement du cancer. Cette mort cellulaire importante serait à l'origine du vieillissement prématuré observé chez les patients CS.

Des tests de complémentation ont permis de mettre en évidence 2 gènes responsables du CS : CSA et CSB (Mayne and Lehmann, 1982) (Tableau 13). Cette découverte est d'ailleurs à l'origine de l'identification de la voie TCR.

Dans certains cas, les mutations de CSA ou CSB ne mènent pas au CS mais au syndrome de sensibilité aux UVs (UVSS) qui ne se caractérise qu'en une hypersensibilité aux UVs, sans symptômes associés (Horibata et al., 2004; Nardo et al., 2009). Chez ces patients, les cellules sont très sensibles aux dommages oxydants l'ADN (D'Errico et al., 2007). Les protéines CSA et CSB seraient donc impliquées dans la suppression des lésions oxydatives et des mutations affectant ce processus, sans bloquer la TCR, seraient à l'origine du phénotype observé chez les patients UVSS.

3. Trichothiodystrophie (TTD)

La trichothiodystrophie (TTD) est une maladie autosomique récessive rare caractérisée par des cheveux courts, épars et fragiles et des ongles anormaux dus à un défaut de synthèse des sulfures dans la kératine. Les patients présentent également un retard mental et un défaut de croissance et dans la moitié des cas, une hypersensibilité aux UVs.

Les cas de TTD avec photosensibilité présentent une mutation dans une des sous-unités XPB, XPD ou p8 de TFIIH. Ces mutations déstabilisent le complexe, causant une insuffisance de transcription avant la différenciation terminale des kératinocytes (Faghri et al., 2008; Stefanini et al., 2010).

RESULTATS ET DISCUSSION

ARTICLE I : ROLE DU MEDIATEUR DANS LA MISE EN PLACE DU PIC *IN VIVO*

I. CONTEXTE DU TRAVAIL

Les modèles d'initiation de la transcription basés sur les études *in vitro* décrivent une séquence d'assemblage linéaire des GTFs et de la Pol II, débutant par le recrutement de TFIID sur le promoteur, suivi par l'assemblage de TFIIA et TFIIB, puis la Pol II liée à TFIIF et, enfin, TFIIH et TFIIIE (Sikorski and Buratowski, 2009). Cependant, ce modèle ne tient pas compte des co-activateurs, tels que le Médiateur, dont la présence pourrait influencer l'assemblage du PIC *in vivo*.

Le Médiateur est un complexe multiprotéique qui permet de moduler le niveau de transcription en réponse aux signaux de régulation. *In vitro*, il est nécessaire à l'activation de la transcription en réponse aux activateurs spécifiques, il augmente le niveau de la transcription basale et stimule l'activité kinase de TFIIH (Kim et al., 1994). On sait désormais que le Médiateur établit de nombreux contacts avec la machinerie de base de la transcription, notamment avec TBP (Larivière et al., 2006), TFIIB (Kang et al., 2001), TFIIF (Takagi et al., 2006), TFIIIE (Sakurai and Fukasawa, 2003), TFIIH (Esnault et al., 2008) et la Pol II (Soutourina et al., 2011). Toutes ces observations suggèrent que le Médiateur influence la mise en place du PIC en interagissant avec les GTFs et la Pol II pour permettre leur recrutement sur l'ADN. *In vivo*, le Médiateur est requis pour la transcription de la quasi-totalité des gènes de *S. cerevisiae* (Holstege et al., 1998; Thompson and Young, 1995). De plus, des études de génomique fonctionnelle par ChIP et/ou ChIP-on-chip sur l'ensemble du génome de *S. cerevisiae* ont permis de déterminer que ce complexe est présent sur les régions régulatrices de la majorité des gènes de classe II, actifs ou inactifs, constitutifs ou inductibles (Andrau et al., 2006; Ansari et al., 2009).

Bien que le Médiateur ait fait l'objet de nombreuses études ces dernières années, sa complexité et notamment la présence de nombreuses sous-unités essentielles, ont empêché de parvenir à une compréhension détaillée des mécanismes mis en jeu dans le contrôle de l'expression génique. C'est pourquoi nous étudions ce complexe dans notre laboratoire, l'objectif principal étant de comprendre les mécanismes moléculaires de l'activation de la transcription par le Médiateur *in vivo*. Plus particulièrement nous souhaitons appréhender comment les interactions protéine-protéine au sein du Médiateur et entre le Médiateur et ses partenaires

permettent la transmission des signaux de régulation des activateurs vers la Pol II et les GTFs et contribuent au recrutement des divers composants du PIC.

Un crible double-hybride nous a déjà permis de décrire en détail les interactions protéine-protéine qui existent entre les sous-unités du Médiateur (Guglielmi et al., 2004). Ce complexe, conservé de la levure à l'homme, se compose chez *S. cerevisiae* de 25 sous-unités réparties en 4 modules : la tête, le milieu, la queue et le module Cdk8. 10 de ces sous-unités sont essentielles chez *S. cerevisiae*. Ce crible double-hybride a également identifié toute une série d'interactions entre le Médiateur et des protéines appartenant aux différents complexes nucléaires. Une de ces interactions, entre la sous-unité Med11 du Médiateur et Rad3 de TFIIH, a été analysée avec succès et a permis de montrer que le Médiateur favorise indépendamment le recrutement de la Pol II, de TFIIE et de TFIIH, ainsi que la stabilisation du module kinase de TFIIH (Esnault et al., 2008). Ces résultats ont abouti à la construction d'un modèle ramifié d'assemblage du PIC, différant fortement du modèle séquentiel proposé *in vitro* en absence de co-activateur, confirmant le rôle majeur du Médiateur pour coordonner le recrutement des divers composants du PIC.

Mon travail de thèse porte sur la sous-unité Med17, qui joue un rôle central au sein du module de tête du Médiateur. Des expériences de photo-pontage *in vivo* réalisées dans notre laboratoire ont permis de montrer que Med17 interagit directement avec la sous-unité Rpb3 de la Pol II. Des mutations de Med17 induisent une co-létalité avec le mutant *rpb3-2*, dans lequel le contact entre Med17 et Rpb3 est affecté. De plus, l'association de la Pol II à l'ADN est fortement diminuée dans les mutants *med17ts*. Ces résultats indiquent que l'interaction entre Rpb3 et Med17 est essentielle au recrutement de la Pol II sur l'ADN *in vivo* (Soutourina et al., 2011). Un mutant conditionnel *med17-138 (srb4-138ts)* a permis de montrer que le Médiateur est requis pour la transcription de l'ensemble des gènes de classe II (Thompson and Young, 1995 ; Holstege et al., 1998). L'étude de ce même mutant a également démontré que le Médiateur est essentiel à la transcription basale par la Pol II (Takagi and Kornberg, 2006). Cependant, cette mutation induit une dissociation du module de tête du reste du complexe Médiateur et conduit donc à la perte de fonction du Médiateur. L'étude d'autres mutants de Med17 avec des phénotypes différents, et notamment dont la structure du Médiateur est peu modifiée, pourrait nous permettre d'en apprendre davantage sur le rôle de cette sous-unité essentielle et par extrapolation sur celui du Médiateur dans la transcription des gènes de classe II. L'importance de la sous-unité Med17 est également soulignée par son implication dans le cancer. En effet, une diminution de l'expression de *MED17* dans les cellules du cancer de la prostate inhibe la prolifération cellulaire, ralentit la progression du cycle cellulaire et augmente l'apoptose (Vijayvargia et al., 2007). Cette sous-unité pourrait donc être une cible potentielle des traitements anti-cancéreux.

La stratégie suivie pour ce projet est celle mise au point par notre laboratoire pour l'étude de la sous-unité Med11 (Esnault et al., 2008). Elle consiste à mutagéniser aléatoirement un gène codant pour une sous-unité essentielle du complexe Médiateur afin d'obtenir des mutants thermosensibles (ts), et de rechercher ensuite les causes de ce phénotype. Avant mon arrivée au laboratoire, une collection de 29 mutants *med17ts* avait ainsi été obtenue. 5 de ces mutants montrent une co-létalité avec le mutant de la Pol II *rpb3-2* et ont été utilisés pour confirmer le rôle essentiel du contact Med17-Rpb3 *in vivo*. Dans la méthodologie suivie par C. Esnault, la caractérisation des mutants débute par la recherche de pertes d'interaction de la protéine mutagénisée avec ses partenaires par la méthode du double-hybride. La découverte d'un défaut d'interaction spécifique avec un seul de ses partenaires devient alors un atout pour la compréhension du rôle du Médiateur dans un nouveau processus. Cependant, après de nombreux essais infructueux, nous avons constaté l'impossibilité d'utiliser le double-hybride pour étudier les interactions de Med17 dans les mutants *med17ts*. Nous avons donc dû nous baser sur d'autres méthodes afin de sélectionner les mutants les plus intéressants pour la suite de l'étude.

Comme nous cherchons à étudier des mutants n'affectant pas la structure globale du Médiateur, nous avons commencé par vérifier, grâce à des expériences de co-IP, que dans la majorité des mutants de Med17, l'arrangement du complexe Médiateur en 3 modules principaux est conservé. Pour la suite de l'étude nous nous sommes concentrés sur 6 des 29 mutants de Med17, sélectionnés pour leur Médiateur intègre, leur faible nombre de mutations et leurs phénotypes variés.

Au même moment, l'équipe d'O. Elpeleg publiait un article impliquant la sous-unité Med17 dans l'atrophie cérébrale infantile. La mutation détectée (L371P) se trouvant dans une région très conservée de la protéine, il est possible de produire la mutation équivalente (M504P) chez la levure *S. cerevisiae*. La transformation de la souche thermosensible *med17-138* avec un plasmide exprimant la protéine mutée *med17-M504P* ne permet pas de compléter le phénotype ts, démontrant un défaut fonctionnel du Médiateur porteur de la mutation M504P (Kaufmann et al., 2010). Nous avons construit la souche *med17-M504P* et montré que cette mutation confère un phénotype ts. Nous avons donc décidé d'inclure ce mutant à notre étude afin de décrypter les mécanismes moléculaires pouvant mener à la pathologie observée chez l'homme.

Notre but étant d'étudier l'effet des mutations de Med17 sur le recrutement des partenaires connus du Médiateur, il nous faut partir d'un complexe dont la fonctionnalité globale n'est pas affectée et notamment qui conserve sa capacité à s'associer à la chromatine. C'est pourquoi, en premier lieu, nous avons vérifié par ChIP sur 3 gènes modèles que les mutants étudiés n'affectent pas l'occupation du Médiateur. Les ChIP de la Pol II, de TFIID et de TBP ont, par ailleurs, montré

que les mutations de Med17 affectent différenciellement l'occupation et/ou le recrutement des composants du PIC sur l'ADN. S'il avait déjà été montré avec les travaux sur Med11 que le Médiateur influence le recrutement de TFIIH indépendamment des autres facteurs, cette fonction n'avait pas encore été démontrée pour la TBP. Ces résultats nous ont permis de sélectionner 3 mutants intéressants pour des analyses complémentaires : *med17-98* affecté pour l'occupation de tous les facteurs sauf la TBP, *med17-444* affecté pour l'occupation de tous les facteurs sauf le module cœur de TFIIH et *med17-670* affecté pour l'occupation du module cœur de TFIIH et de la Pol II mais pas de TBP ni du module kinase de TFIIH (Figure 26).

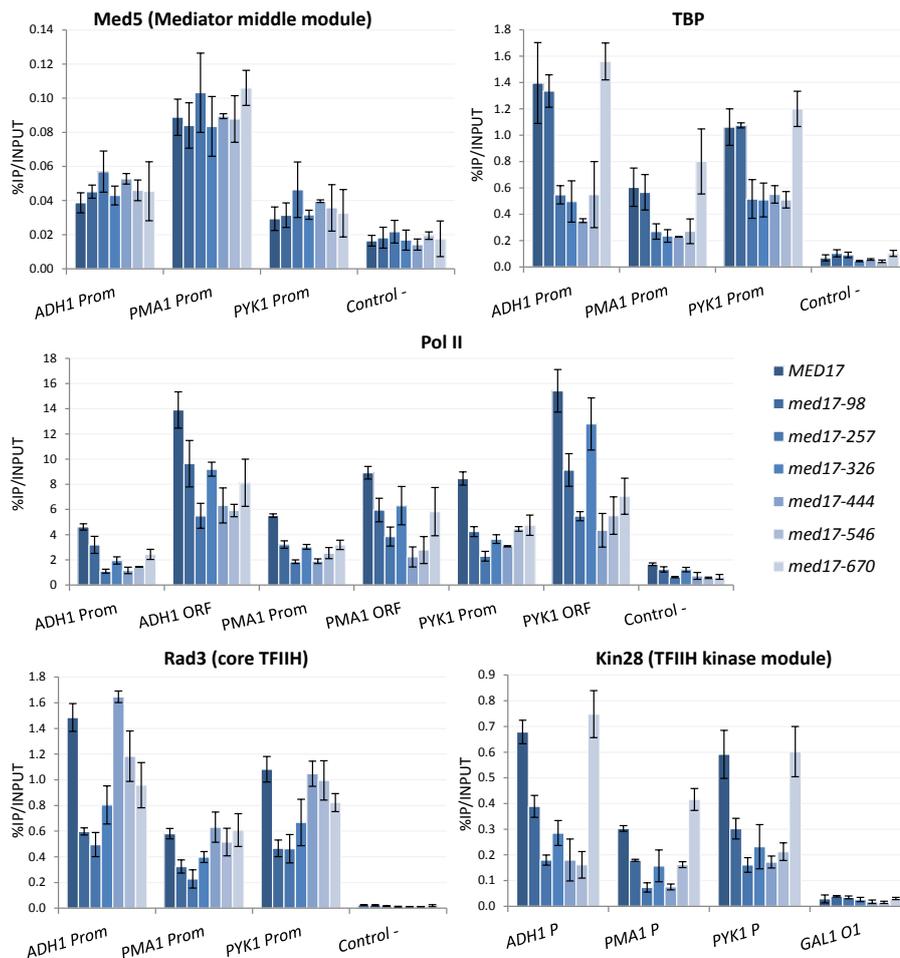


Figure 26 : Influence des mutations *med17ts* sur le recrutement du Médiateur, de la Pol II, de TFIIH et de la TBP.

Pour savoir si les effets observés par ChIP sont gène-spécifiques ou globaux, des expériences de ChIP-sequencing (ChIP-seq) ont été réalisées pour le mutant *med17-M504P* et les 3 mutants sélectionnés précédemment. Après avoir aligné les séquences obtenues sur le génome de la levure, nous avons identifié les régions avec un enrichissement significatif par rapport au contrôle. Cette étape nous a permis de définir les profils de liaison des différents composants du PIC sur la chromatine. Par la suite, la comparaison des densités de lectures dans ces régions enrichies entre les mutants et la souche sauvage pour chacune des protéines étudiées nous a

permis de montrer l'effet global des mutations de Med17 sur l'occupation des composants du PIC. Nos résultats confirment le rôle universel du Médiateur dans la transcription des gènes et sont en accord avec le rôle central de la sous-unité Med17 au sein du complexe Médiateur.

Afin d'expliquer les défauts d'occupation observés par CHIP, nous avons vérifié l'interaction entre le Médiateur et les différents composants du PIC par co-IP. Seul le mutant *med17-M504P* présente une diminution d'interaction entre Med17 et Rpb3, qui pourrait expliquer la diminution de l'occupation des facteurs observée par CHIP. Nos résultats suggèrent que la mutation de Med17 responsable de l'atrophie cérébrale infantile induit une diminution de l'interaction entre le Médiateur et la Pol II conduisant à une déstabilisation de l'ensemble des composants du PIC, et sont donc importants pour notre compréhension de cette pathologie.

L'ensemble de ces résultats confirme que le Médiateur influence l'assemblage du PIC *in vivo* en permettant le recrutement indépendant des GTFs et de la Pol II. Nous savions déjà que le Médiateur est capable d'influencer le recrutement indépendant de TFIIH et TFIIIE, cette étude montre qu'il influence également le recrutement et/ou la stabilisation de TBP indépendamment des autres facteurs. De façon inattendue, nos résultats suggèrent aussi que le Médiateur coordonne le recrutement et/ou la stabilisation des modules cœur et kinase de TFIIH indépendamment. Nos travaux confirment ainsi qu'*in vivo*, l'assemblage du PIC n'est pas séquentiel comme ce qui est observé *in vitro*, et que le Médiateur est important pour orchestrer l'arrivée des différents composants de la machinerie de transcription. Notre travail révèle notamment que le Médiateur, via sa sous-unité centrale Med17, coordonne la formation du PIC, à différents niveaux *in vivo*, comprenant en plus de la Pol II, les deux modules de TFIIH et la TBP.

II. ARTICLE 1

Mediator independently orchestrates multiple steps of preinitiation complex assembly *in vivo*

Fanny Eyboulet^{1,2,3}, Sandra Wydau^{1,2,3§}, Olivier Alibert⁴, Helen Neil^{1,2,3#}, Claire Boschiero^{1,2,3}, Thomas Eychenne^{1,2,3}, Marie-Claire Nevers⁵, Hervé Volland⁵, David Cornu⁶, Virginie Redeker⁶, Michel Werner^{1,2,3} and Julie Soutourina^{1,2,3*}

¹CEA, iBiTec-S, FRE3377, F-91191 Gif-sur-Yvette cedex, France; ²CNRS, FRE3377, F-91191 Gif-sur-Yvette cedex, France; ³Univ Paris-Sud, FRE3377, F-91191 Gif-sur-Yvette cedex, France; ⁴CEA, iRCM, F-91057 Evry cedex, France; ⁵CEA, iBiTecS, Service de Pharmacologie et d'Immunoanalyse, F-91191 Gif sur Yvette cedex, France; ⁶CNRS, Centre de Recherche de Gif, Imagif, F-91198 Gif-sur-Yvette Cedex, France

* Corresponding author. E-mail: julie.soutourina@cea.fr.

§ Present address: *Université Paris Descartes, EA 4065, Faculté des Sciences Pharmaceutiques et Biologiques, 75006 Paris, France*

Present address: *Inserm U1052, Centre de Recherche en Cancérologie de Lyon, F-69008 Lyon, France*

Short title: Mediator orchestrates the PIC formation

Keywords: Mediator, transcription, RNA polymerase II, TBP, TFIID, preinitiation complex, *Saccharomyces cerevisiae*, genome-wide distribution of transcription machinery

Abstract

Mediator is a large multiprotein complex conserved in all eukaryotes. This complex has a crucial coactivator function in transcription by RNA polymerase II (Pol II). It transmits the regulatory signals from the specific transcription factors to Pol II transcription machinery and stimulates the formation of preinitiation complex (PIC) and the phosphorylation of the Pol II CTD by TFIIF. While Mediator has been studied intensively, its complexity has precluded a detailed understanding of the molecular mechanisms of its function *in vivo*. Med17 is an essential and central component of Mediator head module. Previously, we have shown that the direct Mediator-Pol II interaction between Med17 Mediator subunit and the Rpb3 Pol II subunit is essential for global Pol II recruitment *in vivo*. In this work, we took advantage of our large collection of conditional temperature-sensitive *med17* mutants to investigate Mediator role in coordinating PIC formation *in vivo* at the genome level. The effect of a yeast mutation equivalent to human Med17-L371P responsible for an infantile cerebral atrophy was also analyzed. Our ChIP-seq results showed that *med17* mutations differentially affected the global presence of several PIC components including Mediator, TBP, TFIIF modules and Pol II. This work provides a comprehensive genome-wide view of Mediator role in PIC formation indicating that Mediator coordinates multiple steps of a branched PIC assembly pathway.

Introduction

Transcription, the first step of gene expression, is well regulated, ensuring that the eukaryotic cells can realize the genetic programs adapted to the environmental conditions. In eukaryotes, the synthesis of mRNAs and a number of small non-coding RNAs requires the RNA polymerase II (Pol II) and the general transcription factors (GTFs: TFIIA, B, D, E, F, H) that assemble into a large complex on the promoter DNA. This transcription preinitiation complex (PIC) is a key intermediate in Pol II transcription. *In vitro* reconstitution studies of transcription initiation suggested a model in which PIC components assemble in a linear sequence (Buratowski et al. 1989; Ranish and Hahn 1996). TFIID is the first GTF that binds the promoter. Then, TFIIA and TFIIB are recruited followed by Pol II probably in association with TFIIF. Finally, TFIIE and TFIIH complete the formation of a PIC sufficient for basal *in vitro* transcription but unable to drive activated transcription in response to specific activators. Important insights have been made on the PIC architecture in yeast and human systems by biochemical and structural studies (for review (Grunberg and Hahn 2013)). Recently, two structural models of human PIC (He et al. 2013) and yeast PIC (Murakami et al. 2013) have been proposed by cryo-electron microscopy approaches. However, all these systems include only basal transcription components. *In vivo*, a precise map of PIC locations across the yeast genome, including Pol II and GTFs, has been recently obtained permitting to identify TATA-like elements bound by TBP on TATA-less promoters and distinct PICs for divergent transcription (Xu et al. 2009; Rhee and Pugh 2012).

Unlike bacterial or archaeal transcriptional systems, an additional level of Pol II transcription regulation is brought about by multiprotein complexes, coactivators and corepressors, which modify the chromatin structure or directly contribute to PIC formation. Mediator of transcription regulation is one of these coregulators. While Mediator has been studied intensively, its complexity has precluded a detailed understanding of the molecular

mechanisms of its function *in vivo*. Mediator is a large multiprotein complex that transmits the regulatory signals from the specific transcription factors to Pol II transcription machinery (Kornberg 2005; Ries and Meisterernst 2011). Together with the GTFs, this coactivator promotes PIC formation and transcription initiation. Mediator also stimulates the phosphorylation of the Pol II carboxy-terminal domain (CTD) by TFIIF. This coactivator is generally required for Pol II transcription (Holstege et al. 1998; Andrau et al. 2006; Zhu et al. 2006). Mediator is a huge complex (1.5 MDa) composed of 25-30 subunits that are organized into four structural modules: head, middle, tail and Cdk8 kinase modules. Mediator subunits are engaged in numerous contacts within the complex (Guglielmi et al. 2004; Koschubs et al. 2010; Lariviere et al. 2012). The large size, complexity and extreme flexibility have precluded the determination of a complete crystallographic structure of Mediator complex. Recently, structural data have been obtained for the yeast Mediator head module, the largest Mediator subcomplex characterized to date at high resolution (Lariviere et al. 2012; Robinson et al. 2012).

Mediator complex is conserved in all eukaryotes (Bourbon 2008) supporting a very ancient eukaryotic origin for this four-module complex. Given its central role in transcription regulation, it is not unexpected that Mediator has been implicated in numerous developmental processes in animals and mutations in Mediator subunits are involved in a number of pathologies (Risheg et al. 2007; Schwartz et al. 2007). For example, a mutation in Med23 subunit cosegregates with intellectual disability (Hashimoto et al. 2011). In line with the fact that oncogenesis results from gene dysregulation, Mediator has been found to be involved in several cancers (Zhang et al. 2005; Vijayvargia et al. 2007; Firestein et al. 2008; Gade et al. 2009; Li et al. 2010; Kuuselo et al. 2011).

Mediator complex is at the heart of PIC assembly helping to regulate PIC activity and interacting with Pol II. Mediator is capable of receiving the activation signals through protein-

protein interactions with the hundreds of transcription activators of the eukaryotic cell, integrates these signals and activates transcription by triggering the recruitment of Pol II and possibly some of the GTFs. For example, Mediator cooperates with TFIID in yeast and human cells to stabilize the PIC (Johnson et al. 2002; Lariviere et al. 2006; Cai et al. 2010; Takahashi et al. 2011). Previously, we have identified a direct contact between the Med11 Mediator head subunit and the Rad3 TFIID subunit and showed that this contact is essential for the recruitment of the GTF to the PIC independently of Pol II (Esnault et al. 2008). Recently, we showed that the Rpb3 Pol II subunit and the Med17 Mediator subunit are in direct contact to position Pol II on promoters and activate transcription strongly supporting a notion that the Mediator-Pol II interaction is generally required for Pol II recruitment and transcription of class II genes in eukaryotes (Soutourina et al. 2011). We showed that Mediator plays an essential role in PIC assembly at the level of Pol II, TFIID and TFIIE recruitment (Esnault et al. 2008; Soutourina et al. 2011).

Med17 is one of the ten essential yeast Mediator subunits. *MED17* was initially isolated as *SRB4*, one of the *SRB* genes whose mutations suppressed the growth phenotype of truncations of the Pol II Rpb1 CTD (Thompson et al. 1993). The general requirement of Mediator for Pol II transcription has been shown using *med17-138* mutant that generally affects Pol II transcription in a manner comparable with that of the *rpb1-1* Pol II mutant (Holstege et al. 1998). This classical *med17* allele causes dissociation of the head module from the rest of Mediator leading to a loss of Mediator function (Linder et al. 2006; Takagi et al. 2006; Takagi and Kornberg 2006). The central role of Med17 subunit in Mediator function is also consistent with a central and extended positioning of this subunit within the Mediator head structural model (Lariviere et al. 2012). This Mediator head subunit is directly involved in Mediator-Pol II contact essential for global Pol II transcription (Soutourina et al. 2011). An important role of Med17 subunit has also been highlighted by the fact that a mutation in this

subunit has been associated with infantile cerebral atrophy (Kaufmann et al. 2010) and by the involvement of this subunit in cancer (Vijayvargia et al. 2007).

In this work, we aimed at obtaining a detailed genome-wide view of the Mediator role in PIC assembly *in vivo* by characterizing our large collection of conditional temperature-sensitive (ts) mutants in Med17 Mediator head subunit. A mutant of yeast Med17 equivalent to human Med17-L371P responsible for the infantile cerebral atrophy having a severe ts phenotype was also included in our analysis. Our ChIP-seq results show that *med17* mutations differentially affect the global presence of several PIC components including Mediator, TBP, TFIID modules (TFIID core and TFIID kinase modules) and Pol II. This work provides a comprehensive genome-wide view of Mediator Med17 subunit role in PIC formation indicating that Mediator differentially coordinates multiple steps of a branched assembly pathway. We also suggest that an infantile cerebral atrophy could result from a defect in Mediator-Pol II contact and a global destabilization of PICs.

Results

Collection of conditional med17 Mediator mutants

Previously, to investigate the *in vivo* role of Med17 in Pol II recruitment, we have selected a large collection of conditional *med17* mutants (Soutourina et al. 2011). *med17* mutants with different thermosensitive phenotypes, from slow growth to lethal at non-permissive temperature (37°C), were selected by random mutagenesis (**Figure S1, Table S1**). We separated some multiple mutations to obtain 5 additional point mutants by site-directed mutagenesis. A L371P mutation in Med17 subunit has been associated with infantile cerebral atrophy (Kaufmann et al. 2010). A corresponding M504P mutation in the homologous *S. cerevisiae* Med17 inactivated the protein at 37°C. To study the molecular consequences of *med17* mutation associated with infantile cerebral atrophy, we constructed a *med17-504* yeast

strain carrying *med17-M504P* mutation and investigated the growth phenotype of this mutant. **Figure 1A** shows that *med17-504* leads to a slow growth phenotype at 30°C and a strong thermosensitive phenotype at 37°C. In total, we obtained a collection of 30 conditional mutants in Med17 subunit. The integrity of the three main Mediator modules in all *med17* mutants has been verified by CoIP experiments (**Table S2**).

Effects of med17 mutations on Mediator stability, interaction with Pol II and mRNA level

Four mutants were selected for further study based essentially on their growth phenotypes (normal or slightly slower growth at 30°C combined with pronounced temperature-sensitive phenotype at 37°C), Mediator integrity and low mutation number: *med17-98* (*L86Q, E186G, E448Q*), *med17-444* (*Q444P*), *med17-504* (*M504P* corresponding to *L371P* human mutation) and *med17-670* (*V670E*) (**Figure 1A**). All *med17* mutations are localized within or near the conserved domains of this subunit defined by a comparative genomics approach (**Figure S2A**) (Bourbon 2008). We mapped, when possible, the mutated amino acid residues of *med17* mutants on the structural model of Mediator head module (Lariviere et al. 2012) (**Figure S2B**) and showed that they are often located close to the module surface.

The effect of *med17* mutations on the integrity of the three main Mediator modules was first verified by CoIP experiments (**Figure S3** and **Table S2**). HA-tagged Med5 Mediator subunit was immunoprecipitated with anti-HA antibody and coimmunoprecipitated Med17 Mediator head subunit and Med14 Mediator subunit bridging tail and middle modules were detected by Western blotting with antibodies produced in house. We observed that Mediator head and tail modules coimmunoprecipitated with the Med5 middle module subunit in all four mutants without any significant difference compared to the wild-type strain.

To investigate in more detail Mediator integrity in selected *med17* mutants, Mediator was immunoprecipitated from crude extracts of Med5-HA tagged strains and immunoprecipitated proteins were analyzed by mass spectrometry approach (**Figure 1B**). To improve Mediator subunit identification by mass spectrometry, the immunoprecipitated proteins were separated on gel and the bands corresponding to IgG heavy and light chains were cut out. This procedure thus excluded some Mediator subunits with molecular weights close to those of IgG chains. A strain carrying untagged Med5 subunit was used as a negative IP control. We identified up to 19 Mediator subunits belonging to all four Mediator modules. 13 Mediator subunits of head, middle, tail and Cdk8 modules, indicated in red on Figure 1B, were present for all *med17* mutants and the wild type strain. Mediator subunits with low molecular weight (Med9, 10, 11, 21, 22) could not be detected in all samples. Our mass spectrometry analysis indicated the presence of Mediator head, middle, tail and Cdk8 modules for all *med17* mutants suggesting that mutations did not affect Mediator stability.

Med17 Mediator subunit directly interacts *in vivo* with Rpb3 Pol II subunit (Soutourina et al. 2011). We examined if this contact was modified in *med17* mutants. For these experiments, we constructed yeast strains carrying EGFP-tagged Med17 subunit and Myc-tagged Rpb3 subunit in wild-type and *med17* mutant contexts. Proteins were cross-linked with formaldehyde, Med17-EGFP-tagged subunit was immunoprecipitated under denaturing conditions and cross-linked Rpb3-Med17 band was detected with anti-Myc antibody (**Figure 1C**, red arrow). Our results demonstrated that Rpb3-Med17 contact was reduced in the *med17-504* mutant, but was not affected in *med17-98*, *-444* and *-670* compared to the wild-type.

We investigated the effect of *med17* mutations after a shift to 37°C on mRNA abundance of *ADH1*, *PYK1* and *PMA1* genes (**Figure S4**). *med17-98*, *444* and *607* mutants affected the level of these mRNAs. The effect of *med17-98* was accentuated at 37°C and that

of other mutations was essentially similar at 30°C (T0) and 37°C. It should be noted that *med17-504* shows little, if any, effect on steady state levels of the three mRNAs. The measurement of mRNA levels being a cumulative result of RNA synthesis and RNA processing, we decided to measure Pol II occupancy by ChIP since it directly reflects transcriptional activity in yeast.

Mutations in Med17 differentially affect chromatin association of PIC components on selected constitutively expressed genes

To analyze the molecular effect of *med17* mutations on PIC formation, we determined the occupancy of Pol II, GTFs and Mediator by ChIP experiments on selected constitutively expressed genes. Yeast strains grown in YPD rich medium and transferred to non-permissive temperature (37°C) for 45 min were used for chromatin preparation. The association of Mediator was measured by ChIP experiments with three different Mediator subunits: Med6-HA (head), Med15-HA (tail) and Med5-HA (middle) (**Figure 2A, B, C**). The Mediator occupancy of *ADH1*, *PYK1* and *PMA1* promoter regions remained largely unchanged except for Med15 Mediator occupancy that was reduced in *med17-444* mutant and to some extent for other mutants on *PMA1* promoter region. In contrast, all examined mutations led to a decrease in Pol II association with promoter and transcribed regions of class II genes compared to the wild type (**Figure 2D**), *med17-444* and *med17-504* showing the most pronounced effect, reflecting the effects on Pol II transcription.

Mediator head module directly interacting with TFIID and TBP, we wondered if *med17* mutations influenced occupancy of these GTFs *in vivo*. Two TFIID modules, TFIID core (Rad3-HA subunit) and TFIID kinase module (Kin28-HA) were ChIPed (**Figure 2E and F**). Interestingly, we found that mutations in Med17 Mediator subunit affected differentially the presence of the two TFIID modules on *ADH1*, *PYK1* and *PMA1* promoters. In *med17-98*

and *med17-504* mutants, Med17 mutations led to a decrease in occupancy of both TFIID modules. *med17-444* mutation specifically affected Kin28 occupancy. More unexpectedly, a decrease in Rad3 occupancy without any effect on Kin28 occupancy was observed in *med17-670*.

We also analyzed TBP occupancy in *med17* mutants (**Figure 2G**). As we previously observed for *med11* Mediator mutants, TBP occupancy was similar to that of the wild-type strain in *med17-98* and *-670* mutants. In contrast, TBP association to *ADH1*, *PYK1* and *PMA1* gene promoters was reduced in *med17-444* and *med17-504* indicating that Mediator could affect the recruitment or stability of this GTF *in vivo*.

Effects of med17 mutations on chromatin occupancy of PIC components on inducible GAL1 gene

To address the recruitment of PIC components on an inducible yeast gene, we performed ChIP experiments of Mediator, Pol II and GTFs in *med17* mutants on *GAL1* gene. The association of Mediator (Med5 subunit), Pol II, TFIID core (Rad3 subunit), TFIID kinase module (Kin28 subunit) and TBP were analyzed by ChIP at 45 min of galactose induction (data not shown). The association of PIC components was affected generally in the same manner than for the three constitutively expressed genes, *med17-504* being an exception with reduced Mediator recruitment and unmodified Rad3 and TBP recruitment on *GAL1* promoter region at 45 min of galactose induction.

Two *med17* mutants were chosen for further experiments to address the kinetics of *GAL1* induction. *med17-444* mutant is the only one that affected TBP recruitment on *GAL1* promoter and *med17-504* effects on *GAL1* gene differ from those on *ADH1*, *PYK1* and *PMA1* genes for several PIC components. Yeast cells were grown in YP raffinose medium at 30°C, then galactose was added at the 0 time point and cultures were transferred to 37°C. Samples

were collected at 0, 20, 40 and 60 min of galactose induction and a shift to 37°C. We analyzed the kinetics of Mediator, Pol II and TBP recruitment upon galactose induction in *med17-444* and *-504* mutants compared to the wild-type (**Figure 3**). Mediator recruitment was decreased in *med17-504* for all kinetics points, showing a particular Mediator behavior on *GALI* gene in this mutant (**Figure 3A**). In *med17-444* mutant, Mediator occupancy of *GALI* gene promoter reached the similar to the wild-type level, even though the kinetics of Mediator recruitment was slowed by this mutation (20-min point). Pol II occupancy generally followed that of Mediator (**Figure 3B**). We confirmed that *med17-444* mutation slowed the TBP recruitment during galactose induction. *med17-504* showed a TBP recruitment kinetics similar to that of the wild-type strain (**Figure 3C**).

Taken together, our ChIP results demonstrated that mutations in Med17 Mediator subunit could differentially affect the recruitment of PIC components and also pointed out some gene-specific effects suggesting that regulatory mechanisms might differ depending on particular Pol II-transcribed genes.

Genome-wide location analysis of Mediator, Pol II, TBP and TFIID modules

To extend our ChIP analysis of PIC assembly to the whole yeast genome, we performed ChIP-seq experiments for Mediator (Med15 subunit), Pol II and GTFs (TBP and two TFIID subunits: Kin28 and Rad3). Input DNA and DNA from ChIP with an untagged strain were used as negative controls. **Figure 4A** shows examples of ChIP-seq density distributions using the Integrated Genome Browser (IGB) for all the proteins at selected class II genes in wild-type strains. To correct for chromatin-state bias corresponding to the highly expressed genomic loci observed for control samples (Chen et al. 2012; Eyboullet et al. 2013; Teytelman et al. 2013; Jeronimo and Robert 2014), we subtracted the normalized signal of untagged strain sample from the ChIP samples for each protein as described previously

(Eyboulet et al. 2013) and in Materials and Methods. The ChIP signals of Mediator inside ORFs most likely resulted from chromatin-state bias and are mostly corrected by untagged strain signal subtraction.

We compared the location of Mediator, Pol II and GTFs peaks around transcription start sites (TSS) in wild-type strains in a metagene analysis (**Figures 4B**). To consider only Pol II transcribed gene promoters, intergenic regions encompassing Pol III-transcribed genes were excluded. We also excluded from the analysis divergent genes with double peaks for GTFs. Mean tag density (1 bp bin) was determined for each protein in a 1600-bp window centered on the TSS. These analyses show that ChIP-seq distributions of TFIID subunits and TBP superposed with each other close to TSS. As expected, Pol II is distributed inside transcribed regions. Mediator distribution was located upstream of TSS on regulatory regions. We also calculated the frequency of the transcription factor binding sites upstream of TSS and showed that Mediator and transcription factor binding profiles overlapped indicating Mediator association to the UAS (**Figure S5**).

med17 mutations can globally destabilize the PIC in vivo

To investigate whether PIC assembly is generally affected in *med17* mutants or in a gene-specific manner, we analyzed the genome-wide occupancy of Mediator, Pol II and GTFs by ChIP-seq experiments in mutants compared to the wild-type (WT) strains. The normalization step is essential for ChIP-seq comparative analysis, but widely used normalization methods like a quantile normalization step cannot be applied in situations where occupancy increases or decreases genome-wide. To be able to detect potential global effects of Mediator mutations on PIC component occupancy, we performed the following normalization step. A count of reads was determined on promoter regions of Pol II-transcribed genes for Med15, Rad3, Kin28 and TBP, and on Pol II-transcribed genes ORFs

for Pol II and then normalized relative to qPCR data on a set of selected regions, as described in Materials and Methods. Regression analysis of PIC components binding in WT versus each *med17* mutant was systematically performed.

We analyzed genome-wide Mediator occupancy only for Med15 Mediator subunit. It was impossible for us to obtain a good quality ChIP-seq results for Med5 Mediator middle module subunit or Med6 and Med8 Mediator head module subunits. Low ChIP signals were obtained for Med8-HA subunit. High background levels in ChIP-seq experiments for Med5-HA and Med6-HA subunits precluded the accurate enrichment detection. It should be noted that Mediator enrichment levels in ChIP experiments are relatively low, depending on the Mediator subunit used, possibly due to a low efficiency of formaldehyde crosslinking and/or a dynamic association of Mediator to chromatin. Chromatin associated with Med17 Mediator head subunit can be ChIPed efficiently. However, we have observed on several occasions that the addition of tags to mutant subunits could lead to an aggravation of growth phenotypes. For this reason, the Med17 occupancy was not investigated in mutant strains.

Our ChIP results on the selected class II genes showed the most pronounced effects on Pol II, TBP and Kin28 occupancy for *med17-444* mutant whereas Rad3 occupancy remained unaffected in this mutant (**Figure 2, 3**). Genome-wide analysis of *med17-444* mutant revealed a global 1.8-fold decrease (the slope of regression line, shown in red, equal to 0.55) in Med15 Mediator occupancy with a high correlation coefficient (R^2 equal to 0.93) (**Figure 5A**). As a consequence, the presence of other PIC components like Pol II, TFIIH and TBP was also globally reduced to the different extent. We observed a global decrease in genome-wide occupancy for Pol II and Kin28 TFIIK subunit (2.6 and 2.7-fold, respectively). TBP association to chromatin was also globally affected with a 1.6-fold decrease. Rad3 TFIIH core subunit binding was affected but only slightly (1.2-fold decrease), consistent with ChIP results on the *ADHI*, *PYK1* and *PMA1* gene promoters. Taken together, our genome-wide

results on *med17-444* mutant confirmed a central role for Med17 Mediator subunit in PIC formation and suggested that Mediator tail module destabilization on the chromatin had a global effect on *in vivo* binding of several PIC components but only marginally affected Rad3 occupancy.

We showed that *med17-504* mutation, equivalent to a human mutation responsible for an infantile cerebral atrophy, impaired Pol II-Mediator interaction (**Figure 1C**) and led to reduced binding of Pol II and GTFs to *ADH1*, *PYK1* and *PMA1* genes (**Figure 2**) and Mediator, Pol II and some GTFs on *GAL1* gene (**Figure 3** and data not shown). To determine how global were the effects of *med17-504*, we compared the genome-wide Mediator, Pol II, Rad3, Kin28 and TBP occupancy in this mutant and the WT (**Figure 5B**). The genome-wide binding of Mediator Med15 subunit in *med17-504* was similar to that of the wild-type strain (the slope of regression line equal to 0.93 and R^2 equal to 0.98). In contrast, the association of Pol II and GTFs was globally diminished in this mutant compared to the WT, suggesting that the impaired Pol II-Mediator interaction leads to a general PIC assembly defect. Our genome-wide analysis of *med17-444* and *-504* mutants demonstrated that a destabilization of Mediator binding to chromatin or an impaired interaction between Mediator and Pol II globally destabilize the PIC assembly *in vivo*.

Mutations in Med17 Mediator subunit specifically affect the genome-wide association of TFIIH core subcomplex

Previously, we have shown that mutations in Mediator Med11 subunit affecting its interaction with Rad3 TFIIH core subunit can selectively impair the binding of TFIIH kinase module (Esnault et al. 2008). Unexpectedly, our ChIP experiments suggested that some *med17* mutations like *med17-670* could lead to a selective dissociation of Rad3 TFIIH core subunit. We therefore addressed the question of the TFIIH core genome-wide binding in *med17*

mutants. Our ChIP-seq results revealed that the genome-wide association of this TFIH subcomplex was affected independently of TFIK subcomplex in *med17-670* (**Figure 6A**) and -98 mutants (**Figure 6B**). We observed a global 2-fold decrease of the Rad3 TFIH core subunit binding in these mutants with the slope of regression line equal to 0.55 and 0.51 for *med17-670* and -98, respectively, with high correlation coefficients. In contrast, the association of Kin28 TFIK subunit was similar to the WT level (the slope of regression line equal to 1.01 and 0.94, R^2 equal to 0.98 and 0.94 for *med17-670* and -98, respectively). The presence of Med15 Mediator subunit was mostly unchanged or only slightly decreased (the slope of regression line equal to 0.87 and 0.80, R^2 equal to 0.93 and 0.94 for *med17-670* and -98, respectively). The genome-wide binding of TBP was not reduced and even slightly increased (the slope of regression line equal to 1.25 and 1.14, R^2 equal to 0.92 and 0.96 for *med17-670* and -98, respectively). In *med17-670* mutant, a global 1.4-fold decrease in Pol II occupancy was observed compared to the WT strain with the slope of regression line equal to 0.72 and a high correlation coefficient equal to 0.97. Pol II association was also reduced in *med17-98* mutant showing 1.2-fold global decrease with a reduced correlation coefficient of 0.83 that suggested a more complex transcriptional effect for this mutation. Our genome-wide analysis of PIC components of *med17-670* and -98 mutants suggests that Mediator can selectively affect the TFIH core subcomplex binding to the yeast genome without impairing that of the TFIH kinase module or TBP.

Discussion

In this work, we have performed a genome-wide analysis of PIC assembly *in vivo* taking advantage of Mediator mutants of essential Med17 head module subunit. We propose that Mediator head module in general, and Med17 in particular, independently coordinates the PIC assembly *in vivo* at different levels. (i) Our results support a central role of Mediator

Med17 subunit in Mediator organization and function, since *med17* mutants globally affect Pol II transcription and specific mutations in Med17 could result in a global destabilization of PIC assembly *in vivo*. (ii) We show that Mediator recruitment stabilizes TFIIK kinase and TFIIH core modules independently, since *med17* mutations can differentially affect the genome-wide occupancy of TFIIH modules. We therefore propose that the recruitment or the stability of TFIIH modules is regulated independently on yeast genome. (iii) Our ChIP and ChIP-seq results show that Mediator selectively contributes to TBP recruitment or stabilization to chromatin, supporting Mediator head-TBP functional contact. (iv) This work suggests a mechanism for an infantile cerebral atrophy caused by a mutation in Med17. We propose that the disease results from a defect in Mediator-Pol II contact that leads to a global destabilization of PICs.

Previously, we proposed a model of PIC assembly through multiple pathways (Esnault et al. 2008; Soutourina et al. 2011). Our results suggested several pathways for PIC formation that may coexist on the promoters of Pol II-transcribed genes challenging the linear model of PIC assembly based on *in vitro* studies and expanding the regulatory mechanisms of gene transcription. We proposed that Mediator association to regulatory regions (**Figure 7A**) leads to PIC formation (**Figure 7E**) via preinitiation intermediates (**Figure 7C, C', D**) suggesting two hypothetical intermediates (**Figure 7B, B'**). In line with this model, the present work suggests the existence of new intermediates in PIC formation (**Figure 7**). The model from Esnault et al. (Esnault et al. 2008) was significantly enriched and consolidated considering a direct *in vivo* Mediator-Pol II interaction (Soutourina et al. 2011) (**Figure 7B**) and the present work that suggests the existence of Mediator-TFIIK and Mediator-TBP PIC intermediates (**Figure 7B'', B''', C''**). TFIIIE occupancy was not analyzed in this study, that is why no information on this GTF is present on the B'', B''' and C'' cartoons. We also incorporate TBP recruitment to our current model.

Mediator head module directly interacts with TFIID complex via Med11-Rad3 contact and this interaction is important for the recruitment of this GTF to promoters (Esnault et al. 2008). Mutations in Med11 subunit differentially affected the promoter occupancy of TFIID core and TFIID kinase modules. A *med11* mutation that impaired Mediator-TFIID contact destabilized TFIID on a group of promoters. Our genome-wide analysis of *med17* mutants supports a functional role of Mediator in the independent recruitment of TFIID core (Rad3 subunit) and TFIID kinase (Kin28 subunit) modules. Indeed, *med17-504* mutation reduced the global association of Rad3 and Kin28 to a similar extent. On the contrary, *med17-444* mutation, while affecting the occupancy of both TFIID modules, displayed a particularly pronounced decrease in Kin28 occupancy. These results show that TFIID module was further destabilized in *med17-444* mutant, similarly to *med11-T47A* mutant (**Figure 7C**). Vice versa, we also observed that the genome-wide association of TFIID core subcomplex could be affected independently of TFIID subcomplex in *med17-670* and *med17-98* mutants. A large decrease of Rad3 genome-wide occupancy was observed in these mutants, whereas the association of Kin28 TFIID subunit remained unchanged. Taken together, our results suggest that Mediator differentially affects the recruitment or the stabilization of TFIID modules and that the stability of TFIID modules is regulated independently within the PICs. We show that two biochemically separable TFIID modules have independent *in vivo* behavior. The recruitment of core TFIID module can be selectively affected, independently of TFIID and TBP (**Figure 7C''**). This intermediate presumes the potential existence of Mediator-TFIID-containing subcomplex in the absence of core TFIID (**Figure 7B'''**). In line with this suggestion, one of the Mediator functions consists in stimulating the CTD phosphorylation by Kin28 TFIID kinase (Kim et al. 1994). Moreover, Kin28 was reported to phosphorylate Med4 and Med14 Mediator subunits *in vitro* (Liu et al. 2004) and Kin28 phosphorylation of Med4 occurs *in vivo* (Guidi et al. 2004). Recently, Kin28 TFIID kinase has been reported to regulate

the transient association of Mediator to core promoters and a release of Mediator from the PIC and the promoter escape (Jeronimo and Robert 2014; Wong et al. 2014), highlighting reciprocal links between Mediator and TFIID kinase module.

TBP directly binds to the Mediator head module via Med8-Med18-Med20 subcomplex (Lariviere et al. 2006; Cai et al. 2010). Previously, human Mediator and TBP-containing TFIID complexes have been shown to assemble cooperatively on promoter DNA *in vitro* (Johnson et al. 2002). In accordance with *in vitro* data, our *in vivo* genome-wide analysis of yeast Mediator mutants showed that TBP chromatin binding could be affected by Mediator mutations (*med17-444* and *-504*). It should be noted that the loss of Mediator function in *med17-138* mutant has been previously shown to reduce TBP occupancy of several class II promoters (Kuras and Struhl 1999; Li et al. 1999). TBP association depended on the Mediator mutant examined in this study. The genome-wide binding of this GTF did not decrease or even slightly increased in *med17-670* and *med17-98* mutants. In this work, we identified mutations in Med17 subunit that affected global TBP binding to chromatin further supporting a coordinated behavior with Mediator. Interestingly, even though the effects of *med17* mutations are global, the range of the effects can depend on specific genes. For example, *med17-504* mutation resulted in a global decrease in TBP occupancy, but did not affect TBP recruitment to *GALI* gene promoter upon transcription activation. Conversely, *med17-98* mutation did not reduce the genome-wide TBP or TFIID occupancy. However, in this mutant the TFIID binding together with that of Pol II and core TFIID module was reduced on the three model promoters, but not that of TBP. These results suggest the existence of a Mediator-TBP intermediate in PIC assembly (**Figure 7B''**), consistent with a direct interaction between Mediator head and TBP protein from *in vitro* experiments (Lariviere et al. 2006; Cai et al. 2010).

In *med17-444* and *-504* mutants, PIC assembly was destabilized leading to a decrease

of Pol II, Rad3 TFIID core subunit, Kin28 TFIID subunit and TBP presence. The following mechanistic explanations specific for each mutant can be proposed. The presence of Mediator tail module (Med15 subunit) on chromatin was globally affected in *med17-444* mutant. We propose that, as a consequence of this effect on tail module occupancy, overall PIC assembly is destabilized with a large decrease in Pol II, Kin28 and TBP occupancy and a moderate effect on Rad3 occupancy. In contrast, the binding of Mediator remained globally unaffected in *med17-504* mutant. This mutation led also to a global decrease in Pol II and GTFs occupancy. We propose that an impaired interaction between Mediator (Med17 subunit) and Pol II (Rpb3 subunit) in this mutant destabilized the global PIC assembly *in vivo*, as was observed for Pol II *rpb3-2* mutant with reduced Mediator-Pol II interaction (Soutourina et al. 2011) highlighting functional importance of Mediator-Pol II contact (**Figure 7B**). Taken together, our results revealed that specific point mutations in Med17 Mediator head subunit (*med17-Q444P*, *med17-M504P*) could result in a global destabilization of PIC assembly *in vivo* through two different mechanisms, a destabilization of Mediator binding to chromatin or an impaired interaction between Mediator and Pol II, consistent with a central role of Med17 subunit in Mediator function.

Among *med17* mutants examined, only *med17-444* mutation reduced the global Mediator tail (Med15 subunit) occupancy. Other *med17* mutations did not affect or only slightly reduced the level of genome-wide Med15 binding to the chromatin. Our ChIP experiments on *ADH1*, *PYK1* and *PMA1* genes did not reveal any decrease in Mediator head or middle occupancy in *med17* mutants. Moreover, Mediator stability in *med17* mutants examined in this study is supported by mass spectrometry and CoIP experiments indicating the presence of all Mediator modules. A decrease in ChIP signal for Mediator tail module in *med17-444* might reflect a conformational change in Mediator complex bound to chromatin leading to reduced cross-linking efficiency or Mediator tail destabilization on chromatin.

Mediator is conserved in all eukaryotes. The mutated residues in *med17* mutants characterized in this study correspond to the same (human Q310 for yeast Q444P in *med17-444* and human E314 for yeast E448Q, one of the three mutations in *med17-98*) or functionally close (human L371 for yeast M504P in *med17-504*) amino acids in human Med17 protein (**Figure S6**). Given the conservation of the Mediator subunits, the molecular mechanisms of Mediator function in PIC assembly are likely to be conserved from yeast to human and may extend to all eukaryotes. We showed that *med17-M504P* mutation, equivalent to human mutation responsible for an infantile cerebral atrophy, led to a severe temperature-sensitive phenotype in yeast, impaired Pol II-Mediator interaction and resulted in a decrease of genome-wide occupancy of all PIC components tested except Mediator itself. Our results suggest that infantile cerebral atrophy is the direct consequence of a defect in Mediator-Pol II contact and a global destabilization of PICs. Recently, we have shown that Mediator, in addition to its coactivator role, also can serve as a link between transcription and DNA repair (Eyboulet et al. 2013). Several *med17* mutants were UV-sensitive in a global-genome repair deficient context. The *med17-504* and other mutants analyzed in this study do not show any increase in UV sensitivity (data not shown), suggesting that these mutations led to transcriptional defects only.

One of the questions that we addressed analyzing the impact of conditional mutations in Med17 Mediator subunit on genome-wide PIC assembly was its gene-specific versus global nature. A high correlation between genome-wide densities of PIC components points out global effects of *med17* mutations. It should be noted that the range of the effects could still be gene-specific. Consistent with a central role of Med17 subunit in Mediator function, all *med17* mutations reduced the global Pol II chromatin binding, but to a different extent. The global decrease of Pol II occupancy in *med17* mutants compared to the wild type was associated with a high correlation coefficient, indicating that most of the genes were affected

to a similar extent. *med17-98* mutant was an exception since it showed a global decrease of Pol II presence but with a reduced correlation coefficient (R^2 equal to 0.83) suggesting a more complex and gene-specific transcriptional effect for this mutation. The occupancy of Kin28 TFIIF subunit remained globally unchanged in *med17-98* mutant compared to the wild type, but Kin28 ChIP signals were decreased on *ADH1*, *PYK1* and *PMA1* gene promoter. *med17-444* mutation did not modify the presence of Rad3 TFIIF core subunit on the three constitutively expressed genes or *GALI* inducible gene, but our genome-wide analysis revealed a slight decrease in global Rad3 occupancy. *med17-504* mutant had a particular effect on *GALI* gene showing a reduced Mediator association and unmodified Rad3 and TBP recruitment, whereas the genome-wide Mediator occupancy was similar to that of the wild-type and occupancy of all other PIC components was globally reduced. Our results suggest that specific PIC assembly pathways could be active on specific genes and/or that multiple pathways could operate on the same gene.

In conclusion, we suggest that Mediator differentially coordinates multiple steps of a branched assembly pathway *in vivo* at the level of independent recruitment or stabilization of several PIC components including Pol II, TBP and two TFIIF modules. The global effects of Mediator mutations on PIC formation are consistent with a central role of Med17 subunit in Mediator function and organization. Additional PIC components, for example, GTFs that were not examined in this study might contribute to PIC assembly. Together with Mediator, other coactivators like SAGA or TFIID complexes modify the PIC assembly and activity. In the future, it would be relevant to consider the dynamics of PIC assembly to improve our understanding of *in vivo* transcription mechanisms at work in eukaryotes.

Materials and Methods

Strains and plasmids

All *S. cerevisiae* strains are described in **Table S3**. All plasmids are listed in **Table S4**. The oligonucleotides used in this study can be found in **Table S5**.

ChIP and ChIP-seq

ChIP experiments were performed as described (Ghavi-Helm et al. 2008). 100 ml cell cultures were grown to exponential phase in YPD medium at 30°C, shifted for 45 minutes to 37°C and cross-linked with 1% formaldehyde for 10 min. The 3HA-tagged proteins were immunoprecipitated with 12CA5 antibody and Pol II with 8WG16 anti-CTD antibody, bound to IgG magnetic beads (Dynabead). ChIP-seq experiments were performed as described (Eyboulet et al. 2013). Chromatin preparation for ChIP-seq experiments was performed as described for conventional ChIP, except that an additional sonication step with Bioruptor (Diagenode, 6 cycles of 30s with medium intensity setting) was included to generate DNA fragments of approximately 200 bp mean size. DNA sequencing of 40 nt-tags was performed on Solexa genome analyzer GA-IIx or Hi-Seq using the procedures recommended by the manufacturer (Illumina). Input DNA and DNA from ChIP with untagged strain were used as negative controls.

Coimmunoprecipitation experiments

Whole yeast extracts were prepared from 100 ml of cells growing to exponential phase in YPD medium at 30°C. Whole yeast extract preparation, immunoprecipitation in IP buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 20% glycerol, 1 mM dithiothreitol, 0.5 mM EDTA, 0.05% NP-40) supplemented with a protease inhibitor cocktail (Complete, Roche) and 1 mM PMSF and Western blotting were performed as described previously (Soutourina et al. 2006).

The 12CA5 anti-HA antibodies were used against HA- or Myc-tagged proteins and rabbit polyclonal anti-Med14 and anti-Med17 antibodies against corresponding N-terminal peptides were used to detect Med14 and Med17 Mediator subunits, respectively.

Formaldehyde cross-linking experiments were performed as described previously (Soutourina et al. 2011). Briefly, cells were cross-linked with 1% formaldehyde for 10 min, as described for ChIP. Lysis was performed in 0.5 ml FA/SDS buffer in the presence of glass beads (0.75 ml, 425 to 600 μ m) by vortexing for 45 min at 4°C. Crude extract preparation and immunoprecipitations were performed as described for conventional CoIP except that the immune complexes were washed as described for ChIP. A sonication step was added to recover chromatin-associated proteins. Denaturing conditions (8 M urea for 20 min at room temperature, cooled to 4°C and diluted with lysis buffer to final urea concentration of 2 M) were used before immunoprecipitation to dissociate the subunits of Pol II and of Mediator.

Mass spectrometry analysis

To analyze the Mediator integrity in *med17* mutants, Mediator was immunoprecipitated from crude extracts of Med5-HA tagged strains as described for CoIP experiments except that protein G magnetic beads were used instead of IgG magnetic beads (Dynabead) and the elution was performed for 20 min at 65°C instead of 2 min at 85°C. *MED17* strain carrying non-tagged Mediator subunit was used as a negative control. Proteins from each eluate were separated by SDS-PAGE in order to fractionate the protein samples into three fractions and exclude the regions corresponding to heavy and light IgG chains. The gels were stained with Coomassie Blue and each fraction was cut in bands of about 2 mm and subjected to in-gel trypsin digestion with the Progest robot (Genomic Solutions, Chemsford, MA) using standard conditions including reduction and alkylation. After overnight tryptic digestion, peptides were extracted with 60% acetonitrile and 0.1% (v/v) formic acid. Tryptic

peptides extracted from different bands of three gel regions were pooled, vacuum dried and resuspended in 0.1% (v/v) formic acid prior to nanoLC-MS/MS mass spectrometry analyses. The same cutting pattern of the SDS-PAGE lane was performed for each eluate.

NanoLC-MS/MS analyses were performed with the Triple-TOF 4600 mass spectrometer (AB Sciex) coupled to the nanoRSLC ultra performance liquid chromatography (UPLC) system (Thermo Scientific) equipped with a trap column (Acclaim PepMap100C18, 75 μ m i.d. \times 2 cm, 3 μ m) and an analytical column (Acclaim PepMapRSLCC18, 75 μ m i.d. \times 15 cm, 2 μ m, 100 Å). Peptide separation was performed with a 5 to 35% solvent B gradient in 40 min. Solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in 100% acetonitrile. NanoLC-MS/MS experiments were conducted in a Data Dependent acquisition method by selecting the 20 most intense precursors for CID fragmentation with Q1 quadrupole set at low resolution for better sensitivity. Raw data were processed with MS Data Converter software (AB Sciex) for generating .mgf data files and protein identification were performed using the MASCOT search engine (Matrix science, London, UK) against the Swissprot database (release 2014_08) with carbamidomethylation of cysteines and oxidation of methionines set as fixed and variable modifications respectively. Peptide and fragment tolerance were respectively set at 20 ppm and 0.05 Da. Results were analysed with Scaffold 3.6.5 software (Proteome Software). Only proteins with at least two unique peptides and 95 % probability for both peptides and proteins identifications were considered.

Quantitative RT-PCR analysis

RNA was extracted with hot acidic phenol following a protocol derived from Schmitt et al. (Schmitt et al. 1990). Reverse transcription of 0.5 μ g RNA samples was performed using iScript cDNA synthesis Kit (BIO-RAD) with a mix of oligo(dT) and random hexamers for priming. Quantitative PCR results were normalized using 25S rRNA, 18S rRNA and *SCR1*

RNA as internal controls. Values represent the average of three independent experiments, and error bars indicate standard deviations.

Data analysis

ChIP-seq data were analyzed essentially as described (Eyboulet et al. 2013). The sequences were aligned on *S. cerevisiae* genome (UCSC version sacCer3) with Bowtie version 0.12.7 (Langmead et al. 2009). Only uniquely mapped tags were used, and a maximum of two mismatches was allowed. Conversions to different file formats were performed using Samtools version 0.1.16. To avoid possible sequencing artifacts, reads with the best-quality scores were conserved for each position. The number of mappable tags for each ChIP-seq experiment is indicated in **Table S6**.

To calculate ChIP-seq density, reads were extended to an assumed fragment length of 150 to 180 nt, and a count of reads was determined per 1-bp bin using Bedtools version 2.15.0. ChIP-seq density profiles were displayed using IGB yeast genome browser. We used input DNA and DNA from ChIP with untagged strain as negative controls. The ChIP sample from the untagged strain generally shows very low tag density on the genome, except some regions representing <1% of the genome that display an apparent enrichment. It should be noted that the majority of these regions were located inside the highly transcribed class II genes. To correct for this non-random background distribution, we subtracted the normalized signal of untagged strain sample from the ChIP samples for each protein. The ChIP signals of untagged strain have different impacts on the ChIP signals for different proteins. For this reason, the subtraction step was preceded by a normalization of the ChIP signal of an untagged strain compared with the corresponding ChIP sample based on the qPCR analyses of nontranscribed control regions.

The significantly enriched regions were identified using peak calling MACS2 software version 2.0.10.20120703 (Zhang et al. 2008). To avoid possible sequencing artifacts, reads beyond five repetitions at the same position of the genome were removed. A minimum false discovery rate (q-value) cutoff peak detection of 0.01 was used for Med15 and 0.05 for Pol II, Rad3, Kin28 and TBP. This step was followed by PeakSplitter step of PeakAnalyzer software for subdivision of ChIP-seq regions into discrete peaks (Salmon-Divon et al. 2010). The significantly enriched regions were also identified in control ChIP sample with untagged strain using Bayespeak version 1.10.0 (Spyrou et al. 2009) with a threshold of 0.9. Around 200 peaks were identified by this method in control sample and were used to correct peakcalling data from Med15, Rad3, Kin28 and TBP ChIP-seq.

GTFs, Mediator and Pol II significantly enriched peaks were annotated relative to the intergenic region or the ORF in which they were detected using BedTools version 2.15.0. The nomenclature used for intergenic regions was as follows: tandem regions were named relative to the downstream-located gene, divergent and convergent regions were named relative to the two genes surrounding the region.

To compare read counts in WT and *med17* mutant ChIP-seq data, a count of reads was determined on promoter regions of Pol II-transcribed genes for Rad3, Kin28, TBP and Med15, and on Pol II-transcribed genes ORFs for Pol II, using Bedtools. For this step, all mapped reads were conserved to allow quantitative comparison between the different samples. Read numbers were normalized relative to qPCR data on a set of selected regions using the ratio between WT and mutant strains. The ratio of ChIP-qPCR values between WT and mutant strains was calculated for each selected region. In parallel, the ratio of ChIP-seq reads between WT and mutant strains was determined for the same region. The median of normalization coefficients between ChIP-qPCR and reads ratios was used for data normalization.

The coordinates of transcription factor binding sites (TFBS) were retrieved from the UCSC transRegCode genome browser track. These data were then filtered to keep only those having ChIP evidence for DNA binding (9777 TFBS). We then use BedTools to associate the TFBS with the closest TSS (2337 TSS). The frequency of TFBS was finally computed in a 3000-bp windows centered on the TSS using R software with 1-bp bin. Mediator, Pol II, TBP, Rad3 and Kin28 tag density was determined around the same 2337 TSS and mean tag density for each nucleotide position was then calculated and plotted over the 3000-bp window.

Data access

The ChIP-seq data have been deposited to the Array Express under accession numbers E-MTAB-2961 for the mutant strains and E-MTAB-1595 for the wild-type strain.

Acknowledgements

We thank the SPI (CEA/Saclay) for monoclonal antibodies. This work has benefited from the facilities and expertise of the high throughput-sequencing and of the SICaPS mass spectrometry platforms of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr). This work was supported by the Agence Nationale de la Recherche (Grant number ANR_11_BSV8_020_01). F.E. was supported by a grant from the Fondation ARC (Grant number DOC20130606697).

Figure Legends

Figure 1. *med17* thermosensitive mutants.

(A) Med17 mutant phenotypes. Cultures of wild-type (WT) and mutant *med17* yeast strains were serially diluted, spotted on YPD agar plates and incubated for 3 days at permissive (30°C) or non-permissive (37°C) temperature.

(B) Mass spectrometry analysis of Mediator integrity in *med17* mutants. Mediator was immunoprecipitated through Med5-HA from crude extracts using magnetic protein G beads coupled to anti-HA antibodies. Mediator subunits identified by mass spectrometry analysis in a wild-type and *med17* mutants are indicated. *MED17* strain carrying non-tagged Mediator subunit was used as a negative control (Control IP). Mediator subunits identified in all samples are shown in red.

(C) Interaction between Pol II and Mediator in *med17* mutants. Rpb3-Myc Med17-EGFP strains were grown at 30°C in YPD medium and cross-linked or not with formaldehyde (FA), as indicated. Med17-EGFP was immunoprecipitated (IP) with anti-EGFP antibody from crude extracts (Input) and analyzed by western blotting with anti-Myc antibody (CoIP) against Rpb3. The cross-linked Rpb3-Med17 band is indicated in red. The position of unidentified cross-linked proteins with the tagged Med17 or Rpb3 subunits is indicated by a vertical bar.

Figure 2. Effect of *med17* mutations on Mediator, Pol II, TFIID, and TBP occupancy.

Cells were grown to exponential phase at 30°C on YPD medium and then transferred for 45 min to 37°C. Quantitative ChIP experiments were performed using anti-HA antibody (12CA5) against Mediator subunits Med6-HA (A), Med15-HA (B) or Med5-HA (C), Rad3-HA (E), Kin28-HA (F), TBP-HA (G) or anti-Rpb1 Pol II antibody (8WG16) (D). Immunoprecipitated DNA was amplified with primers corresponding to *ADHI*, *PMAI* and *PYK1* ORF (O) or promoters (P). *GALI* ORF was used as a negative control. Mean values and

standard deviation (indicated by error bars) of three independent experiments are shown.

Figure 3. Kinetics of galactose induction in *med17-444*, *504* and wild-type strains.

Yeast strains were grown in a raffinose-supplemented medium at 30°C, then galactose was added, cultures were transferred at the same time to 37°C and samples were collected for ChIP experiments at indicated time points upon galactose induction (T0, T20, T40 and T60 min). Quantitative ChIP experiments were performed using anti-HA antibody (12CA5) against Mediator subunit Med5-HA (A), TBP-HA (C) or anti-Rpb1 Pol II antibody (8WG16) (B). Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to *GALI* gene promoter (*GALI P*) or ORF (*GALI O*). A nontranscribed region on chromosome V was used as a negative control.

Figure 4. Enrichment profiles of Mediator, Pol II, TFIIH and TBP on yeast genome.

(A) Examples of Mediator, Pol II, TBP, and TFIIH ChIP-seq enrichment profiles on selected class II genes. Densities of sequence tags were assessed from ChIP-seq experiments performed with Med15-HA (Mediator), TBP-HA, Rad3-HA (TFIIH core), and Kin28-HA (TFIIK) strains using anti-HA antibody. Pol II was immunoprecipitated using anti-Rpb1 antibody. ChIP-seq density profiles are displayed using IGB yeast genome browser. Input DNA and DNA from ChIP with an untagged strain were used as negative controls. Densities of sequence tags were displayed after subtraction of normalized control of untagged strain.

(B) Distribution of Mediator, TBP, TFIIH and Pol II ChIP-seq densities around TSS. Intergenic regions encompassing Pol III-transcribed genes and divergent genes were excluded. The tag density was determined for each protein in a 1600-bp window centered on the TSS from (Xu et al. 2009; Rhee and Pugh 2012). Mean tag density for each nucleotide position

was then calculated and plotted over the window.

Figure 5. Correlation of PIC components density between WT and *med17-444* and *-504* mutant strains.

The density of sequence tags in Med15, Rad3, Kin28 and TBP ChIP-seq experiments has been calculated for promoter regions of Pol II-transcribed genes. The density of sequence tags in Pol II ChIP-seq experiments has been calculated for Pol II-transcribed genes. 3852 Pol II-enriched genes were used for these analyses. Tag densities have been normalized relative to qPCR data on a set of selected genes. Each point on the plot corresponds to one promoter region or one ORF. Promoter regions correspond to intergenic regions in tandem or in divergent orientation, excluding intergenic regions encompassing Pol III-transcribed genes. A linear regression (red line) for ChIP-seq density in WT versus ChIP-seq density in mutant and a R^2 correlation coefficient are indicated.

(A) Med15, Pol II, Rad3, Kin28 and TBP ChIP-seq density in WT versus ChIP-seq density in *med17-444* on class II promoter regions or ORFs (for Pol II).

(B) Med15, Pol II, Rad3, Kin28 and TBP ChIPseq density in WT versus ChIP-seq density in *med17-504* on class II promoter regions or ORFs (for Pol II).

Figure 6. Correlation of PIC components density between WT and *med17-670* and *-98* mutant strains.

The density of sequence tags in Med15, Rad3, Kin28 and TBP ChIP-seq experiments has been calculated for promoter regions of Pol II-transcribed genes as in Figure 5.

(A) Med15, Pol II, Rad3, Kin28 and TBP ChIP-seq density in WT versus ChIP-seq density in *med17-670* on class II promoter regions or ORFs (for Pol II).

(B) Med15, Pol II, Rad3, Kin28 and TBP ChIPseq density in WT versus ChIP-seq density in *med17-98* on class II promoter regions or ORFs (for Pol II).

Figure 7. Model of multiple steps of PIC assembly differentially orchestrated by Mediator.

Different preinitiation intermediates in multistep pathways of PIC assembly are indicated and connected by arrows. Selected PIC components are shown in the figure. A model from Esnault et al. (Esnault et al. 2008) was enriched considering a direct *in vivo* Mediator-Pol II interaction (Soutourina et al. 2011) (B) and the present work (B'', B''', C''). The predicted intermediates are indicated by the translucent colors.

References

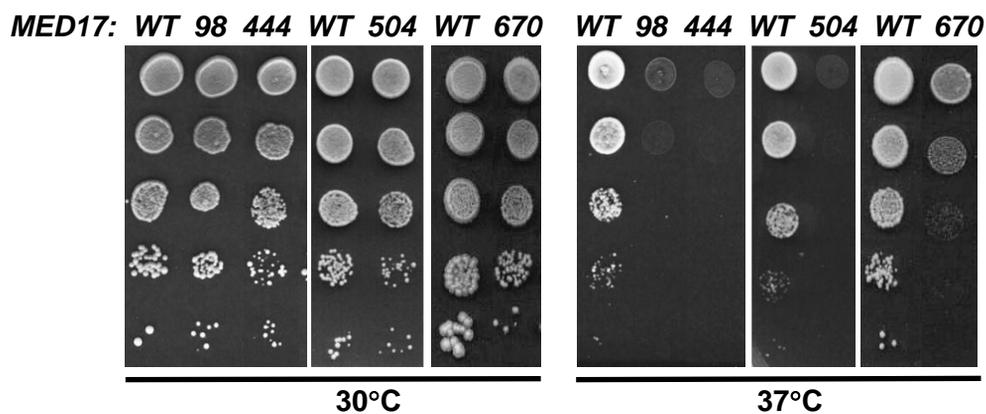
- Andrau JC, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M, Holstege FC. 2006. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* **22**(2): 179-192.
- Bourbon HM. 2008. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional Mediator complex. *Nucleic Acids Res* **36**(12): 3993-4008.
- Buratowski S, Hahn S, Guarente L, Sharp PA. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**(4): 549-561.
- Cai G, Imasaki T, Yamada K, Cardelli F, Takagi Y, Asturias FJ. 2010. Mediator head module structure and functional interactions. *Nat Struct Mol Biol* **17**(3): 273-279.
- Chen Y, Negre N, Li Q, Mieczkowska JO, Slattery M, Liu T, Zhang Y, Kim TK, He HH, Zieba J et al. 2012. Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat Methods* **9**(6): 609-614.
- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M. 2008. Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Mol Cell* **31**(3): 337-346.
- Eyboulet F, Cibot C, Eychenne T, Neil H, Alibert O, Werner M, Soutourina J. 2013. Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment. *Genes Dev* **27**(23): 2549-2562.
- Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S et al. 2008. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* **455**(7212): 547-551.
- Gade P, Singh AK, Roy SK, Reddy SP, Kalvakolanu DV. 2009. Down-regulation of the transcriptional mediator subunit Med1 contributes to the loss of expression of metastasis-associated dapk1 in human cancers and cancer cells. *Int J Cancer* **125**(7): 1566-1574.
- Ghavi-Helm Y, Michaut M, Acker J, Aude JC, Thuriaux P, Werner M, Soutourina J. 2008. Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes Dev* **22**(14): 1934-1947.
- Grunberg S, Hahn S. 2013. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* **38**(12): 603-611.

- Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M. 2004. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* **32**(18): 5379-5391.
- Guidi BW, Bjornsdottir G, Hopkins DC, Lacomis L, Erdjument-Bromage H, Tempst P, Myers LC. 2004. Mutual targeting of mediator and the TFIID kinase Kin28. *J Biol Chem* **279**(28): 29114-29120.
- Hashimoto S, Boissel S, Zarhrate M, Rio M, Munnich A, Egly JM, Colleaux L. 2011. MED23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science* **333**(6046): 1161-1163.
- He Y, Fang J, Taatjes DJ, Nogales E. 2013. Structural visualization of key steps in human transcription initiation. *Nature* **495**(7442): 481-486.
- Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**(5): 717-728.
- Jeronimo C, Robert F. 2014. Kin28 regulates the transient association of Mediator with core promoters. *Nat Struct Mol Biol* **21**(5): 449-455.
- Johnson KM, Wang J, Smallwood A, Arayata C, Carey M. 2002. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev* **16**(14): 1852-1863.
- Kaufmann R, Straussberg R, Mandel H, Fattal-Valevski A, Ben-Zeev B, Naamati A, Shaag A, Zenvirt S, Konen O, Mimouni-Bloch A et al. 2010. Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am J Hum Genet* **87**(5): 667-670.
- Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**(4): 599-608.
- Kornberg RD. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**(5): 235-239.
- Koschubs T, Lorenzen K, Baumli S, Sandstrom S, Heck AJ, Cramer P. 2010. Preparation and topology of the Mediator middle module. *Nucleic Acids Res* **38**(10): 3186-3195.
- Kuras L, Struhl K. 1999. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**(6736): 609-613.
- Kuuselo R, Savinainen K, Sandstrom S, Autio R, Kallioniemi A. 2011. MED29, a component of the mediator complex, possesses both oncogenic and tumor suppressive characteristics in pancreatic cancer. *Int J Cancer*.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.
- Lariviere L, Geiger S, Hoepfner S, Rother S, Strasser K, Cramer P. 2006. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* **13**(10): 895-901.
- Lariviere L, Plaschka C, Seizl M, Wenzek L, Kurth F, Cramer P. 2012. Structure of the Mediator head module. *Nature* **492**(7429): 448-451.
- Li LH, He J, Hua D, Guo ZJ, Gao Q. 2010. Lentivirus-mediated inhibition of Med19 suppresses growth of breast cancer cells *in vitro*. *Cancer Chemother Pharmacol* **68**(1): 207-215.
- Li XY, Virbasius A, Zhu X, Green MR. 1999. Enhancement of TBP binding by activators and general transcription factors. *Nature* **399**(6736): 605-609.
- Linder T, Zhu X, Baraznenok V, Gustafsson CM. 2006. The classical *srb4-138* mutant allele causes dissociation of yeast Mediator. *Biochem Biophys Res Commun* **349**(3): 948-953.

- Liu Y, Kung C, Fishburn J, Ansari AZ, Shokat KM, Hahn S. 2004. Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol Cell Biol* **24**(4): 1721-1735.
- Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ et al. 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**(6159): 1238724.
- Ranish JA, Hahn S. 1996. Transcription: basal factors and activation. *Curr Opin Genet Dev* **6**(2): 151-158.
- Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**(7389): 295-301.
- Ries D, Meisterernst M. 2011. Control of gene transcription by Mediator in chromatin. *Semin Cell Dev Biol* **22**(7): 735-740.
- Risheg H, Graham JM, Jr., Clark RD, Rogers RC, Opitz JM, Moeschler JB, Peiffer AP, May M, Joseph SM, Jones JR et al. 2007. A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat Genet* **39**(4): 451-453.
- Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. 2012. Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A* **109**(44): 17931-17935.
- Salmon-Divon M, Dvinge H, Tammaja K, Bertone P. 2010. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**: 415.
- Schmitt ME, Brown TA, Trumpower BL. 1990. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* **18**(10): 3091-3092.
- Schwartz CE, Tarpey PS, Lubs HA, Verloes A, May MM, Risheg H, Friez MJ, Futreal PA, Edkins S, Teague J et al. 2007. The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. *J Med Genet* **44**(7): 472-477.
- Soutourina J, Bordas-Le Floch V, Gendrel G, Flores A, Ducrot C, Dumay-Odelot H, Soularue P, Navarro F, Cairns BR, Lefebvre O et al. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**(13): 4920-4933.
- Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**(6023): 1451-1454.
- Spyrou C, Stark R, Lynch AG, Tavares S. 2009. BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics* **10**: 299.
- Takagi Y, Calero G, Komori H, Brown JA, Ehrensberger AH, Hudmon A, Asturias F, Kornberg RD. 2006. Head module control of mediator interactions. *Mol Cell* **23**(3): 355-364.
- Takagi Y, Kornberg RD. 2006. Mediator as a general transcription factor. *J Biol Chem* **281**(1): 80-89.
- Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, Szutorisz H, Swanson SK, Martin-Brown S, Washburn MP et al. 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**(1): 92-104.
- Teytelman L, Thurtle DM, Rine J, van Oudenaarden A. 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A* **110**(46): 18602-18607.
- Thompson CM, Koleske AJ, Chao DM, Young RA. 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**(7): 1361-1375.
- Vijayvargia R, May MS, Fondell JD. 2007. A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. *Cancer Res* **67**(9): 4034-4041.

- Wong KH, Jin Y, Struhl K. 2014. TFIIF Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol Cell* **54**(4): 601-612.
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**(7232): 1033-1037.
- Zhang X, Krutchinsky A, Fukuda A, Chen W, Yamamura S, Chait BT, Roeder RG. 2005. MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol Cell* **19**(1): 89-100.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.
- Zhu X, Wiren M, Sinha I, Rasmussen NN, Linder T, Holmberg S, Ekwall K, Gustafsson CM. 2006. Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell* **22**(2): 169-178.

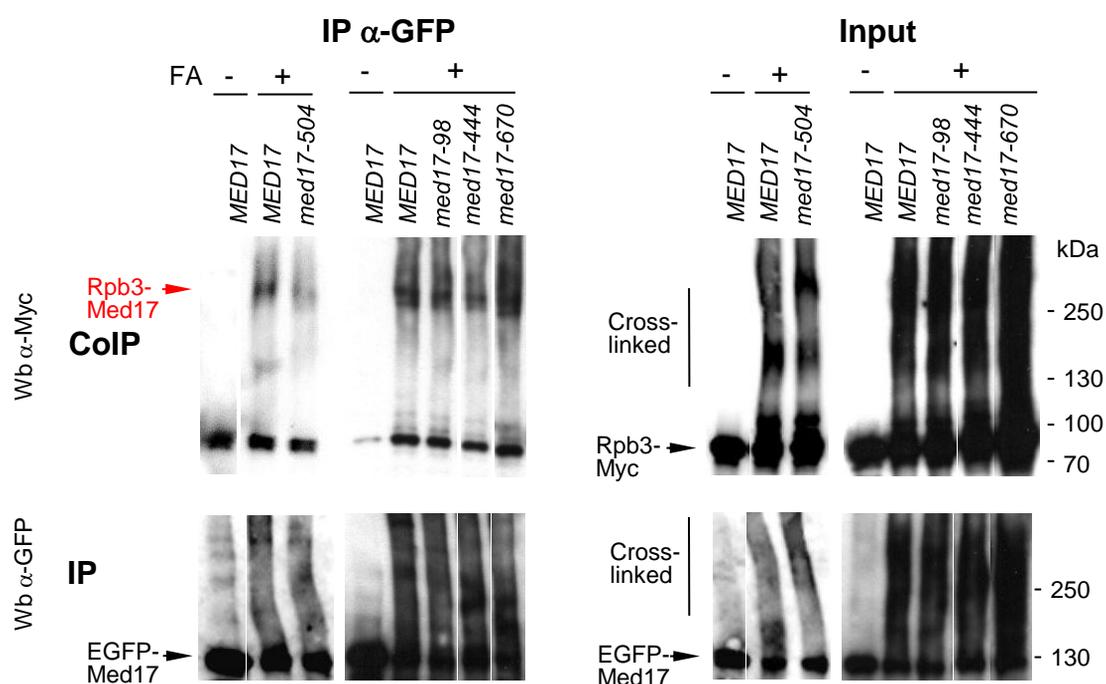
A

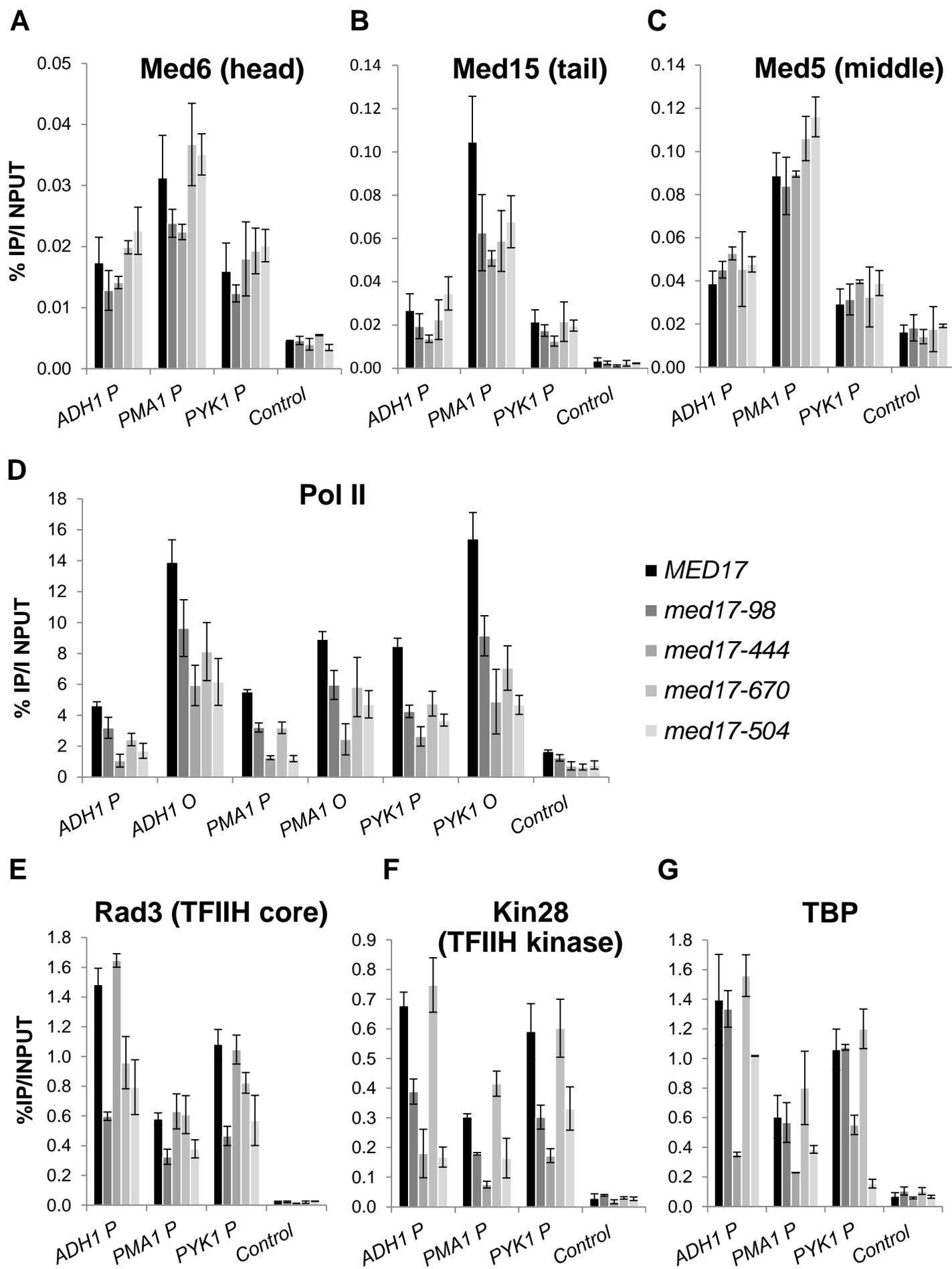


B

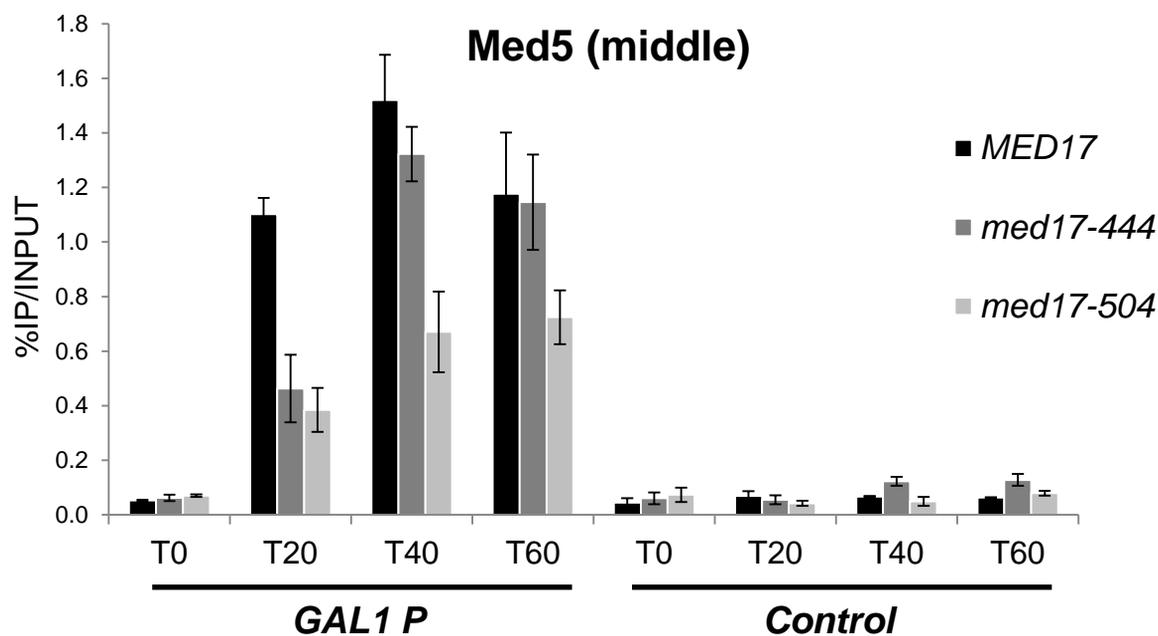
Yeast protein	Module	MW, kDa	Number of unique peptides					Control IP
			IP Med17-WT	IP Med17-98	IP Med17-444	IP Med17-504	IP Med17-670	
MED12	Cdk8	167	5	4	2	5	13	0
MED13	Cdk8	160	3	6	5	3	10	0
MED17	Head	78	10	12	11	13	13	0
MED18	Head	34	5	1	0	1	4	0
MED6	Head	33	5	1	1	1	3	0
MED20	Head	23	6	5	3	6	5	0
MED22	Head	14	2	1	0	1	2	0
MED11	Head	13	1	1	0	1	2	0
MED5	Middle	129	35	34	30	36	38	3
MED1	Middle	64	12	11	7	9	14	0
MED4	Middle	32	8	6	7	4	9	0
MED7	Middle	26	3	4	1	3	4	0
MED9	Middle	17	0	0	0	0	2	0
MED10	Middle	18	2	0	1	4	1	0
MED21	Middle	16	3	0	0	0	2	0
MED14	Tail	123	22	17	17	22	25	0
MED15	Tail	120	10	9	11	12	16	0
MED16	Tail	111	37	39	32	40	44	0
MED3	Tail	43	3	2	2	2	5	0

C

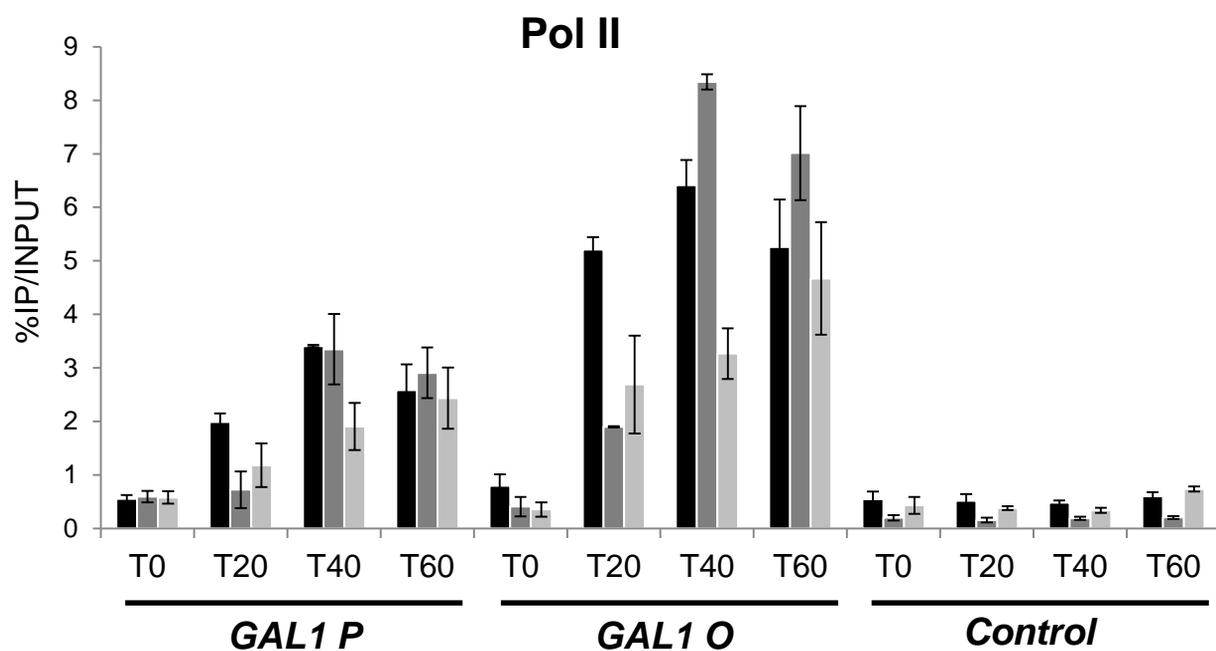




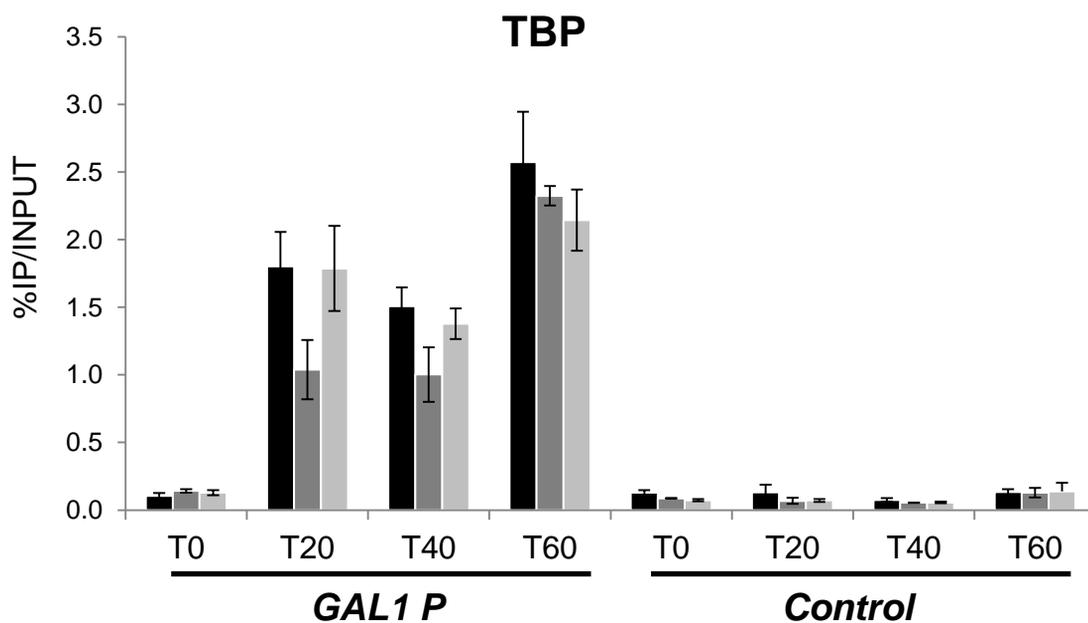
A

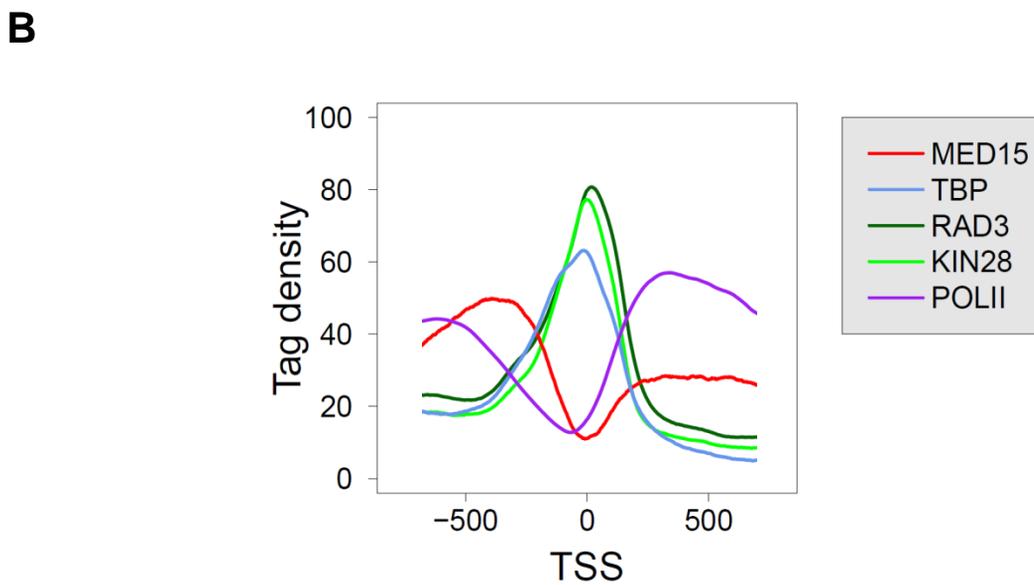
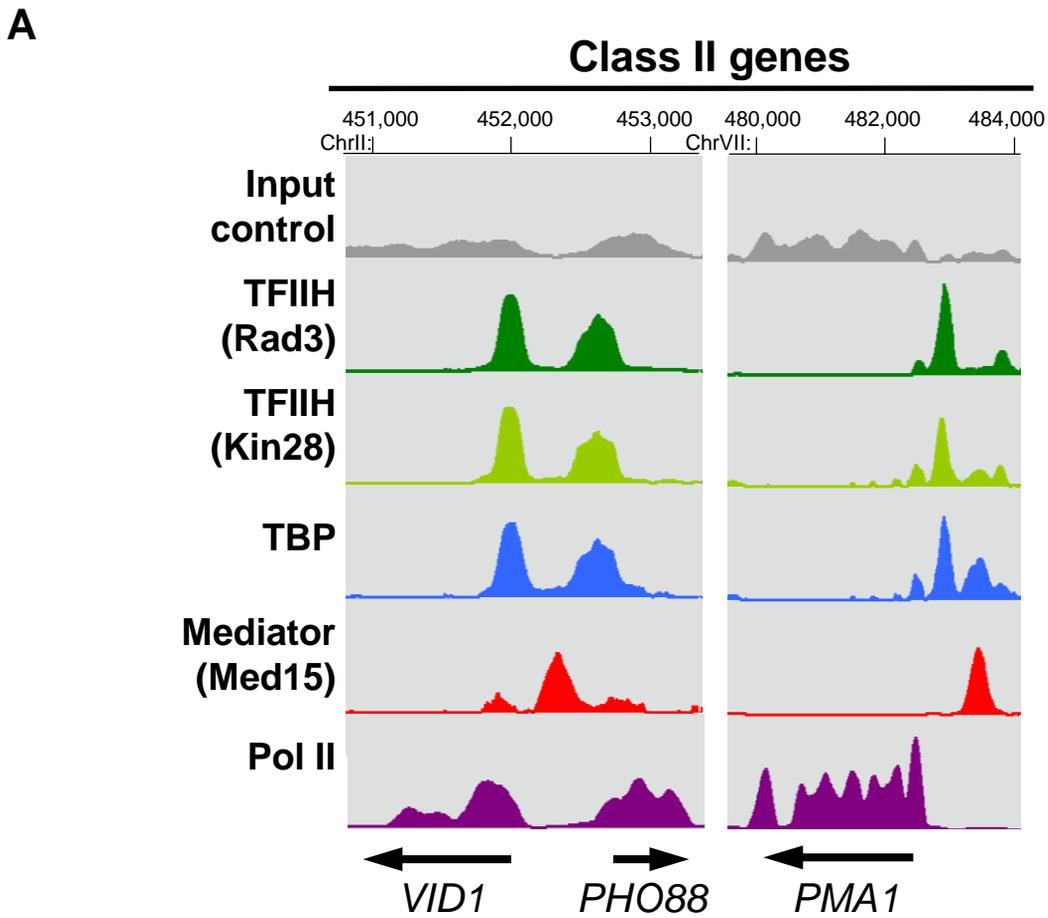


B



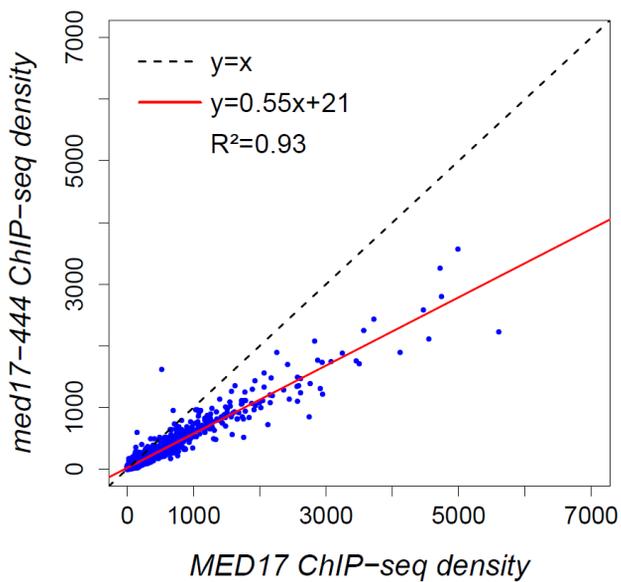
C



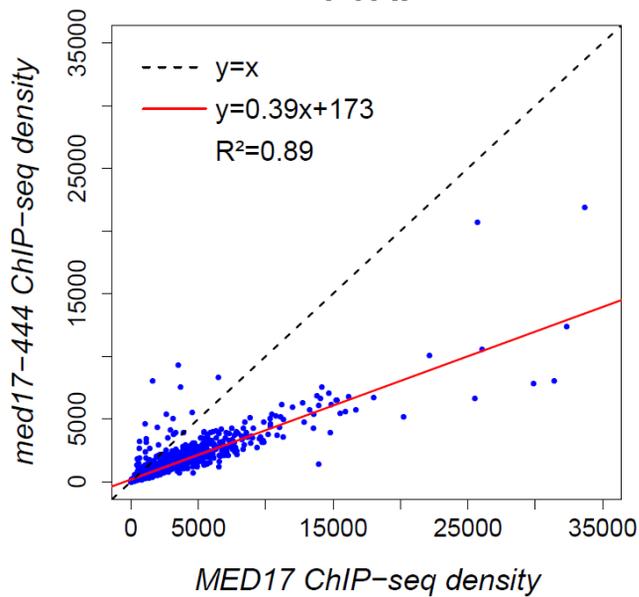


A

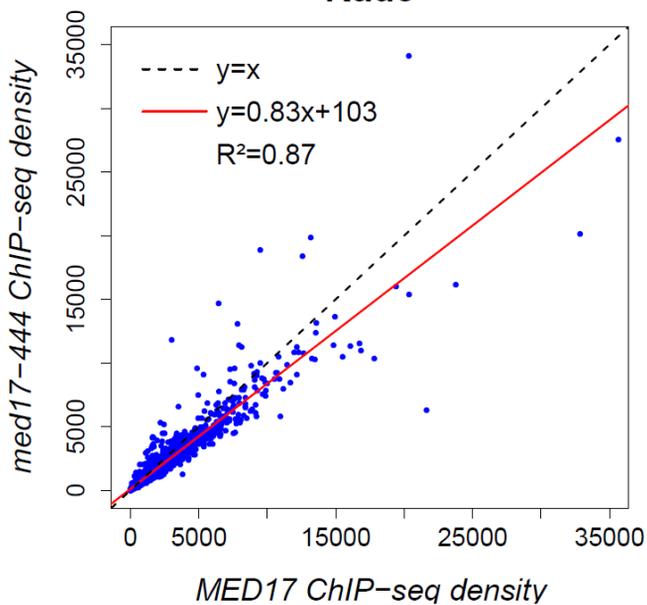
Med15



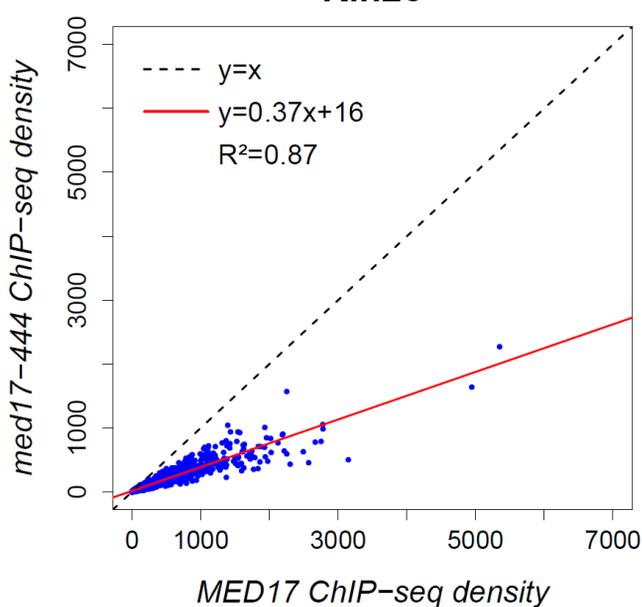
Pol II



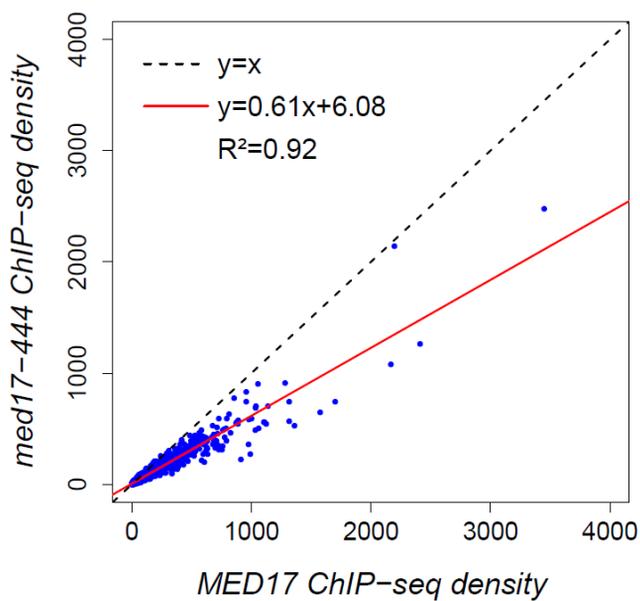
Rad3



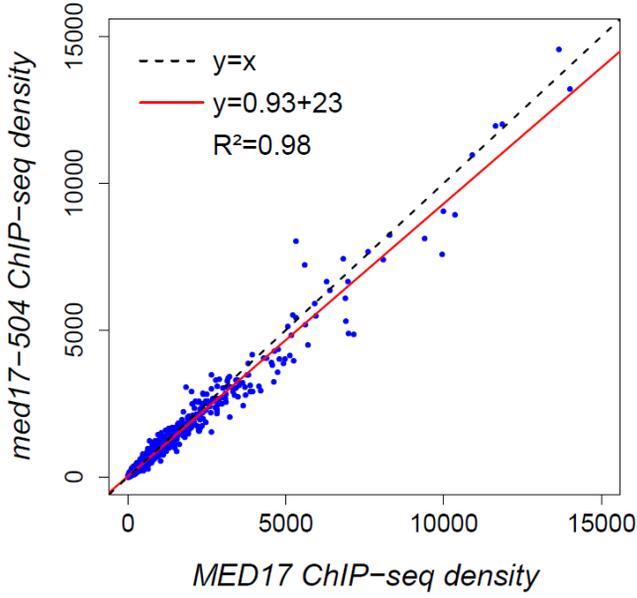
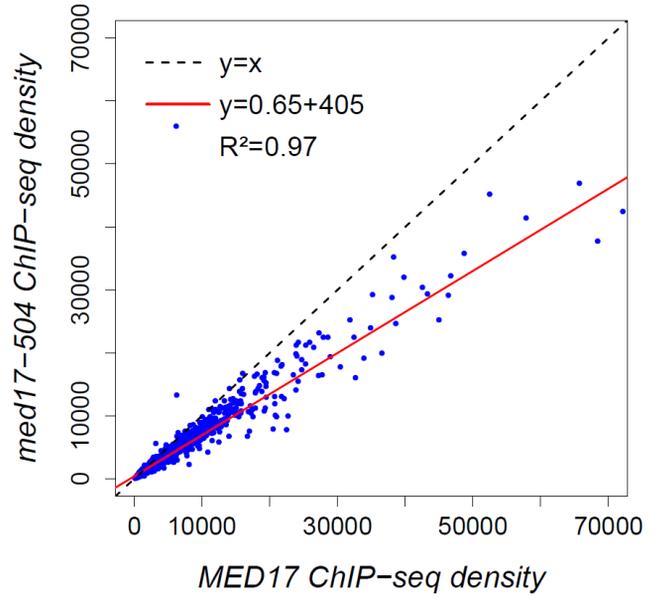
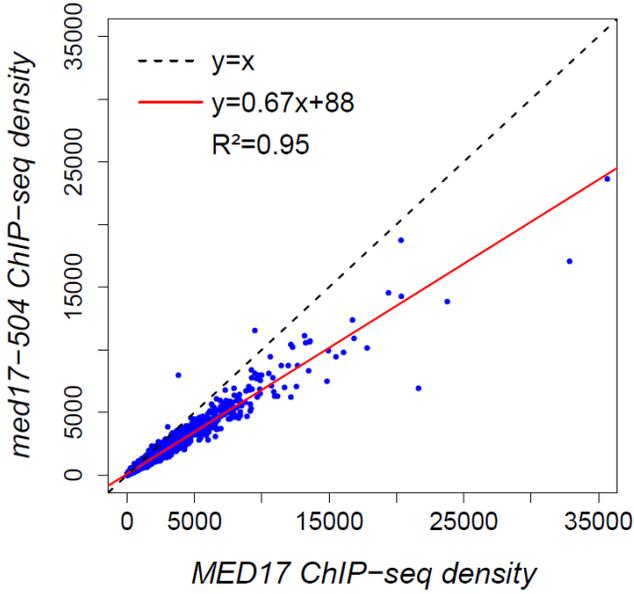
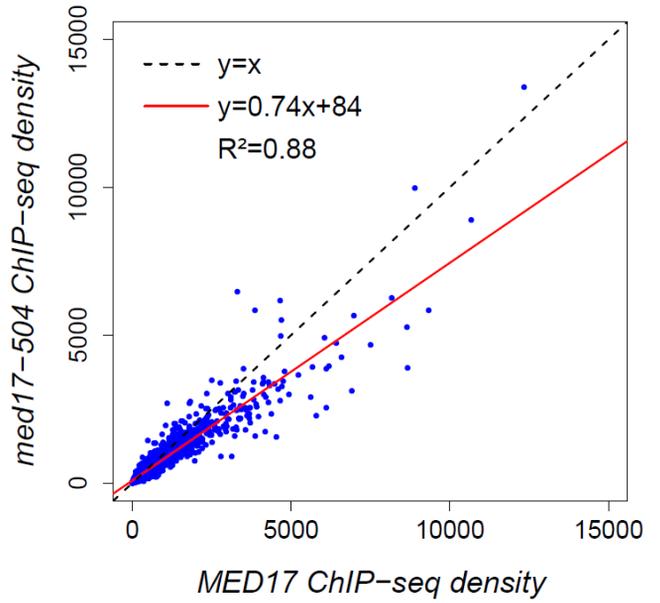
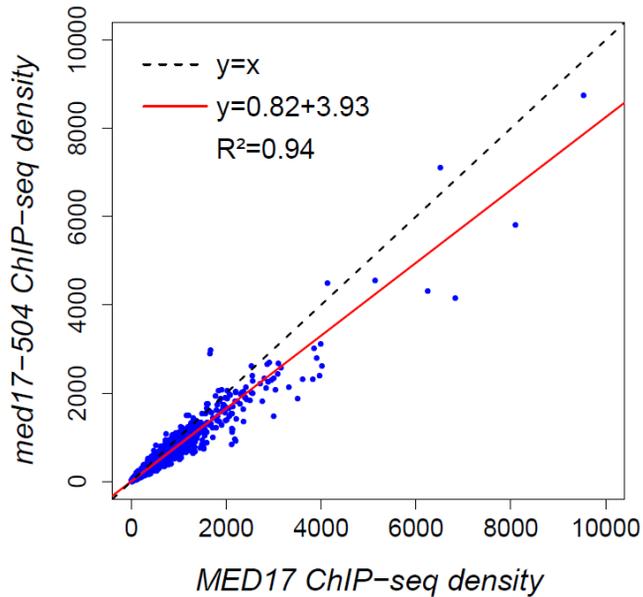
Kin28



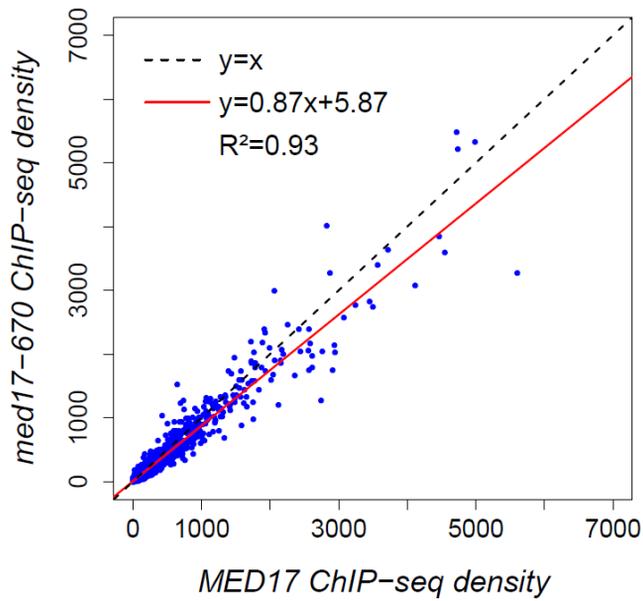
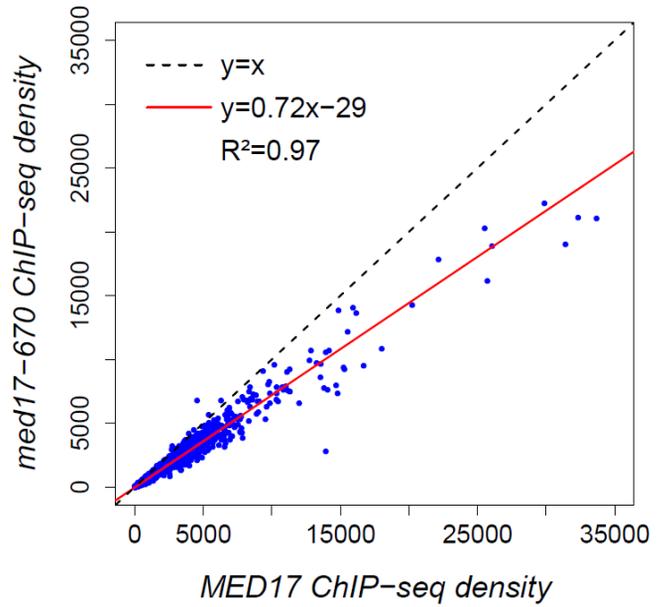
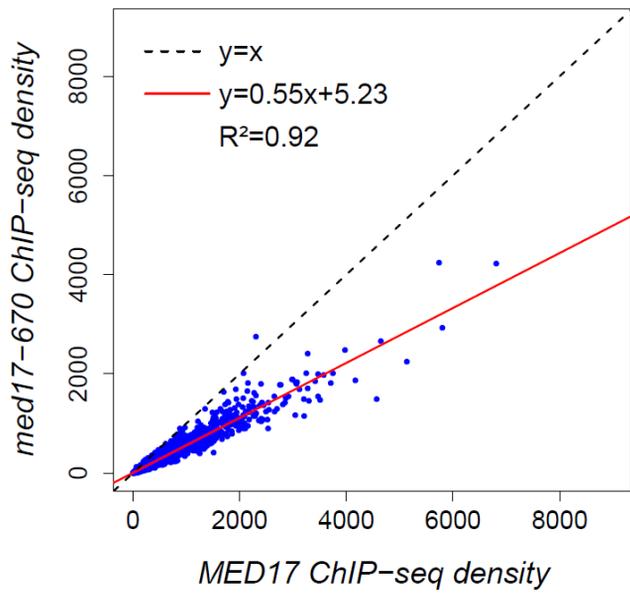
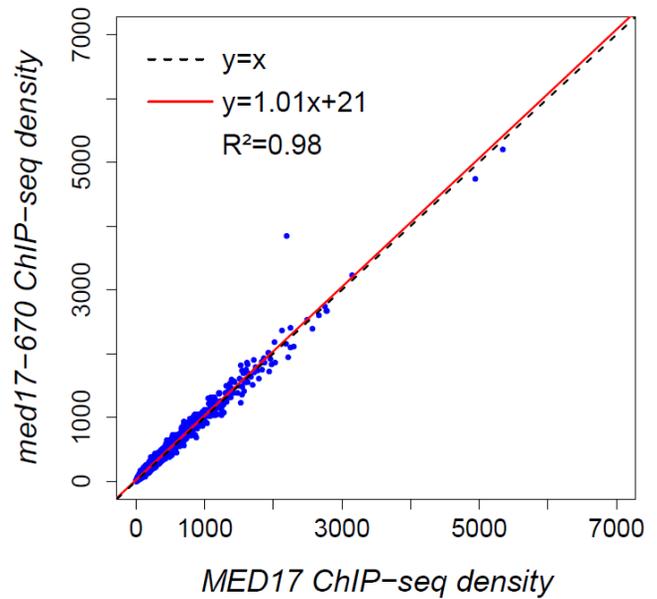
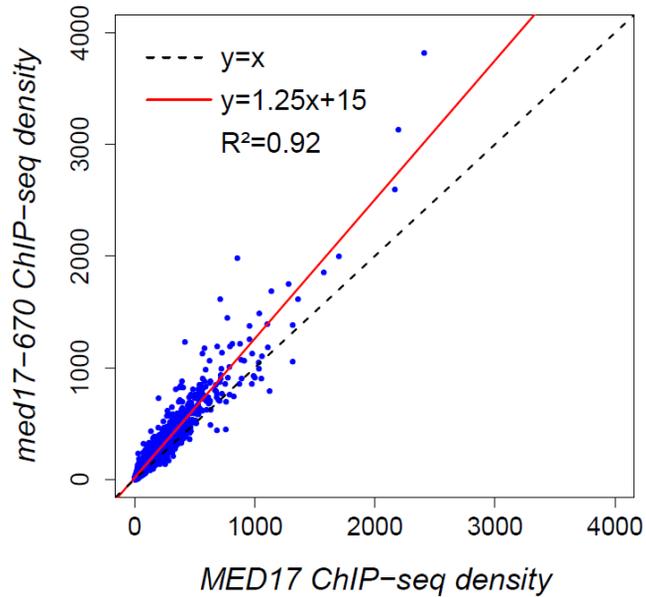
TBP



med17-444

B**Med15****Pol II****Rad3****Kin28****TBP***med17-504*

A

Med15**Pol II****Rad3****Kin28****TBP****med17-670**

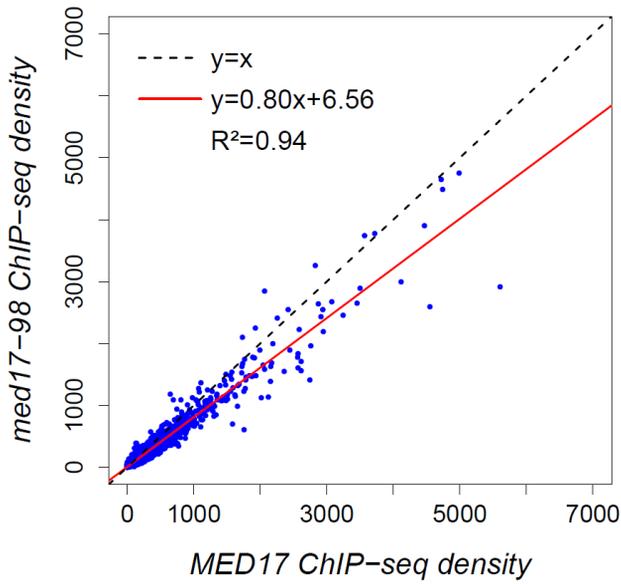
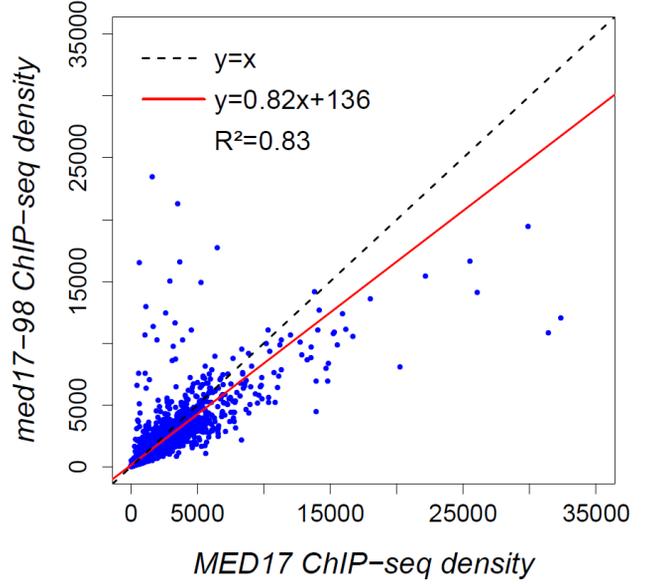
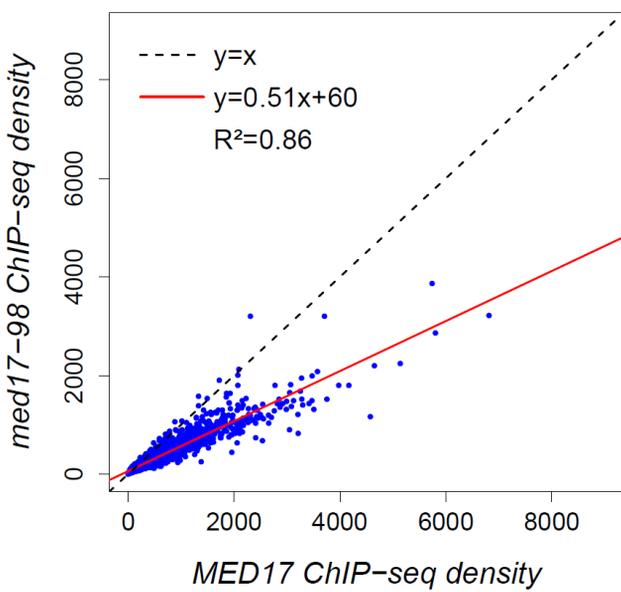
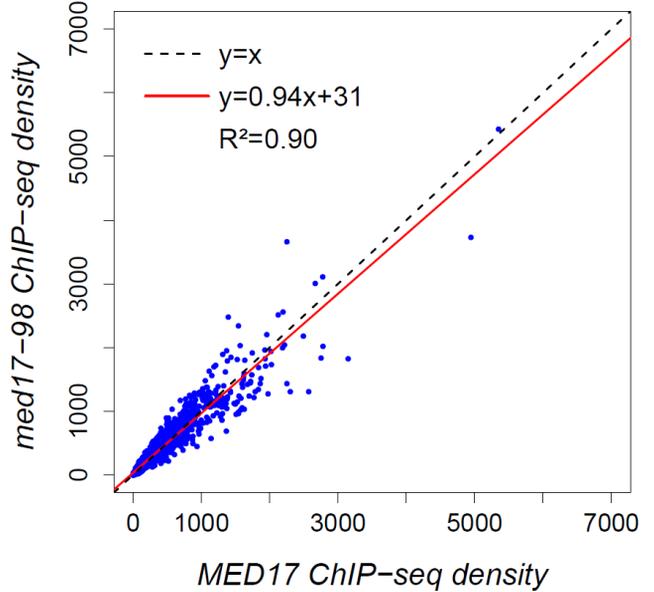
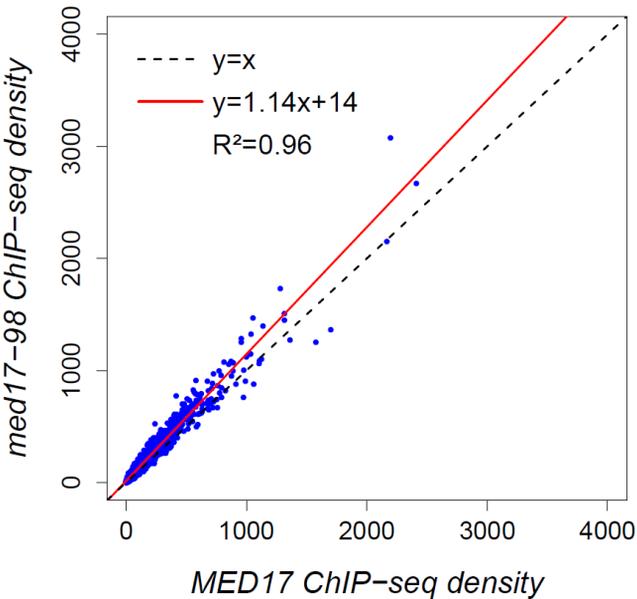
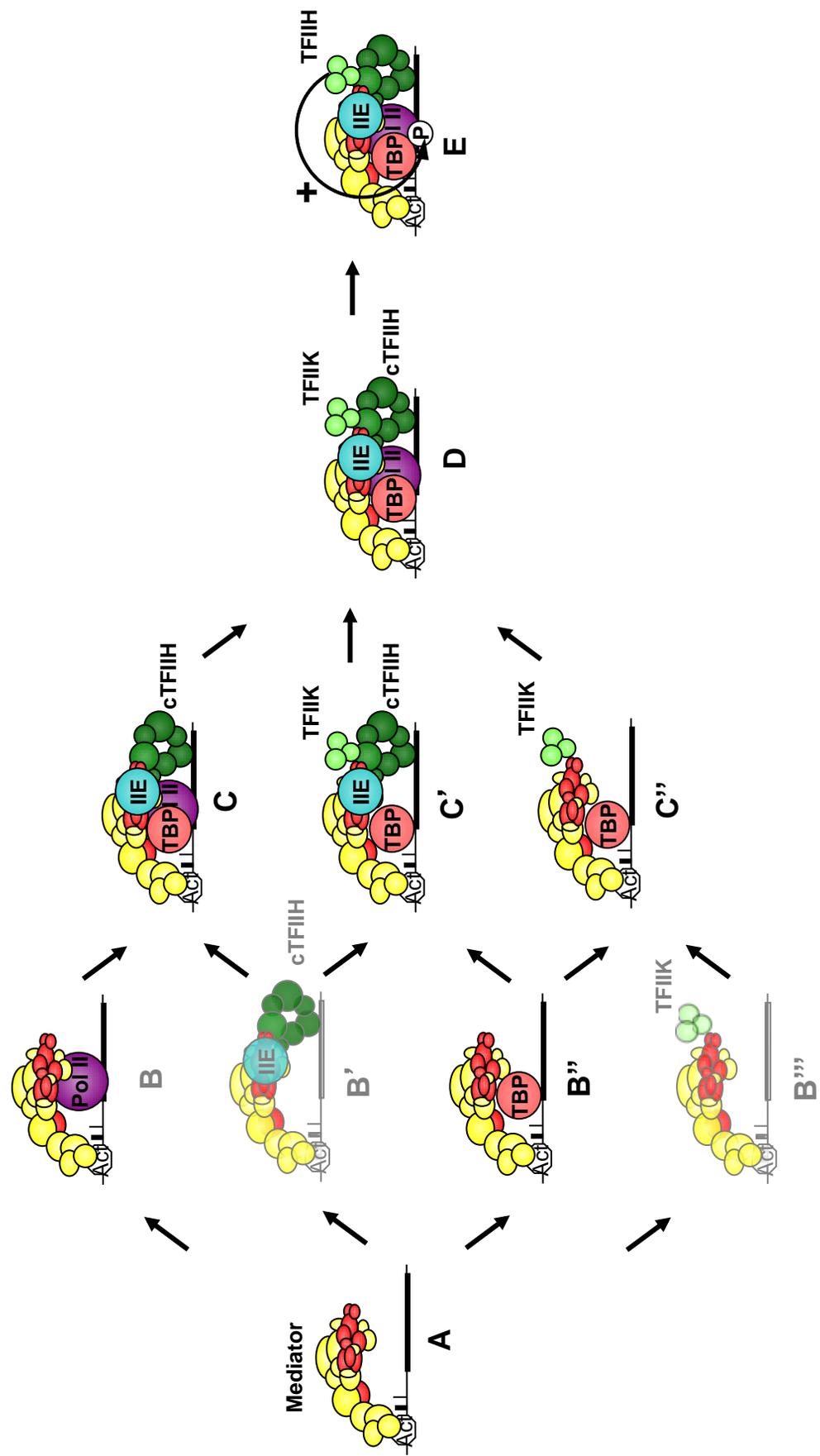
B**Med15****Pol II****Rad3****Kin28****TBP***med17-98*

Figure 7



Supplementary data

Supplementary Materials and Methods

Plasmid constructions and cloning

Plasmids are listed in supplementary Table S4. All cloning experiments were done using the Gateway Invitrogen cloning method. Wild-type *MED17* gene was amplified from YPH499 genomic DNA using oligonucleotides matching the gene sequence initiation codon and following codons for the 5' forward primer and the stop codon and preceding codons for the 3' reverse primer. The oligonucleotides were flanked with *attB1* or *attB2* sequences, respectively. The amplified sequence was cloned into pDONR201 (Invitrogen) using standard BP reaction. The recombinant plasmid was sequence verified. The cloned sequence was then transferred into pVV208 (CEN URA3), pVV204 (CEN TRP1) vector (Van Mullem et al. 2003) or pAG414-GPD-EGFP (Alberti et al. 2007) by the LR reaction.

Point mutations and M504P mutation in Med17 equivalent to human L371P mutation responsible for the infantile cerebral atrophy were obtained using QuikChange Lightning site-directed mutagenesis kit (Agilent), cloned using the Gateway standard method and transferred into pVV204.

Yeast strains

Strains are listed in supplementary Table S3. *MED17* was deleted in YPH499 (*MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1*) complemented with pVV208-*MED17* and replaced by a KanMX6 marker using the standard one-step method (Longtine et al. 1998). *med17* conditional mutants were obtained as previously described (Soutourina et al. 2011).

Med5-HA, Med6-HA, Med15-HA, Rad3-HA and Kin28-HA strains carrying C-terminal HA-tagged version of Med5, Med6 and Med15 Mediator subunits, Rad3 and Kin28 TFIIF

subunits were constructed by inserting 3HA epitopes followed by His3MX6 marker using the standard one-step methods. The HA-TBP strain expressing N-terminal HA-tagged version of TBP was obtained by inserting 3HA epitopes preceded by *LEU2* marker.

Med17-EGFP Rpb3-Myc double-tagged strains were derived from Y4945 (*med17* Δ /pVV208-MED17) strain that was first tagged for Rpb3-Myc by insertion of 13Myc epitopes followed by a HISMX marker using the standard one-step method, then transformed by pAG414-GPD-EGFP-MED17 plasmid. Finally, *MED17 URA3* plasmid was chased on 5-FOA medium.

Unless otherwise stated, yeast strains were grown at 30°C in YPD rich medium containing 2% of glucose. For galactose induction, yeast cells were grown in YP Raffinose rich medium containing 2% of raffinose. Galactose was then added to reach 2% final concentration.

Supplementary Table S1. *med17* mutants and their phenotypes

Name	Mutations	Phenotype ¹
Simple mutants		
<i>med17-96</i>	L674P	N 30 / S 37
<i>med17-249</i>	L558P	S 30 / S 37
<i>med17-370</i> ^{*2}	P370S	N 30 / S 37
<i>med17-444</i> [*]	Q444P	S 30 / L 37
<i>med17-504</i> [*]	M504P	S 30 / L 37
<i>med17-510</i> [*]	L510P	S 30 / L 37
<i>med17-546</i> [*]	L546P	S 30 / S 37
<i>med17-670</i> [*]	V670E	S 30 / S 37
Double mutants		
<i>med17-8</i>	E123G L510P	N 30 / L 37
<i>med17-68</i>	P370S L441P	S 30 / L 37
<i>med17-95</i>	P104T L510P	S 30 / L 37
<i>med17-140</i>	M442L Q444P	S 30 / L 37
<i>med17-241</i>	F295L L546P	N 30 / S 37
<i>med17-326</i>	F388L L546P	N 30 / S 37
<i>med17-327</i>	M442V V670E	S 30 / L 37
Triple mutants		
<i>med17-98</i>	L86Q E186G E448Q	N 30 / S 37
<i>med17-218</i>	I397N Y554F L617P	N 30 / L 37
<i>med17-247</i>	S230P K433I L508W	N 30 / S 37
<i>med17-233</i>	K387I A450T L674P	S 30 / L 37
<i>med17-257</i>	L520S I541K E669G	S 30 / L 37
<i>med17-248</i>	I393T L469Q I647N	N 30 / S 37
<i>med17-299</i>	N76I I467T D568Y	S 30 / S 37
<i>med17-348</i>	D498N I545F V670E	S 30 / L 37
Multiple mutants		
<i>med17-157</i>	R352K C449F K517R N534Y K618Q	N 30 / L 37
<i>med17-158</i>	F159Y S226T K280M K377N E438G V465E	S 30 / L 37
<i>med17-187</i>	N35S K247E L254I K266R N356D E400G L617P	N 30 / S 37
<i>med17-208</i>	L441P N611H A655T E669D	S 30 / L 37
<i>med17-239</i>	M207K N276Y L302S R543C	N 30 / L 37
<i>med17-396</i>	S12G I60V S283G Y348F I476T L609Q	N 30 / L 37
<i>med17-404</i>	I404V I429N E472V K517E	N 30 / L 37

¹ Normal (N), slow growth (S) and lethal (L) phenotypes at 30°C (30) or 37°C (37) are indicated.

² Mutants obtained by site-directed mutagenesis are indicated by asterisk. The four mutants chosen for a detailed analysis are shown in red.

Supplementary Table S2. Mediator integrity in *med17* mutants¹

Mutant	Head					Middle	Tail			
Co-IP	Med6	Med20	Med8-HA			Med5-HA	Med14	Med2-HA		
IP	Med8-HA	Med5-HA	Med8-HA	Med2-HA	Med8-HA	Med5-HA	Med5-HA	Med2-HA	Med8-HA	Med2-HA
<i>med17-8</i>	+	+	+	+	+	+	+	+	+/-	+
<i>med17-68</i>		+/-		+	+	+/-	+	++		+
<i>med17-95</i>	+	-	+	+	+	+	+	+	+	+
<i>med17-96</i>	+	+/-	+/-	+	+	+	+	+	+	+
<i>med17-98</i>	+	+	+	+	+	+	+	+	++	+
<i>med17-140</i>	+/-	+/-	+	+	+/-	+	+	+	+	+
<i>med17-157</i>	+/-	+/-	+	+	+	+	+	+	+	+
<i>med17-158</i>	++	+/-	+	+	+	+/-	+	++	++	+
<i>med17-187</i>	+	+	+	+	+	+	+	++	+	+
<i>med17-208</i>		+		+	+	+	+	+		+
<i>med17-218</i>	+	+/-	+	+	+	+/-	+	+	+	+
<i>med17-233</i>	+/-	+/-	+	+	+	+/-	+/-	+/-	+/-	+/-
<i>med17-239</i>	+/-	+/-	+	+	+/-	+	+	+/-	+	+/-
<i>med17-241</i>	+/-	+/-	+/-	+/-	+/-	+	+	+	+	+
<i>med17-247</i>	+	+/-	+/-	+	+	+	+	+	++	+
<i>med17-248</i>	+	+/-	+	+	+	+/-	+	+	++	+
<i>med17-249</i>	+	+/-	+	+	+	+	+	+/-	+	+
<i>med17-257</i>	-	+/-	+/-	+	+/-	+	+	+/-	+/-	+
<i>med17-299</i>	+/-	+	+/-	+	+/-	+	+	+/-	+	+
<i>med17-326</i>	+	+/-	+	+	+	+	+	+	++	+
<i>med17-327</i>	-	+/-	+/-	+	+/-	+	+	+/-	+	+
<i>med17-348</i>	-	-	+/-	+/-	+/-	+	+	+/-	+	+
<i>med17-370</i>	++	+	+	+	++	+	+	+	+	+
<i>med17-396</i>	+	+/-	+	+	+	+	+	+	+	+
<i>med17-404</i>	+/-	+/-	+	+	+	+/-	+	+	+	+
<i>med17-444</i>	+	+/-	+	+	+	+	+	+	+	+
<i>med17-510</i>	+	+	+	+	+	+	+	+	+	+
<i>med17-546</i>	+	+/-	+	+	+	+	+	+	+	+
<i>med17-670</i>	+	+/-	+	+	++	+	+	+/-	+	+

¹ Yeast strains carrying a HA-tag on Med8, Med5 or Med2 Mediator subunits were used for CoIP experiments. Med6, Med20, Med8-HA, Med5-HA, Med14 and Med2-HA subunits were detected by western blotting with corresponding antibodies. Levels of western blotting signals compared to the control IP with a wild-type strain are indicated as follows: similar (+), slightly decreased (+/-), decreased (-) and slightly increased (++) levels. It should be noted that lower IP signals could explain lower CoIP signals.

Supplementary Table S3. Yeast strains

Name	Tagged or Mutant protein	Genotype	Reference
YPH499	WT	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1</i>	(Sikorski and Hieter 1989)
Y4945	<i>MED17</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 ΔMED17::KanMX6 / MED17 CEN URA3</i>	(Soutourina et al. 2011)
Y5084	<i>MED17</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // MED17 CEN TRP1</i>	(Soutourina et al. 2011)
Y 5304	<i>med17-8</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-8 CEN TRP1</i>	This work
Y 5305	<i>med17-68</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-68 CEN TRP1</i>	This work
Y 5306	<i>med17-95</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-95 CEN TRP1</i>	This work
Y 5307	<i>med17-96</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-96 CEN TRP1</i>	This work
Y5308	<i>med17-98</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5309	<i>med17-140</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-140 CEN TRP1</i>	(Eyboulet et al. 2013)
Y5310	<i>med17-157</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-157 CEN TRP1</i>	This work
Y5311	<i>med17-158</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-158 CEN TRP1</i>	This work
Y5312	<i>med17-187</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-187 CEN TRP1</i>	This work
Y5313	<i>med17-208</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-208 CEN TRP1</i>	This work
Y5314	<i>med17-218</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-218 CEN TRP1</i>	This work
Y5315	<i>med17-233</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-233 CEN TRP1</i>	(Eyboulet et al. 2013)
Y5316	<i>med17-239</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-239 CEN TRP1</i>	This work
Y5317	<i>med17-241</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-241 CEN TRP1</i>	This work
Y5318	<i>med17-248</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-248 CEN TRP1</i>	This work
Y5319	<i>med17-247</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-247 CEN TRP1</i>	This work
Y5320	<i>med17-249</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-249 CEN TRP1</i>	This work
Y5321	<i>med17-257</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-257 CEN TRP1</i>	(Eyboulet et al. 2013)
Y5322	<i>med17-299</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-299 CEN TRP1</i>	This work
Y5323	<i>med17-326</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5327	<i>med17-327</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-327 CEN TRP1</i>	This work
Y5328	<i>med17-348</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-348 CEN TRP1</i>	This work
Y5324	<i>med17-370</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-370 CEN TRP1</i>	This work
Y5329	<i>med17-396</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-396 CEN TRP1</i>	This work

Y5331	<i>med17-404</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-404 CEN TRP1</i>	This work
Y5325	<i>med17-444</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y6106	<i>med17-504</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5326	<i>med17-510</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-510 CEN TRP1</i>	This work
Y5327	<i>med17-546</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-546 CEN TRP1</i>	(Eyboulet et al. 2013)
Y5328	<i>med17-670</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y5485	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 / MED17 CEN TRP1 MED5::3HA::HIS3</i>	This work
Y5773	Med6-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 / MED17 CEN TRP1 MED6::3HA::HIS3</i>	This work
Y5540	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 / MED17 CEN TRP1 MED15::3HA::HIS3</i>	(Eyboulet et al. 2013)
Y5551	HA-TBP	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5557	Rad3-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5545	Kin28-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5637	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5643	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5538	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5639	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5640	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y5641	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6063	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5491	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5489	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-258 CEN TRP1</i>	This work
Y5520	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5494	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5496	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y5521	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6060	Med5-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 med17::KanMX6 // med17-504 CEN TRP1</i>	This work

Y5774	Med6-HA <i>med17-98</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5779	Med6-HA <i>med17-257</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5775	Med6-HA <i>med17-326</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5776	Med6-HA <i>med17-444</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5777	Med6-HA <i>med17-546</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y7778	Med6-HA <i>med17-670</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6061	Med6-HA <i>med17-504</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED6::3HA::HIS3 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5558	Rad3-HA <i>med17-98</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5602	Rad3-HA <i>med17-257</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5559	Rad3-HA <i>med17-326</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5560	Rad3-HA <i>med17-444</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5561	Rad3-HA <i>med17-546</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y5562	Rad3-HA <i>med17-670</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6025	Rad3-HA <i>med17-504</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5546	Kin28-HA <i>med17-98</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5600	Kin28-HA <i>med17-257</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5547	Kin28-HA <i>med17-326</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5548	Kin28-HA <i>med17-444</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5549	Kin28-HA <i>med17-546</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y5550	Kin28-HA <i>med17-670</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6066	Kin28-HA <i>med17-504</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5552	HA-TBP <i>med17-98</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-98 CEN TRP1</i>	This work
Y5601	HA-TBP <i>med17-257</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5553	HA-TBP <i>med17-326</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-326 CEN TRP1</i>	This work
Y5554	HA-TBP <i>med17-444</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-444 CEN TRP1</i>	This work
Y5555	HA-TBP <i>med17-546</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-546 CEN TRP1</i>	This work
Y5556	HA-TBP <i>med17-670</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-670 CEN TRP1</i>	This work
Y6064	HA-TBP <i>med17-504</i>	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // med17-504 CEN TRP1</i>	This work
Y5563	Rpb3-Myc MED17	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX MED17 CEN URA3</i>	This work

Y5651	Rpb3-Myc MED17- GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-MED17 CEN TRP1</i>	This work
Y5739	Rpb3-Myc med17-98- GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-98 CEN TRP1</i>	This work
Y5652	Rpb3-Myc med17- 257-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-257 CEN TRP1</i>	This work
Y5740	Rpb3-Myc med17- 326-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-326 CEN TRP1</i>	This work
Y5741	Rpb3-Myc med17- 444-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-444 CEN TRP1</i>	This work
Y5742	Rpb3-Myc med17- 546-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-546 CEN TRP1</i>	This work
Y5743	Rpb3-Myc med17- 670-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-670 CEN TRP1</i>	This work
Y6341	Rpb3-Myc med17- 504-GFP	<i>MATa ura3-52 his3- Δ200 ade2-101uaa trp1- Δ63 lys2-801uag leu2-Δ1 Δmed17::KanMX6 RPB3::13Myc::HIS3MX EGFP-med17-504 CEN TRP1</i>	This work

Supplementary Table S4. Plasmids

Name	Genotype	Reference
pVV208-MED17	<i>Amp CEN URA3 MED17</i>	(Soutourina et al. 2011)
pVV204-MED17	<i>Amp CEN TRP1 MED17</i>	(Soutourina et al. 2011)
pVV204-med17-8	<i>Amp CEN TRP1 med17-E123G, L510P</i>	This work
pVV204-med17-68	<i>Amp CEN TRP1 med17-P370S, L441P</i>	(Soutourina et al. 2011)
pVV204-med17-95	<i>Amp CEN TRP1 med17-P104T, L510P</i>	This work
pVV204-med17-96	<i>Amp CEN TRP1 med17-L674P</i>	This work
pVV204-med17-98	<i>Amp CEN TRP1 med17-L86Q, E186G, E448Q</i>	This work
pVV204-med17-140	<i>Amp CEN TRP1 med17-M442L, Q444P</i>	(Soutourina et al. 2011)
pVV204-med17-157	<i>Amp CEN TRP1 med17-R352K, C449F, K517R, N534Y, K618Q</i>	This work
pVV204-med17-158	<i>Amp CEN TRP1 med17-F159Y, S226T, K280M, K377N, E438G, V465E</i>	(Soutourina et al. 2011)
pVV204-med17-187	<i>Amp CEN TRP1 med17-N35S, K247E, L254I, K266R, N356D, E400G, L617P</i>	This work
pVV204-med17-208	<i>Amp CEN TRP1 med17-L441P, N611H, A655T, E669D</i>	(Soutourina et al. 2011)
pVV204-med17-218	<i>Amp CEN TRP1 med17-I397N, Y554F, L617P</i>	This work
pVV204-med17-233	<i>Amp CEN TRP1 med17-K387I, A450T, L674P</i>	(Soutourina et al. 2011)
pVV204-med17-239	<i>Amp CEN TRP1 med17-M207K, N276Y, L302S, R543C</i>	This work
pVV204-med17-241	<i>Amp CEN TRP1 med17-F295L, L546P</i>	This work
pVV204-med17-248	<i>Amp CEN TRP1 med17-I393T, L469Q, I647N</i>	This work
pVV204-med17-247	<i>Amp CEN TRP1 med17-S230P, K433I, L508W</i>	This work
pVV204-med17-249	<i>Amp CEN TRP1 med17-L558P</i>	This work
pVV204-med17-257	<i>Amp CEN TRP1 med17-L520S, I541K, E669G</i>	(Soutourina et al. 2011)
pVV204-med17-299	<i>Amp CEN TRP1 med17-N76I, I467T, D568Y</i>	This work
pVV204-med17-326	<i>Amp CEN TRP1 med17-F388L, L546P</i>	This work
pVV204-med17-327	<i>Amp CEN TRP1 med17-M442V, V670E</i>	(Soutourina et al. 2011)
pVV204-med17-348	<i>Amp CEN TRP1 med17- D498N, I545F, V670E</i>	This work
pVV204-med17-370	<i>Amp CEN TRP1 med17-P370S</i>	This work
pVV204-med17-396	<i>Amp CEN TRP1 med17-S12G, I60V, S283G, Y348F, I476T, L609Q</i>	This work
pVV204-med17-404	<i>Amp CEN TRP1 med17- I404V I429N E472V K517E</i>	This work
pVV204-med17-444	<i>Amp CEN TRP1 med17-Q444P</i>	This work
pVV204-med17-504	<i>Amp CEN TRP1 med17-M504P</i>	This work
pVV204-med17-510	<i>Amp CEN TRP1 med17-L510P</i>	This work
pVV204-med17-546	<i>Amp CEN TRP1 med17-L546P</i>	(Soutourina et al. 2011)
pVV204-med17-670	<i>Amp CEN TRP1 med17-V670E</i>	This work
pAG414GPD-EGFP-MED17	<i>AmpR MED17 CEN TRP1</i>	This work
pAG414GPD-EGFP-med17-98	<i>AmpR med17-98 CEN TRP1</i>	This work
pAG414GPD-EGFP-med17-444	<i>AmpR med17-444 CEN TRP1</i>	This work
pAG414GPD-EGFP-med17-670	<i>AmpR med17-670 CEN TRP1</i>	This work
pAG414GPD-EGFP-med17-504	<i>AmpR med17-504 CEN TRP1</i>	This work

Supplementary Table S5. Oligonucleotides

Name	Forward	Reverse
ATP1-O	TCTTGCAGTCGGTGATGGTA	TATCGGAACCGAAAAGAACG
ALD6-P	AGCCGACAAAAGAAAAACGA	CGTTCCCAAGAGGAGATCAA
ALD6-O	CCCATTTGGTCTTTGACGAT	GCCTTGAAAGCAGCCAATAG
SRM1-Pr	GGCAAGGCAGGATGAATAAA	ATTGGTGGCGACTGTTCTTT
PHO84-O	TTGGGTTGGGTACCGTTCTA	TTCGCCATCTTGTTCTTG TG
QCR6-O	GATGACGATAACGAGCAGCA	TACCCTCCTCCGTGTTCTTG
PIL1-O	GATTCCTTTGGGAAGGGTGGT	ATCGTCATCGTTTTCCAACC
MGR1-P	TGCATGGCATCGTCTTTTAC	GTATGGTCGGTGGAATTGCT
RPA34-Pr	CATCAGGGCCAATCAATTCT	TTTCGACTTCAGCATTGCAC
PSA1-Pr	CTGCACCCGATCCTTCTTAC	TTCTGTTTTCGCGTTCTTCA
ALD6-P	ATGCGATATAGCACCGACCA	CTTGACCTCGAACGGTGTTT
ACT1-P	TCCACGTCCTCTTGATAAAA	GGTTTGAGTAGAAAGGGGAAGG
APA1-P	GAGCGGCCTGAAATACTGTC	AACGTTCCGAAACAGGACAC
PIM1-P	GCCACCGGAAGGAAATAAGT	GTTCAAGCACCTCGAAAACC
STI1-P	CCAAAAGTCTGCTCCCAAAT	TGCAGCGTTACCTTGTTGTT
YIP5-P	CAACGTCTACGCTCAAGGT	CCCTCGAGATCATCGTCAAT
PSA1-P1	CTGCACCCGATCCTTCTTAC	TTCTGTTTTCGCGTTCTTCA
PSA1-P2	GGAGCCACCACGTTTATTTT	GTTTACAAGAGCGCCACGAG
PRB1-P	CCGGGTTAGCAGAGTAGCAG	CTCTTTATCGCGGGTGTGTT
ARG3-P	GTGAATTTTCGAGGGTCACG	AAAATTTGTCCGGGCATTTT
FLC1-P	GCGGAGAGGAGGAGAGAGAC	CGTAGGAAACGCATGTGGTA
HSP42-P	GGGAGGCCTCTGTGAAGTTA	GCCTGAACGTGTCCCTATGT
HSP42-O	CACCTGGTAACGCAAGAACA	CATCGTTCGCTCACCTACAA
HSP150-P1	GGAACACTTGAAGTCTAACGACA	AAGCAGAGGCAACCAAAGTC
HSP150-P2	ATTATCCTGGCCGCTAATCC	GTGGCAAAGCAGTGAGATTG
HSP150-O	CCGTTCAACCAATCTCCACT	GGCAGCGGTACTCTGGTAG
HSP104-P1	CCCATCTCAAAGAACTGCAA	GGAACAAGTGACAAAGGAACG
ADH1-Prom1	ATAGGCGCATGCAACTTCTT	CATCAGCTCTGGAACAACGA
ADH1-Prom2	TTCTTCATTACGCACACT	AGGGAACGAGAACAATGACG
ADH1-ORF	GGGTATTGACGGTGGTGAAG	AAACGTTGATGACACCGTGA
PYK1-Prom1	CGCACCGTCACAAAGTGTT	TGGGAAGGAAAGGAAATCAC
PYK1-Prom2	CCTTTCCTTCCCATATGATGC	ACTTTGAAAGGGGACCATGA
PYK1-ORF	TGCTTTGAGAAAGGCTGGTT	TCTGATTTCTGGACCCTTGG
PYK1-ATG	GGTCCCCTTTCAAAGTTATTCTCT	CTTGAGAAGAACCCTCCATCATTG
PMA1-ORF	GGTTTTGGTCATTGCCACTT	ACGGCCATAGTGGTGGTAAC
PMA1-Prom1	AACAAACCCGGTCTCGAAG	GAAGTGCCGCATTAGGAAAT
PMA1-Prom2	GATGGTGGGTACCGCTTATG	TTGGTGTTATAGGAAAGAAAGAGA
GAL1-Prom	ACGCTTAACTGCTCATTGCT	TGTTCCGGAGCAGTGC GGCGC
GAL1-ORF1	AAAGAACTTGCACCGGAAA	GGCCCATATTCGCTTAAACA
GAL1-ORF2	ACATTTCCACACCCTGGAAC	TTCTTCGCGAGAACAATTCA
rDNA-25S	GGTTATATGCCGCCCGTCTTGA	CCCAACAGCTATGCTCTTACTC
rDNA-18S	GGTCTGTGATGCCCTTAGACG	AGTTTCACAAGATTACCAAGACCTC
SCR1	GTCCTGGGCAGAGCTGTCT	AAGGTGGAGCCCCTAAGGA
RPN2	GCGGATACAGGCACATTGGATAC	TGTTGCTACCTTCTCTACCTCCTTAC
TAF10	ATATTCCAGGATCAGGTCTTCCG	GTAGTCTTCTCATTCTGTTGATGTTG
PMA1-ORF2	TCTCAAAGCCCGTTAAATG	CCGTTTATAGCACCGAAGTT

Supplementary Table S6. Total number of mapped reads for ChIP-seq experiments

Sample	Million of mapped reads
INPUT-MED17	2.9
INPUT-med17-98	3.0
INPUT-med17-444	4.0
INPUT-med17-670	3.5
POL2-MED17	7.9
POL2- med17-98	8.7
POL2- med17-444	10.4
POL2- med17-670	7.8
TBP-WT	2.8
TBP- med17-98	2.8
TBP- med17-444	3.2
TBP- med17-670	3.6
RAD3-WT (core TFIIH)	5.0
RAD3- med17-98	3.4
RAD3- med17-444	11.3
RAD3- med17-670	3.9
KIN28-WT (TFIIH kinase module)	2.9
KIN28- med17-98	3.0
KIN28- med17-444	4.0
KIN28- med17-670	3.5
MED15-WT (Mediator)	3.5
MED15- med17-98	9.2
MED15- med17-444	3.0
MED15- med17-670	3.7
INPUT- med17-504	15.2
POL2-WT	18.7
POL2- med17-504	16.2
TBP-WT	9.9
TBP- med17-504	11.7
RAD3-WT	10.6
RAD3- med17-504	8.0
KIN28-WT	5.0
KIN28- med17-504	3.2
MED15-WT	6.3
MED15- med17-504	3.5
NT-WT	1.7

Supplementary figures legends

Figure S1. Growth phenotypes of *med17* mutants.

Cultures of wild-type (WT) and mutant *med17* yeast strains were serially diluted, spotted on YPD agar plates and incubated for 3 days at permissive (30°C) or non-permissive (37°C) temperature. *med17* mutants selected for a detailed analysis are indicated in red.

Figure S2. Schematic representation of the localization of selected Med17 mutations.

(A) Localization of Med17 mutations is indicated for selected mutants: *med17-98* (L86Q, E186G, E448Q) in blue, *med17-444* (Q444P) in orange, *med17-504* (M504P) in green, *med17-670* (V670E) in red.

The grey boxes indicate the positions of the conserved domains (Bourbon 2008).

(B) Location of sites of *med17* (red spheres) mutations is indicated on the model of *S. cerevisiae* Mediator head module from Larivière et al. (Lariviere et al. 2012) for *med17-98*, -444, -504 and -670 when the corresponding residues are present on the model. The residues L86 and E186 mutated in *med17-98* are absent in this model.

Figure S3. Mediator integrity in *med17* mutants.

Wild-type (WT) and mutant strains were grown to exponential phase at 30°C. Mediator was immunoprecipitated (IP) through Med5-HA from crude extracts (Input) of wild-type and mutant strains using magnetic beads coupled to anti-HA antibodies. *MED17* strain carrying non-tagged Mediator subunit was used as a negative control. Coimmunoprecipitated Mediator subunits were detected by western blotting using anti-Med14 and -Med17 antibodies.

Figure S4. *med17* mutations influence mRNA abundance after a shift at 37°C.

Yeast strains were grown at 30°C in YPD medium and then shifted for the indicated time at 37°C. mRNA levels were determined by quantitative RT-PCR. qPCR results were normalized to the WT signal. *25S*, *18S rRNA* and *SCR1* genes were used as internal controls for normalization.

Figure S5. Frequency of transcription factor binding sites (TFBS) around TSS.

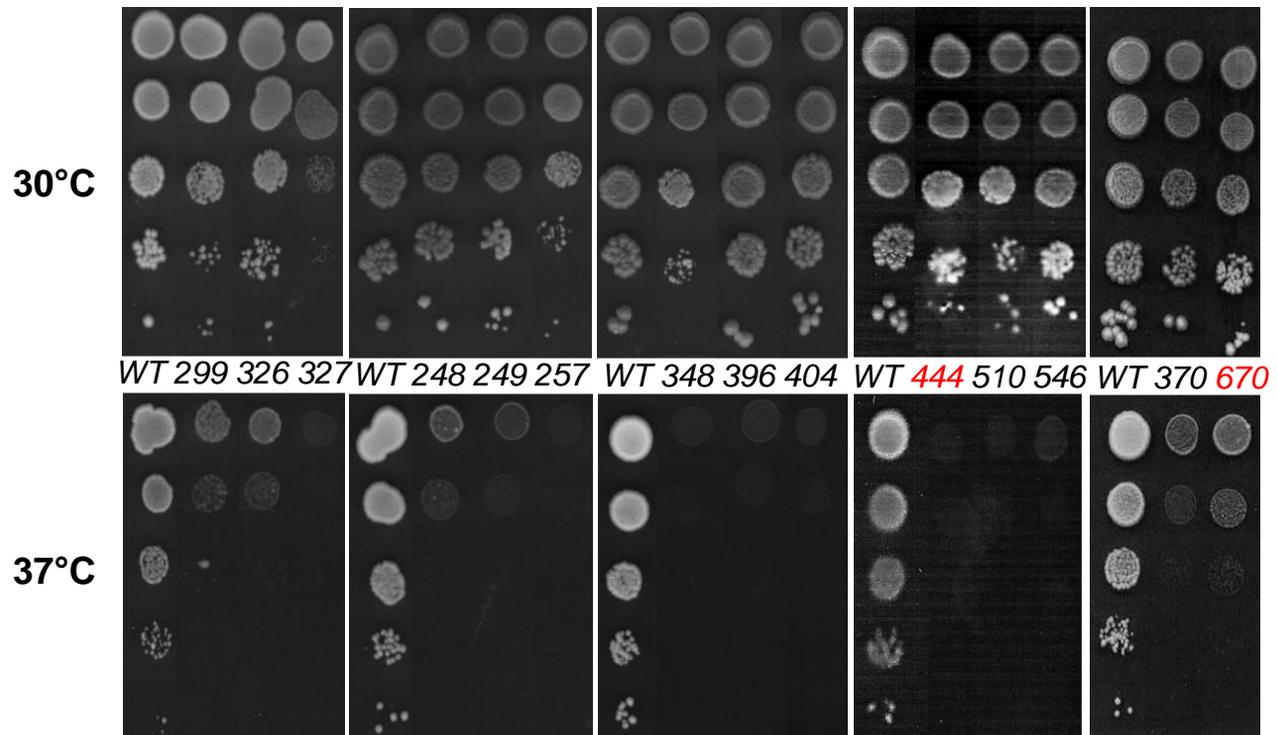
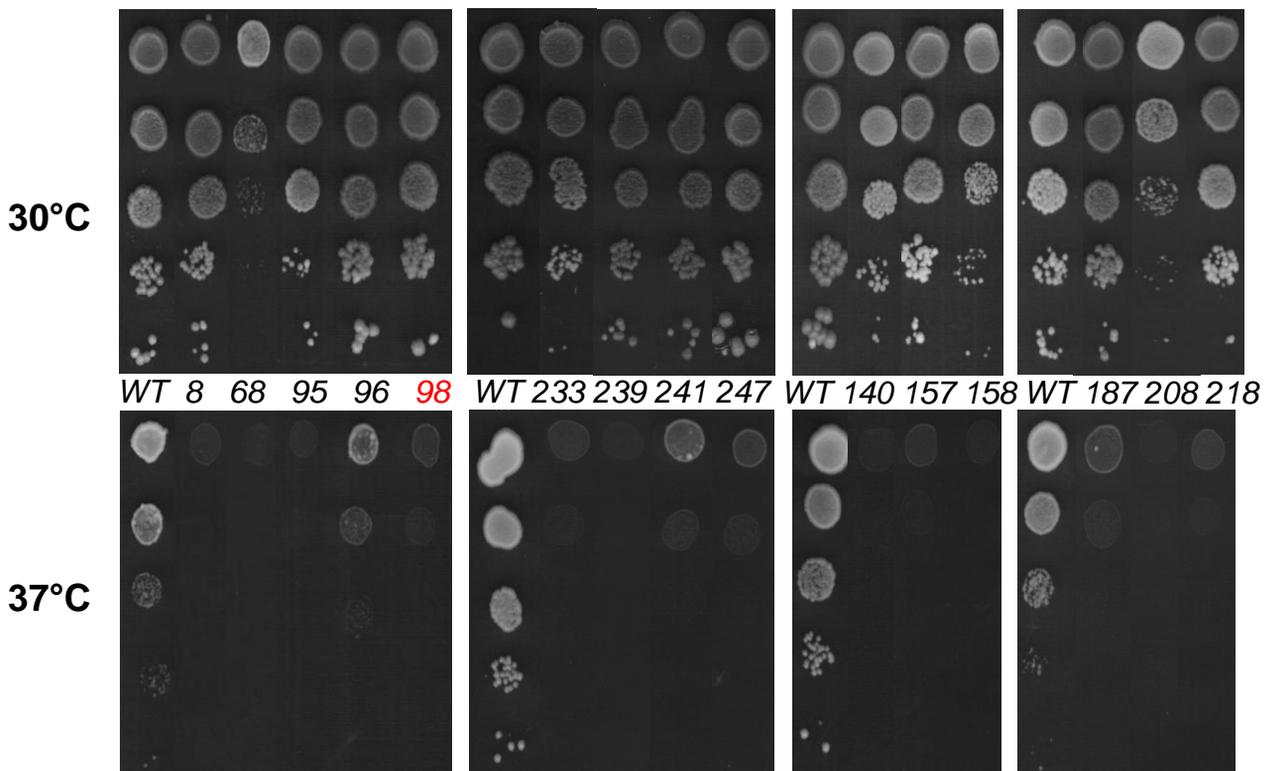
The frequency of TFBS was computed in a 3000-bp window centered on the TSS with 1-bp bin. Mediator, Pol II, TBP, Rad3 and Kin28 tag density was determined around TSS and mean tag density for each nucleotide position was then calculated and plotted over the window.

Figure S6. Conservation of Med17 Mediator subunit from yeast *S. cerevisiae* to human.

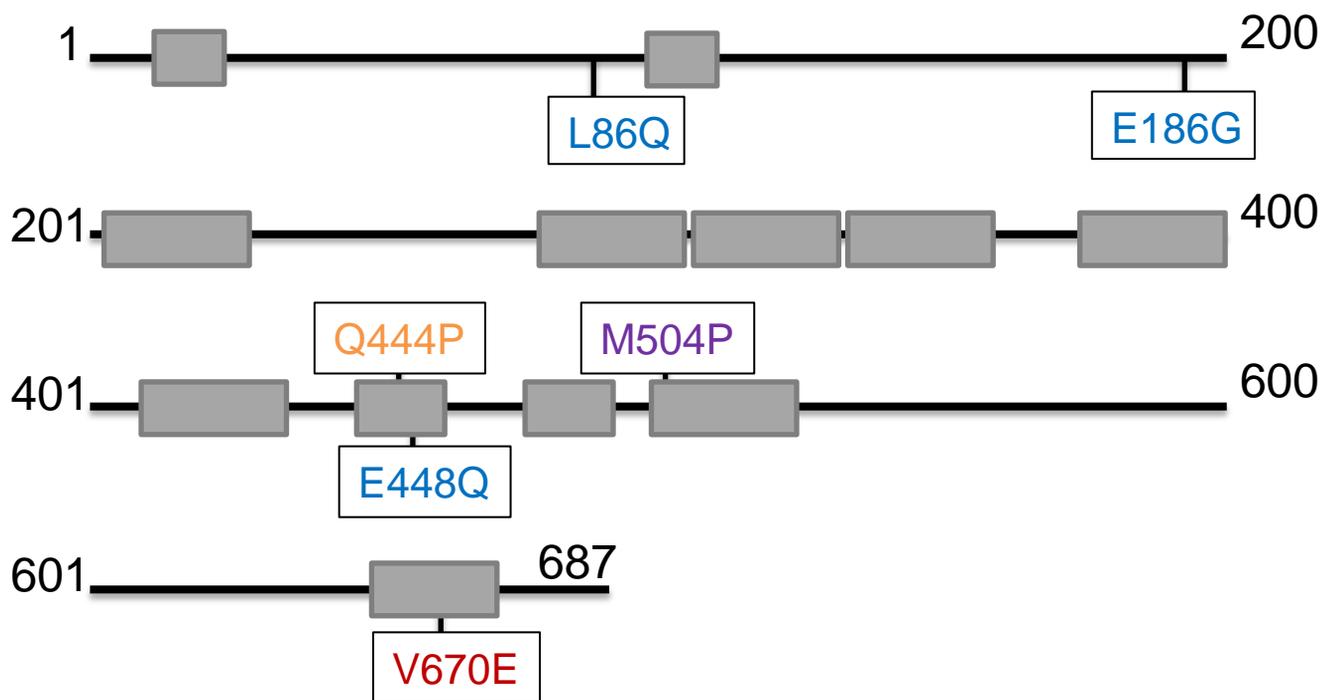
Alignment of Med17 Mediator subunit from yeast *S. cerevisiae* (Sc) and human (Hs) was modified from (Bourbon 2008). Localization of Med17 mutations is indicated for mutants characterized in this study: *med17-98* (L86Q, E186G, E448Q), *med17-444* (Q444P), *med17-504* (M504P), *med17-670* (V670E).

Supplementary references

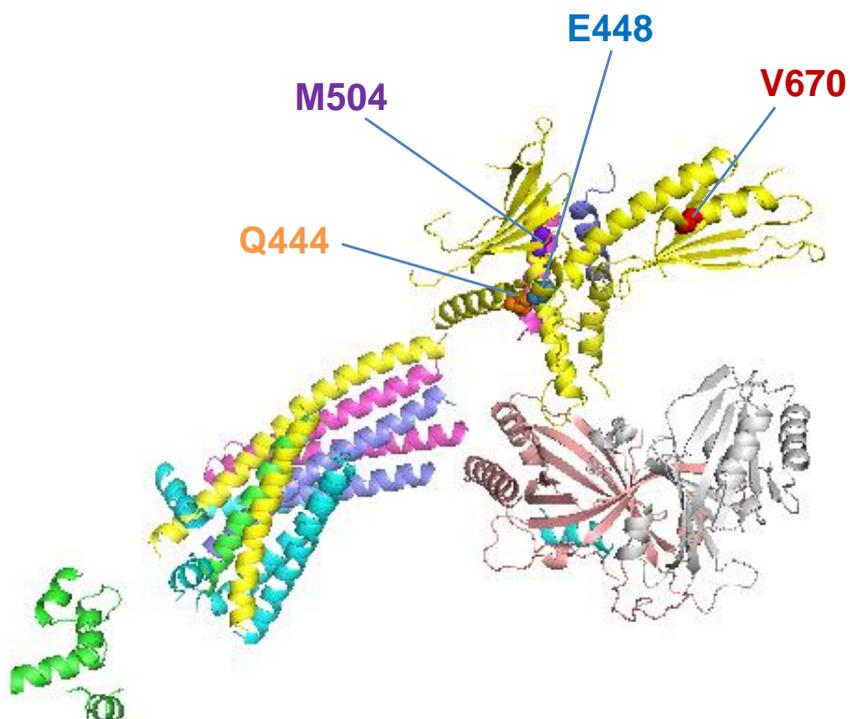
- Alberti S, Gitler AD, Lindquist S. 2007. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**(10): 913-919.
- Bourbon HM. 2008. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res* **36**(12): 3993-4008.
- Eyboulet F, Cibot C, Eychenne T, Neil H, Alibert O, Werner M, Soutourina J. 2013. Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment. *Genes Dev* **27**(23): 2549-2562.
- Lariviere L, Plaschka C, Seizl M, Wenzek L, Kurth F, Cramer P. 2012. Structure of the Mediator head module. *Nature* **492**(7429): 448-451.
- Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**(10): 953-961.
- Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**(1): 19-27.
- Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. 2011. Direct interaction of RNA polymerase II and Mediator required for transcription *in vivo*. *Science* **331**(6023): 1451-1454.
- Van Mullem V, Wery M, De Bolle X, Vandenhoute J. 2003. Construction of a set of *Saccharomyces cerevisiae* vectors designed for recombinational cloning. *Yeast* **20**(8): 739-746.



A



B



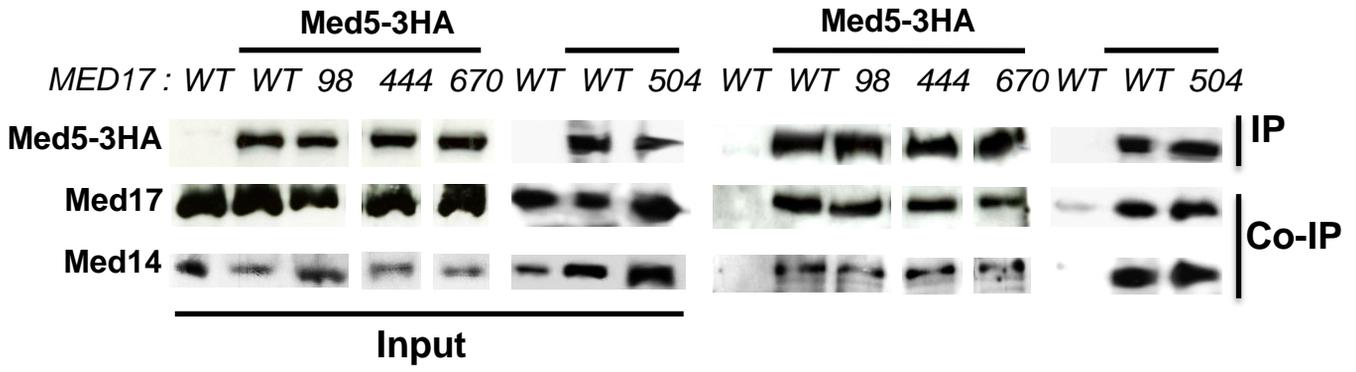
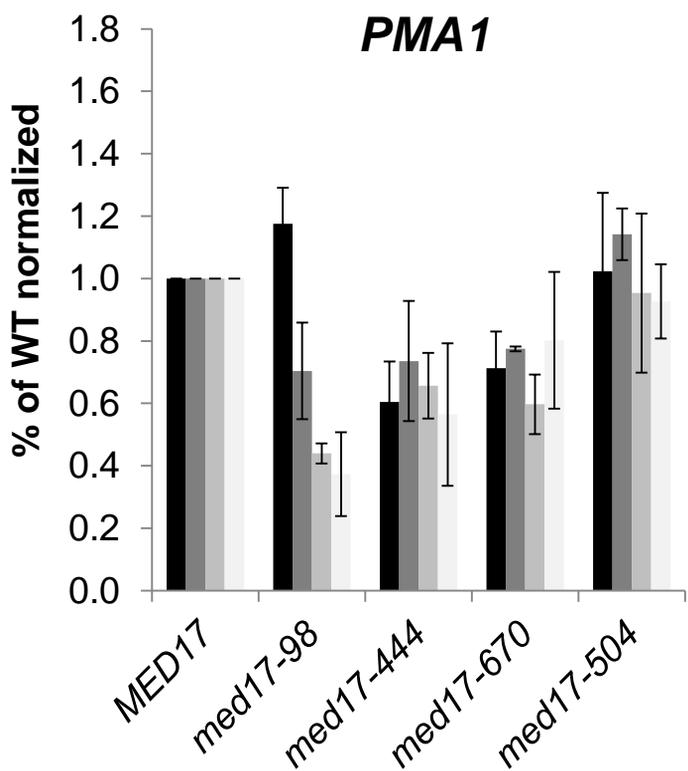
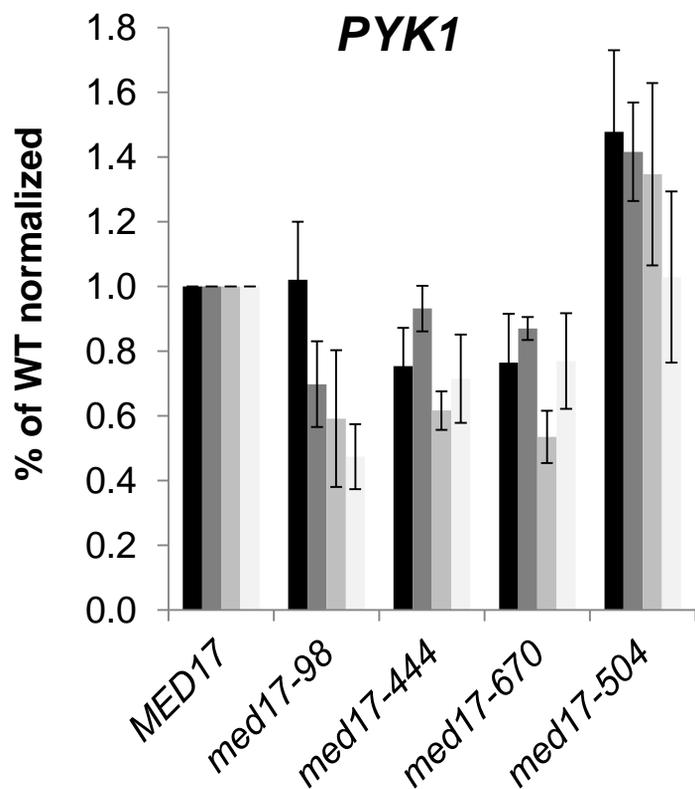
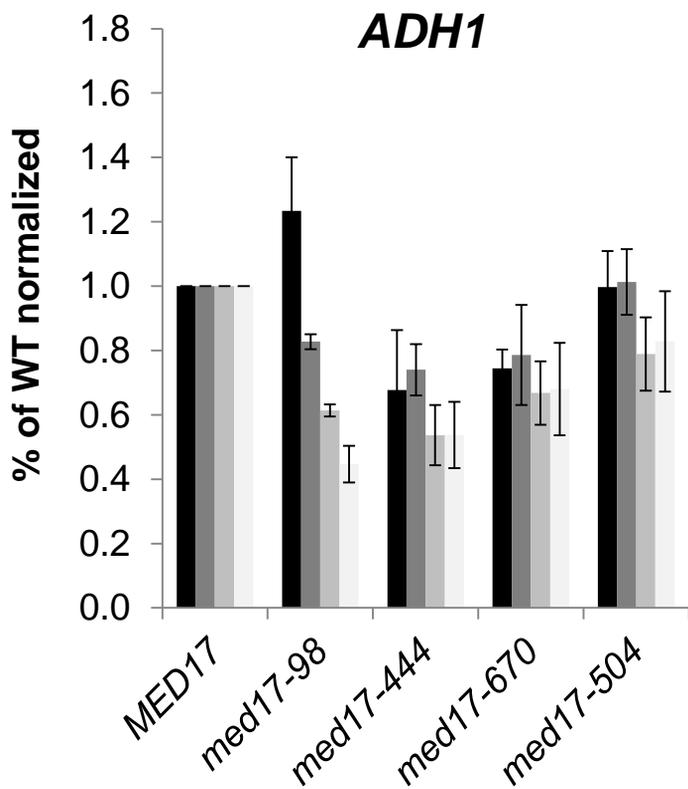
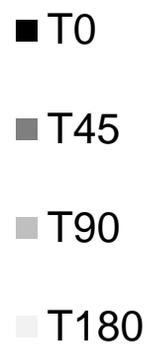


Figure S4



Time at 37°C (min)



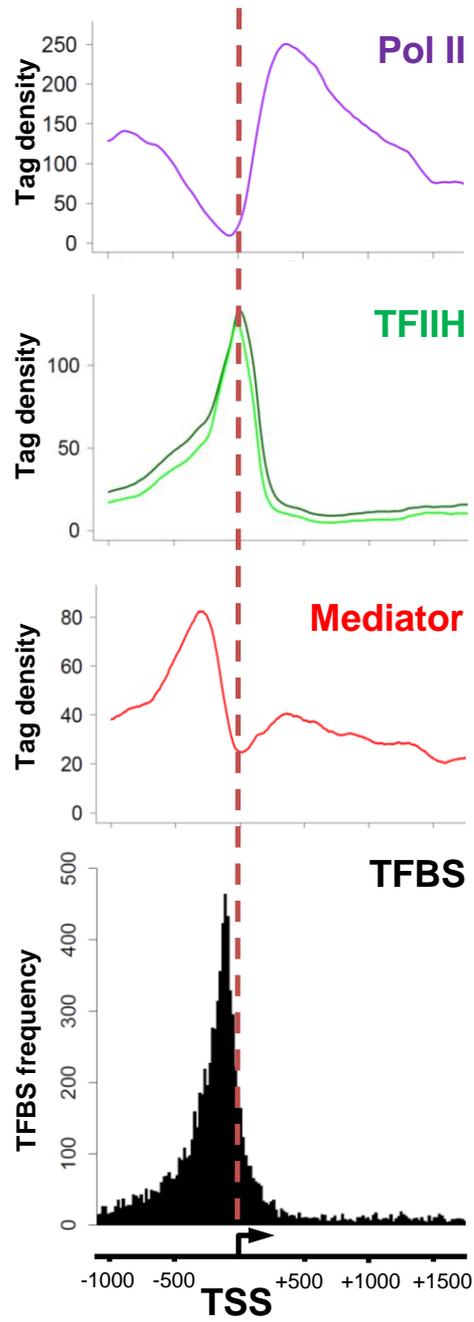


Figure S6



ARTICLE II : ROLE DU MEDIATEUR DANS LA REPARATION DE L'ADN

I. CONTEXTE DU TRAVAIL

Les mécanismes de réparation de l'ADN diffèrent selon le type de lésion. Le système de réparation par excision de nucléotides (NER) présent à la fois chez les procaryotes et les eucaryotes, est le plus versatile des systèmes de réparation en termes de reconnaissance des lésions. Il permet l'élimination des dommages qui génèrent une distorsion de la double hélice d'ADN, tels que les CPD (« cyclobutane pyrimidine dimers »), générés par les UVs. Ce mécanisme de réparation a été largement étudié, notamment *in vitro*, à la fois chez la levure (Prakash and Prakash, 2000) et chez les mammifères (Park and Choi, 2006) et consiste en 4 grandes étapes : (1) reconnaissance de la lésion, (2) ouverture de la double hélice d'ADN, (3) excision du fragment d'ADN lésé, et (4) re-synthèse du fragment manquant. Il existe 2 voies différentes dans le système NER : (1) celle réparant les lésions sur l'ensemble du génome, appelé GGR (« global genome NER ») et (2) celle réparant les lésions bloquant la transcription, appelé TCR (« transcription coupled NER »).

Dans le processus TCR, la réparation de l'ADN est couplée au mécanisme de la transcription. Ce système répare efficacement les dommages bloquant la progression de la Pol II sur le brin transcrit pour permettre une reprise rapide de l'activité transcriptionnelle. Bien que les composants requis pour la reconnaissance de la lésion et sa réparation par le système TCR soient désormais bien compris *in vitro*, de nombreuses questions demeurent quant au couplage de la réparation avec la transcription. Certaines protéines, initialement identifiées comme des composants de la machinerie de transcription ou de la machinerie de réparation de l'ADN peuvent jouer un rôle dans les deux processus. C'est notamment le cas du facteur TFIIH, qui intervient dans la transcription des gènes par la Pol II en tant que GTF mais également dans le système de réparation NER (Compe and Egly, 2012). Les complexes PAF et Ccr4-Not qui jouent un rôle dans l'étape d'élongation par la Pol II seraient également requis dans la TCR chez la levure (Gaillard et al., 2009).

TFIIH contient deux hélicases, Rad3/XPD et Ssl2/XPB, qui servent à ouvrir la double hélice d'ADN durant l'initiation de la transcription et la NER. Des mutations dans XPD sont responsables de la trichothiodystrophie (TTD) et du Xeroderma pigmentosum (XP) qui se caractérisent par une hypersensibilité de la peau aux UVs et une prédisposition au cancer. Les endonucléases Rad2/XPG

et Rad1-Rad10/XPF-ERCC1 permettent de cliver le dommage respectivement en 3' et 5' dans les deux voies du système NER. Rad2/XPG interviendrait également dans la reconnaissance de la lésion avant la mise en place du complexe de réparation (Sarker et al., 2005). Des mutations dans le gène *XPG* humain conduisent à une pathologie associant Xeroderma pigmentosum et syndrome de Cockayne (XP/CS) (Clarkson, 2003; Scharer, 2008). En plus, un rôle transcriptionnel de Rad2 a été suggéré chez la levure *S. cerevisiae* (Lee et al., 2002) et des expériences de ChIP ont montré la présence de XPG sur le promoteur et les régions distales des gènes dépendants des récepteurs nucléaires, en absence de stress génotoxique (Le May et al., 2010). XPG aurait également un rôle dans la formation des boucles de chromatine entre le promoteur et le terminateur du gène *RAR β 2* activé, et cette fonction serait importante pour sa régulation (Le May et al., 2012).

Les sous-unités du Médiateur sont impliquées dans de nombreux contacts au sein du complexe lui-même mais également avec les composants de la machinerie transcriptionnelle. Le crible double-hybride réalisé dans notre laboratoire a ainsi permis d'identifier un contact entre Med17 et Rad2, homologue chez la levure de la protéine XPG humaine. Cette interaction suggère deux possibilités : (1) Rad2 pourrait être impliqué dans la transcription en influençant la mise en place du PIC ou sa stabilisation sur le promoteur ; et/ou (2) le Médiateur pourrait jouer un rôle dans la réparation, par exemple, en permettant le recrutement de Rad2 sur l'ADN. Bien que le Médiateur se voit attribuer de plus en plus de rôles au-delà de l'initiation de la transcription (Kagey et al., 2010; Takahashi et al., 2011; Wang et al., 2005; Zhu et al., 2011b), son intervention dans la réparation de l'ADN n'a encore jamais été démontrée. Afin de répondre à cette question, nous avons décidé d'explorer le lien fonctionnel existant entre le Médiateur et Rad2/XPG en utilisant notre collection de mutants ts obtenue pour la sous-unité Med17 du Médiateur.

Afin d'étudier la localisation de Rad2 à l'échelle du génome de la levure, nous avons réalisé des expériences de ChIP-on-chip qui nous ont permis de révéler la présence de ce facteur à la fois sur le promoteur et les ORFs des gènes de classe II, mais également sur les gènes de classe III et les régions télomériques. Mon implication dans ce travail a alors consisté en l'analyse de la localisation de la protéine Rad2 sur le génome de *S. cerevisiae* par ChIP-seq, en la comparant avec celles du Médiateur, de la Pol II et des GTFs. Nous avons ainsi mis en évidence une forte corrélation entre l'occupation de Rad2 et du Médiateur sur les régions promotrices des gènes de classe II, confirmant un lien fonctionnel entre ces deux facteurs.

Nous avons ensuite montré que le recrutement de Rad2 est dépendant de la transcription. Ainsi, Rad2 est recruté sur le gène *GAL1* lors de son induction par le galactose. En utilisant un mutant de la Pol II présentant un défaut de transcription à 37°C, nous avons montré qu'une

diminution de la présence de la Pol II sur le génome est corrélée avec une diminution de la présence de Rad2.

Afin de vérifier l'implication suggérée de Rad2 dans la transcription, nous avons délété le gène correspondant qui n'est pas essentiel à la viabilité de *S. cerevisiae*. L'occupation du Médiateur et de la Pol II, ainsi que la quantité d'ARNm, est inchangée dans la souche *rad2Δ* par rapport à la souche sauvage. De plus, aucun phénotype de croissance à l'exception de la forte sensibilité aux UVs, attendue pour un mutant de la réparation, n'a pu être mis en évidence pour la souche *rad2Δ*. Ces résultats suggèrent que Rad2 ne joue pas de rôle majeur dans la transcription et soulève donc l'hypothèse d'un rôle du Médiateur dans la réparation de l'ADN.

Pour vérifier cette hypothèse, nous avons utilisé notre collection de mutants de Med17 et nous avons montré que, dans un contexte où la voie GGR est déficiente (*rad7Δ*), certaines mutations de Med17 conduisent à une sensibilité aux UVs comparable à celle observée dans le double mutant *rad26Δ rad7Δ*, déficient pour les 2 voies GGR et TCR. Ces observations indiquent que le Médiateur est impliqué dans la réparation et plus particulièrement dans la voie TCR. Des analyses complémentaires ont par ailleurs démontré un défaut de recrutement de Rad2 sur l'ADN et un défaut d'interaction entre Med17 et Rad2 dans les mutants *med17ts* sensibles aux UVs.

Nos résultats, obtenus en utilisant la levure comme modèle, suggèrent un rôle fonctionnel du Médiateur dans la réparation de l'ADN via un contact direct entre le Médiateur et Rad2. Durant cette étude, nous avons montré que Rad2 est présent sur les gènes de classe II, en absence de stress génotoxique, dépendamment de la Pol II et que sa localisation génomique sur les régions promotrices corrèle avec celle du Médiateur. Nous avons ainsi démontré que le Médiateur est important pour le recrutement de Rad2, suggérant un nouveau rôle pour ce complexe dans la réparation par la voie TCR, en plus de son rôle de co-activateur dans la transcription par la Pol II.

II. ARTICLE 2

Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment

Fanny Eyboulet,^{1,2,3} Camille Cibot,^{1,2,3,5} Thomas Eychenne,^{1,2,3} Helen Neil,^{1,2,3,6} Olivier Alibert,⁴ Michel Werner,^{1,2,3} and Julie Soutourina^{1,2,3,7}

¹FRE3377, Institut de Biologie et de Technologies de Saclay (iBiTec-S), Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), F-91191 Gif-sur-Yvette cedex, France; ²FRE3377, Centre National de la Recherche Scientifique (CNRS), F-91191 Gif-sur-Yvette Cedex, France; ³FRE3377, Université Paris-Sud, F-91191 Gif-sur-Yvette Cedex, France; ⁴Institut de Radiobiologie Cellulaire et Moléculaire (iRCM), CEA, F-91057 Evry Cedex, France

Mediator is a large multiprotein complex conserved in all eukaryotes. The crucial function of Mediator in transcription is now largely established. However, we found that this complex also plays an important role by connecting transcription with DNA repair. We identified a functional contact between the Med17 Mediator subunit and Rad2/XPG, the 3' endonuclease involved in nucleotide excision DNA repair. Genome-wide location analyses revealed that Rad2 is associated with RNA polymerase II (Pol II)- and Pol III-transcribed genes and telomeric regions in the absence of exogenous genotoxic stress. Rad2 occupancy of Pol II-transcribed genes is transcription-dependent. Genome-wide Rad2 occupancy of class II gene promoters is well correlated with that of Mediator. Furthermore, UV sensitivity of *med17* mutants is correlated with reduced Rad2 occupancy of class II genes and concomitant decrease of Mediator interaction with Rad2 protein. Our results suggest that Mediator is involved in DNA repair by facilitating Rad2 recruitment to transcribed genes.

[*Keywords:* Mediator; transcription; DNA repair; Rad2/XPG; RNA polymerase II; *Saccharomyces cerevisiae*]

Supplemental material is available for this article.

Received July 24, 2013; revised version accepted October 23, 2013.

Mediator of transcription regulation is a large multi-protein complex conserved in all eukaryotes. It interacts with RNA polymerase II (Pol II) to transmit signals from specific regulators to the Pol II transcription machinery. Together with the general transcription factors (GTFs; TFIIA, TFIIB, TFIID, TFII E, TFII F, and TFII H), this coactivator stimulates the assembly of the preinitiation complex (PIC) and, ultimately, transcription initiation (Kornberg 2005; Ries and Meisterernst 2011). Mediator is generally required at Pol II-transcribed gene promoters for regulated transcription (Holstege et al. 1998; Andrau et al. 2006; Zhu et al. 2006). As expected given its central role in transcription activation, it has been implicated in numerous developmental processes. Mutations that affect Mediator subunits lead to a number of human pathologies (Risheg et al. 2007; Schwartz et al. 2007). Recently, mutations in the Med17 Mediator subunit have been

associated with infantile cerebral atrophy (Kaufmann et al. 2010), and a mutation in the Med23 subunit cosegregated with intellectual disability (Hashimoto et al. 2011). Since oncogenesis results from gene dysregulation, it is not unexpected that Mediator is involved in several cancers (Zhang et al. 2005; Vijayvargia et al. 2007; Firestein et al. 2008; Gade et al. 2009; Li et al. 2010; Kuuselo et al. 2011). Mediator has at least 25 subunits organized into four modules: head, middle, tail, and Cdk8 kinase modules. The complete crystallographic structure of the Mediator complex has not been determined. However, recent structural data have been obtained for the Mediator head module (Lariviere et al. 2012; Robinson et al. 2012). Mediator subunits are engaged in numerous contacts within the complex (Guglielmi et al. 2004) and probably with partners belonging to other components of the transcription machinery. For example, a cooperation of Mediator with TFIID has been reported in

Present addresses: ⁵Centre de Génétique Moléculaire du CNRS, F-91198 Gif-sur-Yvette, France; ⁶Institut National de la Santé et de la Recherche Médicale (Inserm) U1052, Centre de Recherche en Cancérologie de Lyon, F-69008 Lyon, France.

⁷Corresponding author

E-mail julie.soutourina@cea.fr

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.225813.113>.

© 2013 Eyboulet et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

yeast and humans (Johnson et al. 2002; Lariviere et al. 2006; Cai et al. 2010; Takahashi et al. 2011). Recently, we identified a direct interaction between the Med17 Mediator subunit and the Rpb3 Pol II subunit required for global Pol II transcription in vivo (Soutourina et al. 2011). Previously, we showed that a direct contact between the Med11 subunit of Mediator and the Rad3 subunit of TFIIH is essential for the recruitment of the GTF to the PIC independently of Pol II (Esnault et al. 2008). In addition to the crucial role of Mediator as a link between activators and the basal transcription machinery, recent studies suggest broader Mediator transcriptional functions, notably in the post-recruitment steps of transcription (Wang et al. 2005; Takahashi et al. 2011) and in DNA looping between the enhancers and promoters through a functional connection with cohesin (Kagey et al. 2010).

Transcription is coupled with DNA repair, ensuring the continuity of Pol II progression. DNA lesions impede the elongating transcription machinery and can potentially affect gene expression, leading to problems in development, cell growth, and survival. Proteins first identified as components of transcription or DNA repair machineries may be involved in both processes, operating beyond their initially described pathway. The TFIIH complex is the best-characterized dual factor, playing an essential role in Pol II transcription as a GTF and in nucleotide excision DNA repair (NER) (Compe and Egly 2012). NER is a major and evolutionarily conserved DNA repair pathway that removes DNA lesions such as cyclobutane pyrimidine dimers (CPD), the most abundant photoproduct arising upon UV irradiation. During NER, >30 proteins cooperate to recognize, incise, and excise a damaged oligonucleotide from the genomic DNA to protect the cell from the deleterious effects of DNA damages (Lagerwerf et al. 2011). There are two distinct NER pathways: global genome repair (GGR), which removes DNA lesions in the genome overall, and transcription-coupled repair (TCR), which removes DNA lesions that interfere with the progression of RNA polymerase through actively transcribed genes (Svejstrup 2002, 2007; Hanawalt and Spivak 2008). A single irreversibly stalled Pol II in an essential gene could be enough to lead to cell death. This situation explains why a large number of mechanisms and factors have evolved to protect the cell from stalled Pol II. While the NER reaction and the components required for efficient DNA lesion recognition and repair are fairly well understood, many questions on the active coupling of DNA repair with transcription remain unanswered.

Recent reports suggest that an interplay between transcription and DNA repair is likely to be more complex than previously thought. PAF and Ccr4–Not complexes, known for their role in transcription, have been proposed to be required for efficient TCR in yeast (Gaillard et al. 2009). In addition to TFIIH, other NER factors have been suggested to play a dual role in DNA repair and transcription. Rad2/XPG, the 3' endonuclease implicated in NER, is one of such factor. Mutations in the human XPG gene give rise to a xeroderma pigmentosum (XP) sometimes associated with Cockayne syndrome (CS)

(Clarkson 2003; Scharer 2008). Rad2/XPG, together with the Rad1–Rad10/XPF–ERCC1 5' endonuclease, cleaves the damaged DNA strand in both NER pathways, GGR and TCR, resulting in the release of a DNA fragment containing the DNA lesion. Moreover, a specific non-enzymatic function of XPG in recognition of stalled transcription in TCR initiation has been also proposed (Sarker et al. 2005). In addition, a transcriptional role for the NER factor Rad2/XPG has been suggested in the yeast *Saccharomyces cerevisiae* (Lee et al. 2002). In humans, mutations in XPG result in impaired transactivation (Ito et al. 2007), and XPG binds to the promoters and distal regions of several nuclear receptor (NR)-dependent genes in the absence of any genotoxic stress (Le May et al. 2010). Recently, XPG and XPF endonucleases have been found to be involved in chromatin looping between the promoter and the terminator of the activated RAR β 2 human gene (Le May et al. 2012).

In this study, we identified a functionally important interaction between the essential Med17 Mediator subunit and the Rad2/XPG DNA repair protein, suggesting a functional link between Mediator of transcription regulation and DNA repair. Genome-wide location analyses revealed that Rad2 was associated with Pol II-transcribed genes but also with Pol III-transcribed genes and telomeric regions in the absence of exogenous genotoxic stress. We showed that Rad2 occupancy of Pol II-transcribed genes was transcription-dependent. Moreover, Rad2 occupancy of class II gene promoters was highly correlated with that of Mediator. Several Mediator *med17* mutants were UV-sensitive in a GGR-deficient background, and their UV sensitivity was correlated with reduced Rad2 occupancy of class II genes and a concomitant decrease of the interaction between Mediator and Rad2 protein. We thus uncovered a previously unsuspected role of Mediator in DNA repair via Rad2/XPG recruitment.

Results

Physical interaction between the Med17 Mediator subunit and the Rad2/XPG protein

Previously, we characterized Mediator complex organization by screening the Mediator subunits with a yeast genomic library in a two-hybrid system. Protein–protein interactions within the Mediator complex have been published (Guglielmi et al. 2004). Many interactions that we uncovered in the screen were confirmed crystallographically and served as a basis for other Mediator studies (for example, Lariviere et al. 2012; Robinson et al. 2012). In addition, our two-hybrid screening revealed a large number of interactions between *S. cerevisiae* Mediator and other nuclear proteins (B Guglielmi, NL van Berkum, C Boschiero, FCP Holstege, and M Werner, unpubl.). One particularly interesting interaction is the contact between the Med17 Mediator head module subunit and the Rad2 endonuclease, the yeast homolog of human XPG protein involved in DNA repair. Med17 fused to the Gal4 DNA-binding domain interacted in the two-hybrid assay with the Rad2 fragment (549–857) fused to the Gal4 activation

domain (Fig. 1A). The contact between Mediator and Rad2 was confirmed by coimmunoprecipitation (co-IP) experiments with crude extracts of a yeast strain expressing Med17-Myc and HA-Rad2 from their native promoters (Fig. 1B; Supplemental Fig. S1). Our results show that Mediator coimmunoprecipitates with Rad2 in crude extracts and vice versa. Rad2 was detected by Western blotting when Med17-Myc was used to immunoprecipitate Mediator complex, and the Med17-Myc Mediator subunit was coimmunoprecipitated when Rad2 was immunoprecipitated via the HA tag.

Pol II is the main component of the transcription machinery and the first complex in TCR that recognizes the DNA damage (Svejstrup 2002, 2007; Hanawalt and Spivak 2008). The XPG protein was reported to coimmunoprecipitate with Pol II in crude extracts from undamaged HeLa cells (Sarker et al. 2005). In line with these results, our co-IP experiments also indicated that Rad2 coimmunoprecipitated with Pol II in yeast crude extracts

(Fig. 1C). Coimmunoprecipitated Pol II forms phosphorylated on Ser5 and Ser2 of the Rpb1 C-terminal repeat domain (CTD) were also detected by Western blotting (data not shown).

Genome-wide location analysis of Rad2

Recently, an association of XPG, the human homolog of Rad2, with several NR-dependent genes was shown in the absence of genotoxic stress (Le May et al. 2010, 2012). However, to date, no genomic data are available on the Rad2/XPG occupancy in humans or yeast. To determine the targets of Rad2 in the yeast genome in the absence of exogenous genotoxic stress, we performed Rad2 chromatin immunoprecipitation (ChIP)-chip experiments. Two yeast strains carrying N-terminal or C-terminal HA-tagged version of Rad2 were constructed and used in conventional ChIP experiments. No increase in the UV sensitivity of these strains was observed, suggesting that both tagged versions of Rad2 were functional (data not shown). We found that N-terminally and C-terminally tagged Rad2 were significantly enriched inside three constitutively expressed Pol II-transcribed genes tested: *ADH1*, *PYK1*, and *PMA1* (Supplemental Fig. S2A). The strain carrying an N-terminal tagged version of Rad2 was used for ChIP-chip analysis, since the ChIP signal to noise ratios were higher. We found 2471 oligonucleotides significantly bound by Rad2 (P -value < 0.005; out of 40,000 oligonucleotide probes present on the DNA arrays). The genomic profile of Rad2 occupancy was complex, spanning from Pol II-transcribed genes (promoter and transcribed regions) to, unexpectedly, Pol III-transcribed genes and telomeric/subtelomeric regions. Examples of Rad2 ChIP-chip enrichment profiles illustrate these different genomic region categories (Fig. 2A–C).

To gain in resolution and determine precisely the yeast genomic regions enriched by Rad2 protein, we performed ChIP-seq experiments. In agreement with our Rad2 ChIP-chip results, we confirmed the presence of this protein inside and on the promoter regions of Pol II-transcribed genes as well as on Pol III-transcribed genes. Telomeric regions are highly repetitive and therefore are poorly represented in ChIP-seq analysis. We identified 2600 significantly enriched Rad2 peaks using MACS2 peak calling software; 1203 of these peaks were located on promoter regions of Pol II-transcribed genes, 33 were on snRNA genes, and 268 were on Pol III-transcribed genes. This bioinformatics tool, MACS2, also allowed us to identify some Rad2 enrichment peaks inside the transcribed regions of class II genes but was not particularly adapted for detection of large Rad2 distributions in these regions, missing many Rad2-enriched ORFs. To identify precisely the Rad2 enrichment signals inside ORFs and thus complete the Rad2 ChIP-seq distribution analysis, we compared Rad2 mean densities for each ORF with density distributions of two negative controls: untagged strain and input DNA samples. We observed that 818 regions inside the protein-coding genes were significantly enriched by Rad2.

The results obtained by our genome-wide analyses were confirmed by conventional ChIP on a set of selected

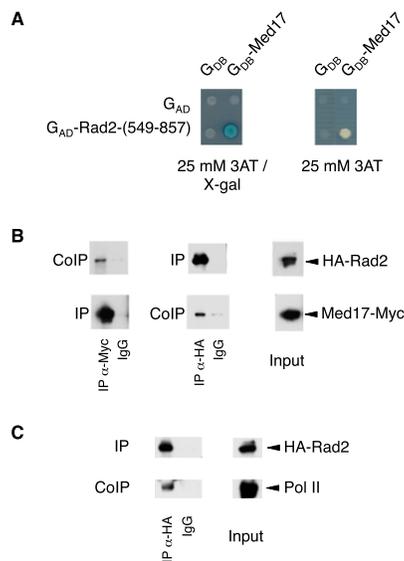


Figure 1. The Med17 Mediator subunit interacts with the Rad2/XPG DNA repair protein. (A) Two-hybrid interaction between the Med17 Mediator subunit and Rad2. Two-hybrid assays were performed in the Y190 strain, and the activation of the *lacZ* or *HIS3* reporters was tested using an X-Gal overlay plate assay or 3AT plate assay (Werner et al. 1993). Med17 was fused to the Gal4 DNA-binding domain (G_{DB} -Med17), and the Rad2-(549–857) fragment was fused to the Gal4 activation domain [G_{AD} -Rad2-(549–857)]. (B) Co-IP of Mediator and Rad2. Mediator was immunoprecipitated through the Med17-Myc subunit with α -Myc antibody from crude yeast extracts (Input; right panel) and analyzed by Western blotting with α -HA antibody (co-IP) against Rad2 (left panel). (Middle panel) Vice versa, Rad2 was immunoprecipitated with α -HA antibody from crude extracts and analyzed by Western blotting with α -Myc antibody (co-IP) against Med17. (IgG) Control immunoprecipitation with IgG magnetic beads only. (C) Co-IP of Rad2 and Pol II. Rad2 was immunoprecipitated with α -HA antibody from crude extracts (Input). Coimmunoprecipitated Pol II was detected by Western blotting with α -Rpb1 antibody (co-IP). (IgG) Control immunoprecipitation with IgG magnetic beads only.

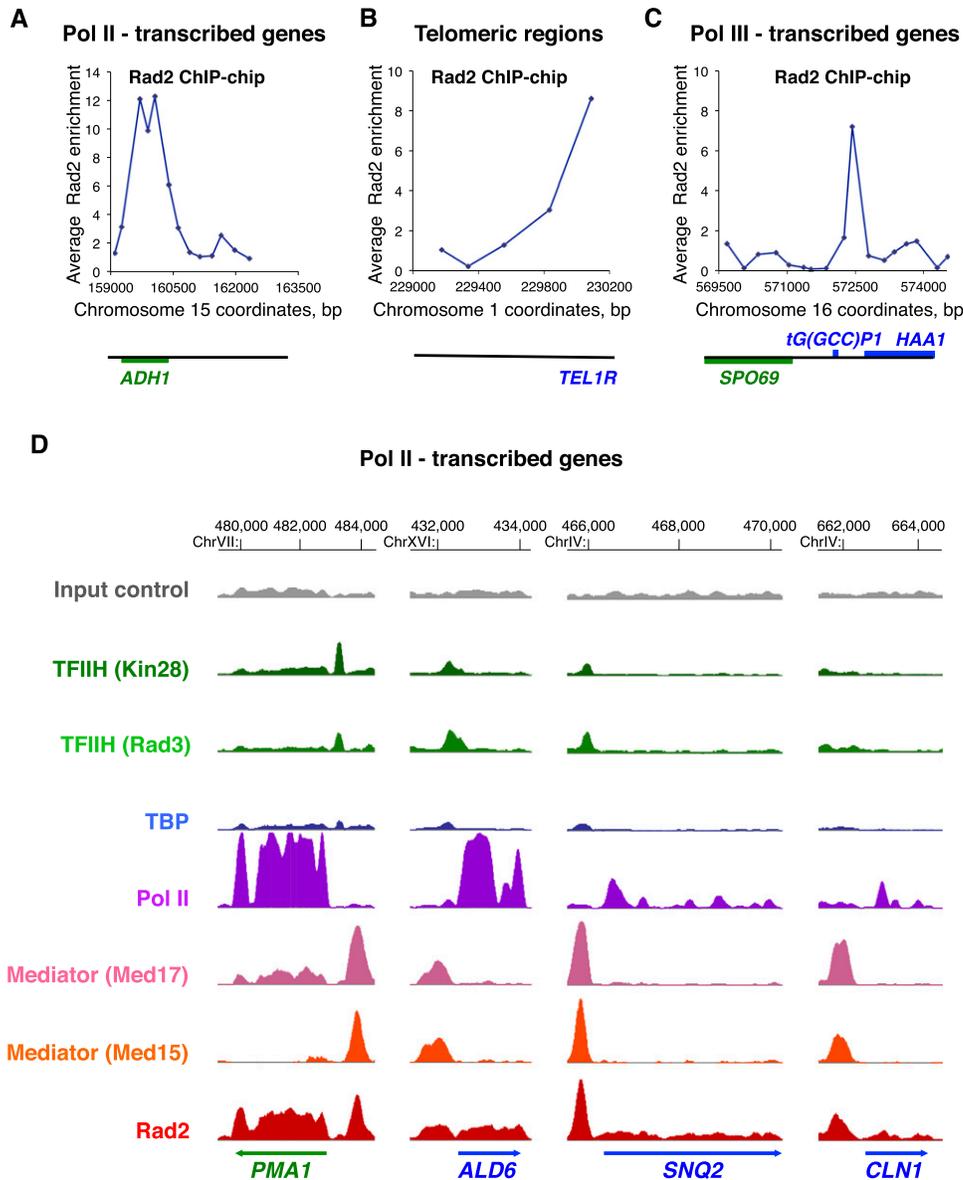


Figure 2. Genome-wide analysis of Rad2, Mediator, Pol II, TBP, and TFIIH. (A–C) Examples of Rad2 ChIP–chip enrichment profiles on Pol II-transcribed genes (A), on a telomeric region (B), and on a Pol III-transcribed gene (C). Rad2 enrichment profiles were assessed from ChIP–chip experiments with HA-Rad2 strain grown at 30°C in YPD medium. Watson strand-transcribed genes are colored in blue, and Crick strand-transcribed genes are colored in green. The genomic positions of probe regions on corresponding chromosomes are indicated along the X-axis. The enrichment ratio is indicated along the Y-axis. (D) Examples of Rad2, Mediator, Pol II, TBP, and TFIIH ChIP-seq enrichment profiles on selected Pol II-transcribed genes. Densities of sequence tags were assessed from ChIP-seq experiments performed with HA-Rad2-tagged, Med15-HA-tagged (Mediator), Med17-HA-tagged (Mediator), HA-TBP-tagged, Rad3-HA-tagged (core TFIIH), and Kin28-HA-tagged (TFIIH kinase module) strains with α -HA antibody. Pol II was immunoprecipitated using α -Rpb1 antibody. ChIP-seq density profiles are displayed using IGB yeast genome browser. Watson strand-transcribed genes are colored in blue, and Crick strand-transcribed genes are colored in green. Input DNA and DNA from ChIP with an untagged strain were used as negative controls. Densities of sequence tags were displayed after subtraction of normalized control of untagged strain as described in the Materials and Methods.

regions representing the categories of Rad2 genomic targets that we identified. Supplemental Figure S2B illustrates significant enrichment of Rad2 on several promoter and transcribed regions of selected class II genes compared with background levels measured on the coding region of the nontranscribed *GAL1* gene. All selected Pol

III-transcribed genes displayed significant Rad2 enrichment (Supplemental Fig. S2C). The spatial Rad2 distribution on the longest class III gene, *SCR1*, was similar to that of Pol III (Supplemental Fig. S2D; Ghavi-Helm et al. 2008). We also confirmed the presence of Rad2 on telomeric regions (Supplemental Fig. S2E). In addition,

a significant Rad2 enrichment was detected on the Pol I-transcribed 35S rDNA region compared with the intergenic *NTS2* region (Supplemental Fig. S2F). Taken together, these results indicated a complex Rad2 distribution on the yeast genome in the absence of exogenous genotoxic stress, spanning the genes of all three transcription classes and telomeric/subtelomeric regions. For further study, we focused particularly on Pol II-transcribed genes.

Genome-wide Rad2 occupancy correlates with Mediator occupancy on promoter regions of Pol II-transcribed genes

To compare Rad2 ChIP-seq profiles with those of Mediator, Pol II, and GTFs, we performed ChIP-seq experiments for Mediator (Med15 and Med17 subunits), Pol II, and GTFs (TBP and two TFIID subunits: the Rad3/XPD subunit of core TFIID and the Kin28/Cdk7 kinase subunit of TFIID). Input DNA and DNA from ChIP with an untagged strain were used as negative controls. Figure 2D and Supplemental Figure S3 show examples of ChIP-seq density distributions using the Integrated Genome Browser (IGB) for all of the proteins on selected class II (Fig. 2D) or class III (Supplemental Fig. S3) genes. Transcription by all three RNA polymerases requires TBP protein. As expected, density peaks of this protein were also detected on the Pol III-transcribed genes. These TBP peaks on class III genes coincided with Rad2 peaks (Supplemental Fig. S3). On the selected promoter regions of Pol II-transcribed genes, there was a particular correlation between Rad2 and Mediator for the location and intensity of peaks (Fig. 2D). Furthermore, genome-wide Mediator and Rad2 occupancies were well correlated. Rad2 and Mediator ChIP-seq densities (25-base-pair [bp] bin) on the yeast genome were computed and compared (Fig. 3A; Supplemental Fig. S4A). As expected, the Rad2 enrichment of several genomic categories, including class II promoter and transcribed regions, made the correlation between Rad2 and Mediator densities complex. Considering only class II gene promoter regions, regression analysis showed a linear correlation with a correlation coefficient of 0.75 for the Med15 Mediator subunit (Fig. 3B) or 0.73 for the Med17 Mediator subunit (Supplemental Fig. S4B). Intergenic regions encompassing Pol III-transcribed genes were excluded from this analysis. The global correlation between Rad2 and Mediator occupancies on promoter regions is consistent with a functional link between Rad2 and the Mediator complex. We determined whether Rad2 association with Pol II-transcribed regions was correlated with Pol II occupancy (Fig. 3C). Considering only Pol II-transcribed regions inside ORFs, encompassing most of the yeast Pol II ChIP signals, no strong correlation could be observed between Rad2 and Pol II ChIP-seq densities (Fig. 3D). Regression analysis showed a lower correlation coefficient of 0.54. In comparison, regression analysis indicated, as expected, a linear correlation between two subunits of the Mediator complex (Med15 and Med17) or TFIID complex (Rad3 and Kin28), with a high correlation coefficient of 0.89 (Sup-

plemental Fig. S4C,D) or 0.88 (Supplemental Fig. S5A,B), respectively, and between TBP and Rad3 or Kin28 TFIID subunits with a high correlation coefficient of 0.75 or 0.81, respectively (Fig. 3E). Pairwise correlations between ChIP-seq densities of Rad2, Mediator, Kin28, Rad3, and TBP showed that the strongest correlation for Rad2 protein occupancy on promoter regions was with Mediator (Fig. 3E). On the other hand, no strong correlation was observed between Rad2 and Kin28 ChIP-seq occupancies, with a correlation coefficient of 0.53 (Supplemental Fig. S5C,D).

The location of Rad2 peaks on the promoter regions of Pol II-transcribed genes was compared with those of Mediator and GTFs. Mean ChIP-seq read densities were calculated for each protein on promoter regions of Pol II-transcribed genes centered on Mediator enrichment peaks (Fig. 3F). To consider only Pol II-transcribed gene promoters, intergenic regions encompassing Pol III-transcribed genes were excluded. These analyses showed that Rad2 peak distribution coincides with that of Mediator (Med15 and Med17 subunits) but not with those of TFIID core (Rad3) or kinase (Kin28) modules. Taken together, our results revealed that genome-wide Rad2 occupancy was highly correlated with that of Mediator on promoter regions of Pol II-transcribed genes and that Rad2 enrichment distribution in these regions superposed with that of Mediator.

Rad2 occupancy on Pol II-transcribed genes is transcription-dependent

An essential question arising from Rad2 recruitment in the absence of exogenous genotoxic stress is whether Rad2 occupancy on class II genes depends on Pol II transcription. To address this question, we examined Rad2 occupancy on the *GAL1* gene by ChIP experiments upon galactose induction. When yeast cells were grown in raffinose-supplemented medium, Rad2 was absent from the *GAL1* promoter and ORF (Fig. 4A). After 60 min of galactose induction, Rad2 was recruited to the promoter and transcribed region of the *GAL1* gene. The constitutively expressed *ADH1* gene was always bound by the Rad2 protein. As expected, the *GAL1* mRNA level was highly induced after galactose induction, while the *ADH1* mRNA level was comparable in raffinose- and galactose-supplemented medium (Fig. 4B). These results indicated that Rad2 recruitment to *GAL1* gene was transcription-dependent.

To address more generally the question of the transcriptional dependency of Rad2 occupancy on Pol II-transcribed genes, we used a Pol II mutant, *rpb1-1*, that rapidly stops transcription after a shift to nonpermissive temperature (Nonet et al. 1987). When cells were incubated for 90 min at 37°C, as expected, Pol II occupancy on Pol II-transcribed genes was strongly reduced in the *rpb1-1* mutant compared with the wild-type strain (Fig. 4C), whereas Pol III occupancy on tested class III genes remained unchanged (Fig. 4D). We also checked that, as expected, RNA levels were strongly decreased in the *rpb1-1* mutant for class II genes (Fig. 4E) but not class III genes (Fig. 4F). Rad2 occupancy on Pol II-transcribed genes followed that of Pol II and was also greatly reduced after incubation at 37°C (Fig. 4G; Supplemental S6A,B).

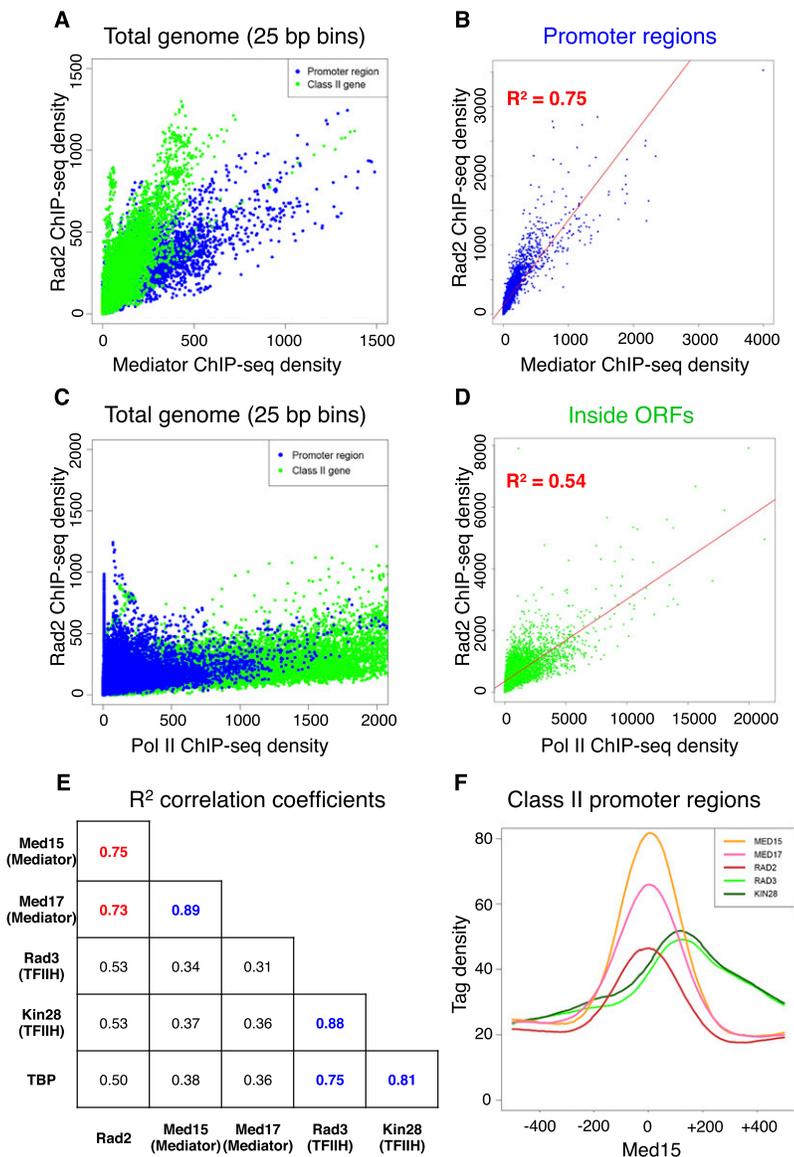


Figure 3. Correlation of Rad2 ChIP-seq density with those of Mediator and Pol II. (A) Rad2 ChIP-seq density versus Mediator (Med15 subunit) ChIP-seq density on total yeast genome. Density of sequence tags in Mediator and Rad2 ChIP-seq experiments has been calculated for each 25-bp bin of yeast genome. Each point on the plot corresponds to one genomic bin. The genomic bins located inside ORFs or class II promoter regions are indicated in green and blue, respectively. Pol III-transcribed genes were excluded from this analysis. (B) Rad2 ChIP-seq density versus Mediator (Med15 subunit) ChIP-seq density on class II promoter regions. The density of sequence tags in Mediator and Rad2 ChIP-seq experiments has been calculated for promoter region of each Pol II-transcribed gene. Each point on the plot corresponds to one promoter region. Promoter regions correspond to intergenic regions for Pol II-transcribed genes in tandem or in divergent orientation, excluding intergenic regions encompassing Pol III-transcribed genes. A linear regression for Rad2 ChIP-seq density versus Mediator ChIP-seq density (red line) and a R^2 correlation coefficient are indicated. (C) Rad2 ChIP-seq density versus Pol II ChIP-seq density on total yeast genome, as described in A. (D) Rad2 ChIP-seq density versus Pol II ChIP-seq density inside ORFs. The density of sequence tags in Rad2 and Pol II ChIP-seq experiments has been calculated for each ORF. Each point on the plot corresponds to one ORF. A linear regression for Rad2 ChIP-seq density versus Pol II ChIP-seq density (red line) and a R^2 correlation coefficient are indicated. (E) Pairwise correlation coefficients between ChIP-seq densities of Rad2, Med15 (Mediator), Med17 (Mediator), Rad3 (TFIIH core module), Kin28 (TFIIH kinase module), and TBP on promoter regions. R^2 correlation coefficients for each pairwise combination are shown. A correlation coefficient for Mediator versus Rad2 is highlighted in red, and pairwise correlation coefficients for different GTF (Rad3, Kin28, and TBP) or Mediator subunit combinations (Med15 and Med17) are shown in blue. (F) Distribution of Mediator, Rad2/XPG in transcription, since such a role has been previously suggested (Lee et al. 2002; Le May et al. 2010, 2012). In *S. cerevisiae*, the *RAD2* gene is not essential for viability. We started by examining whether, under standard growth conditions (rich glucose medium, YPD), any effect of *rad2* deletion on Mediator or Pol II occupancy could be identified. We did not observe any difference between Mediator occupancy of the *rad2Δ* strain compared with the wild-type strain on constitutively expressed gene promoters (Fig. 5A; Supplemental Fig. S7A,B). In addition, Pol II occupancy on selected genes and the corresponding mRNA level in the *rad2Δ* strain were similar to those of the wild-type strain (Fig. 5B,C; Supplemental Fig. S7C). Similarly, no effect of *rad2* deletion on Mediator or Pol II

Rad2, and TFIIH ChIP-seq densities on Pol II-transcribed gene promoters. Tag densities are shown relative to the localization of the maximum of Mediator enrichment peaks. Intergenic regions encompassing Pol III-transcribed genes were excluded. The tag density was determined for each protein in a 1000-bp window centered on the maximum of Mediator enrichment peaks per each nucleotide position. Mean tag density was then calculated and plotted over the window.

Rad2 occupancy decrease was observed for Pol II-transcribed gene promoters and ORFs. This decrease of Rad2 occupancy was not related to Mediator occupancy, since it did not change in the *rpb1-1* mutant (Supplemental Fig. S6C–E). Conversely, Rad2 occupancy on class III genes or telomeric regions remained unchanged under the same conditions (Fig. 4G). Thus, the presence of Rad2 on class II genes is specifically dependent on Pol II transcription.

UV irradiation sensitivity is the only observed phenotype of rad2 deletion

One hypothesis for the functional interplay between Mediator and Rad2 consists of a possible implication of

Rad2/XPG in transcription, since such a role has been previously suggested (Lee et al. 2002; Le May et al. 2010, 2012). In *S. cerevisiae*, the *RAD2* gene is not essential for viability. We started by examining whether, under standard growth conditions (rich glucose medium, YPD), any effect of *rad2* deletion on Mediator or Pol II occupancy could be identified. We did not observe any difference between Mediator occupancy of the *rad2Δ* strain compared with the wild-type strain on constitutively expressed gene promoters (Fig. 5A; Supplemental Fig. S7A,B). In addition, Pol II occupancy on selected genes and the corresponding mRNA level in the *rad2Δ* strain were similar to those of the wild-type strain (Fig. 5B,C; Supplemental Fig. S7C). Similarly, no effect of *rad2* deletion on Mediator or Pol II

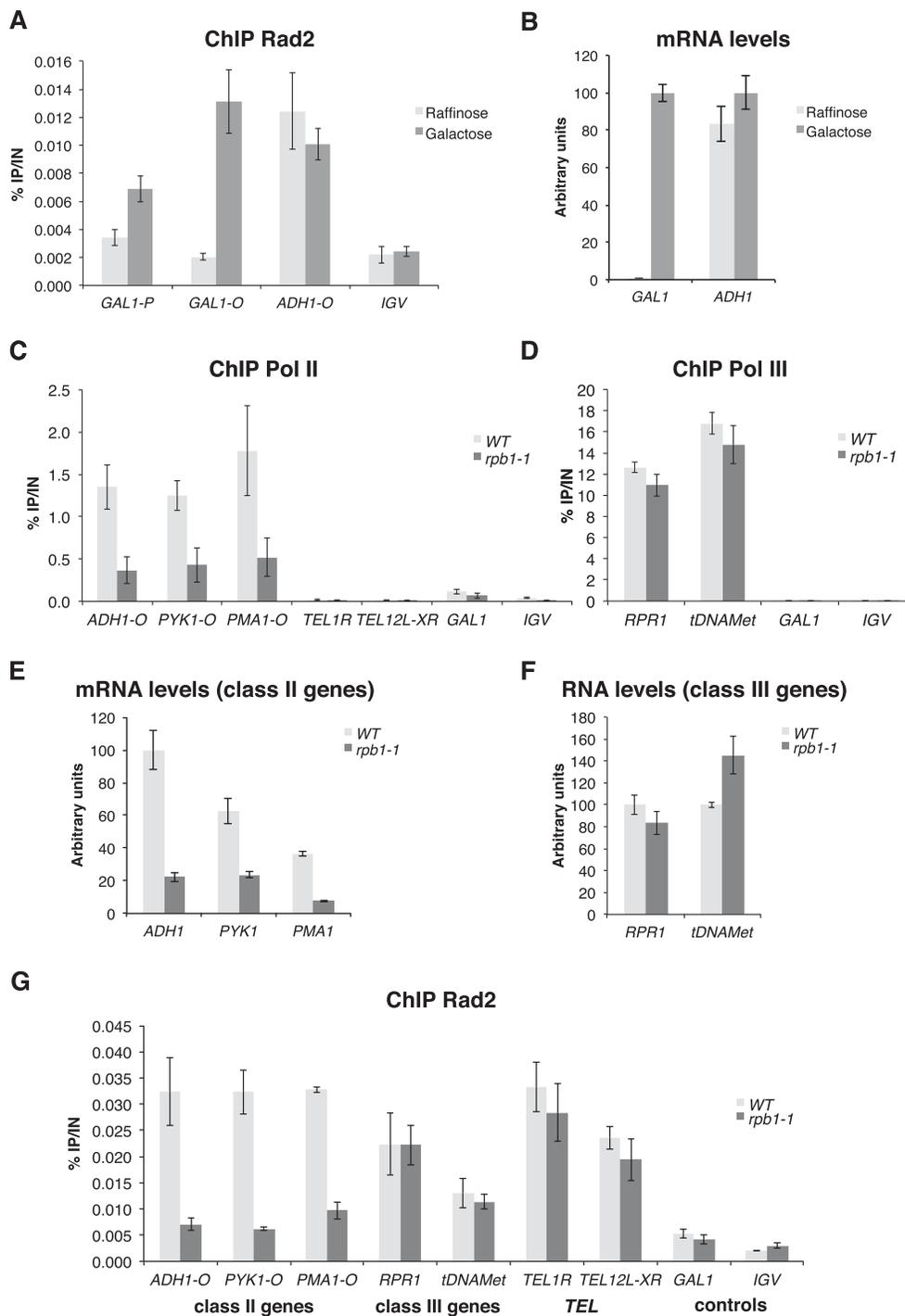


Figure 4. Transcription dependency of Rad2 occupancy on Pol II-transcribed genes. (A) Quantitative ChIP analysis of Rad2 occupancy on the *GAL1* gene before and after galactose induction. The HA-Rad2 tagged strain was grown in a raffinose-supplemented medium, and then galactose was added for 60 min. Quantitative ChIP experiments were performed with α -HA antibody against HA-Rad2. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to the *GAL1* gene promoter (*GAL1 P*) or ORF (*GAL1 O*). *ADH1* ORF represents a constitutively expressed gene control. A nontranscribed region on chromosome V (*IGV*) was used as a negative control. (B) mRNA level of *GAL1* and *ADH1* genes before and after galactose induction. Yeast cells were grown in a raffinose-supplemented medium, and then galactose was added for 60 min. RNA levels for *GAL1* and *ADH1* genes were quantified by RT-qPCR and normalized to *25S* rRNA, *18S* rRNA, and *SCR1* levels. Values are expressed in arbitrary units representing a percentage of the value in galactose-supplemented medium. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. (C,D,G) Effect of *rpb1-1* mutation on Rad2, Pol II, and Pol III occupancies on selected regions. Quantitative ChIP assays were performed using α -Rpb1 antibody (Pol II) (C), α -Myc antibody against Rpc160-Myc (Pol III) (D), and α -HA antibody against HA-Rad2 (G). Cells were grown in selective SD medium complemented with amino acids at 25°C and then shifted for 90 min at 37°C. A *GAL1* ORF and a nontranscribed region on chromosome V (*IGV*) were used as negative controls. (E,F) Effect of *rpb1-1* mutation on RNA level of class II (E) or class III (F) genes. Cells were grown in selective SD medium complemented with amino acids at 25°C and then shifted for 90 min at 37°C. RNA levels for selected genes were quantified by RT-qPCR and normalized to *25S* rRNA, *18S* rRNA, and *SCR1* levels. Values are expressed in arbitrary units representing a percentage of the value in the wild-type strain for the *ADH1* gene (E) or *RPR1* or *tDNAMet* genes (F). Mean values and standard deviation (indicated by error bars) of three independent experiments are shown.

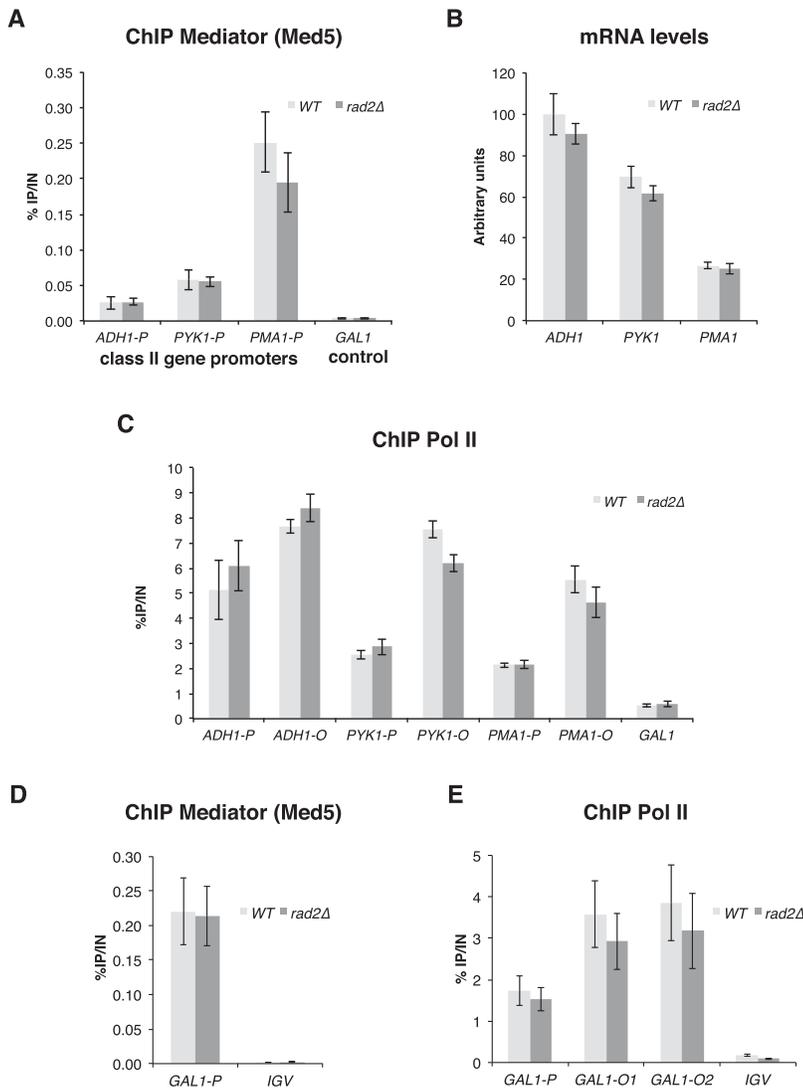


Figure 5. Mediator and Pol II occupancies in the *rad2* deletion strain on selected regions. (A,C) Quantitative ChIP analysis of Mediator and Pol II occupancies under standard growth conditions. *rad2Δ* and wild-type strains were grown in YPD medium at 30°C. Immunoprecipitations were performed using α -HA antibody against Med5-HA (Mediator) (A) and α -Rpb1 antibody (Pol II) (C). (A) Immunoprecipitated fragments from Mediator ChIP experiments were amplified with primers corresponding to selected class II gene promoters. (C) Immunoprecipitated fragments from Pol II ChIP experiments were amplified with primers corresponding to selected class II gene promoters (P) or ORFs (O). Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. A *GAL1* ORF was used as a negative control. (B) Effect of *rad2Δ* mutation on mRNA levels of class II genes. *rad2Δ* and wild-type strains were grown in YPD medium at 30°C. RNA levels for selected genes were quantified by RT-qPCR and normalized to 25S rRNA, 18S rRNA, and *SCR1* levels. Values are expressed in arbitrary units representing a percentage of the value in the wild-type strain for *ADH1* gene. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. (D,E) Quantitative ChIP analysis of Mediator and Pol II occupancies upon galactose induction. *rad2Δ* and wild-type strains were grown in a raffinose-supplemented medium, and then galactose was added for 60 min. Quantitative ChIP experiments were performed with α -HA antibody against Med5-HA (Mediator) (D) and α -Rpb1 antibody (Pol II) (E). Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to *GAL1* gene promoter (*GAL1-P*) or ORF (*GAL1-O1* and *GAL1-O2*). A nontranscribed region on chromosome V (*IGV*) was used as a negative control.

occupancy on the *GAL1* gene was observed in galactose-inducing medium (Fig. 5D,E). Previously, Lee et al. (2002) have stated that *rad2* deletion slowed down the kinetics of galactose-inducible genes. To check this possibility, the Pol II recruitment to the *GAL1* gene promoter and transcribed region and the *GAL1* mRNA level were examined at different times after galactose induction by ChIP and quantitative RT-PCR (qRT-PCR) experiments. However, there was no difference between the *rad2Δ* mutant and the wild-type strain (Supplemental Fig. S8).

To identify growth phenotypes of *rad2* deletion strains and therefore a possible transcriptional role for Rad2, we tested different growth conditions in two backgrounds: BY4741 and YPH499. No growth difference between *rad2Δ* and wild-type strains was observed on YPD medium under different temperature conditions (16°C, 25°C, 30°C, or 37°C), on different carbon sources (YP medium supplemented with galactose, lactate, ethanol, or glycerol), and on drugs to test for multidrug resistance transcrip-

tional response (cycloheximide and ketoconazole) or for NTP depletion conditions (6-azauracil and mycophenolic acid) (Supplemental Fig. S9; data not shown). The only phenotype of the *rad2* deletion strains that we observed is a high UV sensitivity, as expected for a DNA repair protein. Taken together, our results strongly suggest that Rad2 does not play a major role in the transcriptional process in yeast, at least in the two backgrounds examined here.

Physiological significance of Mediator–Rad2 contact in DNA repair

We could not identify a major implication of Rad2 in yeast Pol II transcription. This led to the hypothesis that Mediator might play a role in DNA repair via its interaction with the Rad2 protein. To investigate the functional relationship between Mediator and Rad2, we used our collection of 29 conditional mutants for the Med17 essential Mediator subunit (Soutourina et al.

2011). The integrity of Mediator in *med17* mutants has been verified by co-IP experiments (Soutourina et al. 2011). To directly address a possible role of Mediator in DNA repair, a UV sensitivity phenotype was systematically investigated for *med17* mutants. No UV sensitivity was found for *med17* mutants in the context where both NER pathways, GGR and TCR, were functional (Supplemental Fig. S10). Next, we used a GGR-deficient context (*rad7Δ* strains) to analyze specifically the potential implication of Med17 in the TCR pathway by UV sensitivity spot assays. Several Mediator mutants among our collection of *med17-ts* were UV-sensitive in a *rad7Δ* GGR-deficient context (Fig. 6A). To refine the UV sensitivity analysis, we determined UV survival curves for several *med17* mutants (three UV-sensitive *med17-140*, *med17-233*, and *med17-257* mutants and one UV-insensitive *med17-L546P* mutant) alone or in combination with *rad7Δ* mutation (Fig. 6B). The UV dose response analysis further confirmed UV sensitivity spot assays and clearly indicated that the combination of *rad7* deletion with several *med17* mutations leads to an increased UV sensitivity. An extremely UV-sensitive *rad2* deletion mutant was also included for comparison. Interestingly, the UV sensitivity of *med17-ts rad7Δ* mutants was comparable with that of the *rad26Δ rad7Δ* double mutant

deficient in both GGR (*rad7Δ*) and TCR (*rad26Δ*) NER pathways (Fig. 6B). Rad26 is the yeast homolog of the TCR-specific human CSB protein. In yeast, *RAD26* deletion does not lead to UV sensitivity but significantly increases the UV sensitivity of GGR-deficient mutants. Since *med17* mutants were UV-sensitive in a GGR-deficient context, we wondered whether these mutants could be epistatic with the TCR-deficient *rad26* deletion mutant. No UV sensitivity increase was found for *med17* mutants in the *rad26Δ* context (Supplemental Fig. S11), demonstrating that, indeed, *med17* mutants are epistatic with the *rad26Δ* mutant. Taken together, these results are consistent with a specific implication of Mediator in the TCR pathway. Our results also indicated that the UV sensitivity of *med17-ts* mutants was not a consequence of a transcriptional effect, since this phenotype was observed only in a GGR-deficient context and not in a context where both NER pathways, GGR and TCR, were functional or in a TCR-deficient context. Three of the UV-sensitive *med17-ts* mutants were selected for further molecular study (*med17-257*, *med17-140*, and *med17-233*) (the results on the two latter mutants are shown in Supplemental Fig. S12). Rad2 occupancy in these mutants was determined by ChIP experiments for selected Pol II-transcribed genes. Figure 7A and Supplemental Figure

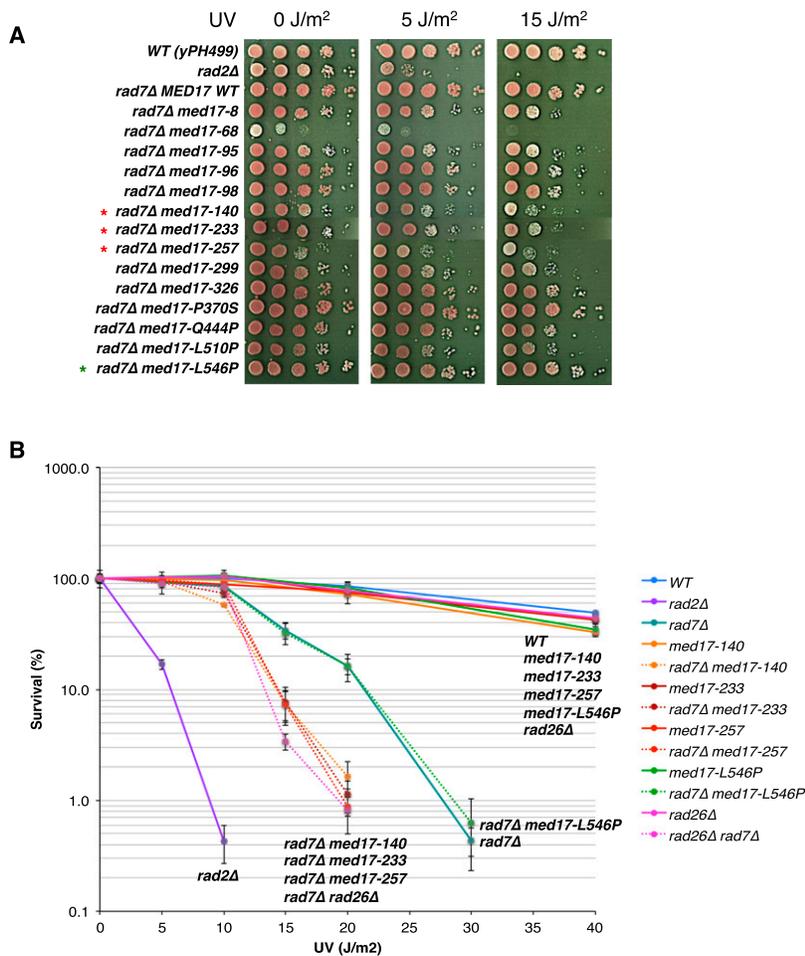


Figure 6. UV sensitivity of *med17* mutants. (A) UV sensitivity of *med17* mutants in a GGR-deficient *rad7Δ* context by spot assays. Cells were serially diluted, spotted on YPD agar plates, treated or not with the indicated doses of UV (UV Stratalinker 1800), and incubated for 3 d at 30°C. Red asterisks mark selected UV-sensitive *med17* mutants, and a green asterisk marks a selected UV-insensitive *med17* mutant. (B) UV survival curves of *med17* and *rad26Δ* mutants alone and in a GGR-deficient *rad7Δ* context. Cells were grown in YPD medium to a mid-exponential phase, diluted, plated on YPD agar, and irradiated with the different doses of UV (UV Stratalinker 1800). Plates were incubated for 3 d at 30°C in the dark before colonies were counted and survival was calculated. Mean values and standard deviation (indicated by error bars) of at least three independent experiments are shown.

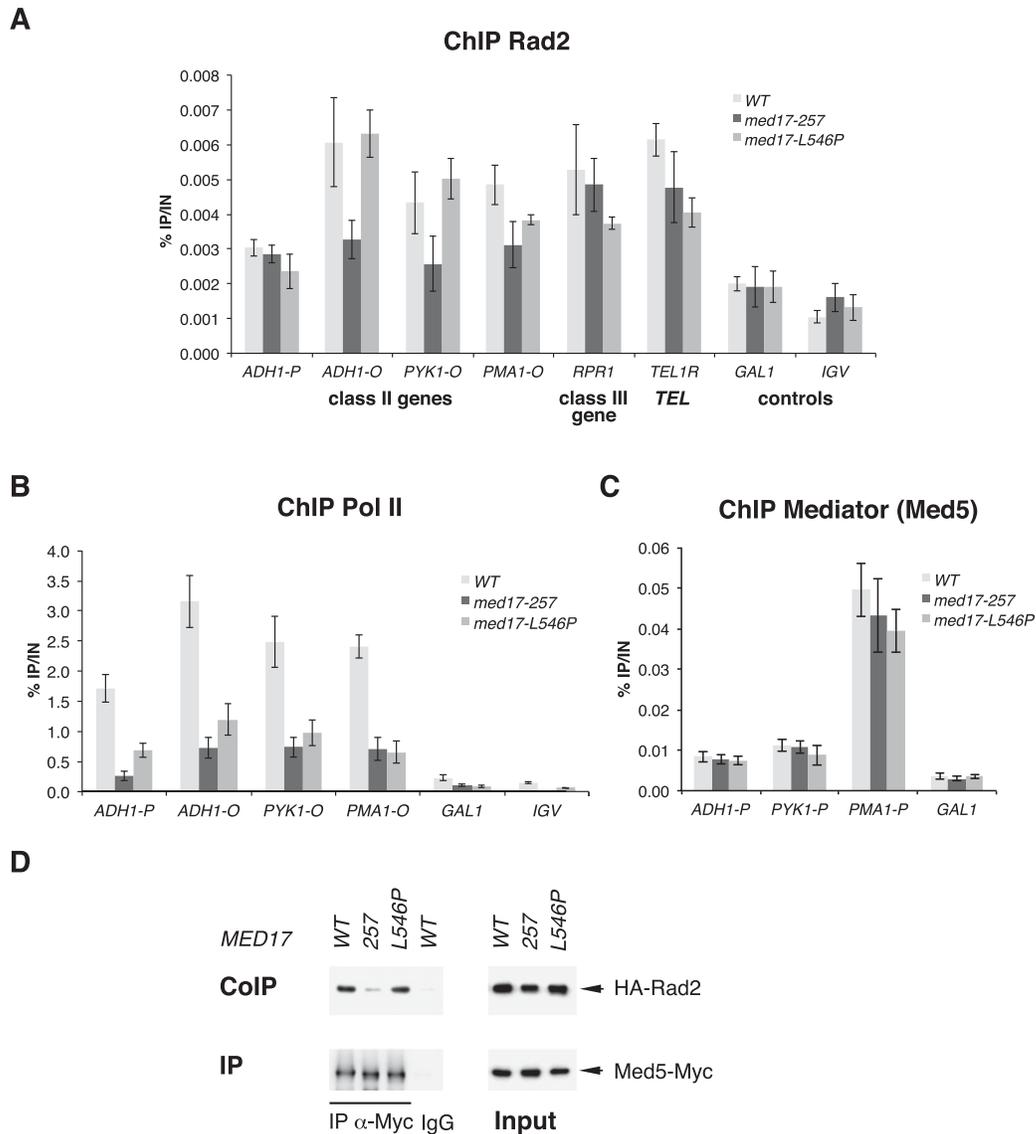


Figure 7. Effects of *med17* mutants on Rad2, Pol II, and Mediator occupancies and on Rad2–Mediator interaction. (A–C) Quantitative ChIP analysis of Rad2, Pol II, and Mediator occupancies on selected regions. ChIP assays were performed using α -HA antibody against HA-Rad2 (A), α -Rpb1 antibody (Pol II) (B), and α -HA antibody against Med5-HA (Mediator) (C). Cells were grown in YPD medium at 30°C and then shifted for 45 min at 37°C. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to selected class II gene promoters (P) or ORFs (O), a selected class III gene, and a telomeric region. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. A *GAL1* ORF and a nontranscribed region on chromosome V (*IGV*) were used as negative controls. (D) Co-IP of Mediator and Rad2. Mediator was immunoprecipitated through a Med5-Myc subunit with a α -Myc antibody from crude extracts (Input; right panel) and analyzed by Western blotting with α -HA antibody (co-IP) against Rad2 (left panel). (IgG) Control immunoprecipitation with IgG magnetic beads only.

S12A show a significant decrease in Rad2 association in UV-sensitive Mediator mutants compared with the wild-type strain and *med17-L546P* mutant, one of the UV-insensitive *med17-ts* mutants. Rad2 occupancy to a class III gene and a telomeric region was not affected in these mutants. All tested *med17* mutations, including *med17-L546P*, which is UV-insensitive, led to a marked decrease in Pol II association with Pol II-transcribed genes but did not modify the Mediator association to corresponding gene promoters (Fig. 7B,C; Supplemental Fig. S12B,C). A

decrease in Pol II occupancy was not responsible for Rad2 occupancy defects of UV-sensitive mutants (*med17-257*, *med17-140*, and *med17-233*) or their UV sensitivity, since a reduction of Pol II occupancy in the *med17-L546P* mutant, similar to that of the UV-sensitive *med17-ts* mutants, was not accompanied by Rad2 occupancy decrease and UV sensitivity. To determine the effects of *med17-ts* mutations on Mediator contact with the Rad2 protein, we performed co-IP experiments and showed that Mediator–Rad2 interaction was reduced in UV-sensitive

mutants *med17-257*, *med17-140*, and *med17-233* compared with the wild-type strain and the *med17-L546P* UV-insensitive mutant (Fig. 7D; Supplemental Fig. S12D). These results indicated that Mediator is involved in Rad2 recruitment and therefore suggest that it is involved in DNA repair.

Discussion

Substantial evidence has been obtained on the essential function of the Mediator complex in transcription activation. However, our experiments indicate that this complex plays an important role in linking transcription with other nuclear processes. In this study, we identified a functionally important contact between the Med17 Mediator subunit and the Rad2 DNA repair protein, the *S. cerevisiae* homolog of human XPG. Mediator might play a role in DNA repair by loading Rad2/XPG to active genes, or Rad2/XPG could be implicated in transcription, influencing PIC assembly or stability on several gene promoters or both. Our results suggest that Mediator is involved in DNA repair through a functional link with the Rad2/XPG protein. (1) The essential Med17 Mediator subunit interacts with Rad2/XPG DNA repair protein. (2) We showed that Rad2 is located on Pol II-transcribed genes (promoter and transcribed regions) in the absence of exogenous genotoxic stress. (3) Genome-wide Rad2 occupancy of class II gene promoters is well correlated with that of Mediator. (4) Moreover, Rad2 distribution on these regions overlapped with that of Mediator but not of TFIIF, further highlighting a functional cooperation between Mediator and Rad2/XPG. It should be noted that Mediator and Rad2 ChIP-seq peaks on class II gene promoters are generally located on the upstream regulating regions, whereas TFIIF and other GTFs peaks are located on core promoter regions close to transcription start sites. (5) Rad2 occupancy of Pol II-transcribed genes is transcription-dependent, since *rpb1-1* Pol II mutation reduces Rad2 occupancy after a shift to nonpermissive temperature when transcription is rapidly inhibited, and since Rad2 is recruited to *GAL1* gene upon galactose induction of this gene. (6) However, no growth phenotypes (except UV sensitivity) or transcriptional effects were observed in the *rad2Δ* context, suggesting that Rad2 does not play a major role in the transcriptional process in yeast, at least in our two backgrounds. (7) On the contrary, Mediator *med17-ts* mutants are UV-sensitive in a GGR-deficient background (*rad7Δ*) and are epistatic with a TCR-deficient mutant (*rad26Δ*). (8) This UV sensitivity of several *med17* mutants is correlated with reduced Rad2 occupancy of class II genes and concomitant decrease of interaction with Rad2 protein.

We suggest that the contact between Med17 and Rad2 is direct, since our two-hybrid screening permitted us to reveal only specific interactions between different subunits within the Mediator complex (Guglielmi et al. 2004) that were further confirmed as direct by structural analysis (for example, Lariviere et al. 2012; Robinson et al. 2012). Moreover, we demonstrated that several mutations in the Med17 Mediator subunit resulted in a decreased

interaction between Mediator and Rad2, an observation in favor of a direct Mediator–Rad2 contact.

Therefore, Mediator is likely to facilitate Rad2 recruitment to transcribed genes, setting the stage for rapid DNA lesion removal. We propose that Mediator could function in transcription-coupled DNA repair, revealing a previously unknown role of the Mediator complex as a link between transcription and DNA repair. Whether Rad2/XPG is the only NER factor that is recruited through an interaction with Mediator or assembles with other NER factors to remove lesions on transcribed genes remains to be addressed. Further investigations will determine whether Mediator is engaged in functional interplay with other NER factors.

In the light of Mediator and Rad2/XPG conservation from yeast to humans, the molecular events governing the Mediator link with DNA repair are likely to exist in all eukaryotes. In support of this idea, we observed the Mediator–XPG contact by co-IP experiments with crude extracts from HeLa cells (data not shown). The conservation of Mediator connection to DNA repair might give insights into our understanding of human diseases like XP/CS.

Interestingly, both Mediator and Rad2/XPG interact with Pol II, the main component of transcription machinery and the first complex in TCR that recognizes the DNA damage. The mechanisms of functional orchestration of these contacts remain to be investigated. Our genome-wide analyses show that Mediator and Rad2 co-occupy class II gene promoter regions and, moreover, that a high correlation is observed between the two profiles on promoter regions. Rad2 is also bound to Pol II-transcribed regions that are occupied by Pol II. However, a correlation between Rad2 and Pol II on these regions was modest, suggesting a complex relationship between Rad2 and Pol II. At this stage, we propose that Mediator facilitates Rad2 recruitment on class II genes promoters. While we do not know what the precise mechanisms of Rad2 loading on ORFs are, our results suggest that they could entail both the loading of Rad2 on Pol II and another independent mechanism. A modest genome-wide correlation between Rad2 and Pol II contrasts with the good genome-wide correlation between TFIIS and Pol II or Pol III occupancies on Pol II- or Pol III-transcribed genes, respectively (Ghavi-Helm et al. 2008). These observations suggested that TFIIS associates at the 5' end of genes and accompanies Pol II and Pol III but cannot be loaded on traveling enzymes. It should be noted that one of the *med17* mutants used in this study (*med17-257*) that showed a UV-sensitive phenotype was lethal in combination with the *rpb3-2* Pol II mutation (Soutourina et al. 2011) and was affected in Pol II–Mediator interaction (data not shown). However, other *med17* mutations with UV-sensitive phenotypes did not lead to *rpb3-2* lethality phenotypes. Our results showed that Pol II occupancy was reduced in UV-sensitive and -insensitive *med17* mutants. Thus, the lower occupancy of Pol II cannot explain the UV sensitivity of the *med17* mutants that are defective in Rad2 interaction. We also showed that the UV sensitivity of *med17* mutants was not a consequence

of a transcriptional defect, since this phenotype was observed only in a GGR-deficient context and not in a context where both NER pathways were functional or a TCR-deficient context. Taken together, our results suggest that UV sensitivity of Mediator mutants was not an indirect consequence of their transcriptional defects.

A transcriptional role for the NER factor Rad2/XPG was first proposed in the yeast *S. cerevisiae* (Lee et al. 2002). Rad2 was suggested to promote efficient Pol II transcription. However, we did not observe the mild growth phenotypes of *rad2* deletion strains or slower galactose induction kinetics in the absence of the *RAD2* gene reported by Lee et al. (2002). Moreover, we did not find any genetic interaction between *rad2Δ* and *rad26Δ*. Rad26 is the yeast homolog of the TCR-specific human CSB protein. This result might be due to the different genetic backgrounds used by us and Lee et al. (2002). In spite of our considerable efforts, we did not obtain any evidence for a direct implication of Rad2 in transcription. In human cells, it has been demonstrated that mutations in XPG prevent its association with TFIIH, resulting in the dissociation of CAK and XPD/Rad3 from the core TFIIH and in impaired transactivation (Ito et al. 2007). Recently, XPG and other NER factors were shown to be associated with the promoters and distal regions of several NR-dependent genes in the absence of any exogenous genotoxic stress (Le May et al. 2010). The presence of these repair proteins at the promoters of activated genes is sensitive to transcription inhibitors and necessary for optimal DNA demethylation and histone post-translational modifications (H3K4/H3K9 methylation and H3K9/14 acetylation) and thus for efficient transcription. Deficiencies in some NER factors impede the recruitment of others and affect NR transactivation. Recently, XPG and XPF endonucleases were found to be involved in chromatin looping between the promoter and the terminator of the activated *RARβ2* human gene (Le May et al. 2012). It should be noted that NRs do not exist in *S. cerevisiae*, which moreover has no DNA methylation system. It remains to be determined whether XPG and other NER factors are generally important for transcription activation in mammalian cells or whether their action is limited to specific pathways such as NR-mediated transcription.

We showed that in addition to Pol II-transcribed genes, Rad2 is also located on Pol III-transcribed genes and telomeric regions in the absence of exogenous genotoxic stress, indicating a complex Rad2 distribution on the yeast genome and opening interesting perspectives of possible novel Rad2 functions. Recently, a role of Mediator on telomeres was proposed, since this complex influences telomeric silencing, cellular life span, and telomere heterochromatin maintenance (Zhu et al. 2011; Peng and Zhou 2012). In yeast, Mediator is bound to the telomeric regions. Our results revealed that Rad2 is also enriched in these regions. The colocalization of Mediator and Rad2 on telomeres raises the possibility that Mediator function on the telomeres could be related to Rad2.

In conclusion, our results suggest an intriguing model that Mediator might play more roles in nuclear processes

than previously assumed by serving as an assembly platform or a regulatory element linking transcription with DNA repair and possibly other chromatin-related processes.

Materials and methods

Strains and plasmids

All *S. cerevisiae* strains are described in Supplemental Table S1. All plasmids are listed in Supplemental Table S2. The oligonucleotides used in this study can be found in Supplemental Table S3.

ChIP, ChIP-chip, and ChIP-seq

ChIP and ChIP-chip experiments were performed as described (Ghavi-Helm et al. 2008). In ChIP experiments, we considered a significant enrichment compared with negative controls or a significant difference between ChIP occupancies when *P*-value was <0.05 in a Student's *t*-test. Chromatin preparation for ChIP-seq experiments was performed as described for conventional ChIP, except that an additional sonication step with Bioruptor (Diagenode; six cycles of 30 sec with medium intensity setting) was included to generate DNA fragments of ~200-bp mean size. DNA sequencing of 40-nucleotide (nt) tags was performed on Solexa genome analyzer GA-IIx using the procedures recommended by the manufacturer (Illumina). Input DNA and DNA from ChIP with an untagged strain were used as negative controls. The ChIP-chip and ChIP-seq data have been deposited to the Array Express under accession numbers E-MEXP-3875 and E-MTAB-1595.

Co-IP experiments

Whole-yeast extract preparation and immunoprecipitation in immunoprecipitation buffer (50 mM HEPES at pH 7.5, 100 mM NaCl, 20% glycerol, 1 mM dithiothreitol, 0.5 mM EDTA, 0.05% NP-40) supplemented with a protease inhibitor cocktail (Complete, Roche) and 1 mM PMSF and Western blotting were performed as described previously (Soutourina et al. 2006). The 12CA5 anti-HA and 9E10 anti-Myc antibodies were used against HA- or Myc-tagged proteins, and the 8WG16 anti-Rpb1 antibody was used against Pol II.

qRT-PCR analysis

RNA was extracted with hot acidic phenol following a protocol derived from Schmitt et al. (1990). Reverse transcription of 0.5-μg RNA samples was performed using iScript cDNA synthesis kit (Bio-Rad) with a mix of oligo(dT) and random hexamers for priming. qPCR results were normalized using 25S rRNA, 18S rRNA, and *SCR1* RNA as internal controls. Values represent the average of three independent experiments, and error bars indicate standard deviations.

Data analysis

ChIP-seq data were analyzed using the following bioinformatics tools. The sequences were aligned on *S. cerevisiae* genome (University of California at Santa Cruz [UCSC] version sacCer3) with Bowtie version 0.12.7 (Langmead et al. 2009). Conversions to different file formats were performed using Samtools version 0.1.16. Only uniquely mapped tags were used, and a maximum of two mismatches was allowed. To avoid possible sequencing

artifacts, reads beyond five repetitions at the same position of the genome were removed, and reads with the best-quality scores were conserved for each position. The number of mappable tags for each ChIP-seq experiment is indicated in Supplemental Table S4. To calculate ChIP-seq density, reads were extended to an assumed fragment length of 180 nt, and a count of reads was determined per 1-bp bin using Bedtools version 2.15.0. ChIP-seq density profiles were displayed using UCSC or IGB yeast genome browsers.

We used input DNA and DNA from ChIP with untagged strain as negative controls. The ChIP sample from the untagged strain generally shows very low tag density on the genome, except some regions representing <1% of the genome that display an apparent enrichment in this negative control. It should be noted that the majority of these regions were located inside the highly transcribed class II genes. To correct for this nonrandom background distribution, we subtracted the normalized signal of untagged strain sample from the ChIP samples for each protein. The ChIP signals of untagged strain have different impacts on the ChIP signals for different proteins. For this reason, the subtraction step was preceded by a normalization of the ChIP signal of an untagged strain compared with the corresponding ChIP sample based on the qPCR analyses of nontranscribed control regions.

The significantly enriched regions were identified using peak calling MACS2 software version 2.0.10.20120703 (Zhang et al. 2008) with a minimum false discovery rate (q -value) cutoff of 0.01 for peak detection followed by PeakSplitter step of PeakAnalyzer software for subdivision of ChIP-seq regions into discrete peaks (Salmon-Divon et al. 2010). Significantly enriched Rad2 peaks were annotated on the yeast genome using ChIPpeakAnno R script from the Bioconductor project (Zhu et al. 2010). To identify the Rad2 enrichment signals inside protein-coding genes, Rad2 mean densities were computed for each ORF and compared with mean densities of two negative controls: the untagged strain sample, normalized as described above, and the input DNA sample, normalized using NCIS R script version 1.0.1 (Liang and Keles 2012). We required that enriched ORFs have the mean density in Rad2 sample higher than that in the untagged strain control plus two standard deviations and, at the same time, in the input DNA control plus two standard deviations.

Acknowledgments

We thank N. van Berkum for two-hybrid screening with the Med17 Mediator subunit, S. Wydau for yeast strains, the SPI (CEA/Saclay) for monoclonal antibodies, A. Polesskaya for advice in human cell culture, and C. Mann and S. Un for critical reading of the manuscript. This work has benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGIF (Centre de Recherche de Gif). This work was supported by the Agence Nationale de la Recherche (grant no. ANR_11_BSV8_020_01). F.E. was supported by a grant from the Fondation ARC (grant no. DOC20130606697).

References

- Andrau JC, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M, Holstege FC. 2006. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* **22**: 179–192.
- Cai G, Imasaki T, Yamada K, Cardelli F, Takagi Y, Asturias FJ. 2010. Mediator head module structure and functional interactions. *Nat Struct Mol Biol* **17**: 273–279.
- Clarkson SG. 2003. The XPG story. *Biochimie* **85**: 1113–1121.
- Compe E, Egly JM. 2012. TFIIF: When transcription met DNA repair. *Nat Rev Mol Cell Biol* **13**: 343–354.
- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M. 2008. Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol Cell* **31**: 337–346.
- Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, et al. 2008. CDK8 is a colorectal cancer oncogene that regulates β -catenin activity. *Nature* **455**: 547–551.
- Gade P, Singh AK, Roy SK, Reddy SP, Kalvakolanu DV. 2009. Down-regulation of the transcriptional mediator subunit Med1 contributes to the loss of expression of metastasis-associated dapk1 in human cancers and cancer cells. *Int J Cancer* **125**: 1566–1574.
- Gaillard H, Tous C, Botet J, Gonzalez-Aguilera C, Quintero MJ, Viladevall L, Garcia-Rubio ML, Rodriguez-Gil A, Marin A, Arino J, et al. 2009. Genome-wide analysis of factors affecting transcription elongation and DNA repair: A new role for PAF and Ccr4–not in transcription-coupled repair. *PLoS Genet* **5**: e1000364.
- Ghavi-Helm Y, Michaut M, Acker J, Aude JC, Thuriaux P, Werner M, Soutourina J. 2008. Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes Dev* **22**: 1934–1947.
- Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M. 2004. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* **32**: 5379–5391.
- Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**: 958–970.
- Hashimoto S, Boissel S, Zarhrate M, Rio M, Munnich A, Egly JM, Colleaux L. 2011. MED23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science* **333**: 1161–1163.
- Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**: 717–728.
- Ito S, Kuraoka I, Chymkowitz P, Compe E, Takedachi A, Ishigami C, Coin F, Egly JM, Tanaka K. 2007. XPG stabilizes TFIIF, allowing transactivation of nuclear receptors: Implications for Cockayne syndrome in XP-G/CS patients. *Mol Cell* **26**: 231–243.
- Johnson KM, Wang J, Smallwood A, Arayata C, Carey M. 2002. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev* **16**: 1852–1863.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435.
- Kaufmann R, Straussberg R, Mandel H, Fattal-Valevski A, Ben-Zeev B, Naamati A, Shaag A, Zenvirt S, Konen O, Mimouni-Bloch A, et al. 2010. Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am J Hum Genet* **87**: 667–670.
- Kornberg RD. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**: 235–239.
- Kuuselo R, Savinainen K, Sandstrom S, Autio R, Kallioniemi A. 2011. MED29, a component of the mediator complex, possesses both oncogenic and tumor suppressive characteristics in pancreatic cancer. *Int J Cancer* **129**: 2553–2565.
- Lagerwerf S, Vrouwe MG, Overmeer RM, Fousteri MI, Mullenders LH. 2011. DNA damage response and transcription. *DNA Repair (Amst)* **10**: 743–750.

- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lariviere L, Geiger S, Hoepfner S, Rother S, Strasser K, Cramer P. 2006. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* **13**: 895–901.
- Lariviere L, Plaschka C, Seizl M, Wenzek L, Kurth F, Cramer P. 2012. Structure of the Mediator head module. *Nature* **492**: 448–451.
- Lee SK, Yu SL, Prakash L, Prakash S. 2002. Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription: Implications for Cockayne syndrome. *Cell* **109**: 823–834.
- Le May N, Mota-Fernandes D, Velez-Cruz R, Iltis I, Biard D, Egly JM. 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.
- Le May N, Fradin D, Iltis I, Bougneres P, Egly JM. 2012. XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell* **47**: 622–632.
- Li LH, He J, Hua D, Guo ZJ, Gao Q. 2010. Lentivirus-mediated inhibition of Med19 suppresses growth of breast cancer cells in vitro. *Cancer Chemother Pharmacol* **68**: 207–215.
- Liang K, Keles S. 2012. Normalization of ChIP-seq data with control. *BMC Bioinformatics* **13**: 199.
- Nonet M, Scafe C, Sexton J, Young R. 1987. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol* **7**: 1602–1611.
- Peng J, Zhou JQ. 2012. The tail-module of yeast Mediator complex is required for telomere heterochromatin maintenance. *Nucleic Acids Res* **40**: 581–593.
- Ries D, Meisterernst M. 2011. Control of gene transcription by Mediator in chromatin. *Semin Cell Dev Biol* **22**: 735–740.
- Risheg H, Graham JM Jr, Clark RD, Rogers RC, Opitz JM, Moeschler JB, Peiffer AP, May M, Joseph SM, Jones JR, et al. 2007. A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat Genet* **39**: 451–453.
- Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. 2012. Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci* **109**: 17931–17935.
- Salmon-Divon M, Dvinge H, Tammoja K, Bertone P. 2010. PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**: 415.
- Sarker AH, Tsutakawa SE, Kostek S, Ng C, Shin DS, Peris M, Campeau E, Tainer JA, Nogales E, Cooper PK. 2005. Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: Insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell* **20**: 187–198.
- Scharer OD. 2008. XPG: Its products and biological roles. *Adv Exp Med Biol* **637**: 83–92.
- Schmitt ME, Brown TA, Trumpower BL. 1990. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* **18**: 3091–3092.
- Schwartz CE, Tarpey PS, Lubs HA, Verloes A, May MM, Risheg H, Friez MJ, Futreal PA, Edkins S, Teague J, et al. 2007. The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. *J Med Genet* **44**: 472–477.
- Soutourina J, Bordas-Le Floch V, Gendrel G, Flores A, Ducrot C, Dumay-Odelot H, Soularue P, Navarro F, Cairns BR, Lefebvre O, et al. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**: 4920–4933.
- Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**: 1451–1454.
- Svejstrup JQ. 2002. Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* **3**: 21–29.
- Svejstrup JQ. 2007. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci* **32**: 165–171.
- Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, Szutorisz H, Swanson SK, Martin-Brown S, Washburn MP, et al. 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**: 92–104.
- Vijayvargia R, May MS, Fondell JD. 2007. A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. *Cancer Res* **67**: 4034–4041.
- Wang G, Balamotis MA, Stevens JL, Yamaguchi Y, Handa H, Berk AJ. 2005. Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Mol Cell* **17**: 683–694.
- Werner M, Chaussivert N, Willis IM, Sentenac A. 1993. Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem* **268**: 20721–20724.
- Zhang X, Krutchinsky A, Fukuda A, Chen W, Yamamura S, Chait BT, Roeder RG. 2005. MED1/TRAP220 exists predominantly in a TRAP/Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol Cell* **19**: 89–100.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.
- Zhu X, Wiren M, Sinha I, Rasmussen NN, Linder T, Holmberg S, Ekwall K, Gustafsson CM. 2006. Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell* **22**: 169–178.
- Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR. 2010. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**: 237.
- Zhu X, Liu B, Carlsten JO, Beve J, Nystrom T, Myers LC, Gustafsson CM. 2011. Mediator influences telomeric silencing and cellular life span. *Mol Cell Biol* **31**: 2413–2421.

Supplemental Material

Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment

Fanny Eyboulet, Camille Cibot, Thomas Eychenne, Helen Neil, Olivier Alibert, Michel Werner and Julie Soutourina

Supplemental Materials and Methods

Strains and plasmids

All *S. cerevisiae* strains are described in **Table S1**. They all derive from the YPH499/YPH500 standard strains (Sikorski and Hieter 1989) except for BY4741 and BY4741 *rad2* Δ that have been obtained from Euroscarf collection (<http://web.uni-frankfurt.de/fb15/mikro/euroscarf>). YPH499 *rad2* Δ strain was obtained by deleting a chromosomal *RAD2* copy with a *rad2* Δ ::*KanMX6* cassette. Yeast strains carrying N-terminal (HA-Rad2) or C-terminal HA-tagged version of Rad2 (Rad2-HA) were obtained by insertion of 3HA epitopes preceded or followed by *His3MX6* marker, respectively, using the standard one-step methods (Longtine et al. 1998; DeMarini et al. 2001). Double-tagged HA-Rad2 Med17-Myc strain was obtained by inserting 13Myc epitopes followed by *TRP1* marker using the standard one-step method in HA-Rad2 strain. Med5-HA, Med17-HA, Med15-HA, Rad3-HA and Kin28-HA strains carrying C-terminal HA-tagged version of Med5, Med17 and Med15 Mediator subunits, Rad3 and Kin28 TFIIH subunits were constructed by inserting 3HA epitopes followed by *His3MX6* marker. The HA-TBP strain expressing N-terminal HA-tagged version of TBP was obtained by inserting 3HA epitopes preceded by *LEU2* marker. The *rpb1-1* strain was constructed by standard meiotic crosses, then N-terminal 3HA-tag on Rad2 and C-terminal 13Myc-tag on Rpc160 Pol III subunit were inserted by the standard one-step method and the resulting strain was transformed by Yep351 (2 μ LEU2) or pYeB-B220 (RPB1 2 μ LEU2) plasmids to obtain HA-Rad2 Rpc160-Myc *rpb1-1* strain or the isogenic wild-type strain, respectively. The Med5-HA strain carrying C-terminal HA-tagged version of

Med5 Mediator subunit was constructed by insertion of 3HA epitopes followed by *His3MX6* marker. The Med5-HA *rad2* Δ strain was obtained by deleting the chromosomal *RAD2* copy with a *rad2* Δ ::*KanMX6* cassette and by inserting 3HA epitopes followed by *His3MX6* marker on *MED5* gene. *med17* conditional mutants were obtained as previously described (Soutourina et al. 2011). The *rad7* Δ strain was obtained by deleting the chromosomal *RAD7* copy with a *rad7* Δ ::*HIS3* cassette. The *rad26* Δ strain was obtained by deleting the chromosomal *RAD26* copy with a *rad26* Δ ::*KanMX6* cassette. The *rad26* Δ *rad7* Δ double mutant was obtained by deleting the chromosomal *RAD26* copy with a *rad26* Δ ::*KanMX6* cassette and then by deleting the chromosomal *RAD7* copy with a *rad7* Δ ::*HIS3* cassette. Double-tagged HA-Rad2 Med5-Myc strain was obtained first by deleting the chromosomal *MED17* copy with a *med17* Δ ::*ADE2* cassette in the strain complemented by wild-type *MED17* borne on a *URA3* vector (pVV208-MED17), then by insertion of 13Myc epitopes followed by *KanMX6* marker using the standard one-step method. *med17* conditional mutants in *rad7* Δ , *rad26* Δ , HA-Rad2, Med5-HA and HA-Rad2 Med5-Myc contexts were obtained by transformation with *TRP1* plasmids (pVV204-med17ts) of the corresponding strains carrying a *med17* deletion complemented by wild-type *MED17* borne on a *URA3* vector (pVV208-MED17) followed by a counter-selection of pVV208-MED17 plasmid on 5-FOA medium. YPH499 *URA3 rad2* Δ strain was obtained by deleting a chromosomal *RAD2* copy with a *rad2* Δ ::*HIS3MX6* cassette in a strain carrying a *med17* deletion complemented by wild-type *MED17* borne on a *URA3* vector (pVV208-MED17).

Unless otherwise stated, yeast strains were grown at 30°C in YPD rich medium containing 2% of glucose. For galactose induction, yeast cells were grown in YPRaff rich medium containing 2% of raffinose and galactose was then added to reach 2% final concentration. Growth phenotype assays are described in legend of **Figure S9**. UV sensitivity spot assays and UV survival assays are described in legends of **Figures 6, S10 and S11**.

Supplemental Tables

Table S1. Yeast strains

Name	Tagged or mutant protein	Genotype	Reference
Y190		<i>MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3,112 URA3::GAL1::lacZ LYS2::GAL4(UAS)::HIS3 cyhRY</i>	(Harper et al. 1993)
YPH499	WT	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1</i>	(Sikorski and Hieter 1989)
YPH499 rad2Δ	Rad2Δ	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 rad2::KanMX6</i>	This work
BY4741	WT	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	Euroscarf
BY4741 rad2Δ	Rad2Δ	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad2Δ::KanMX4</i>	Euroscarf
Y5386	HA-Rad2	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // * MED17 CEN URA3</i>	This work
Y5387	Rad2-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD2::3HA::HIS3 med17::KanMX6 // MED17 CEN URA3</i>	This work
Y5592	HA-Rad2	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2</i>	This work
Y5781	HA-Rad2 Med17-Myc	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED17::13Myc::TRP1</i>	This work
Y5540	Med15-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5551	HA-TBP	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 LEU2::3HA::SPT15 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5557	Rad3-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 RAD3::3HA::HIS3 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5545	Kin28-HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 KIN28::3HA::HIS3 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5084	WT	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5787	Rpb1 HA-Rad2 Rpc160-Myc	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1 HIS3::3HA::RAD2 RPC160::13Myc::KanMX6 // RPB1 2μ LEU2</i>	This work
Y5788	Rpb1-1 HA-Rad2 Rpc160-Myc	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1 HIS3::3HA::RAD2 RPC160::13Myc::KanMX6 // 2μ LEU2</i>	This work
Y5588	Med5-HA WT	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3</i>	This work
Y5596	Med5-HA Rad2Δ	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS3 rad2Δ::KanMX6</i>	This work
Y5321	Med17-257	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5327	Med17-L546P	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-L546P CEN TRP1</i>	This work
Y5309	Med17-140	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-140 CEN TRP1</i>	This work
Y5315	Med17-233	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // med17-233 CEN TRP1</i>	This work
Y5886	Rad7Δ Med17 Ura3	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // MED17 CEN URA3</i>	This work
Y5919	Rad7Δ Med17	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // MED17 CEN TRP1</i>	This work
Y5937	Rad7Δ Med17-257	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // med17-257 CEN TRP1</i>	This work
Y5943	Rad7Δ Med17-L546P	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // med17-L546P CEN TRP1</i>	This work

Y5925	Rad7Δ Med17-140	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // med17-140 CEN TRP1</i>	This work
Y5931	Rad7Δ Med17-233	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad7::HIS3 // med17-233 CEN TRP1</i>	This work
Y5879	Rad2-HA Med17	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5880	Rad2-HA Med17-257	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5881	Rad2-HA Med17-L546P	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // med17-L546P CEN TRP1</i>	This work
Y6137	Rad2-HA Med17-140	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // med17-140 CEN TRP1</i>	This work
Y6138	Rad2-HA Med17-233	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 med17::KanMX6 // med17-233 CEN TRP1</i>	This work
Y5485	Med5-HA Med17	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS med17::KanMX6 // MED17 CEN TRP1</i>	This work
Y5489	Med5-HA Med17-257	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS med17::KanMX6 // med17-257 CEN TRP1</i>	This work
Y5496	Med5-HA Med17-L546P	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED5::3HA::HIS med17::KanMX6 // med17-L546P CEN TRP1</i>	This work
Y6095	HA-Rad2 Med5-Myc Med17	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED5::13Myc::KanMX6 med17::ADE2 // MED17 CEN TRP1</i>	This work
Y6097	HA-Rad2 Med5-Myc Med17-257	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED5::13Myc::KanMX6 med17::ADE2 // med17-257 CEN TRP1</i>	This work
Y6096	HA-Rad2 Med5-Myc Med17-L546P	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED5::13Myc::KanMX6 med17::ADE2 // med17-L546P CEN TRP1</i>	This work
Y6140	HA-Rad2 Med5-Myc Med17-140	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED5::13Myc::KanMX6 med17::ADE2 // med17-140 CEN TRP1</i>	This work
Y6141	HA-Rad2 Med5-Myc Med17-233	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 HIS3::3HA::RAD2 MED5::13Myc::KanMX6 med17::ADE2 // med17-233 CEN TRP1</i>	This work
Y4945	WT Ura3	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 // MED17 CEN URA3</i>	This work
Y5605	Rad2Δ Ura3	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::KanMX6 rad2::HIS3 // MED17 CEN URA3</i>	This work
Y6244	Rad26Δ	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 rad26::KanMX6</i>	This work
Y6245	Rad7Δ	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 rad7::HIS3</i>	This work
Y6247	Rad26Δ Rad7Δ	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 rad26::KanMX6 rad7::HIS3</i>	This work
Y6246	Rad26Δ Med17 Ura3	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 med17::ADE2 rad26::KanMX6 // MED17 CEN URA3</i>	This work
Y5863	Med17-HA Med10	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED17::3HA::HIS3 med10::KanMX6 // MED10 CEN TRP1</i>	This work
Y6298	Med17-HA	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED17::3HA::HIS3</i>	This work
Y6299	Med15-HA	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3</i>	This work
Y6301	Med17-HA Rad2Δ	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED17::3HA::HIS3 rad2::KanMX6</i>	This work
Y6302	Med15-HA Rad2Δ	<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 MED15::3HA::HIS3 rad2::KanMX6</i>	This work
Y6305	Rpb1 Med17-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1 MED17::3HA::HIS3 // RPBI 2μ LEU2</i>	This work

Y6306	Rpb1-1 Med17-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1</i> <i>MED17::3HA::HIS3 // 2μ LEU2</i>	This work
Y6307	Rpb1 Med15-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1</i> <i>MED15::3HA::HIS3 // RPB1 2μ LEU2</i>	This work
Y6308	Rpb1-1 Med15-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1</i> <i>MED15::3HA::HIS3 // 2μ LEU2</i>	This work
Y6309	Rpb1 Med5-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1</i> <i>MED5::3HA::HIS3 // RPB1 2μ LEU2</i>	This work
Y6310	Rpb1-1 Med5-HA	<i>MATα ura3-52 his3-Δ200 trp1-Δ63 leu2-Δ1 rpb1-1</i> <i>MED5::3HA::HIS3 // 2μ LEU2</i>	This work

* Plasmid descriptions are separated from chromosomal genotype by //.

Table S2. Plasmids

Name	Genotype	Reference
Yep351	<i>Amp 2μ LEU2</i>	(Hill et al. 1986)
pYeB-B220	<i>Amp 2μ LEU2 RPB1</i>	(Voutsina et al. 1999)
pVV208-MED17	<i>Amp CEN URA3 MED17</i>	(Soutourina et al. 2011)
pVV204-MED17	<i>Amp CEN TRP1 MED17</i>	(Soutourina et al. 2011)
pVV204-med17-257	<i>Amp CEN TRP1 med17-L520S, I541K, E669G</i>	(Soutourina et al. 2011)
pVV204-med17-L546P	<i>Amp CEN TRP1 med17-L546P</i>	(Soutourina et al. 2011)
pVV204-med17-140	<i>Amp CEN TRP1 med17-M442L, Q444P</i>	(Soutourina et al. 2011)
pVV204-med17-233	<i>Amp CEN TRP1 med17-K387I, A450T, L674P</i>	(Soutourina et al. 2011)

Table S3. Oligonucleotides

Name	Forward	Reverse
<i>GAL1 P</i>	ACGCTTAACTGCTCATTGCT	TGTTCCGAGCAGTGC GGCGC
<i>GAL1 O1</i>	AAAGAAACTTGCACCGGAAA	GGCCCATATTCGCTTTAACA
<i>GAL1 O2</i>	ACATTTCCACACCCTGGAAC	TTCTTCGCGAGAACAATTCA
<i>IGV</i>	TTCATTTGCAATTTGCAGTTCA	CAGCCAGGAAGAATCTCACAA
<i>ADH1 P1</i>	ATAGGCGCATGCAACTTCTT	CATCAGCTCTGGAACAACGA
<i>ADH1 P2</i>	TTCCTTCAATCACGCACACT	AGGGAACGAGAACAATGACG
<i>ADH1 O</i>	GGGTATTGACGGTGGTGAAG	AAACGTTGATGACACCGTGA
<i>PYK1 P1</i>	CGCACCGTCACAAAGTGTT	TGGGAAGGAAAGGAAATCAC
<i>PYK1 P2</i>	CCTTTCCTTCCCATATGATGC	ACTTTGAAAGGGGACCATGA
<i>PYK1 O</i>	ATGGTTGCCAGAGGTGACTT	TCTGGTTGGTCTTGGGTTGT
<i>PMA1 P1</i>	AACAAACCCGGTCTCGAAG	GAAGTGCCGCATTAGGAAAT
<i>PMA1 P2</i>	GATGGTGGGTACCGCTTATG	TTGGTGTATAGGAAAGAAAGAGAAAA
<i>PMA1 O</i>	ATGGATCCACCAAGAGACGA	TGTTGGTACCCAAACCCAAT
<i>RPR1</i>	ATGGTACGCTGTGGTGCTC	CCATAGGTGGGGATCCTTCT
<i>tDNAMet</i>	GCTTCAGTAGCTCAGTAGGAA	TGCTCCAGGGGAGGTTTC
<i>tDNAPhe</i>	GACGCTTGGACCATTTATAAAGCAC	CCATAAGAGAAGGAGCAGTCAAGTTCA
<i>5S</i>	TCGCGTATGGTCAACCACTAC	GTTGCGCCATATCTACCAGA
<i>NTS2</i>	CCGTTATTGGTAGGAGTGTGGTG	ACGGAAATACGCTTCAGAGACC
<i>35S-P</i>	GTGTGAGGAAAAGTAGTTGGGAGGTA	GACGAGGCCATTTACAAAAACATAAC
<i>35S-1430</i>	GCGAACCAGGACTTTTACTTTGA	AACAAAAATAGAACCAAACGTCTATTCC
<i>35S-3551</i>	GGTGGTAAATTCATCTAAAGCTAAATATT	CACGTACTTTTTCACTCTCTTTTCAAAA
<i>35S-4925</i>	GGAGGAGTTATCTTTTCTTCTAACAGCT	AAGGTGCTGGCCTCTTCCA
<i>SCR1-1</i>	CCAGGACATCCATAGCTTGTG	ATGAAAAGTTCTGGCGATG
<i>SCR1-2</i>	CATCGCCAGGAACCTTTTCAT	ACAGCCTAGCACAATTGGAA
<i>SCR1-3</i>	CTTTCTGGTGGGATGGGATA	TTTACGACGGAGGAAAGACG
<i>SCR1-4</i>	GTCCTGGGCAGAGCTGTCT	AAGGTGGAGCCCTAAGGA
<i>SCR1-5</i>	ACCGCTGTTAGGGGAGTTTT	CCAAATTAACCGCCGAAG
<i>SCR1-6</i>	CGGTGCCATCAGGATTTACT	CTCCAACATCCCTCATTGG
<i>SCR1-7</i>	TTTTCGAATATAAATGACGATTGG	TGTCGCTACTCACTCTACAACCA

<i>TEL1R</i>	GTGGAGTGGGGGAATGAGAC	TCAACTACCCTCCCTCTCCAT
<i>TEL12L XR</i>	ACCCCTAATCCAACCCTGAT	GGGAGTGGTACGGTTGAATG
<i>TELR YP2</i>	GGTTGAACATCCGGGTAAAGA	CGGCTTTCTGACCGAAATTA
<i>TEL1R XC1</i>	ACGTGGTAGATGGGGATTGT	CATCTATCCCCTGCCCAATA
<i>TEL1L XR</i>	AATCTAACCCCTGGCCAACCT	TGGTTCGGAGTGGTATGGTT
<i>CLN1 P1</i>	ATTCCTTGTTCGCAACACT	CGCGGGGTTGTAGTAGGTAA
<i>CLN1 P2</i>	GGTGGCAGTGCAGCAATTA	AGCGAACGAAGAACTCCTTG
<i>CLN1 O</i>	GATGCCACAGTGGAGAACA	TTGCCAACAGCTCAGATCAA
<i>TPI1 P</i>	TGGTGGAAAGATTACCCGTTT	CGAGGTGGTATCCGAGAGAA
<i>TPI1 O</i>	TAACGTCGTTGTGCTTACG	CCAACCTGGAAGCCAAGAAC
<i>PCL1 P1</i>	GGCTCTAGTGTCTGCGAATG	ACATTTTCGAGTCGCTTTTGC
<i>PCL1 P2</i>	CTACAAGGCTGTGGGCAGAT	GGATGATTGCTGATGTACGC
<i>PCL1 O</i>	GTCGCCATCGCTAATGACTT	AGAAAATGCGATGGATGTC
<i>PDR5 P1</i>	AAGACTGCCCTCTCTTTCC	ACGGAATCACGAGTCCAAAG
<i>PDR5 P2</i>	GTCACGCAAAGTTGCAAACA	GTCTTCGAACGAGCGGATA
<i>PDR5 O</i>	CTCCAGGCTATGACCCAAAA	GTTCTTCCAAGCGCAACCTA
<i>SNQ2 P1</i>	TCACCCCATTTGAAAAAGAA	TAGTCATATGTGCGCGGAAC
<i>SNQ2 P2</i>	GCCCATTTCCGTTTAAATCC	GAGGGAGGGGCACTACTCAT
<i>SNQ2 O</i>	GTTTCAGAGGCGATCAATGGT	ATCGAAGGCAATTGGATCAT
<i>RAD2 P</i>	AGTGGTGTGACCCCAATACC	ATATGTGGCGGCATCTTTTC
<i>RAD2 O</i>	TGAGTCTTCTAACGCGACGA	TAAGGCAACCTTGACGTTT
<i>ADR1 P</i>	CCGGAATTTACCAACGAAGA	GTTACACAACGCAAGCACA
<i>ADR1 O</i>	TGCCAGCTGTGAGTGAATTG	GCGTTTTTCAGCGTCATCATA

Table S4. Mapped tag total numbers for ChIP-seq experiments

Sample	Million of mapped tags*
Rad2	4.6
Med15 (Mediator)	1.84
Med17 (Mediator)	3.38
Pol II	6.62
TBP	1.64
Kin28 (TFIIH kinase module)	2.8
Rad3 (core TFIIH)	4.54
Input	5.55
Untagged	1.65

* When multiple reads mapped to the same genomic position, a maximum of 5 duplicated reads were considered to remove possible sequencing artefacts.

Supplemental References

- DeMarini DJ, Carlin EM, Livi GP. 2001. Constitutive promoter modules for PCR-based gene modification in *Saccharomyces cerevisiae*. *Yeast* **18**: 723-728.
- Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**: 805-816.
- Hill JE, Myers AM, Koerner TJ, Tzagoloff A. 1986. Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. *Yeast* **2**: 163-167.
- Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**: 953-961.
- Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**: 19-27.

- Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**: 1451-1454.
- Voutsina A, Riva M, Carles C, Alexandraki D. 1999. Sequence divergence of the RNA polymerase shared subunit ABC14.5 (Rpb8) selectively affects RNA polymerase III assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **27**: 1047-1055.

Supplemental Figure legends

Figure S1. Coimmunoprecipitation of Mediator and Rad2 with IgG and untagged strain controls.

Mediator was immunoprecipitated through Med17-Myc subunit with α -Myc antibody from crude yeast extracts (Input, right panel) and analyzed by Western blotting with α -HA antibody (CoIP) against Rad2 (left panel). The presence of HA- or Myc-tagged alleles is indicated above each lane. IgG indicates control IP with IgG magnetic beads only. *HA-RAD2* strain was used as untagged control for immunoprecipitation with α -Myc antibody.

Figure S2. ChIP analysis of Rad2 occupancy on selected class II genes, class III genes, telomeric regions and rDNA locus.

(A) Quantitative ChIP analysis of Rad2 occupancy on selected constitutively expressed class II genes. Two yeast strains carrying N-terminal (HA-Rad2) or C-terminal HA-tagged versions of Rad2 (Rad2-HA) were grown in YPD medium at 30°C. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to selected class II gene promoters (*P*) or ORFs (*O*). A *GAL1* ORF and a nontranscribed region on chromosome V (*IGV*) were used as negative controls.

(B, C, D, E, F) Quantitative ChIP experiments were performed with α -HA antibody against Rad2. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. The yeast strain carrying N-terminal HA-tagged version of Rad2 was grown in YPD medium at 30°C. Immunoprecipitated fragments from ChIP experiments were

amplified with primers corresponding to selected regions. Background is represented as a dotted line. A *GALI* ORF (**B**, **C**, **D**, **E**) or the intergenic *NTS2* region (**F**) were used as negative controls.

(B) Quantitative ChIP analysis of Rad2 occupancy on selected class II gene promoters and ORFs. Representative class II gene promoters (*P*) and ORFs (*O*) were selected for conventional ChIP experiments based on genome-wide analyses.

(C) Rad2 enrichment on selected class III genes.

(D) Rad2 enrichment on the *SCRI* Pol III-transcribed gene. The location of the PCR fragments amplified in ChIP analyses is indicated by gray and black lines. Black boxes represent TATA, A and B boxes, and Terminator (T).

(E) Rad2 enrichment on selected telomeric regions.

(F) Rad2 enrichment on rDNA locus. Schematic organization of rDNA locus with *5S* and *35S* rDNA genes is shown. The location of the PCR fragments amplified in ChIP analyses is indicated by black lines.

Figure S3. Genome-wide analysis of Rad2, Mediator, Pol II, TBP and TFIIH.

Examples of Rad2, Mediator, Pol II, TBP and TFIIH ChIP-seq enrichment profiles on selected Pol III-transcribed genes. Densities of sequence tags were assessed from ChIP-seq experiments performed with HA-Rad2, Med15-HA (Mediator), Med17-HA (Mediator), HA-TBP, Rad3-HA (core TFIIH) and Kin28-HA (TFIIH kinase module) tagged strains with α -HA antibody. Pol II was immunoprecipitated using α -Rpb1 antibody. ChIP-seq density profiles are displayed using IGB yeast genome browser. Watson strand-transcribed genes are colored in blue and Crick strand-transcribed genes are colored in green. Input DNA and DNA from ChIP with untagged strain were used as negative controls. Densities of sequence tags were displayed after subtraction of normalized control of untagged strain as described in

Materials and Methods.

Figure S4. Correlation of Med17 ChIP-seq density with those of Med15 Mediator subunit and Rad2.

(A) Rad2 ChIP-seq density versus Med17 Mediator subunit ChIP-seq density on total yeast genome. Density of sequence tags in ChIP-seq experiments for Rad2 and Med17 has been calculated for each 25-bp bin of yeast genome. Each point on the plot corresponds to one genomic bin. The genomic bins located inside ORFs or class II promoter regions are indicated in green and blue, respectively. Pol III-transcribed genes were excluded from this analysis.

(B) Rad2 ChIP-seq density versus Med17 Mediator subunit ChIP-seq density on class II promoter regions. Density of sequence tags in ChIP-seq experiments for Rad2 and Med17 has been calculated for promoter region of each Pol II-transcribed gene. Each point on the plot corresponds to one promoter region. A linear regression for Rad2 ChIP-seq density versus Med17 ChIP-seq density (red line) and a R^2 correlation coefficient are indicated.

(C) Med17 Mediator subunit ChIP-seq density versus Med15 Mediator subunit ChIP-seq density on total yeast genome, as described in (A).

(D) Med17 ChIP-seq density versus Med15 ChIP-seq density on class II promoter regions, as described in (B).

Figure S5. Correlation of Kin28 TFIIH subunit ChIP-seq density with those of Rad3 TFIIH subunit and Rad2.

(A) Kin28 TFIIH kinase subunit ChIP-seq density versus Rad3 core TFIIH subunit ChIP-seq density on total yeast genome. Density of sequence tags in ChIP-seq experiments for two TFIIH subunits (Kin28 and Rad3) has been calculated for each 25-bp bin of yeast genome. Each point on the plot corresponds to one genomic bin. The genomic bins located inside

ORFs or class II promoter regions are indicated in green and blue, respectively. Pol III-transcribed genes were excluded from this analysis.

(B) Kin28 ChIP-seq density versus Rad3 ChIP-seq density on class II promoter regions. Density of sequence tags in ChIP-seq experiments for two TFIIH subunits (Kin28 and Rad3) has been calculated for promoter region of each Pol II-transcribed gene. Each point on the plot corresponds to one promoter region. A linear regression for Kin28 ChIP-seq density versus Rad3 ChIP-seq density (red line) and a R^2 correlation coefficient are indicated.

(C) Rad2 ChIP-seq density versus Kin28 TFIIH kinase subunit ChIP-seq density on total yeast genome, as described in **(A)**.

(D) Rad2 ChIP-seq density versus Kin28 ChIP-seq density on class II promoter regions, as described in **(B)**.

Figure S6. Effect of *rpb1-1* mutation on Rad2, Pol II and Mediator occupancies on selected regions.

Quantitative ChIP assays were performed using α -HA antibody against HA-Rad2 **(A)**, α -Rpb1 antibody (Pol II) **(B)** and α -HA antibody against HA-tagged Med17, Med15 and Med5 Mediator subunits **(C, D, E)**. Cells were grown in selective SD medium complemented with amino acids at 25°C and then shifted for 90 min at 37°C. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to selected class II gene promoters (*P*) or ORFs (*O*), or selected class III genes or telomeric regions. A *GAL1* ORF and a nontranscribed region on chromosome V (*IGV*) were used as negative controls.

Figure S7. Mediator and Pol II occupancies in *rad2* deletion strain on selected regions.

Quantitative ChIP analysis of Mediator and Pol II occupancies under standard growth conditions. *rad2Δ* and wild-type strains were grown in YPD medium at 30°C. Immunoprecipitations were performed using α -HA antibody against Med17-HA (Mediator) (A), Med15-HA (Mediator) (B) and α -Rpb1 antibody (Pol II) (C). Immunoprecipitated fragments from Mediator ChIP experiments were amplified with primers corresponding to selected class II gene promoters (A, B). Immunoprecipitated fragments from Pol II ChIP experiments were amplified with primers corresponding to selected class II gene promoters (P) or ORFs (O) (C). Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. A *GALI* ORF was used as a negative control.

Figure S8. Kinetics of galactose induction in *rad2* deletion and wild-type strains.

rad2Δ and wild-type strains were grown in a raffinose-supplemented medium at 30°C, then galactose was added and samples were collected for ChIP or quantitative RT-PCR experiments at indicated time points upon galactose induction (0, 10, 20, 30 and 60 min).

(A) Quantitative ChIP analysis of Pol II occupancy was performed using α -Rpb1 antibody. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to *GALI* gene promoter (*GALI P*) or ORF (*GALI O1*, *O2*). A nontranscribed region on chromosome V (*IGV*) was used as a negative control.

(B) *GALI* mRNA level was quantified by RT-qPCR and normalized to 25S rRNA, 18S rRNA and *SCR1* levels. Values are expressed in arbitrary units representing a percentage of the value in the wild-type strain. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown.

Figure S9. Growth phenotypes of *rad2* deletion strains.

Wild-type and isogenic *rad2* deletion strains in BY4741 and YPH499 contexts were serially diluted, spotted on indicated agar plates, and incubated for 3 days at 30°C. For UV-sensitivity assays, cells were treated or not with 40J/m² of UV (UV Stratalinker 1800). Rich media supplemented with 2 or 0.05% glucose (YPD and YPD 0.05%, respectively), 1 or 2% ethanol (YPE 1% and 2%, respectively), 2% galactose (YPGal) or 2% glycerol (YPG), YPD medium containing indicated concentrations of cycloheximide (CHX) or ketoconazole (KET), minimal SC medium containing indicated concentrations of mycophenolic acid (MPA), SD medium supplemented with casamino acids, adenine and tryptophan (uracil-lacking CAW medium) containing indicated concentrations of 6-azauracil were used. The only observed phenotype of *rad2* deletion strains (UV sensitivity) is highlighted in red.

Figure S10. Growth of *med17* mutants after UV treatment.

UV sensitivity of *med17* mutants was assayed in a context where both NER pathways, GGR and TCR, were functional. Cells were serially diluted, spotted on YPD agar plates, treated or not with indicated doses of UV (UV Stratalinker 1800), and incubated for 3 days at 30°C.

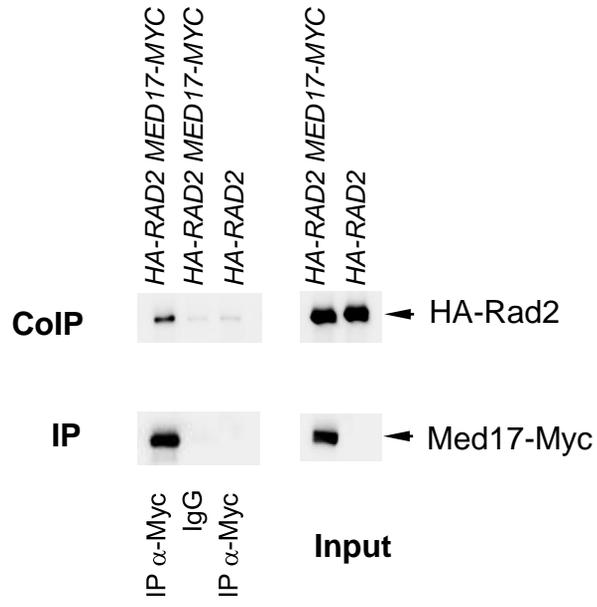
Figure S11. Epistatic analysis of *med17* mutants with a TCR-deficient mutant (*rad26*Δ).

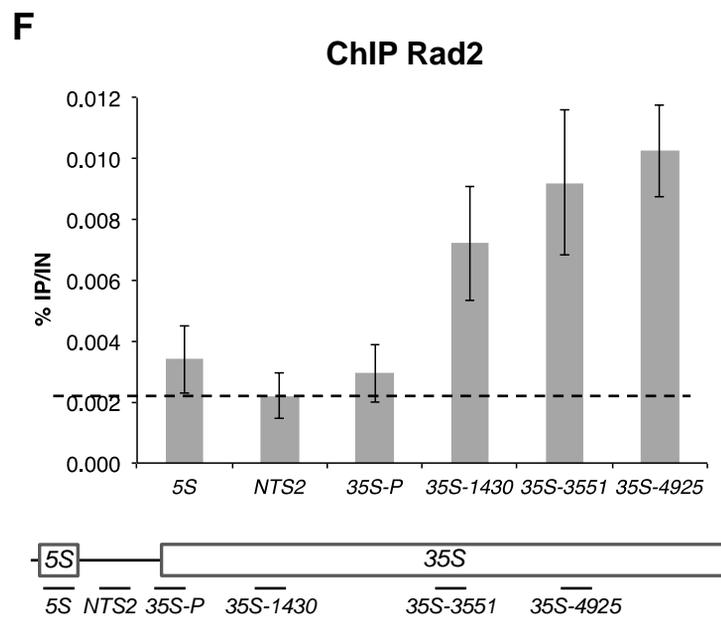
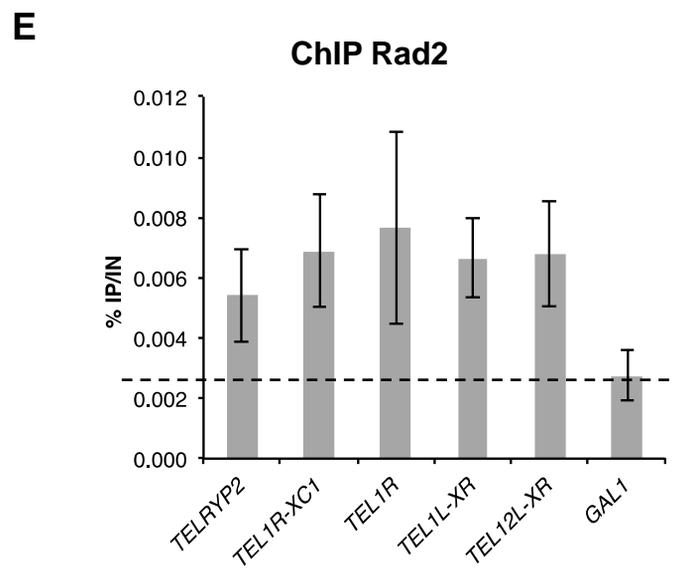
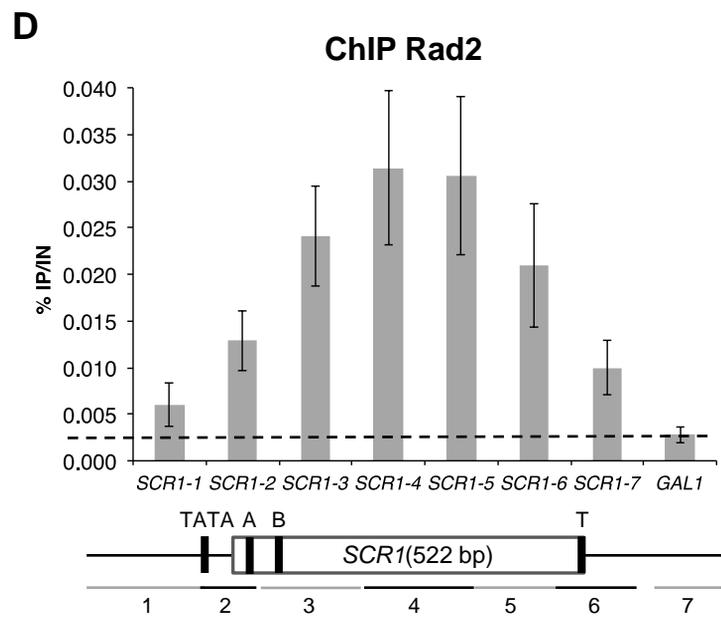
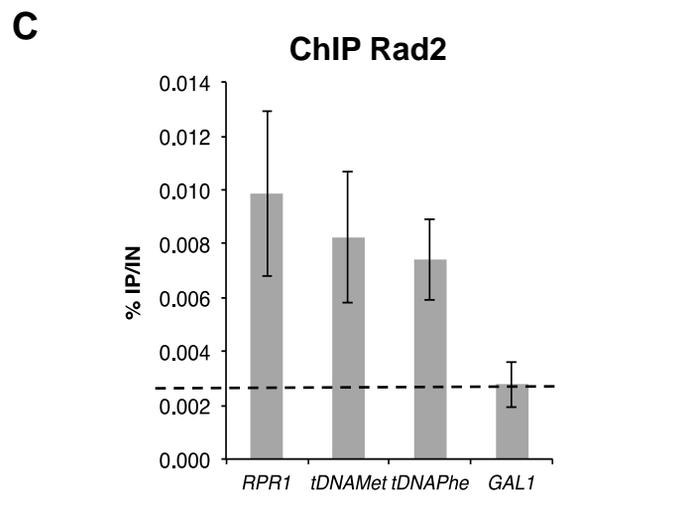
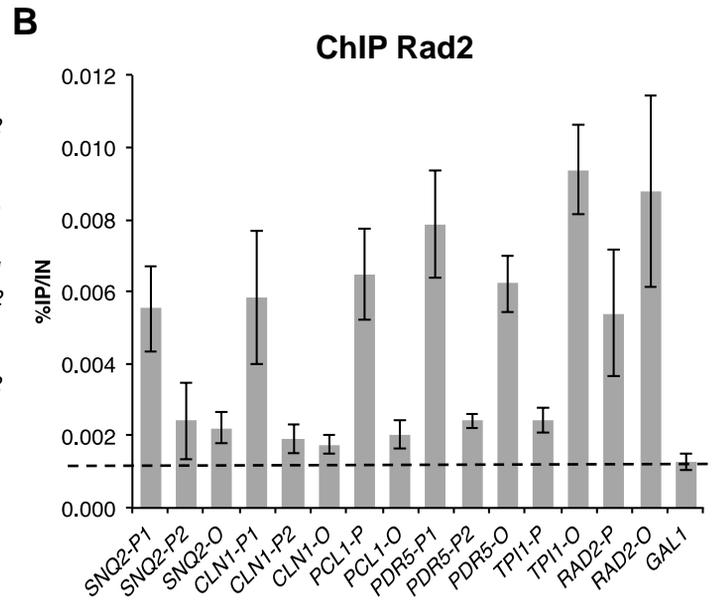
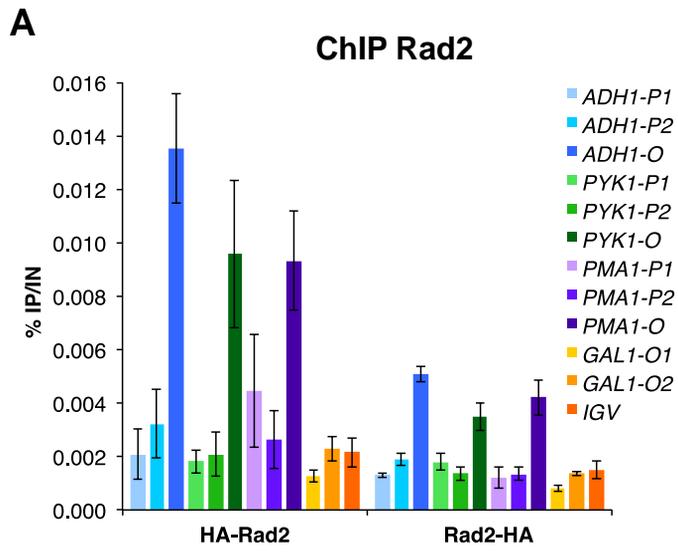
UV sensitivity of *med17* mutants was assayed in a context where the TCR pathway was deficient. Cells were serially diluted, spotted on YPD agar plates, treated or not with indicated doses of UV (UV Stratalinker 1800), and incubated for 3 days at 30°C.

Figure S12. Effects of UV-sensitive *med17* mutants on Rad2, Pol II and Mediator occupancies, and on Rad2-Mediator interaction.

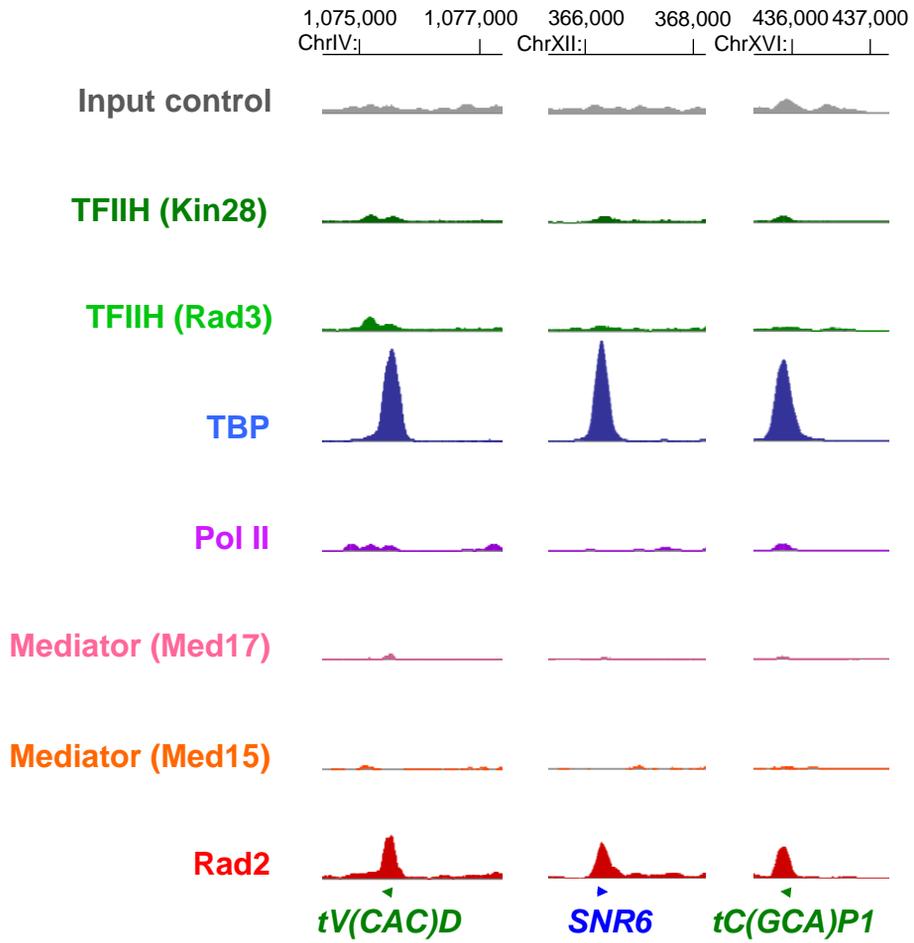
(A, B, C) Quantitative ChIP analysis of Rad2, Pol II and Mediator occupancies on selected regions. ChIP assays were performed using α -HA antibody against HA-Rad2 (**A**), α -Rpb1 antibody (Pol II) (**B**), and α -HA antibody against Med5-HA (Mediator) (**C**). Cells were grown in YPD medium at 30°C and then shifted for 45 min at 37°C. Immunoprecipitated fragments from ChIP experiments were amplified with primers corresponding to selected class II gene promoters (*P*) or ORFs (*O*), to a selected class III gene and a telomeric region. Mean values and standard deviation (indicated by error bars) of three independent experiments are shown. A *GALI* ORF and a nontranscribed region on chromosome V (*IGV*) were used as negative controls.

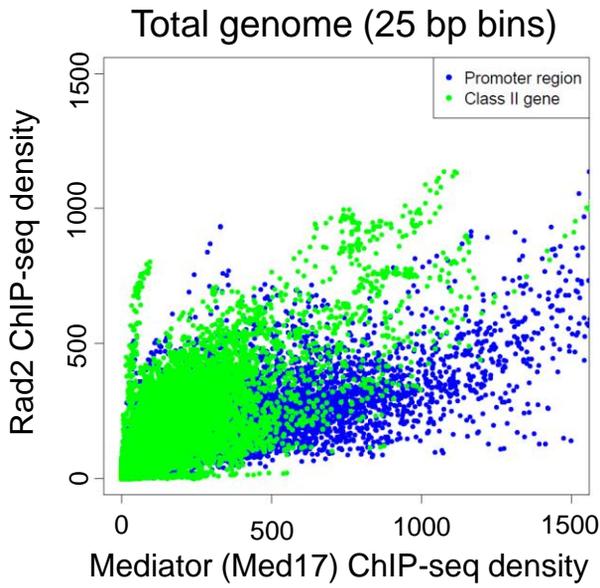
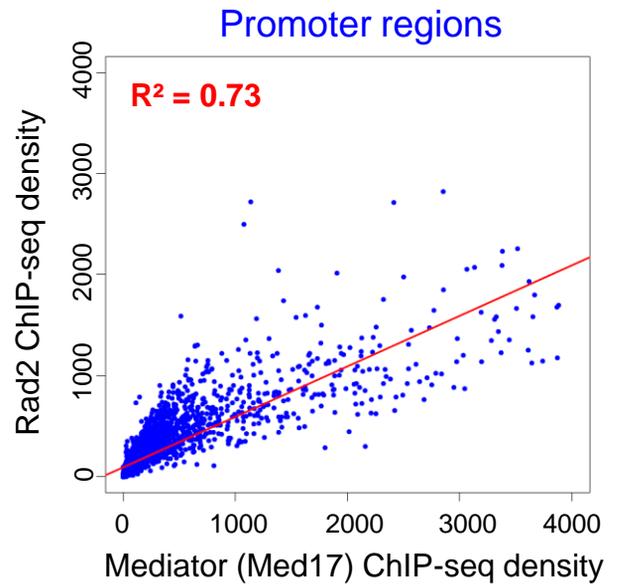
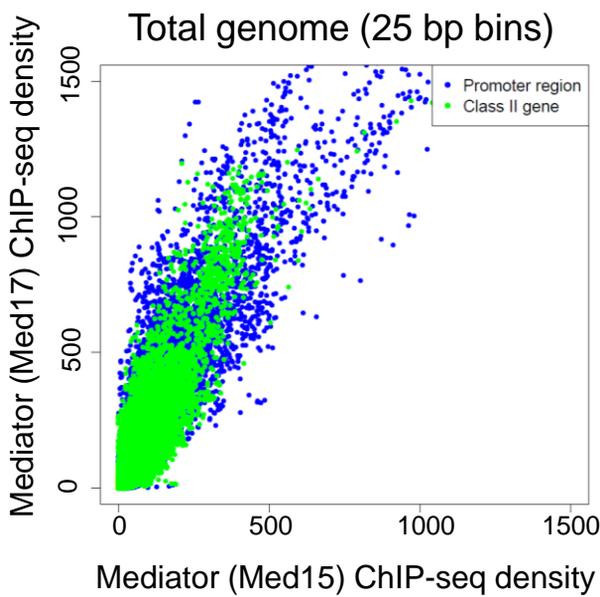
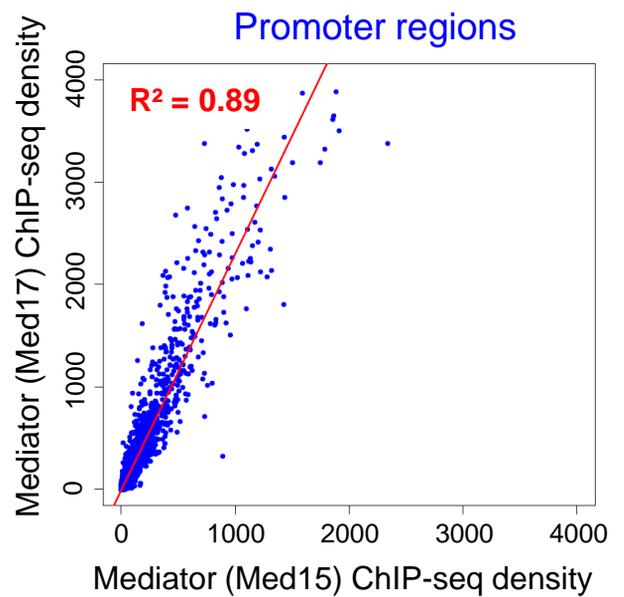
(D) Coimmunoprecipitation of Mediator and Rad2. Mediator was immunoprecipitated through Med5-Myc subunit with α -Myc antibody from crude extracts (Input, right panel) and analyzed by Western blotting with α -HA antibody (CoIP) against Rad2 (left panel). IgG indicates control IP with IgG magnetic beads only.

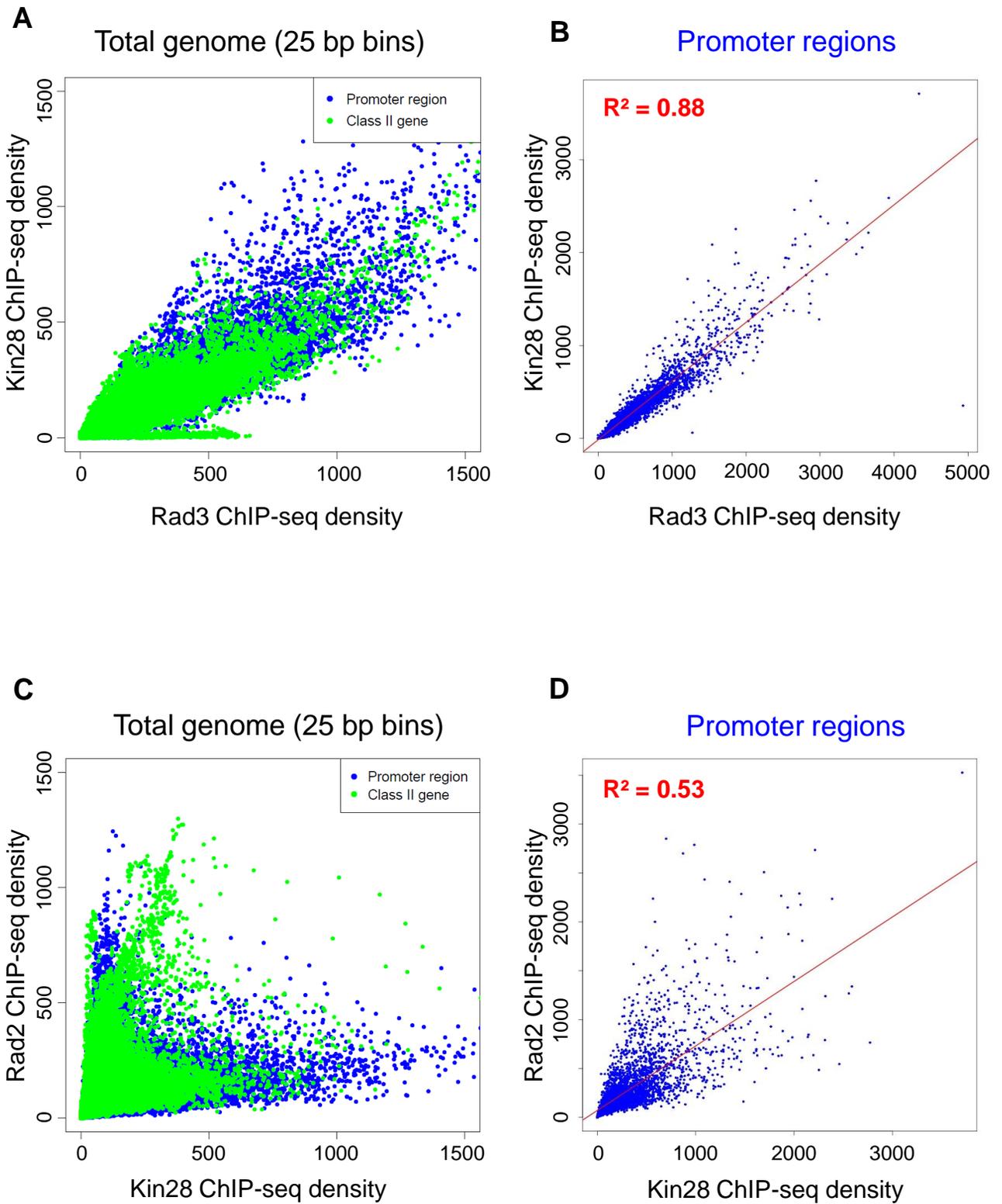


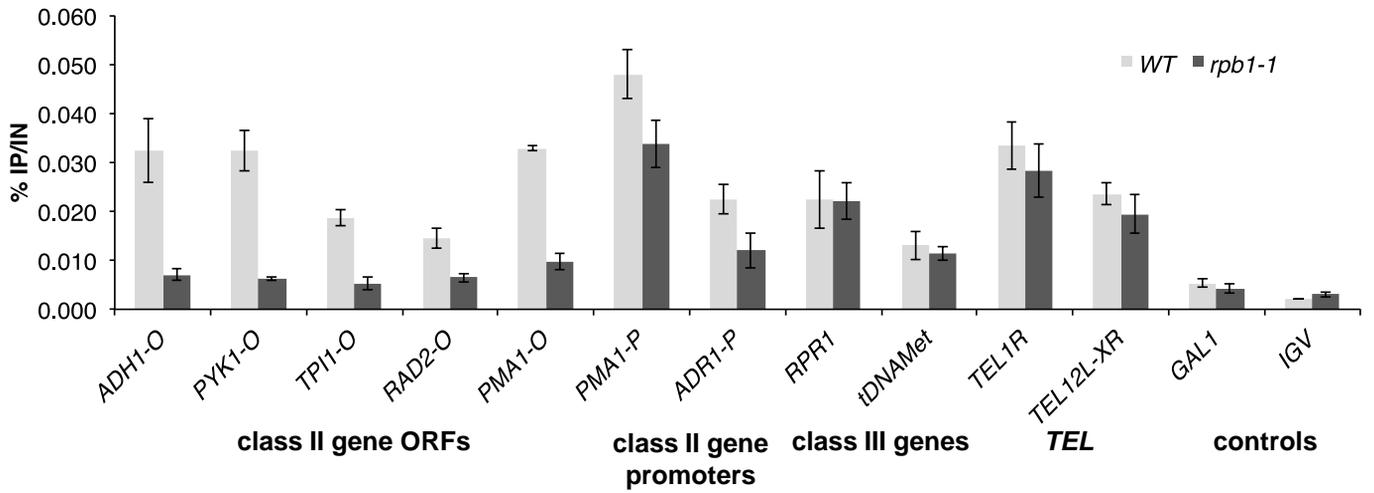
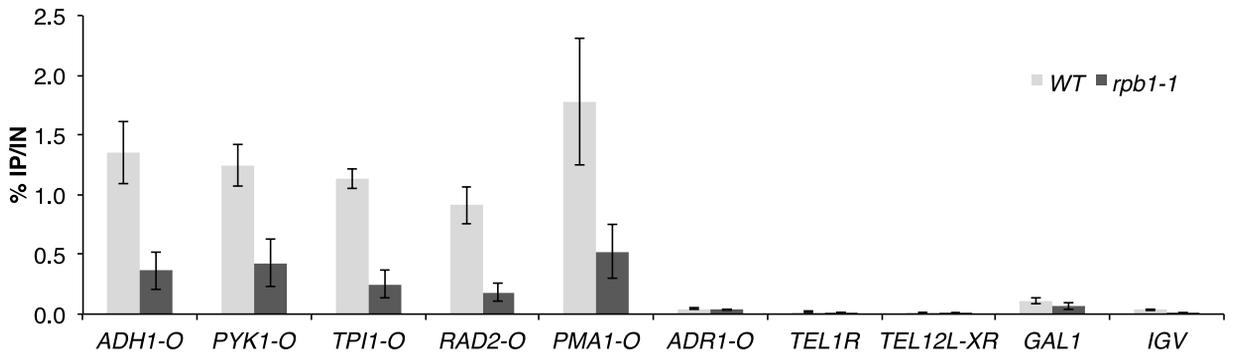
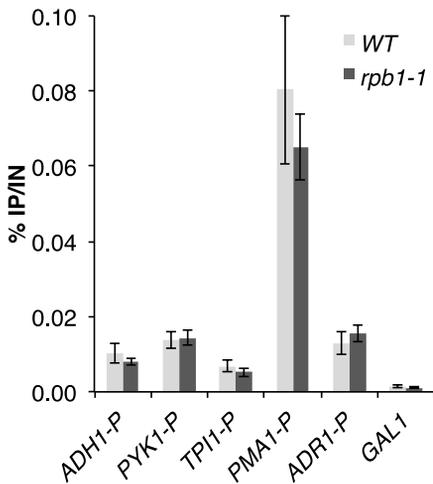
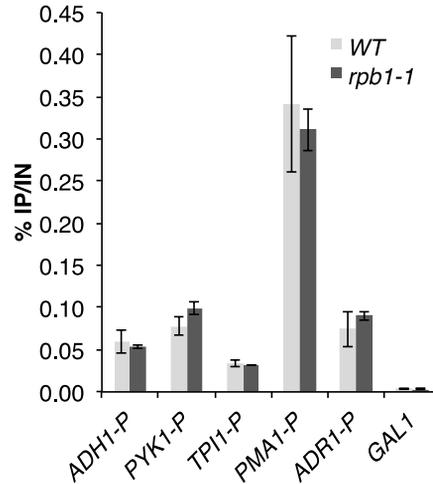
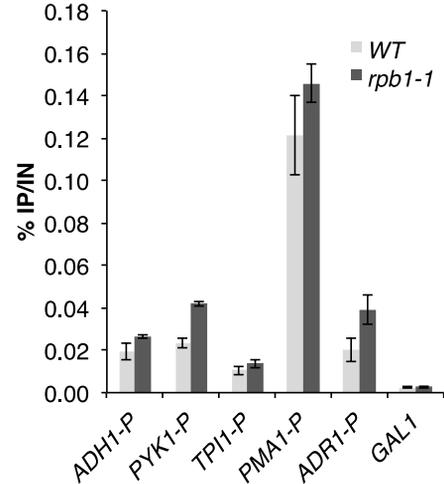


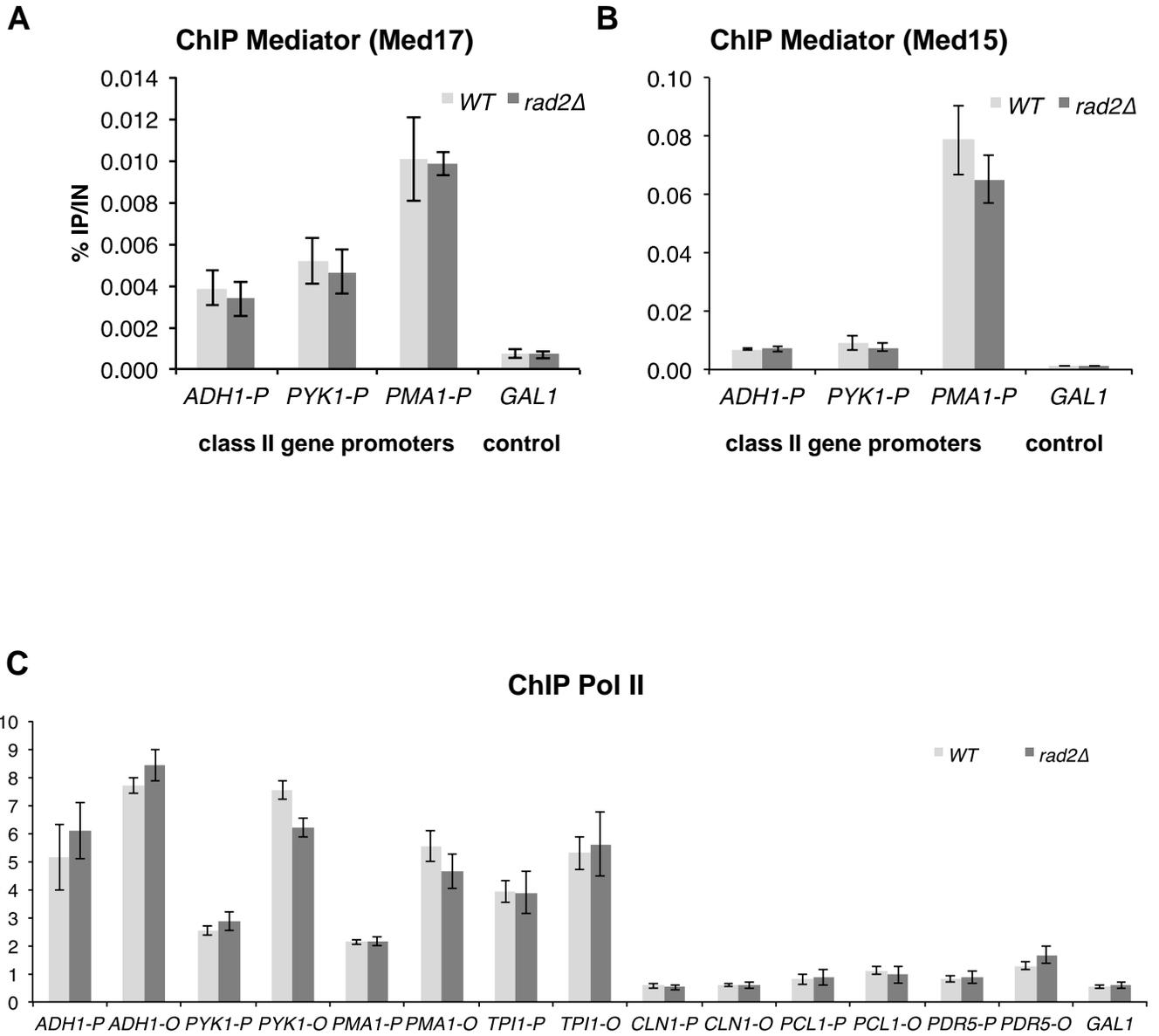
Pol III-transcribed genes

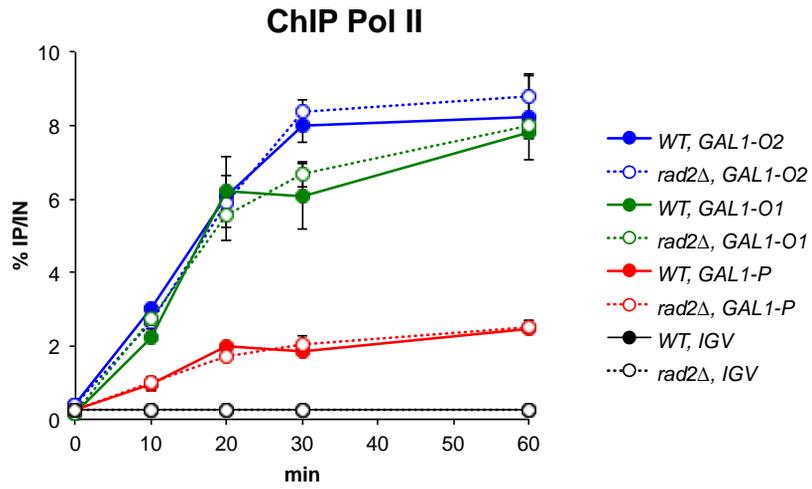
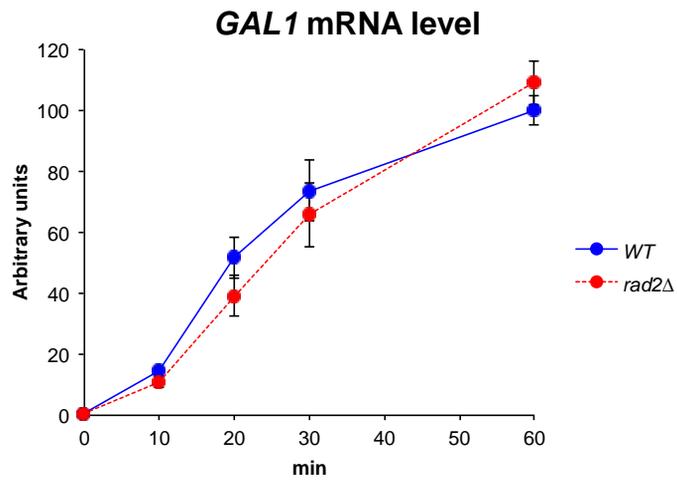


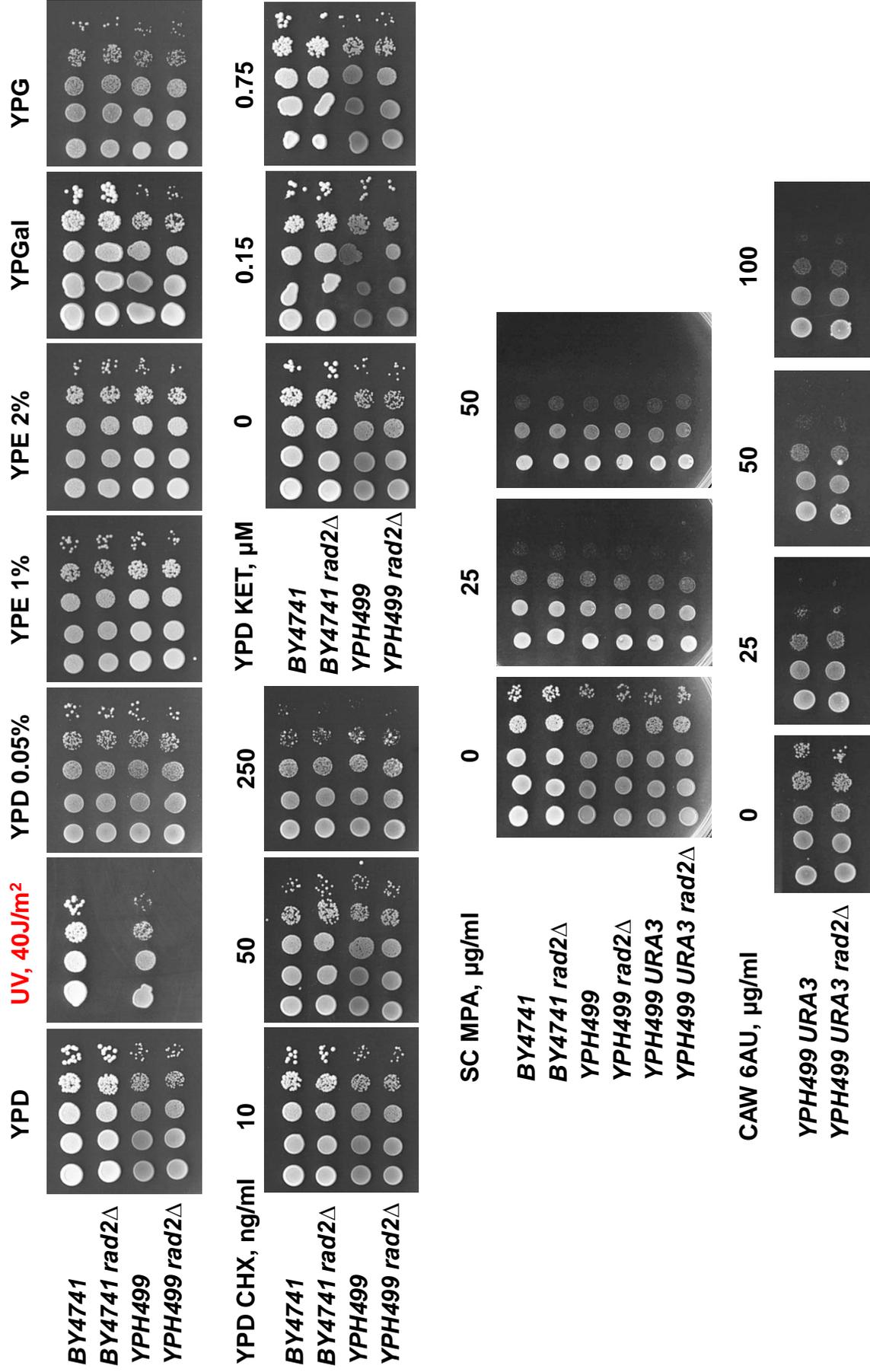
A**B****C****D**

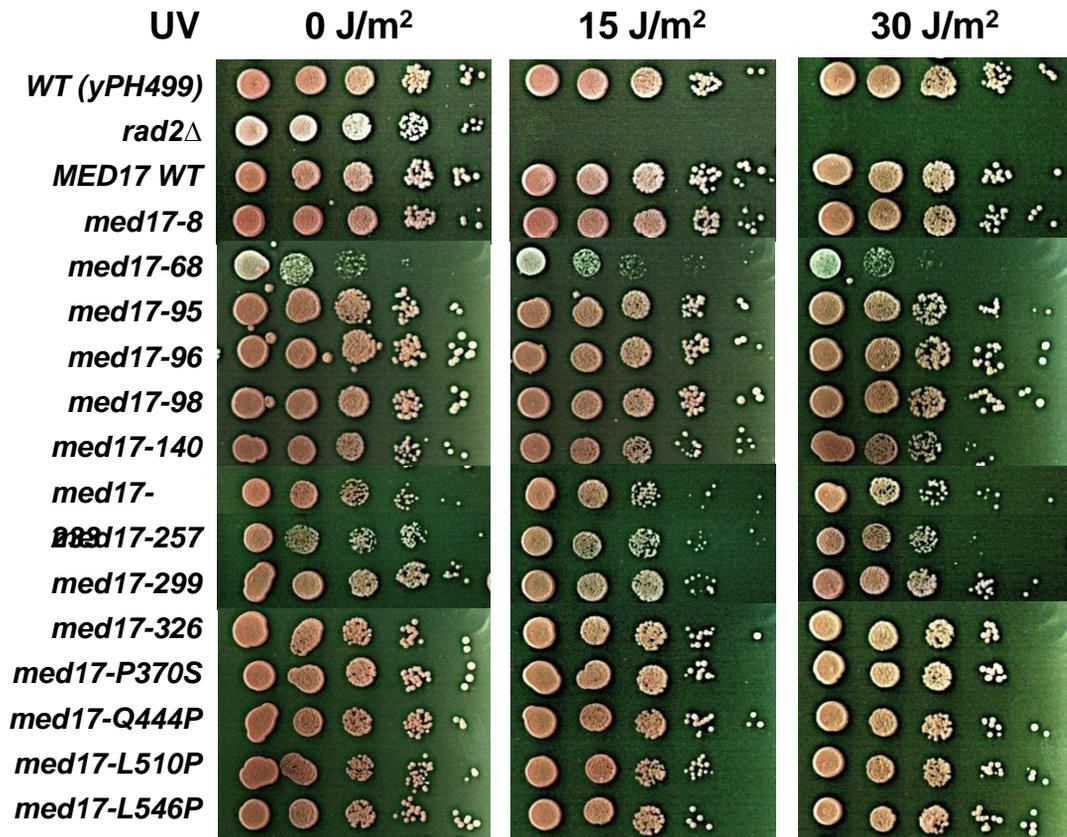


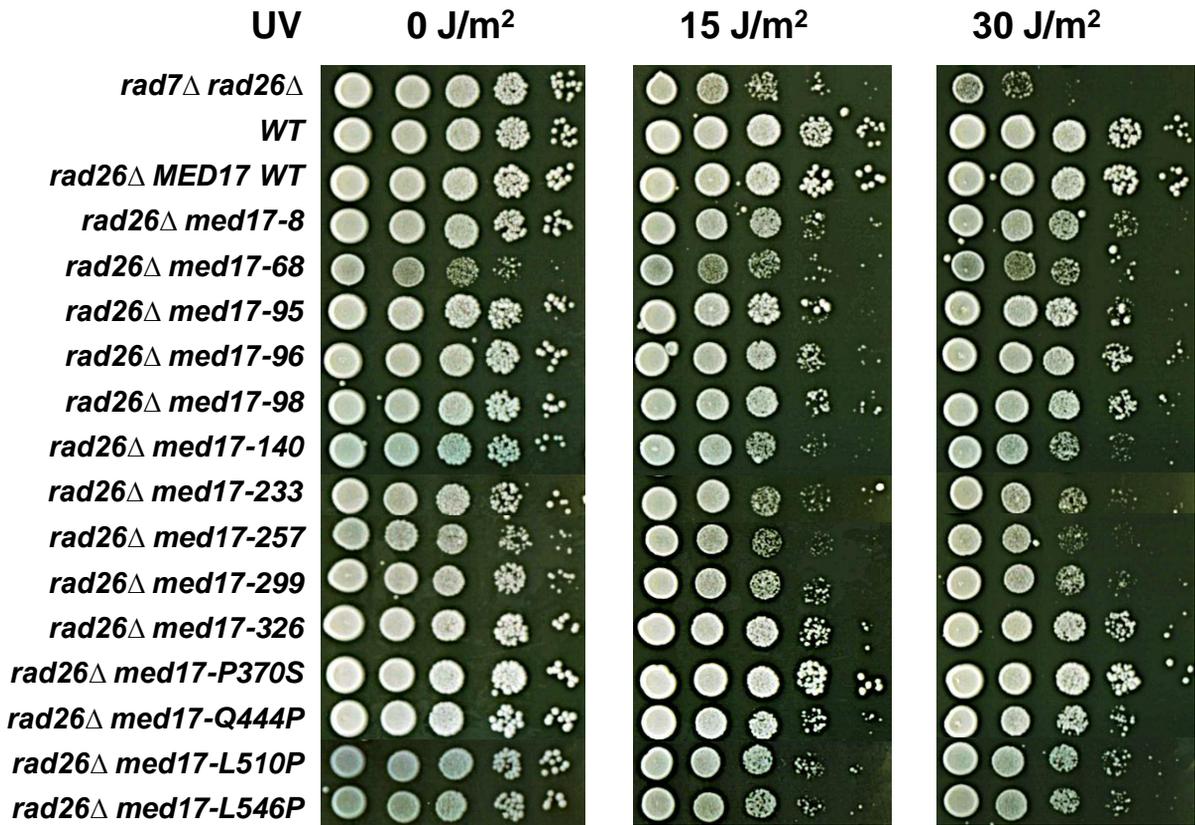
A**ChIP Rad2****B****ChIP Pol II****C****ChIP Mediator (Med17)****D****ChIP Mediator (Med15)****E****ChIP Mediator (Med5)**



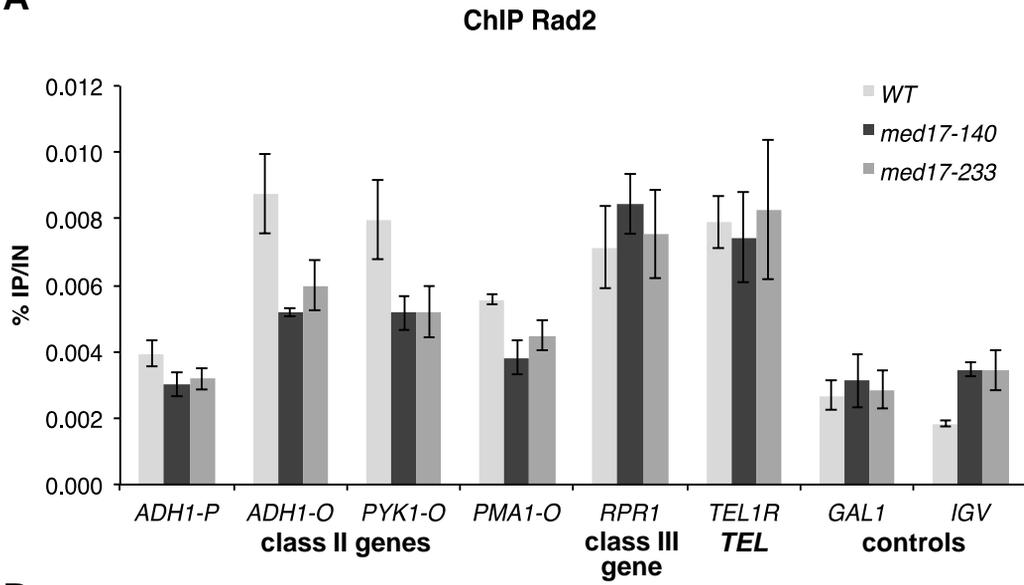
A**B**



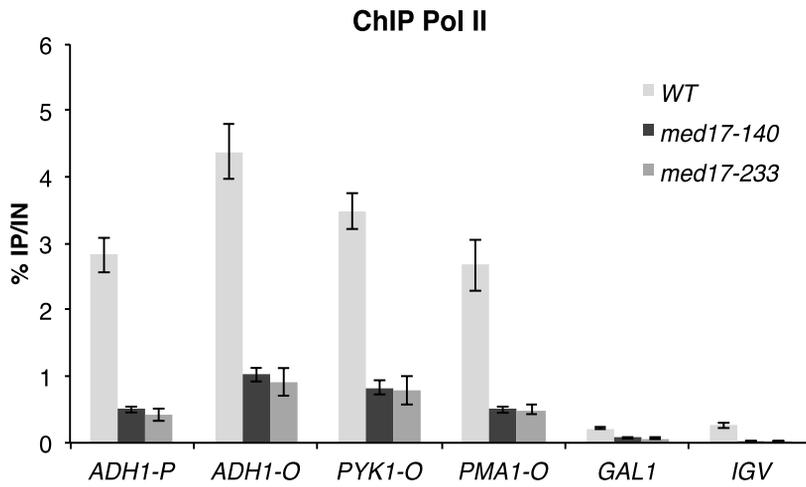




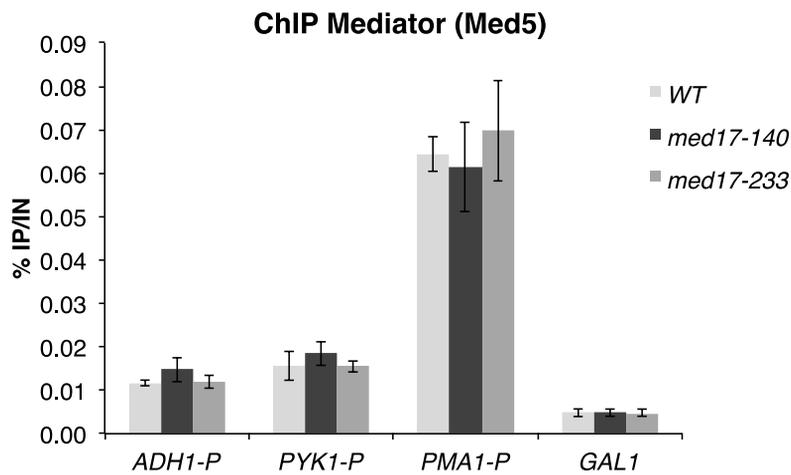
A



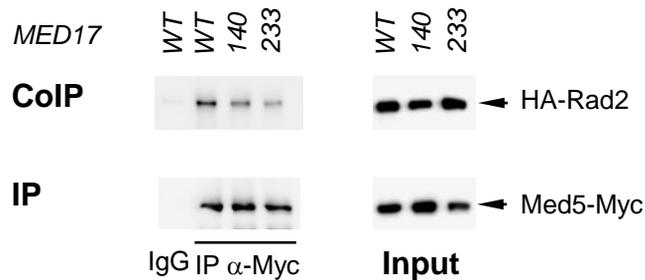
B



C



D



CONCLUSION ET PERSPECTIVES

I. OBTENTION D'UNE COLLECTION DE MUTANTS THERMOSENSIBLES

Le Médiateur de *S. cerevisiae* possède 10 sous-unités essentielles. Notre but est d'étudier le rôle de chacune d'entre elles en utilisant la méthodologie mise au point pour l'étude de Med11. Une collection de mutants thermosensibles a déjà été obtenue pour les sous-unités Med7, Med10, Med11, Med14, Med17 et Med21, en collaboration avec le laboratoire de L. Kuras. La construction est en cours pour les sous-unités Med4, Med6, Med8 et Med22 restantes. Dans le cas de mutants multiples, les mutations ont été séparées afin d'obtenir des mutants ponctuels dont le phénotype a ensuite été vérifié, ceux présentant un phénotype ts étant alors rajoutés à la collection. L'étude des sous-unités Med7 et Med10 du module du milieu est en cours et nous avons montré que certaines mutations de Med7 entraînent divers défauts d'interactions du Médiateur avec ses partenaires (thèse de Thomas Eychenne, résultats non publiés). Nous espérons que l'étude des différentes sous-unités essentielles du Médiateur nous permettra de définir de nouvelles interactions de ce complexe avec les composants de la machinerie de base de la transcription et ainsi d'élucider les fonctions du Médiateur dans la régulation de la transcription.

Au cours de ma thèse, j'ai appliqué cette méthodologie à l'étude de la sous-unité Med17 (anciennement Srb4). Plusieurs contacts de cette sous-unité avec différents complexes nucléaires ont déjà été mis en évidence, notamment avec Med11, qui appartient également au module de tête du Médiateur (Guglielmi et al., 2004) ; Rad2, une endonucléase qui intervient dans la réparation de l'ADN par le système NER (Eyboulet et al., 2013) (cf. III.) et avec la Pol II (Soutourina et al., 2011). Med17 joue également un rôle important au sein du module de tête puisque dans le mutant *srb4-138ts* le complexe Médiateur est totalement dissocié, aboutissant à une diminution de la transcription de l'ensemble des gènes de *S. cerevisiae*.

La mutagenèse aléatoire de cette sous-unité puis la séparation des mutations multiples pour obtenir des mutants ponctuels par mutagenèse dirigée, a permis d'obtenir une collection de 29 mutants ts. Des tests de croissance à 30 et 37°C nous ont permis de caractériser leurs phénotypes. La majorité ne présente qu'un faible défaut de croissance à 30°C, sauf les mutants *med17-68*, *med17-208* et *med17-327* dont la croissance est fortement ralentie à 30°C et qui sont létaux à 37°C. L'ajout d'une étiquette HA sur la sous-unité Med8 est léthal dans les mutants *med17-68* et *med17-208*. Dans la souche *med17-327*, nos analyses par co-IP ont révélé une diminution ou une perte d'interaction entre plusieurs sous-unités du module de tête.

Une co-létalité avec la mutation *rpb3-2*, et une forte diminution de l'occupation de la Pol II sur l'ADN, est observée dans cinq mutants de Med17 dont les 3 décrits ci-dessus (Soutourina et al., 2011). L'ensemble des résultats de cette étude a permis de conclure à l'importance du contact entre Med17 et Rpb3 dans le recrutement de la Pol II *in vivo*.

Ces travaux démontrent que l'étude des mutants de Med17, et par extension, des mutants des autres protéines essentielles du Médiateur, peut aider à décrypter le fonctionnement de ce complexe dans divers processus. Les deux articles présentés dans ce manuscrit en illustrent deux exemples.

II. ROLE DU MEDIATEUR DANS LA MISE EN PLACE DU PIC *IN VIVO*

La majorité des mutants *med17ts* obtenus conserve un complexe Médiateur intact. L'analyse par CHIP de 6 d'entre eux, dans lesquels le Médiateur conserve sa capacité à lier l'ADN, a révélé de nouveaux rôles pour ce complexe dans la mise en place du PIC.

Les mutants *med17-98* et *med17-670* présentent un défaut d'occupation de la Pol II et du module cœur de TFIID, sans que l'occupation de la TBP soit affectée. Ces deux mutants, bien que très similaires au niveau des effets sur les composants du PIC observés par CHIP, ont des phénotypes à 37°C bien différents, avec une croissance plus ralentie pour le mutant *med17-98*. Les défauts observés dans ce mutant augmentent avec le temps de croissance à température restrictive (37°C), comme le démontrent les résultats de RT-PCR et les analyses de la composition du PIC réalisées par CHIP après différents temps d'incubation à 37°C (90 et 180 min). A l'inverse, dans les mutants *med17-444* et *med17-546*, l'occupation du cœur de TFIID est inchangée, ou faiblement affectée, tandis que celle de la TBP est nettement diminuée, suggérant un rôle pour le Médiateur dans le recrutement de ce facteur. Les CHIP réalisées sur le gène *GAL1* après induction par le galactose confirment que le recrutement de la TBP peut être affecté, indépendamment des autres facteurs et que ce défaut ne provient pas d'une diminution du recrutement du Médiateur lui-même. Une cinétique d'induction du gène *GAL1* en présence de galactose, notamment pour le mutant *med17-444*, pourrait permettre de mieux détailler le mode de recrutement de la TBP par le Médiateur. Les mutants *med17-257*, *med17-326* et *med17-504* sont plus « classiques » puisqu'ils présentent un défaut d'occupation de tous les facteurs étudiés, à l'exception du Médiateur lui-même.

De manière tout à fait inattendue, l'étude des mutants *med17-98* et *med17-670* montre que le module kinase peut être présent sur l'ADN en l'absence du module de cœur de TFIID. Cette

observation est confirmée par l'analyse du recrutement des éléments du PIC sur *GAL1* après induction par le galactose chez ce mutant qui montre un défaut de recrutement du module de cœur de TFIIH mais pas du module kinase. Il serait intéressant d'analyser plus en détail cette particularité des mutants *med17-98* et *med17-670* en recherchant des interactions entre le module kinase de TFIIH et les différents composants du PIC.

Ces résultats permettent de compléter le schéma de mise en place du PIC proposé par C. Esnault et al., en rajoutant la TBP qui peut donc, d'après nos résultats, être recrutée indépendamment des autres facteurs, et TFIK qui peut être présent en absence du module cœur de TFIIH (Figure 27). Dans ce modèle, le Médiateur lie la région régulatrice en amont des gènes en interagissant avec les activateurs spécifiques. Puis il favorise indépendamment l'occupation de la Pol II, des modules cœur et kinase de TFIIH, de la TBP et de TFIIE. Le Médiateur stimule donc la transcription en influençant la mise en place du PIC.

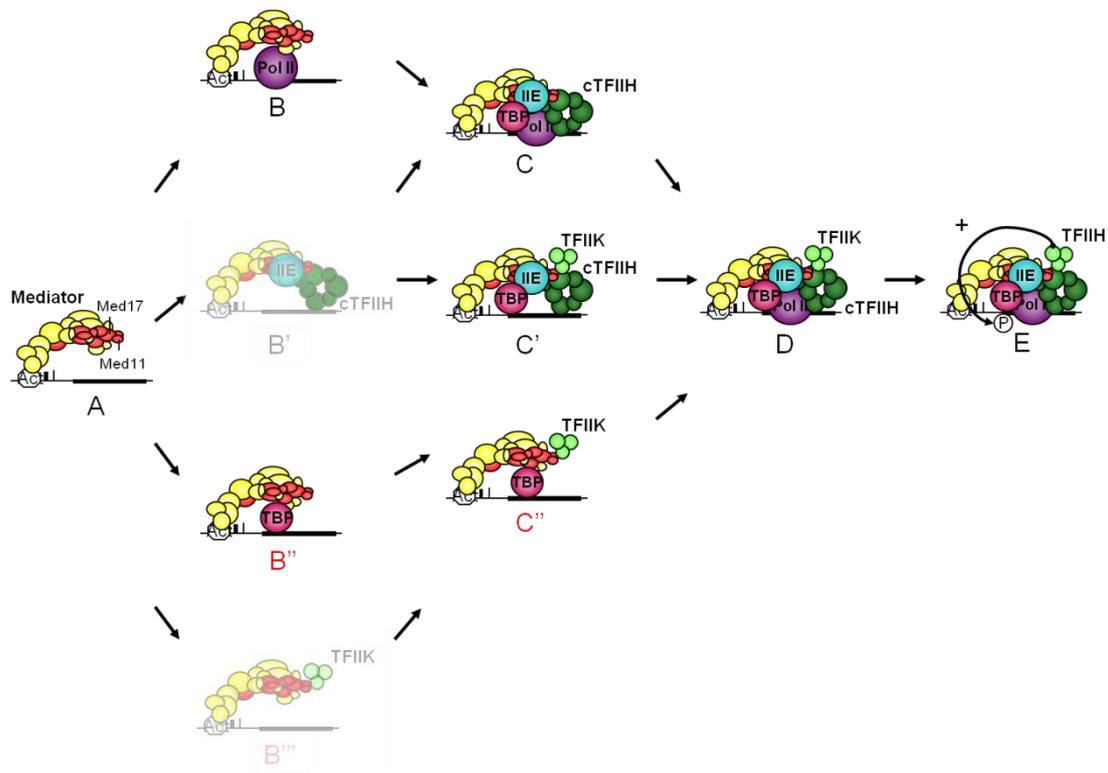


Figure 27 : Modèle de l'assemblage du PIC *in vivo* (figure modifiée d'après Esnault et al., 2008, complété d'après Soutourina et al., 2011 et l'étude des mutants *med17ts* (Article 1)).

Les intermédiaires qui ont été observés sont représentés par des couleurs pleines. Ceux qui sont déduits des intermédiaires observés et d'expériences rapportées dans la littérature, sont indiqués par des couleurs translucides. Les nouveaux intermédiaires, suggérés par l'étude des mutants de Med17 sont indiqués en rouge.

Une analyse à l'échelle du génome de *S. cerevisiae* par ChIP-seq nous a permis de confirmer que le Médiateur se lie préférentiellement sur les régions régulatrices, en amont des GTFs et de la Pol II. Cette dernière est essentiellement détectée sur les ORFs, contrairement à ce qui est observé chez les mammifères où l'existence d'une pause à proximité du promoteur la fait s'accumuler au niveau du TSS. Cette analyse à grande échelle a également permis de déterminer que les effets des mutations de Med17 sur l'occupation des divers composants du PIC, mis en évidence par ChIP, sont essentiellement globaux. La majorité des régions présentant un effet spécifique correspondent à des gènes codants des protéines de choc thermique (HSP) ou d'autres protéines de réponse au stress, ce qui n'est pas étonnant quand on étudie des mutants *ts*.

La liaison des activateurs sur l'UAS, situé en amont du promoteur, est la première étape d'activation de la transcription. Une grande quantité de données est disponible chez *S. cerevisiae* concernant le site de fixation de la plupart des activateurs et leur réseau d'interaction avec d'autres régulateurs (<http://www.yeasttract.com/>). Ainsi, en étudiant le mode d'assemblage du PIC sur les gènes régulés par les mêmes types d'activateurs, nous pourrions déterminer si la combinaison de certains activateurs promeut une séquence d'assemblage du PIC particulière. Or nos analyses par ChIP-seq nous ont permis de détecter quelques cas d'effets spécifiques sur des gènes n'intervenant pas dans la réponse au stress thermique. Si d'autres gènes régulés par les mêmes activateurs présentent le même effet, cela pourrait permettre de décrypter des mécanismes de régulation par le Médiateur activateur-spécifiques.

Chez la levure, ~20% des gènes contiennent une boîte TATA canonique (Basehoar et al., 2004). Ces gènes sont majoritairement dépendants de SAGA pour leur activation, tandis que les gènes n'en contenant pas dépendent plutôt de TFIID. Les travaux du laboratoire de F. Pugh ont permis de décrire la présence ou l'absence de boîte TATA canonique de la majorité des gènes de *S. cerevisiae* (Basehoar et al., 2004; Rhee and Pugh, 2012). A partir de ces annotations, et en utilisant nos données de ChIP-seq, il serait intéressant de regarder les conséquences des mutations *med17ts* sur la composition du PIC sur ces deux types de promoteurs. Si les compositions diffèrent, nous pourrions montrer que le Médiateur coordonne la mise en place du PIC différemment selon que le gène dépend de SAGA ou de TFIID pour son activation. Afin de confirmer ces résultats, nous pourrions alors analyser par ChIP le recrutement de sous-unités spécifiques à SAGA ou à TFIID chez les différents mutants.

Le Médiateur coordonne la mise en place du PIC sur le promoteur, cependant nous ne savons pas dans quelle mesure ce complexe participe à la sélection du site d'initiation de la transcription. Afin de répondre à cette question, nous pourrions regarder la localisation des composants du PIC par rapport au TSS dans nos mutants de Med17. Cependant, la résolution actuelle de nos données

de ChIP-seq ne nous permet pas de détecter des décalages de l'ordre de quelques nucléotides. Afin de gagner en résolution, nous pourrions mettre en place des expériences de ChIP-exo, qui consistent en l'ajout d'une étape de digestion par une exonucléase qui va digérer l'extrémité des fragments d'ADN non protégés par la protéine d'intérêt, et ainsi permettre de détecter l'emplacement exact de cette dernière sur l'ADN. Cette analyse pourrait notamment être effectuée en premier sur le facteur TFIIB, dont on sait qu'il intervient dans la sélection du TSS (Chen and Hampsey, 2004).

Afin d'expliquer les défauts d'occupation observés par ChIP, nous avons regardé par co-IP si l'interaction entre le Médiateur et les différents composants du PIC est conservée. Nous avons déjà pu montrer que l'interaction avec la Pol II est conservée dans les mutants *med17-98*, *med17-444* et *med17-670*. Le mutant *med17-504*, quant à lui, présente une diminution de l'interaction entre Med17 et Rpb3 qui pourrait expliquer la diminution d'occupation de l'ensemble des facteurs étudiés par ChIP. L'analyse de l'interaction du Médiateur avec les GTFs est en cours. Si les résultats s'avèrent corrélés, ceci permettra de faire de nouvelles hypothèses mécanistiques quant au rôle de Med17 dans l'initiation de la transcription.

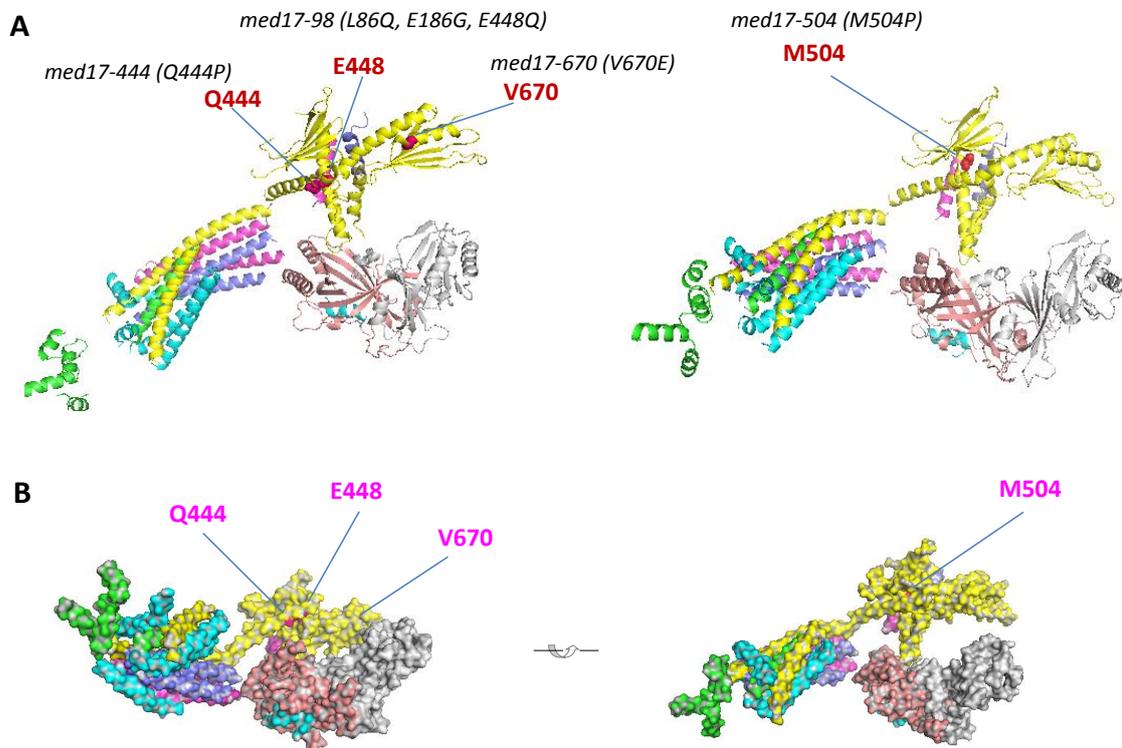


Figure 28 : Localisation des mutations portées par les mutants de Med17 sur la structure du module de tête du Médiateur (Larivière et al., 2012a).

(A) Représentation en rubans du module de tête du Médiateur.

(B) Représentation de la surface du module de tête du Médiateur.

La résolution actuelle de la structure du module de tête ne nous permet pas de placer toutes les mutations portées par les mutants *med17-98*, *med17-444*, *med17-504* et *med17-670*. Cependant, on constate que celles que nous avons pu placer sont situées à la surface ou proches de la surface du complexe, probablement en contact avec d'autres protéines (Figure 28). A l'avenir, il serait intéressant de disposer d'une structure du Médiateur plus détaillée et notamment en interaction avec ses partenaires afin de pouvoir prédire les pertes d'interactions possibles en fonction de la localisation des mutations.

Des expériences de mise en place du PIC *in vitro* ont montré que le premier cycle de la transcription d'un gène est plus lent que les cycles suivants. La rapidité de réinitiation est obtenue par l'existence d'un échafaudage (« scaffold ») de complexes constitué de TFIIA, TFIID, TFIIF, TFIIH et le Médiateur, qui restent associés au promoteur après la première initiation (Liu et al., 2004; Yudkovsky et al., 2000). Cependant, il est important de noter que ce complexe n'a pas encore été identifié *in vivo*. Afin de disséquer plus en détail plusieurs aspects du cycle de transcription, notamment la réinitiation, des expériences de mise en place du PIC *in vitro*, sur matrice immobilisée sont en cours de réalisation selon le protocole établi par le laboratoire de S. Hahn. (Ranish et al., 1999; Yudkovsky et al., 2000). La formation du PIC est observée en incubant un extrait de la souche sauvage ou mutante avec la matrice immobilisée et les différents activateurs. Après 2 minutes, la réaction est arrêtée pour analyser seulement la première initiation. Pour la réinitiation, la réaction n'est arrêtée qu'après 10 minutes de transcription. L'incorporation des GTFs au sein du PIC est ensuite vérifiée par Western Blot. Nous avons également commencé à mettre en place une analyse par spectrométrie de masse, afin de gagner en rapidité et en précision. En réalisant cette analyse sur les mutants ayant des effets différentiels sur la mise en place du PIC *in vivo*, nous devrions parvenir à la caractérisation complète de la fonction du Médiateur dans ce processus. Nous espérons notamment mettre en évidence des mutants n'affectant que la première initiation de la transcription, ou, au contraire, des mutants n'affectant que la réinitiation de la transcription, sans affecter la première initiation. Ces mutants de recyclage n'ont encore jamais été isolés, l'étude de leur phénotype pourrait permettre d'adresser l'importance du recyclage *in vivo*.

A terme, cette étude sera étendue à toutes les sous-unités essentielles du Médiateur. L'association des méthodes d'analyse *in vivo* et *in vitro* permettra une compréhension détaillée des mécanismes par lesquels ces protéines contribuent au processus d'activation de la transcription.

III. RÔLE DU MÉDIATEUR DANS LA RÉPARATION COUPLÉE A LA TRANSCRIPTION

L'étude du réseau d'interactions existant entre les sous-unités du Médiateur Med11 et Med17, la protéine Rad3 du complexe TFIIH, qui intervient à la fois dans la transcription et la réparation, et la protéine Rad2/XPG, impliquée dans le mécanisme de réparation par excision de nucléotides (NER), est d'un intérêt primordial pour comprendre le lien existant entre la réparation et la transcription.

Des expériences de co-IP nous ont tout d'abord permis de confirmer l'interaction entre Med17 et Rad2 détectée par double hybride. Plusieurs arguments suggèrent que ce contact serait direct, notamment le fait qu'il soit affecté par des mutations de Med17. Par la suite, nous souhaitons préciser le ou les domaines de liaison de Med17 sur Rad2 par des expériences de double-hybride chez la levure. Ceux de TFIIH étant déjà connus, ces travaux permettront d'apporter de nouvelles informations utiles à la compréhension du réseau d'interactions existant entre TFIIH, Rad2 et le Médiateur.

Chez l'homme, XPG est associé au promoteur de gènes dépendants des récepteurs nucléaires même en l'absence de stress génotoxique (Le May et al., 2010). Cependant, aucune donnée de localisation de XPG sur l'ensemble du génome n'était jusqu'alors disponible, que ce soit chez l'homme ou chez la levure. C'est pourquoi nous avons mis en place une analyse par CHIP-on-chip puis CHIP-seq afin de déterminer les sites de fixation de Rad2 sur le génome de *S. cerevisiae*. Nos résultats révèlent une distribution très complexe de ce facteur, présent à la fois sur les régions promotrices et les ORFs des gènes de classe II mais également sur les gènes de classe I et III et les régions télomériques.

Le Médiateur et Rad2 interagissent tous deux avec la Pol II, le principal composant de la machinerie de transcription, mais également le premier complexe à reconnaître les dommages de l'ADN dans la voie TCR du système NER. Nos analyses par CHIP-seq révèlent que le Médiateur et Rad2 sont tous deux présents sur les régions promotrices des gènes de classe II et présentent des profils d'occupation fortement corrélés ($R^2=0.75$) confirmant l'existence d'un lien fonctionnel entre ces deux facteurs. Rad2 est également présent sur les ORFs des gènes de classe II et son interaction avec la Pol II ne suffit pas à expliquer cette observation puisque la corrélation entre Rad2 et la Pol II sur ces régions est très modeste. Une autre explication serait que le Médiateur influence également la présence de Rad2 sur les ORFs. Cependant, l'occupation des ORFs par le Médiateur, bien que décrite dans certaines publications, représenterait des régions fortement exprimées, enrichies de façon non spécifique même dans les contrôles de CHIP (Park et al., 2013;

Teytelman et al., 2013). A ce niveau, nous suggérons donc que le Médiateur permet le recrutement de Rad2 sur le promoteur et que la présence de ce dernier facteur également sur les ORFs est liée à son interaction avec la Pol II et à d'autres mécanismes que nous n'avons pas encore mis en évidence.

Nous n'avons pas mis en évidence de défaut de croissance ou de défaut de transcription chez les mutants *rad2Δ* et n'avons donc pas pu conclure à un rôle pour Rad2 dans la transcription tel qu'il l'avait été suggéré par Lee et al. (2002). Par contre, en mettant à profit notre collection de mutants *med17ts*, nous avons montré que, dans un contexte *rad7Δ* où la voie GGR est déficiente, les mutations de Med17 conduisent à une sensibilité aux UVs. Ce phénotype n'est pas une conséquence d'un défaut de transcription puisqu'il n'est pas observé dans un contexte où les deux voies NER sont fonctionnelles. De plus, dans ces mutants, nous avons observé un défaut d'interaction entre Med17 et Rad2, et une diminution de l'occupation de Rad2 sur l'ADN sans défaut d'occupation du Médiateur. Ainsi, nous appuyons notre hypothèse selon laquelle l'interaction entre Med17 et Rad2 serait importante pour le recrutement de Rad2 sur l'ADN et donc que le Médiateur intervient à la fois dans la transcription et dans la réparation de l'ADN. TFIIH jouant également un rôle dans les deux processus, et interagissant à la fois avec Rad2 et le Médiateur, il serait intéressant d'explorer plus en détail le lien fonctionnel existant entre ces trois protéines. Nous pourrions commencer par déterminer si les contacts avec TFIIH sont conservés en présence de mutations du Médiateur ou de TFIIH. Puis, nous pourrions vérifier la présence de ces différents facteurs sur l'ADN après un stress génotoxique, dans différents contextes génétiques afin d'émettre des hypothèses quant à leur mode de recrutement. Pour ce faire nous pourrions mettre en place des expériences de CPD-ChIP qui consistent, après un stress UV, à immunoprécipiter les facteurs de la réparation, puis l'ADN contenant des lésions CPD à l'aide d'anticorps anti-CPD. Cette technique permet ainsi de déterminer spécifiquement la fixation des protéines sur les régions d'ADN endommagées.

Nos analyses par ChIP-on-chip ont révélé que Rad2 est présent sur les régions télomériques. Récemment deux études ont montré que le Médiateur s'associe également aux télomères pour influencer le « silencing » et la maintenance de l'hétérochromatine, régulant ainsi la durée de vie des cellules (Peng and Zhou, 2012; Zhu et al., 2011b). Nous souhaitons maintenant vérifier si la fonction du Médiateur sur les télomères est liée à Rad2. Pour ce faire, nous nous proposons d'analyser l'occupation sur les télomères du Médiateur dans un contexte *rad2Δ* et de Rad2 dans des contextes *med17ts*. Nous étudierons également l'effet d'une délétion de Rad2 sur la longueur des télomères et sur le « silencing ».

A terme, nous prévoyons que l'étude du lien fonctionnel entre le Médiateur, Rad2 et Rad3 nous permettra de décrypter les mécanismes du couplage de la transcription et de la réparation de l'ADN *in vivo*.

IV. CONSERVATION DES FONCTIONS DU MEDIATEUR CHEZ LES MAMMIFERES

Étant donné son rôle central dans l'activation de la transcription, le Médiateur est impliqué dans de nombreuses pathologies humaines et notamment dans différents types de cancers. L'étude du Médiateur est donc indispensable afin de décrypter les mécanismes pouvant mener à ces diverses pathologies et pouvoir les traiter.

Bien que les séquences primaires des sous-unités du Médiateur ne soient pas très conservées chez les eucaryotes, les programmes de prédiction des structures secondaires suggèrent une conservation bien plus importante à ce niveau. Des alignements de séquences basés sur ces structures secondaires ont permis de définir des motifs conservés de la levure à l'homme (Bourbon, 2008). La majorité des mutants ts de Med17, mais également des autres sous-unités étudiées notamment Med10 et Med7, possèdent au moins une mutation dans un de ces motifs conservés (Figure 29).

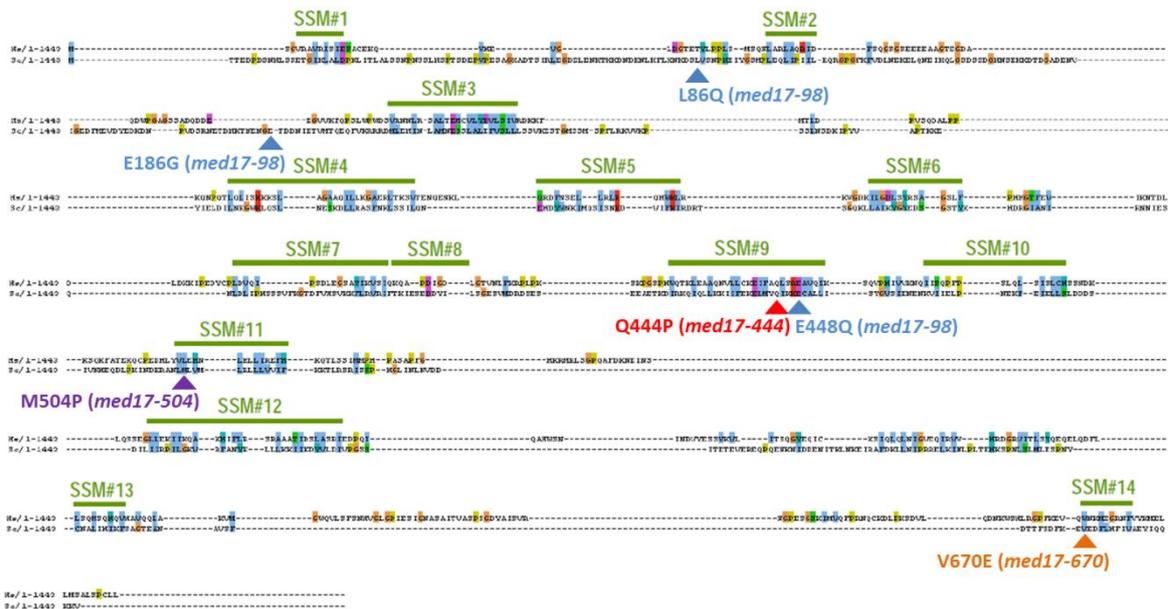


Figure 29 : Localisation des mutations de Med17 sur l'alignement des séquences de l'homme et de la levure (modifié d'après Bourbon, 2008)

Nous allons donc pouvoir bénéficier de cette conservation des séquences et transposer les mutations ts de la levure dans des cellules HeLa pour explorer la fonction du Médiateur chez l'homme. L'effet des mutations sur la prolifération et la survie cellulaire sera analysé en premier lieu. Puis, nous réaliserons des expériences de CHIP et CHIP-seq pour étudier l'impact des mutations sur l'assemblage du PIC et la transcription. Nous serons ainsi en mesure d'adresser la conservation des fonctions du Médiateur de la levure à l'homme. Cette étude ouvrira également la possibilité de répondre aux questions supplémentaires concernant le rôle du Médiateur dans la dérégulation globale de l'expression des gènes qui est une caractéristique majeure des cancers.

Nous avons déjà tiré profit de la conservation des séquences de Med17 entre la levure et l'homme pour transposer dans notre modèle d'étude la mutation responsable de l'atrophie cérébrale infantile L371P (Kaufmann et al., 2010). L'analyse de ce mutant (M504P), qui donne un phénotype ts chez *S. cerevisiae*, a mis en évidence une diminution de l'interaction entre le Médiateur et la Pol II qui pourrait être à l'origine du défaut d'occupation des composants du PIC observé par CHIP et CHIP-seq. L'analyse de l'interaction avec les autres partenaires du Médiateur, et leur conservation dans les cellules humaines, permettra d'en apprendre davantage sur les mécanismes moléculaires à l'origine de la pathologie observée chez l'homme.

Le Médiateur et la protéine de réparation Rad2/XPG étant conservés de la levure à l'homme, nous avons voulu vérifier si l'interaction entre ces deux facteurs, détectée chez *S. cerevisiae*, existe aussi dans les cellules humaines. Des expériences de co-IP réalisées à partir d'extraits bruts de cellules HeLa nous ont permis d'observer un contact entre le Médiateur et XPG. Nous allons par la suite vérifier si le lien entre le Médiateur et la réparation, que nous venons de mettre en lumière, existe également chez l'homme.

Les mutations responsables des syndromes XP et XP/CS sont bien caractérisées (Figure 30). Nous allons regarder par co-IP si le contact entre le Médiateur et XPG est conservé dans les cellules de fibroblastes primaires issues de ces patients. Nous nous attendons, par exemple, à ce que les mutations de tronçatures conduisent à une perte d'interaction entre ces deux protéines. Cela nous apportera des informations importantes pour mieux comprendre les mécanismes moléculaires complexes à l'origine de ces pathologies.

CONCLUSIONS ET PERSPECTIVES

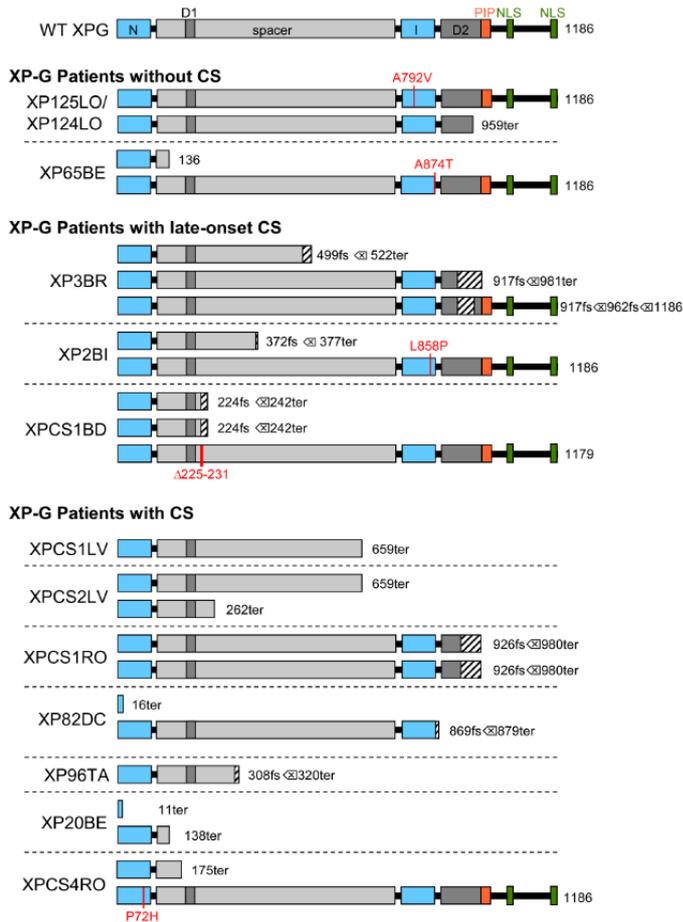


Figure 30 : Mutations de XPG chez les patients XP et XP/CS (Scharer, 2008).

Les mutations ponctuelles ou conduisant à un décalage du cadre de lecture (fs « frameshift ») ou un codon stop (ter) sont indiquées sur le schéma. Les patients sont classés selon qu'ils présentent les pathologies XP, XP et CS à apparition tardive ou XP et CS sévère. Les zones hachurées indiquent des insertions de séquences qui n'appartiennent pas initialement à XPG.

A la suite de nos travaux, nous serons en mesure de proposer des mécanismes permettant de mieux comprendre comment des défauts du Médiateur peuvent mener à certaines pathologies. L'étude des différents rôles du Médiateur dans des processus tels que la transcription ou la réparation nous permettra également d'appréhender comment ils sont coordonnés au sein de la cellule. La compréhension des mécanismes d'activation de la transcription pourrait permettre de proposer d'éventuelles cibles thérapeutiques pour traiter le cancer notamment en inhibant certaines interactions protéine-protéine que nous aurons mises en évidence. L'étude des mécanismes moléculaires du rôle du Médiateur dans la réparation de l'ADN pourra aussi nous permettre de mieux comprendre l'origine et les symptômes des pathologies XPG et XPG/CS.

METHODES

CHAPITRE I : CHIP-SEQ

L'immunoprécipitation de chromatine (ChIP) est la technique de choix pour étudier la distribution, *in vivo*, des protéines sur le génome. Cette technique consiste en une première étape de fixation au formaldéhyde (pontage) qui va permettre de former des liaisons covalentes entre les protéines et l'ADN. La chromatine est ensuite préparée et fragmentée jusqu'à l'obtention de fragments de taille comprise entre 0,2 et 1 kb. Les protéines liées à l'ADN sont ensuite immunoprécipitées en utilisant un anticorps spécifique dirigé contre la protéine d'intérêt. Après réversion du pontage ADN-protéine, l'ADN immunoprécipité est purifié. Les fragments issus du ChIP peuvent ensuite être analysés par qPCR ou à l'échelle du génome par hybridation sur puce à ADN (ChIP-on-chip) ou séquençage à haut débit (ChIP-seq). La technique du ChIP-seq vient peu à peu remplacer son prédécesseur, le ChIP-on-chip, car elle présente de nombreux avantages sur cette dernière : meilleure résolution (localisation plus précise des sites de liaison), meilleure couverture du génome et meilleur rapport signal / bruit de fond.

I. PRINCIPE

Pour ce projet, les échantillons de ChIP ont été séquencés sur 2 types de séquenceurs : « GAIIIX Genome Analyser » ou « HiSeq 2000 » en suivant la procédure recommandée par le fabricant (Illumina). Chez Illumina, l'amplification est réalisée sur une cellule micro-fluidique (« flow-cell »). Des adaptateurs sont ligués aux extrémités des matrices ADN avant leur dénaturation. Ces adaptateurs vont ensuite se fixer de manière covalente sur la « flow-cell » et permettre à l'ADN de s'hybrider sur le support. Plusieurs étapes d'extension et dénaturation vont aboutir à la création d'une population de matrices identiques, regroupées physiquement sur la « flow-cell », formant des colonies (« clusters »). Une amorce peut ensuite s'hybrider aux extrémités 3' de ces matrices pour l'étape du séquençage. Une ADN polymérase liée à la matrice ajoute un seul nucléotide complémentaire de la base de l'ADN matrice. Ce nucléotide est lié à un fluorophore qui va permettre son identification et à un terminateur, qui empêche l'incorporation du nucléotide suivant. Après capture de l'image fluorescente, le fluorophore et le terminateur sont éliminés, un nouveau nucléotide est ajouté et le cycle reprend de la même façon jusqu'à ce que la taille de lecture demandée (généralement 50 nucléotides) ait été atteinte.

II. SCHEMA EXPERIMENTAL

Cette partie sera consacrée à la description de la méthode du ChIP-seq suivie durant ce projet (Figure 31). Seules les étapes nécessitant une attention particulière seront exposées ici, pour le protocole détaillé de la technique de ChIP, se référer aux annexes à la fin de ce manuscrit.

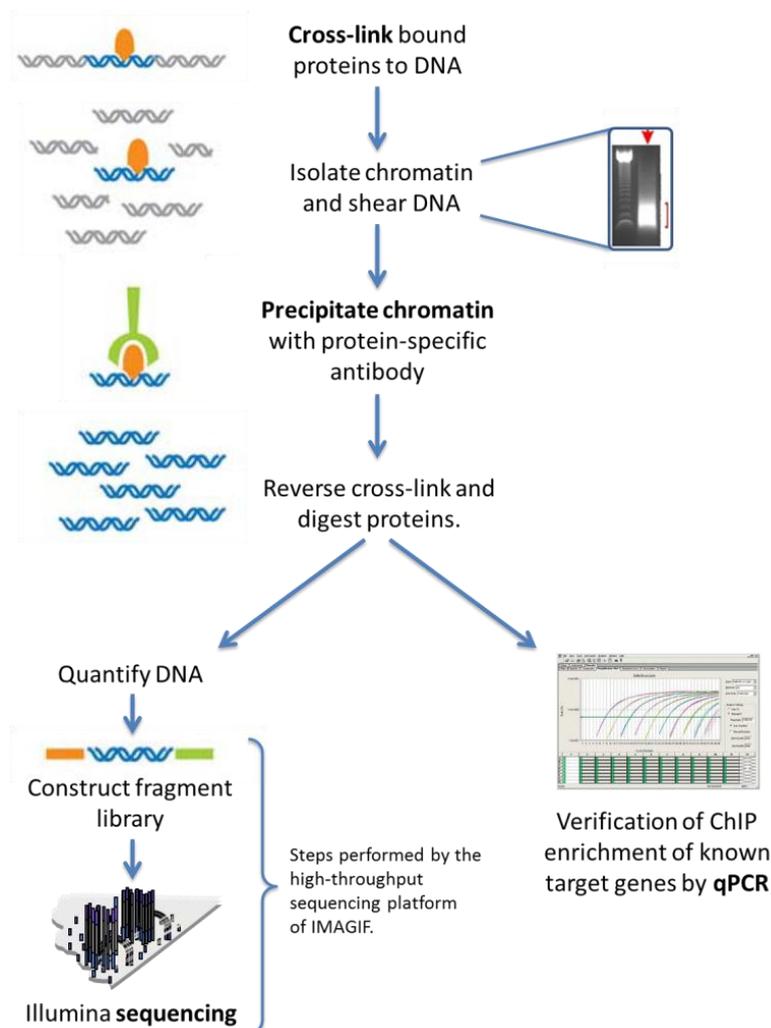


Figure 31 : Méthode du ChIP-seq.

Le schéma ci-contre résume les étapes du ChIP-seq suivies dans ce projet.

1. Fragmentation de la chromatine

Initialement, les échantillons de ChIP ont été préparés en vue d'une analyse qPCR uniquement. Nous avons donc suivi le protocole mis en place dans notre laboratoire et effectué 3 cycles de 40s de sonication à la sonde (Vibra-Cell™, Sonics®) pour obtenir des fragments d'une taille moyenne de ~400 pb. Lorsque nous avons décidé de passer à l'analyse de l'ensemble du génome par ChIP-seq, il nous a fallu fragmenter de nouveau la chromatine pour atteindre la taille moyenne de 200 pb requise pour le séquençage.

L'ajout de cycles à la sonde classique ne nous ayant pas permis de diminuer davantage la taille des fragments, nous avons décidé d'ajouter une étape de fragmentation au Biorupteur

(Bioruptor® UCD-200, Diagenode). Après avoir testé différentes conditions, nous sommes parvenus à une taille moyenne des fragments de ~200 pb (Figure 32) grâce au protocole suivant :

- 3 x 40s avec intervalles de 20s à la sonde classique. Pulse, 60% du cycle actif, puissance 4.
- Séparation des échantillons en plusieurs tubes pour avoir au maximum 300 µL par tube de 2 mL.
- 6 x 30s avec intervalles de 30s au Biorupteur. Puissance moyenne. Au bout de 3 cycles, le bac est refroidi par ajout de glace pendant quelques minutes avant d'effectuer les 3 cycles suivants.

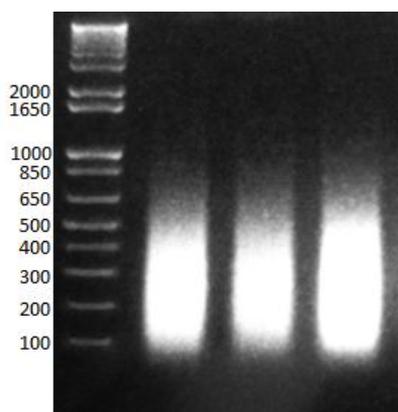


Figure 32 : Taille des fragments obtenus après sonication.

Visualisation sur gel d'agarose des fragments de chromatine après fragmentation en utilisant la sonde classique puis le Biorupteur.

2. Contrôles utilisés en ChIP-seq

La ChIP consiste en un enrichissement, et non pas en une purification des sites liés par un facteur. Selon la taille du génome de l'organisme étudié, et la protéine immunoprécipitée, une fraction plus ou moins importante des fragments d'ADN, et donc des lectures, se trouvera dans le bruit de fond, le reste constituant les fragments spécifiques. C'est pourquoi pour définir un enrichissement dans un échantillon de ChIP, il faut effectuer une comparaison avec un échantillon contrôle. Deux types de contrôles sont communément utilisés dans les expériences de ChIP : l'« Input » constitué de l'ADN total, fragmenté, avant immunoprécipitation et la « mock IP », une IP réalisée sans anticorps ou avec un anticorps non-spécifique ou avec un anticorps dirigé contre une étiquette sur une souche non étiquetée.

Le contrôle de type « Input » est utilisé dans la plupart des analyses ChIP-seq. Il permet de déterminer les zones enrichies de façon non-spécifique, résultant par exemple des variations d'amplification lors de la préparation des banques ou des biais de séquençage. Ce contrôle positif a été utilisé pour identifier les régions significativement enrichies dans nos échantillons de ChIP (cf. Chapitre II. II.5).

Le contrôle de type « mock IP » ne permet d'immunoprécipiter que très peu d'ADN, conduisant à des variations entre les contrôles eux-mêmes. Cependant, il permet de mettre en évidence les enrichissements non spécifiques pouvant apparaître durant l'étape

d'immunoprécipitation du fait de régions plus accessibles aux anticorps. Dans ce projet, nous avons utilisé deux contrôles négatifs :

- IP NT = IP avec un anticorps anti-HA sur une souche non étiquetée HA.
- IGG = IP avec des billes nues, sans ajout d'anticorps.

Nous reviendrons sur l'intérêt de ces contrôles dans le chapitre consacré aux analyses bio-informatiques (cf. Chapitre II. V.).

3. Quantification de l'ADN

Pour la construction des banques pour le séquençage, la quantité minimum d'ADN requise par les plateformes est généralement de 10 ng. Après dosage de l'ADN (Quant-it™ Picogreen®, Life Technologies™), nous avons réalisé que nos échantillons de ChIP contenaient très peu de matériel. Nous avons donc décidé de regrouper plusieurs échantillons (regroupement de 4 à 6 IPs) pour essayer d'atteindre la quantité d'ADN requise par la plateforme. Nous avons ensuite ajouté une étape d'évaporation au lyophilisateur (Speedvac™, Savant™) pendant 30min à 1h suivant le volume à évaporer. Une fois atteint un volume < 30 µL, nous avons complété chaque échantillon avec le volume d'eau nécessaire pour atteindre 30 µL.

Malgré le regroupement de plusieurs IPs effectué, il nous a été très difficile de dépasser les 5 ng de matériel par échantillon. Quelques cycles de PCR supplémentaires ont donc été effectués par la plateforme au moment de la construction des banques pour augmenter la quantité d'ADN. Après cette étape, les quantités de matériel restaient très faible et à l'issue de la préparation des banques, celles-ci étaient contaminées par des dimères d'adaptateurs. Un traitement supplémentaire par purification sur colonne a donc été réalisé afin d'éliminer ces dimères et augmenter le nombre de séquences « utiles » disponibles à l'issue du séquençage.

4. Séquençage

A. Profondeur de séquençage

La qualité du séquençage dépend avant tout de la complexité de la banque générée à partir de l'ADN immunoprécipité. Pour une expérience de ChIP-seq, l'unité de base du séquençage est une ligne de « flow-cell ». Sur un appareil de type GAIIIX, 4 à 6 millions de lectures par ligne peuvent être générées. Les appareils plus récents, de type HiSeq, permettent d'obtenir plus de 30 millions de lectures par ligne.

La profondeur de séquençage correspond au rapport entre la longueur de l'ensemble des séquences lues mises bout à bout et la longueur du génome cible. Par exemple, si l'on séquence 25 millions de bases (Mb) pour un génome de 5 Mb, nous obtenons une profondeur de 5X. Plus la

profondeur est importante, plus nombreuses seront les lectures chevauchantes qui pourront être assemblées, et plus grande sera la fraction du génome couverte. Ceci permet d'obtenir une séquence finale la plus complète possible, avec un minimum de régions non séquencées. Cependant, dans une expérience de ChIP-seq où seules certaines régions du génome sont enrichies, nous ne cherchons pas à recouvrir l'intégralité du génome. Dans ce cas, on considère que la profondeur de séquençage est suffisante quand le « point de saturation » est atteint. Celui-ci correspond au point à partir duquel aucun site de liaison supplémentaire ne sera identifié, malgré une profondeur de séquençage supérieure.

B. Multiplexage

Pour les génomes de taille relativement petite, comme celui de la levure *S. cerevisiae*, le nombre de lectures obtenu pour une ligne de « flow-cell » peut être bien supérieur au nombre de lectures nécessaires pour atteindre la couverture du génome. Les plateformes de séquençage facturant par ligne, il peut être intéressant, économiquement parlant, de séquencer plusieurs échantillons sur une même ligne. Cette technique est appelée multiplexage.

Pour préparer les échantillons pour le multiplexage, une étiquette avec un identifiant unique est ajoutée à chaque banque. Lors du séquençage, cet identifiant unique permet de distinguer la provenance de chaque séquence, et de la ré-attribuer ensuite à l'échantillon d'origine.

C. « Single-read » versus « paired-end »

Les fragments de ChIP-seq sont généralement séquencés à une extrémité seulement (« single read»). Il est cependant possible de les séquencer aux deux extrémités (« paired end »). Ce type de séquençage n'est généralement pas utilisé dans le cas du ChIP-seq, il est plutôt réalisé pour détecter les variations de la structure du génome. Dans le cas du ChIP-seq, il peut cependant être utile pour cartographier les lectures dans les séquences répétées, identifier les réseaux d'interactions de la chromatine (Fullwood et al., 2010) ou détecter précisément les sites de liaison des TFs (Wei et al., 2006).

CHAPITRE II : ANALYSES BIOINFORMATIQUES

Au cours des dernières années, le domaine du séquençage de l'information génétique est rentré dans l'ère du haut débit (HTS « high-throughput sequencing » ou NGS « next generation sequencing »). Cette avancée technologique permet une analyse plus en profondeur de l'ADN et de l'ARN mais engendre d'énormes volumes de données dont l'analyse nécessite l'acquisition de compétences spécifiques dans le domaine de la bioinformatique. Ce chapitre sera consacré à l'analyse des données de CHIP-seq, mais la plupart des informations reste valable pour les autres types de données de NGS.

La majorité des analyses effectuées pour ce projet ont été exécutées sous Unix (système CentOS, Environnement Gnome) directement en lignes de commande Bash. Les représentations graphiques ont été effectuées dans le logiciel R version 2.14.0.

I. FORMAT DES DONNEES

Dans le domaine de la bioinformatique, les formats de fichier sont nombreux et parfois complexes. Les fichiers stockant les données sont organisés selon le type d'information qu'ils contiennent. A chacun est associée une extension particulière permettant de savoir de quel type de donnée il s'agit. Les informations qui suivent sont librement adaptées du site UCSC (<http://genome.ucsc.edu/FAQ/FAQformat>).

1. FastQ

A. Définition

Le format FastQ est un format de fichier texte permettant de sauvegarder à la fois les données biologiques et les scores de qualité associés. En sortie de séquenceur à haut débit, les données sont sauvegardées sous ce format avec une extension .fq ou .fastq.

B. Description

Un fichier FASTQ utilise 4 lignes par séquence :

- La **ligne 1** commence par le caractère « @ » suivi de la référence de la séquence (nom de l'instrument, n° de run, identifiant de la flowcell, n° de ligne, coordonnées du cluster...). Cette ligne peut également contenir l'index de la séquence, dans le cas d'un échantillon multiplexé.
- La **ligne 2** contient la séquence nucléotidique brute (« read »).
- La **ligne 3** commence par le caractère « + », parfois suivi par la répétition de l'identifiant de la séquence et de sa description si celle-ci est présente.

- **Colonne 11** : Qualité de la séquence.

- **Colonne 12 (facultative)** : Informations concernant l'alignement et notamment le code « XA » si la séquence s'aligne sur le génome de référence et « XM » si la séquence ne s'aligne pas.

Exemple :

Séquence non alignée

```
HWUSI-100 4 * 0 0 * * 0 0 GGCAGCAGGCGCG ##### XM:i:1
HWUSI-165 0 chrI 2081 255 20M * 0 0 CCTTTAACAAGCG IHHHHHHHHHIG
XA:i:0 MD:Z:20 NM:i:0
```

Séquence alignée

C. Utilisation

Les fichiers BAM et SAM peuvent être manipulés facilement à l'aide de l'utilitaire SamTools (Li et al., 2009, <http://samtools.sourceforge.net/>) qui permet notamment de les trier et de les indexer. L'outil BamToBed de BedTools (Quinlan and Hall, 2010, <http://bedtools.readthedocs.org/>) permet de convertir les données BAM ou SAM au format BED, afin de s'affranchir des lignes contenant des séquences non alignées.

3. BED/bigBED

A. Définition

Le format BED (« browser extensible data ») est un format de fichier texte tabulé qui permet de sauvegarder des intervalles génomiques telles que des données d'annotation, des transcrits ou des séquences. Ce format permet de représenter graphiquement les éléments du génome. Les données sont sauvegardées sous ce format avec une extension .bed.

Le format bigBED est la version compressée binaire du format BED, qui assure un archivage compact et indexable des données. L'extension pour ce format est .bb.

B. Description

Un fichier BED utilise une ligne par région génomique et de 3 à 12 colonnes séparées par des tabulations. Dans le cas de données d'alignements, le fichier BED ne contient que les éléments alignés. De plus, l'information de la séquence est perdue, il ne reste que sa position.

Les 3 colonnes requises sont :

- « **chrom** » : Nom du chromosome.

- « **chromStart** » : Position de départ de l'élément sur le chromosome. La première base du chromosome est numérotée 0.

- « **chromEnd** » : Position de fin de l'élément sur le chromosome. La position de fin n'est pas comprise dans l'élément.

Les 9 colonnes facultatives sont :

- « **name** » : Nom de l'élément décrit.
- « **score** » : Un score entre 0 et 1000.
- « **strand** » : Brin sur lequel est aligné l'élément (« + » ou « - »).
- « **thickStart** » : Par exemple le codon « start » pour un gène.
- « **thickEnd** » : Par exemple le codon « stop » pour un gène.
- « **itemRgb** » : Une valeur RGB (Rouge vert bleu, abrégé RVB), format de codage des couleurs permettant de représenter sur le « browser » l'élément selon la couleur définie.
- « **blockCount** » : Nombre de blocs (par exemple exon) dans la séquence.
- « **blockSizes** » : Liste de la taille des blocs, séparés par des virgules.
- « **blockStarts** » : Liste de la position du début de chaque bloc, séparée par des virgules.

Exemple 1 : Fichier de séquences.

chrI	26	66	HWUSI-EAS1656:15702	255	+
chrI	30	70	HWUSI-EAS1656:13745	255	-
chrI	32	72	HWUSI-EAS1656:9932	255	-

Exemple 2 : Fichier d'annotation.

chrI	1806	2169	YAL068C	0	-
chrI	2479	2707	YAL067W-A	0	+
chrI	5076	6237	SUT432	0	-

C. Utilisation

Les fichiers BED peuvent être manipulés facilement à l'aide de l'utilitaire BedTools (Quinlan and Hall, 2010, <http://bedtools.readthedocs.org/>) qui permet notamment de chercher des intersections, de fusionner ou de soustraire les intervalles génomiques de plusieurs fichiers. Le format bigBED est obtenu à partir d'un fichier .bed en utilisant la commande bedToBigBed.

4. WIG/bigWIG

A. Définition

Le format wiggle (WIG) est utilisé pour décrire des données continues, cartographiées sur le génome. Il permet de représenter des données de densité, par exemple l'occupation d'une protéine, sous forme de graphique. Les données sont sauvegardées sous ce format avec une extension .wig.

Le format bigWIG est la version compressée binaire du format WIG, qui assure un archivage compact et indexable des données. L'extension pour ce format est .bw.

B. Description

Un fichier WIG est composé de lignes de déclaration et de lignes comportant les données. Il existe deux options pour formater les données : « variableStep » et « fixedStep ». L'option variableStep est utilisée pour les données recouvrant des intervalles irréguliers. La ligne comportant les données contient 2 colonnes séparées par des tabulations contenant la position sur le chromosome et une valeur d'effectif (densité ou autre).

La ligne de déclaration commence avec le mot variableStep ou fixedStep, et est suivie du n° de chromosome. Le paramètre « span » est optionnel et indique le pas entre des positions spécifiées. Il permet d'écrire de façon plus succincte les données composées d'intervalles contiguës avec la même valeur de données. Le même pas est utilisé pour l'ensemble du fichier Wig.

Les lignes de données comprennent la position sur le chromosome, qui débute à 1, suivie de l'effectif correspondant, dans notre cas, la valeur de densité. Les positions n'ayant aucune valeur associée n'apparaissent pas.

Ex : Fichier de densité.

```
variableStep chrom=chrVI span=1
760 1
761 4
762 4
763 4
764 5
```

← Ligne de déclaration

} Lignes de données

C. Utilisation

Les fichiers WIG et bigWIG sont surtout destinés à la visualisation dans un navigateur de génome (« genome browser »). Le format WIG est obtenu à partir d'un fichier .bed en utilisant le programme WigMaker (développé par Arnaud Krebs et Tao Ye, IGBMC). Le format bigWIG est obtenu à partir d'un fichier .wig en utilisant l'utilitaire wigToBigWig (disponible sur le site de l'UCSC).

5. BEDGraph

A. Définition

Le format BEDGraph est utilisé pour décrire des données continues, cartographiées sur le génome. Il est similaire au format wiggle à la différence que les données exportées au format bedGraph gardent leur format original. Les données sont sauvegardées sous ce format avec une extension .bg.

B. Description

Tout comme le format BED, le format BEDGraph utilise une ligne par région génomique. Le fichier se compose de 4 colonnes séparées par des tabulations :

- « **chrom** » : Nom du chromosome.
- « **chromStart** » : Position de départ de l'élément sur le chromosome. La première base du chromosome est numérotée 0.
- « **chromEnd** » : Position de fin de l'élément sur le chromosome.
- « **data value** » : Effectif (valeur de densité ou autre).

Exemple : Fichier de densité.

chrI	0	1	111
chrI	1	5	112
chrI	5	7	113
chrI	7	9	114

C. Utilisation

Les BEDGraph sont surtout destinés à la visualisation dans un navigateur de génome. Le format BEDGraph est obtenu à partir d'un fichier .bed en utilisant la commande `genomeCoverageBed` de l'utilitaire `BedTools` (Quinlan and Hall, 2010, <http://bedtools.readthedocs.org/>). Tout comme les WIG, ils peuvent être compressés au format bigWIG.

II. TRAITEMENT DES DONNEES

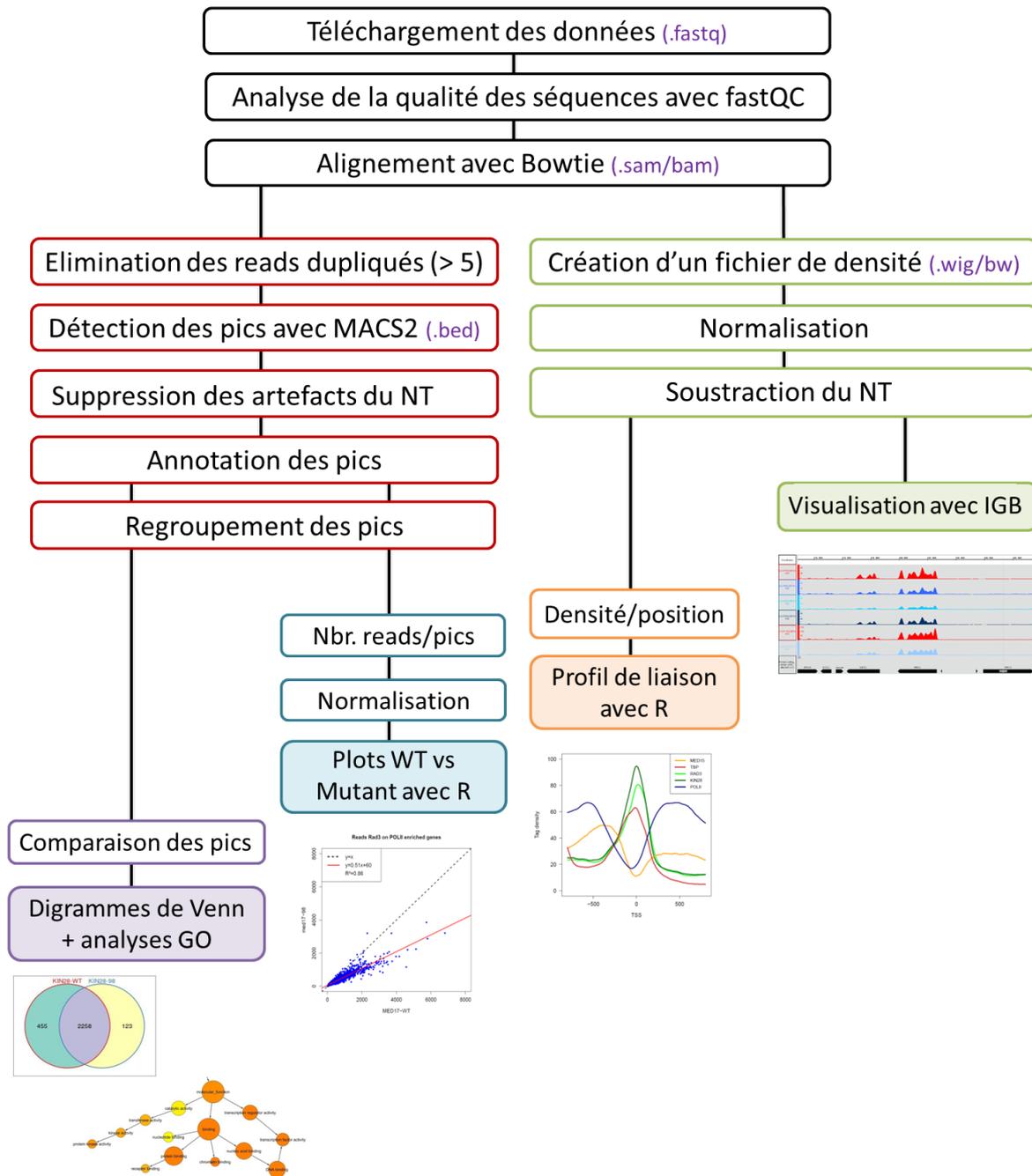


Figure 33 : « Workflow » de l'analyse ChIP-seq suivi pour ce projet.

Ce schéma présente les différentes étapes réalisées pour analyser les données de ChIP-seq et les différents types de résultats auxquels elles ont permis d'aboutir.

1. Analyses fastQC

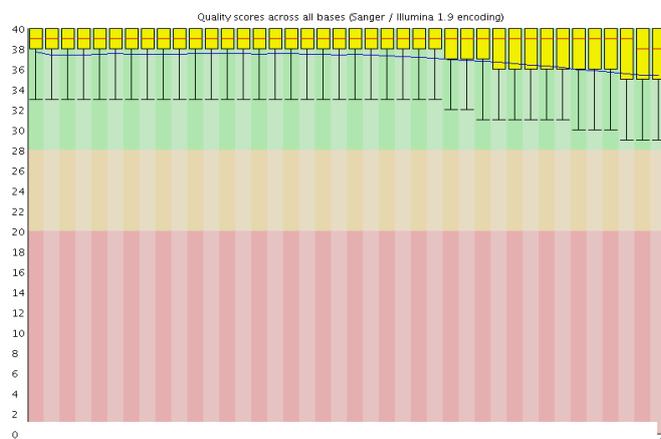
FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) est un programme permettant de contrôler la qualité des données NGS. Les fichiers en entrée peuvent être au format BAM, SAM ou FastQ. L'analyse FastQC permet de mettre rapidement en évidence différents problèmes pouvant apparaître dans les données de séquençage et dont il est important d'avoir connaissance avant de poursuivre l'analyse.

En sortie de FastQC, un rapport au format HTML est généré, qui intègre les résultats des différents tests de qualité. Le rapport contient 11 parties :

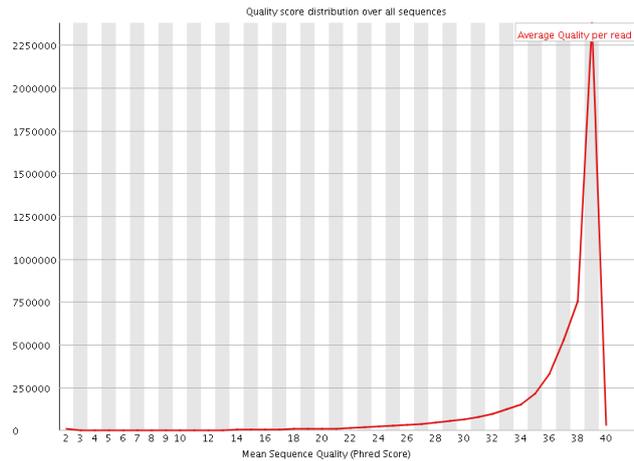
1- « Basic Statistics » : Cette section présente différentes informations regroupées dans un tableau : le nom du fichier d'origine, son type, le format d'encodage de la qualité, le nombre total de séquences analysées, le nombre de séquences filtrées, la longueur des séquences et le % de GC.

Measure	Value
Filename	Werner_120213_ATGTCA_L008.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	5143235
Filtered Sequences	0
Sequence length	40
%GC	43

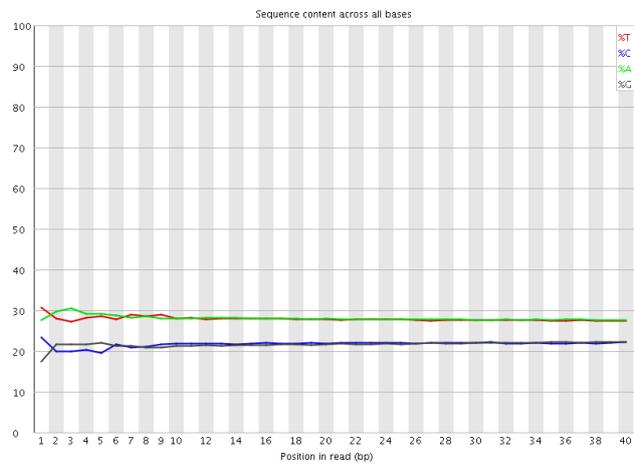
2- « Per Base Sequence Quality » : Ce module présente une vue d'ensemble de la qualité des séquences à chaque position. En général, la qualité du séquençage diminue avec l'augmentation de la durée de lecture. Ainsi pour les « runs » longs, il peut y avoir une diminution de la qualité en fin de séquence. Si la qualité tombe à un niveau trop faible, la façon la plus courante d'y remédier consiste à tronquer les lectures (« trim ») en fonction de leur qualité.



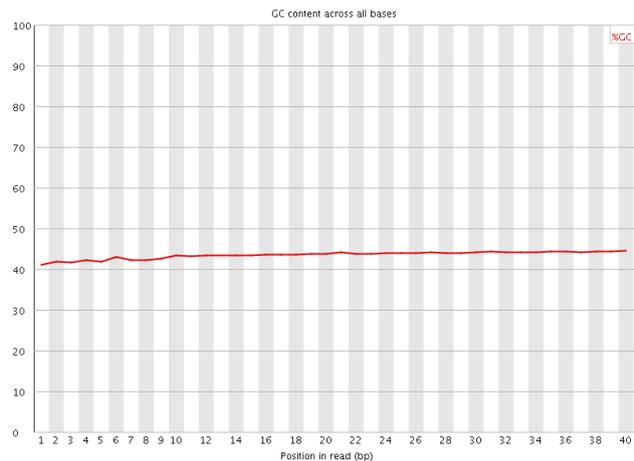
3- « Per Sequence Quality Scores » : Ce module présente la qualité moyenne de l'ensemble des séquences. Pour les « runs » longs, cette qualité peut être améliorée par une troncature des lectures en fonction de la qualité (cf. module précédent).



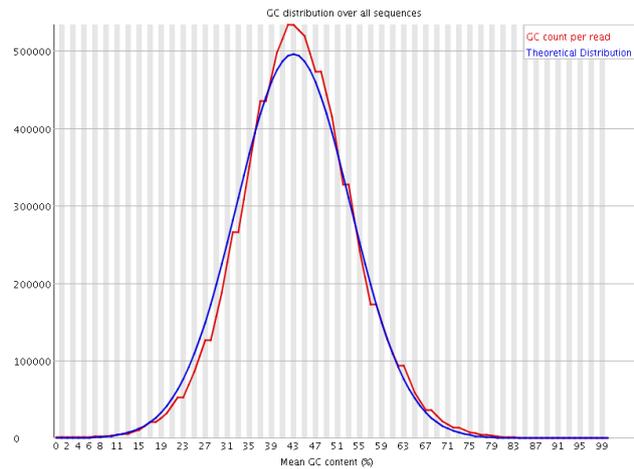
4- « Per Base Sequence Content » : Ce module présente la proportion de chacune des 4 bases de l'ADN à chaque position. Des erreurs peuvent apparaître à ce niveau dans divers cas : séquences surreprésentées (dimères d'adaptateurs), biais de fragmentation, biais de composition de la banque...



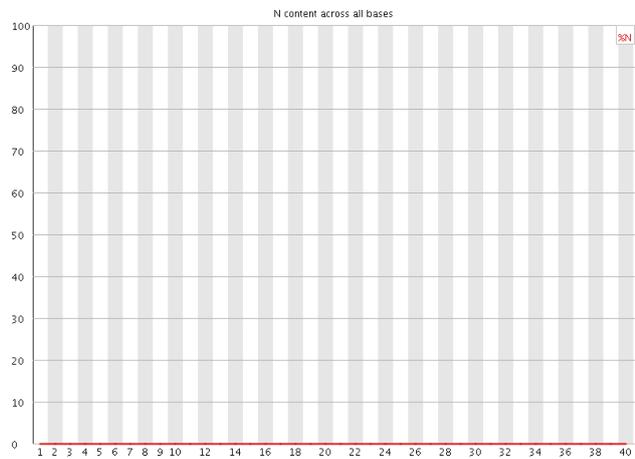
5- « Per Base GC Content » : Ce module présente la composition en GC à chaque position. En présence de séquences surreprésentées, une erreur peut apparaître à ce niveau.



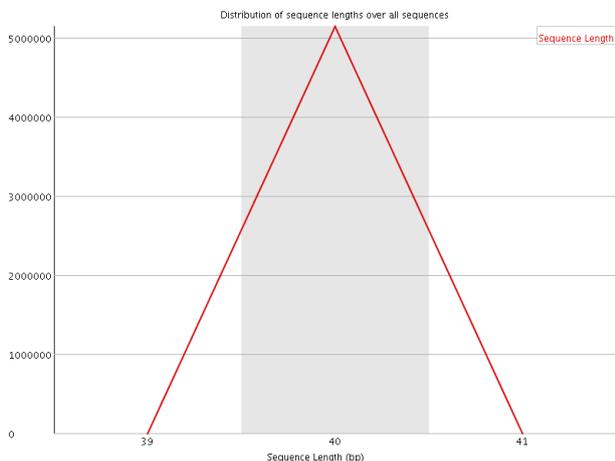
6- « Per Sequence GC Content » : Ce module présente la teneur en GC sur toute la longueur de chaque séquence et le compare à une distribution normale modélisée de la teneur en GC. Comme précédemment, en présence de séquences surreprésentées, une erreur peut apparaître à ce niveau.



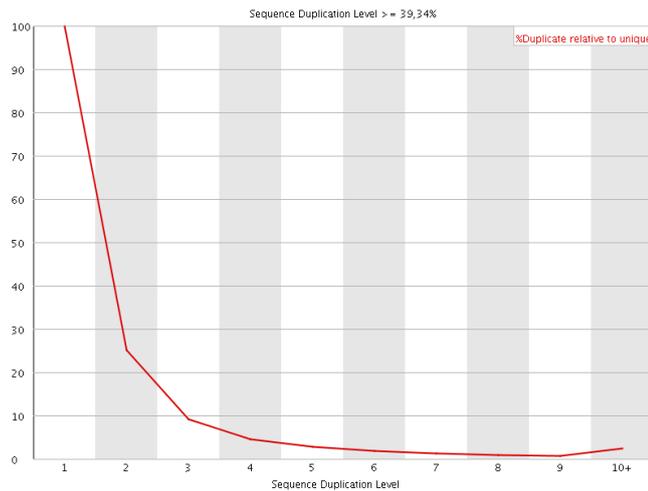
7- « Per Base N Content » : Si le séquenceur n'est pas en mesure d'appeler une base avec une confiance suffisante alors il placera un N au lieu d'une base. Ce module indique le % de N appelé à chaque position pour l'ensemble des séquences. La raison la plus courante de l'augmentation du nombre de N dans une séquence est une perte de qualité, ainsi pour les « runs » longs, les N sont fréquents en fin de séquence. Une troncature des lectures en fonction de leur qualité peut alors aider à résoudre ce problème.



8- « Sequence Length Distribution » : Certains séquenceurs génèrent des séquences de longueurs uniformes, mais d'autres peuvent contenir des lectures de longueurs variables. Ce module génère un graphique montrant la distribution de la taille des fragments. Pour certains séquenceurs, il est normal d'avoir des lectures de longueurs différentes, les erreurs à ce niveau peuvent alors être ignorées.



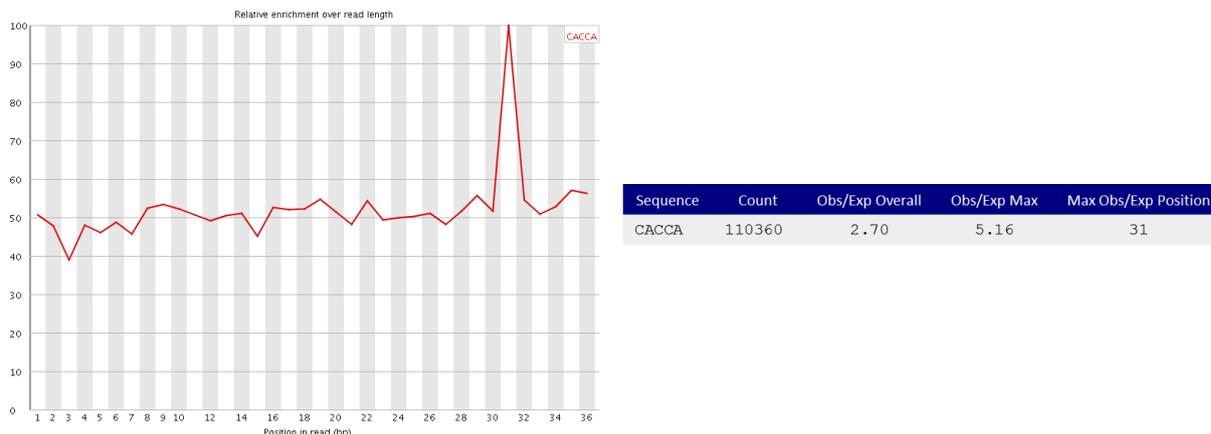
9- « Duplicate Sequences » : Ce module présente le nombre relatif de séquences avec différents degrés de duplication. Dans une bibliothèque diversifiée, la majorité des séquences n'apparaît qu'une fois. Un fort taux de duplication peut indiquer la présence d'empilements artéfactuels (par exemples des biais créés lors de l'amplification par PCR). Cependant, le taux de duplication dépend largement de la complexité du génome étudié et de la profondeur de séquençage. Aussi, dans certains cas, les erreurs à ce niveau peuvent être ignorées.



10- « Overrepresented Sequences » : Une bibliothèque normale contient un ensemble diversifié de séquences, sans séquence individuelle constituant une infime partie de l'ensemble. Si une seule séquence est très surreprésentées dans l'ensemble cela signifie soit qu'elle est biologiquement significative, ou que la bibliothèque est contaminée. Ce module présente l'ensemble des séquences qui représentent plus de 0,1 % du total. Pour chaque séquence surreprésentée le programme cherche des correspondances dans une base de données de contaminants communs et rend compte de la meilleure correspondance trouvée. Il s'agit souvent de séquences d'adaptateurs utilisés lors de la création des banques pour le séquençage.

Sequence	Count	Percentage	Possible Source
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCC	1116057	10.10	TruSeq Adapter, Index 5 (100% over 51bp)
NATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCC	53786	0.48	TruSeq Adapter, Index 5 (98% over 51bp)
GATTGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCC	50336	0.45	TruSeq Adapter, Index 5 (98% over 51bp)
CATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCC	31259	0.28	TruSeq Adapter, Index 5 (98% over 51bp)
GACCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCC	26617	0.24	TruSeq Adapter, Index 5 (98% over 51bp)

11- « Overrepresented K-mers » : Ce module compte l'enrichissement de chaque 5-mer au sein de la bibliothèque. Il calcule un niveau attendu de ce k-mer en fonction du contenu en bases de la bibliothèque dans son ensemble et calcule ensuite un ratio observé / attendu pour ce k-mer. En plus de présenter une liste de hits, ce module trace un graphe pour les 6 meilleurs résultats afin de montrer le modèle d'enrichissement de ce k-mer sur toute la longueur des séquences. Cela montre si l'enrichissement est général, ou s'il existe un motif de biais en différents points sur la séquence.



2. Alignement sur le génome de référence

A. Génération du fichier d'alignement avec Bowtie

La première étape de l'analyse des données de CHIP-seq consiste à identifier, pour chaque lecture issue du séquençage, la ou les localisations sur le génome de référence qui correspond parfaitement, ou presque, à un ou deux mésappariements (« mismatch ») près, à la séquence de la lecture. Plusieurs outils existent, permettant de réaliser cette étape. Nous avons choisi d'utiliser Bowtie (Langmead et al., 2009, <http://bowtie-bio.sourceforge.net>).

Bowtie est un outil d'alignement de lectures de petites tailles (quelques dizaines de nucléotides) sur de larges génomes. L'alignement des lectures est réalisé suivant une combinaison de différentes options permettant de déterminer quels alignements sont autorisés et doivent être reportés.

Usage :

```
bowtie -a --best --strata -v2 -m1 --trim5 5 --trim3 2 -q --sam
/genomes/SacCer/Bowtie/SacCer3 IP-Prot.fastq > IP-
Prot_align.sam
```

Paramètres :

-a : Bowtie reporte tous les alignements valides.

--best --strata : Bowtie reporte le meilleur alignement satisfaisant les critères imposés.

-v2 : Mode « v » pour ne pas tenir compte des valeurs de qualité des lectures, valeur « 2 » pour autoriser jusqu'à 2 discordances.

-m1 : Seules les lectures avec un positionnement unique sur le génome sont retenues.

-q : Spécifie le format du fichier d'entrée, ici FASTQ.

--sam : Spécifie le format du fichier de sortie, ici SAM.

--trim3 5 : Retire 5 nucléotides en 3' de la séquence avant de chercher à l'aligner.

--trim5 2 : Retire 2 nucléotides en 5' de la séquence avant de chercher à l'aligner.

/genomes/SacCer/Bowtie/SacCer3 : Chemin d'accès au génome de référence, ici de *S. cerevisiae* v.3 (avril 2011), indexé pour Bowtie.

En résumé, la combinaison de ces différents paramètres permet de reporter les lectures s'alignant de manière unique, avec 0, 1 ou 2 mésappariements sur le génome de la levure. Le meilleur alignement, contenant le plus faible nombre de mésappariement sera retenu.

A l'issue de l'alignement, Bowtie renvoie dans le terminal un certain nombre de statistiques :

```
# reads processed: 1000000
# reads with at least one reported alignment: 792353 (79.24%)
# reads that failed to align: 74575 (7.46%)
# reads with alignments suppressed due to -m: 133072 (13.31%)
# Reported 792353 alignments to 1 output stream(s)
```

Les paramètres utilisés pour aligner les données de ce projet sont résumés en annexe (cf. Annexe VI. I.).

B. Conversion en fichier BAM

Les fichiers SAM générés à l'issue de l'alignement sont très volumineux. Pour gagner de l'espace disque, ces fichiers sont convertis au format BAM à l'aide de l'utilitaire SamTools (Li et al., 2009, <http://samtools.sourceforge.net/>)

Usage :

```
samtools view -S -b IP-Prot_align.sam > IP-Prot_align.bam
```

Paramètres :

-S : Spécifie le format du fichier d'entrée, ici SAM.
-b : Spécifie le format du fichier de sortie, ici BAM.

Contrairement au format SAM, les fichiers BAM ne sont pas directement visualisables dans un terminal. Leur visualisation nécessite de passer par l'utilitaire SamTools :

```
samtools view IP-Prot_align.bam | more
```

Dans le cas de plusieurs fichiers de séquences que l'on souhaite regrouper en un seul fichier d'alignement, par exemple dans le cas où un seul échantillon a été séquencé en plusieurs fois, l'option « cat » de SamTools peut être utilisée :

```
samtools cat -o IP1+IP2-Prot_align.bam IP1-Prot_align.bam IP2-Prot_align.bam
```

Paramètres :

-o IP1+IP2-Prot_align.bam : Spécifie le nom du fichier de sortie.

Les fichiers BAM et SAM contiennent à la fois les séquences alignées et les séquences non alignées. Afin de connaître le nombre de séquences alignées uniquement, il faut éliminer les séquences notées « XM »:

```
samtools view IP-Prot_align.bam | egrep -v 'XM:' | wc -l
```

Paramètres :

egrep -v 'XM:' : La fonction egrep permet de rechercher une chaîne de caractère, ici « XM: », l'option -v signifie que les lignes contenant cette chaîne sont éliminées.
wc -l : Compte les lignes du fichier spécifié en entrée.

C. Elimination des lectures dupliquées

Avant la construction des banques pour le séquençage, les échantillons d'ADN subissent quelques cycles d'amplification par PCR. Cette étape peut créer des biais, notamment des empilements de lectures à la même localisation sur le génome, qui vont gêner l'analyse des données. Afin d'éviter ces empilements artificiels, les lectures dupliquées doivent être éliminées. Cette étape peut être réalisée à l'aide de l'utilitaire SamTools, à partir d'un fichier BAM trié.

1- Tri :

```
samtools sort IP-Prot_align.bam IP-Prot_align_sorted.bam
```

2- Suppression des lectures dupliquées

```
samtools rmdup -s IP-Prot_align_sorted.bam IP-Prot_rmdup.bam
```

Paramètres :

-s : spécifie que les lectures sont « single end » car par défaut cette commande ne fonctionne qu'avec des lectures « paire-end ».

Cette commande permet de ne conserver qu'une seule lecture avec la même localisation sur le génome (mêmes coordonnées de début et de fin). La lecture avec la meilleure qualité est conservée. A l'issue de la commande « rmdup », SamTools affiche les statistiques concernant le nombre de lectures dupliquées :

```
# 4479408 / 7977387 = 0.5615 in library
```

Cependant, dans le cas du génome de la levure, la suppression de toutes les lectures dupliquées n'est pas justifiée. En effet, ce génome n'étant constitué que de 12 Mb, un séquençage ChIP-seq à 10 millions de lectures aboutira statistiquement à plusieurs lectures à la même localisation. C'est pourquoi nous avons décidé de conserver jusqu'à 5 lectures dupliquées, afin de ne pas perdre en profondeur de séquençage, tout en éliminant les empilements artéfactuels. La commande « rmdup » de SamTools ne permettant pas de sélectionner le nombre de duplicats à conserver, nous avons créé notre propre programme : « rm_dup_reads » (développé par Olivier Alibert, iRCM) :

```
rm_dup_reads -n5 IP-Prot_align_sorted.bam > IP-Prot_rmdup5.sam
```

Paramètres :

-n5 : Permet de spécifier le nombre de lectures à conserver à une même position. Les lectures avec le meilleur score de qualité sont conservées.

Par la suite, afin de pouvoir effectuer une comparaison quantitative du nombre de lectures dans nos différents échantillons, nous avons cependant conservé l'ensemble des séquences, quel que soit leur taux de réplification.

3. Conversion au format BED

Suite à l’alignement, il peut être nécessaire de convertir les données en fichier BED, qui est utilisé en entrée de nombreux programmes, tels que seqMINER (Ye et al., 2011) ou WigMaker (développés par Arnaud Krebs et Tao Ye, IGBMC). Cette conversion peut être réalisée grâce à l’utilitaire BedTools (Quinlan and Hall, 2010, <http://bedtools.readthedocs.org/>).

Usage :

```
bam2bed IP-Prot_rmdup.bam > IP-Prot_rmdup.bed
```

A l’issue de cette étape, il peut arriver que les coordonnées de certaines lectures dépassent la taille du chromosome. La commande « bedClip » permet de corriger ce problème :

```
bedClip IP-Prot_rmdup.bed /chromosomes/sacCer3.chrom.sizes IP-Prot_rmdup_clip.bed
```

Paramètres :

/chromosomes/sacCer3.chrom.sizes : Chemin d’accès au fichier spécifiant la taille des chromosomes.

4. Création d’un fichier de densité

Afin de visualiser les profils de liaison des différents facteurs analysés par CHIP, il faut quantifier leur taux d’occupation par position sur l’ensemble du génome. Cette étape est réalisée par la création d’un fichier de densité WIG ou BEDGraph qui décompte les lectures sur des intervalles définies (« bins ») (Figure 34).

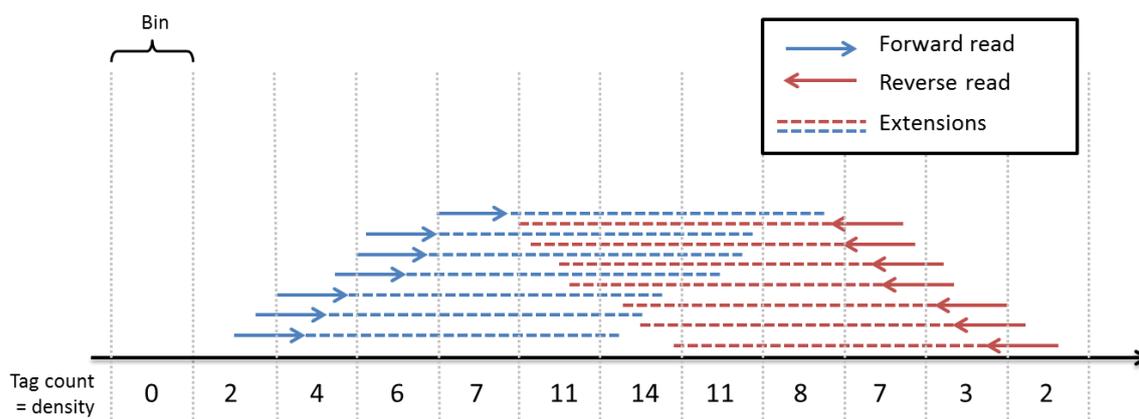


Figure 34 : Description de la création d’un fichier de densité

A. Extension des lectures

Avant séquençage, la chromatine a été fragmentée jusqu'à ce que la majorité des fragments aient une taille d'environ 200 pb. L'ADN a ensuite été sélectionné selon cette taille, or, seule une extrémité de ce fragment a par la suite été séquencée. En étendant la lecture, nous allons donc recréer artificiellement le fragment original. Les lectures sont étendues d'un certain nombre de bases à leur extrémité 3' en tenant compte du brin sur lequel elles sont alignées. La taille de l'extension peut être calculée directement par les logiciels de « peakcalling » tels que MACS (cf. II.5. de ce chapitre) ou à l'aide d'utilitaires tels que HOMER (Heinz et al., 2010, <http://homer.salk.edu/>). L'extension peut ensuite être réalisée avec l'utilitaire Bedtools.

Usage :

```
slopBed -r 150 -l 0 -s -i IP-Prot_clip.bed -g
/chromosomes/sacCer3.chrom.sizes > IP-Prot_clip_ext150.bed
```

Paramètres :

- r : nombre de bases à ajouter au coordonné « end » de la lecture.
- l : nombre de bases à soustraire au coordonné « start » de la lecture.
- s : Spécifie qu'il faut tenir compte du brin pour orienter les lectures.
- i : Fichier d'entrée
- g : Chemin d'accès au fichier indiquant la taille des chromosomes.

B. Création du fichier de densité

a. Format WIG

Le programme « WIGMaker » utilisé pour créer les fichiers de densité au format WIG a été écrit par Tao Ye et Arnaud Krebs de l'IGBMC à Strasbourg. Ce programme utilise en entrée le format BED. Il réalise automatiquement l'extension des lectures décrite à l'étape précédente.

Usage :

```
java -Xmx3000m -jar WigMaker.jar IP-Prot_clip.bed 150 10
```

Paramètres :

- Xmx3000m : Mémoire allouée.
- 150 : Taille de l'extension des lectures
- 10 : Taille des intervalles pour lesquels la densité est calculée.

b. Format BEDGraph

L'utilitaire BedTools peut également créer des fichiers de densité, au format BEDGraph. Ce programme utilise en entrée le format BED. L'extension des lectures doit être réalisée préalablement. L'avantage de ce format par rapport au WIG est qu'il peut également reporter les positions de densité nulle.

Usage :

```
genomeCoverageBed -d -bga -i IP-Prot_clip_ext150.bed -g
/chromosomes/sacCer3.chrom.sizes > TP-Prot_clip_ext150.bg
```

Paramètres :

- i : Fichier d'entrée
- g : **Chemin d'accès au fichier indiquant la taille des chromosomes.**
- d : Reporte la densité pour chaque position du génome (par nucléotide).
- bga : Option « bg » pour avoir la sortie au format BEDGraph et option « a » pour reporter également les positions de densité nulle.

c. Conversion au format bigWIG

Les fichiers WIG et BEDGraph générés à l'issue de l'alignement sont très volumineux. Afin de fluidifier la visualisation dans l'explorateur de génome, ces fichiers sont convertis au format bigWIG.

Usage (à partir d'un fichier .wig) :

```
wigToBigWig -clip wig_for_IP-Prot.wig
/chromosomes/sacCer3.chrom.sizes bigWig_for_IP-Prot.bw
```

Paramètres :

- clip : Permet de raccourcir les coordonnées des lectures en fonction de la taille des chromosomes, si cela n'a pas déjà été fait précédemment.

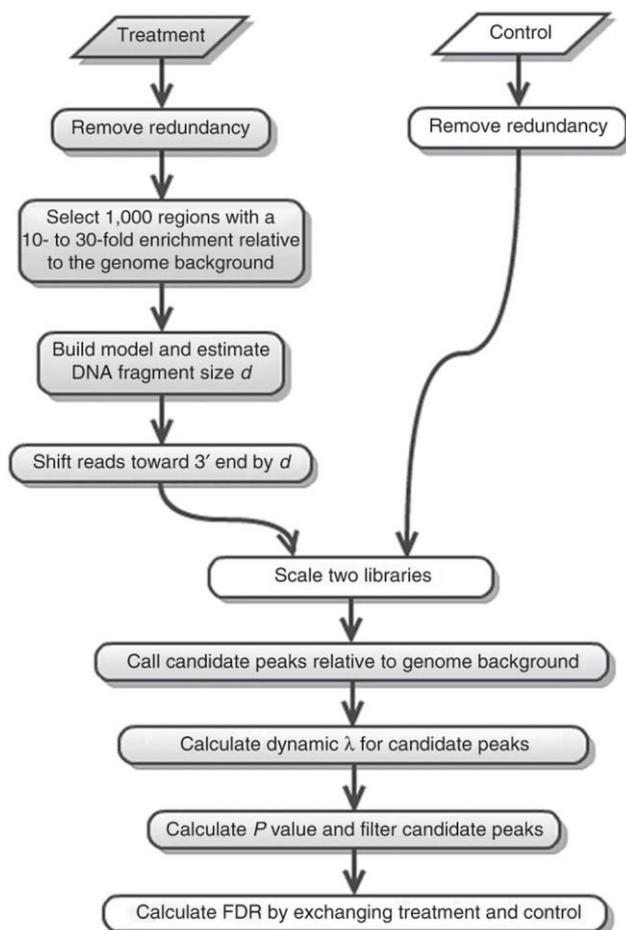
Usage (à partir d'un fichier .bg) :

```
bedGraphToBigWig IP-Prot_clip_ext150.bg /chromosomes/
sacCer3.chrom.sizes IP-Prot_clip_ext150.bg
```

5. Détection des régions liées avec MACS

L'étape suivante de l'analyse des données de ChIP-seq consiste en la recherche des régions enrichies en lectures, issues du séquençage de l'ADN immunoprécipité. Ces régions présentant une forte densité de lectures (pics), par rapport à un contrôle, sont considérées comme liées par le facteur étudié. Différents logiciels permettant de détecter ces régions liées ont été développés (« peakcallers »). Le programme que nous avons choisi d'utiliser pour ce projet est MACS (« Model-based Analysis of ChIP-Seq ») (<https://github.com/taoliu/MACS>) spécialement conçu pour identifier les sites de liaison des facteurs de transcription. Un premier essai avec la version initiale de ce logiciel, la seule disponible à ce moment, nous permettait de détecter seulement quelques pics. Cependant la version 2 sortie entre temps, qui utilise une nouvelle méthode de calcul de l'enrichissement par rapport au contrôle, nous a permis d'obtenir un nombre de pics satisfaisant. Nous avons donc choisi d'utiliser MACS2 pour la détection des régions enrichies en lectures dans nos échantillons.

A. Principe



MACS reçoit en entrée l'échantillon traité (ChIP) et un contrôle facultatif, de type « Input » ou « mock IP ». MACS fonctionne en 4 grandes étapes : (1) suppression des lectures dupliquées, (2) ajustement de la position des lectures, (3) calcul de l'enrichissement dans les pics et (4) estimation du FDR (« false discovery rate »). Le workflow détaillé de MACS est schématisé ci-dessous (Figure 35). La section qui suit sera consacrée à la description des étapes clés de l'algorithme de MACS.

Figure 35: Workflow de MACS (Feng et al., 2012).

1- **Suppression des lectures dupliquées** : Afin d'éviter les empilements de lectures à la même localisation sur le génome, qui pourraient être détectés, à tort, comme des pics, il est indispensable d'éliminer les lectures dupliquées. Par défaut MACS ne retient pas plus d'une lecture par position, cependant l'option « keep-dup » permet de spécifier si l'on souhaite conserver les lectures dupliquées.

2- **Ajustement de la position des lectures en estimant la taille des fragments** : Les lectures « forward » et « reverse » proviennent des extrémités opposées du fragment d'ADN immunoprécipité. Le séquençage étant orienté de l'extrémité du fragment vers le milieu, ceci crée un profil bi-modale avec une sous-représentation des lectures au centre, où se trouve lié le facteur.

Par défaut, MACS scanne le génome avec une fenêtre de deux fois la taille de la chromatine fragmentée (définie par l'option « bandwidth ») et sélectionne 1000 régions présentant un fort taux d'enrichissement par rapport au bruit de fond (entre 10 et 30 fois, défini par l'option « mfold »). MACS crée un profil pour chaque orientation « forward » ou « reverse » de ces 1000 régions et estime ensuite la taille des fragments originaux (appelée d) en calculant la distance entre les sommets des profils « reverse » et « forward » (Figure 36). La moitié de cette valeur correspond au « peak shift ». Cette distance doit être estimée afin de combiner au mieux les deux profils opposés en un seul. Les lectures sont alors étendues en 3' jusqu'à atteindre cette valeur « d ».

Dans la majorité des cas, cette procédure, appelée création du modèle de « shift », permet d'estimer de manière fiable, la taille des fragments. Cependant, quand l'enrichissement est très faible, ou que le facteur étudié se fixe sur une large région d'ADN, MACS peut échouer à construire le modèle. Dans ce cas, la valeur « d », estimée en fonction de la taille des fragments sélectionnés lors de la préparation des banques, peut être saisie manuellement en utilisant l'option « shiftsize ».

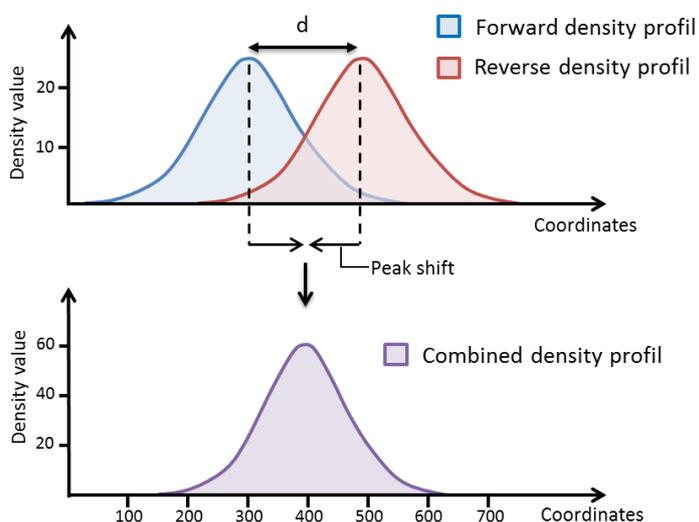


Figure 36 : Illustration de la méthodologie de MACS pour estimer la taille des fragments séquencés.

3- **Calcul de l'enrichissement dans les pics** : En se basant sur les positions ajustées à l'étape précédente, MACS scanne le génome avec une fenêtre de taille 2d pour identifier les régions significativement enrichies par rapport au bruit de fond. Les fenêtres significatives se chevauchant sont alors fusionnées pour former une liste de régions candidates. Par défaut, quand les échantillons ChIP et contrôle ont été séquencés à des profondeurs différentes, MACS abaisse les valeurs de l'échantillon avec la profondeur la plus importante. MACS assigne ensuite une valeur-p à chaque région candidate. Si cette valeur est inférieure au seuil défini par l'utilisateur (option « pvalue »), la région est reportée comme un pic dans le fichier de sortie.

4- **Estimation du FDR** : Quand un échantillon contrôle est disponible, MACS peut estimer le taux de fausse découverte (FDR ou valeur-q) pour chaque pic en échangeant les échantillons contrôle et ChIP. Il va alors identifier le pic dans le contrôle en utilisant les mêmes paramètres que ceux utilisés pour l'échantillon ChIP. Comme le contrôle n'est pas censé présenter d'enrichissement, tout pic trouvé par MACS dans cet échantillon peut être considéré comme un faux positif. Pour un seuil de valeur-p défini, MACS peut ainsi calculer le FDR en divisant le nombre de pics dans le contrôle passant ce seuil par le nombre de pics dans l'échantillon ChIP passant ce même seuil.

B. Usage

```
macs2 callpeak -t IP-Prot.bed -c Input-Prot.bed -n Name -g
12495682 --keep-dup=all -m 2,100 --bw 300 -q 0.05
```

Paramètres utilisés dans ce projet :

-t/--treatment : Fichier traité (ChIP).

-c/--control : Fichier contrôle (Input ou mock).

-n/--name : Nom utilisé pour créer les fichiers de sortie.

-g/--gsize : Taille du génome de référence.

--keep-dup : Contrôle la façon dont MACS gère les lectures dupliquées. L'option « all » conserve toutes les lectures. Si un nombre entier est spécifié, MACS conserve au maximum ce nombre de lecture à chaque position.

--m/--mfold : Défini les taux d'enrichissement minimum et maximum pour la sélection des pics servant à construire le modèle de « shift ». Les valeurs par défaut sont 10 pour le minimum et 30 pour le maximum.

--bw : Largeur de la bande (« bandwidth ») utilisée pour scanner le génome lors de la création du modèle de « shift ».

-q/--qvalue : Seuil de détection (FDR) des pics significatifs. Valeur par défaut : 0.01.

Autres paramètres :

-p/--pvalue : Seuil de détection (valeur-p) des pics significatifs. Dans la version 2 de MACS, si cette valeur est spécifiée, elle est utilisée à la place du FDR, qui n'est alors plus calculé.

--nomodel : Cette option permet d'omettre l'étape de création du modèle de « shift ».

Dans ce cas, la taille du « shift » doit être précisée manuellement.

--shiftsize : Quand l'option « nomodel » est spécifiée, MACS utilise ce paramètre pour ajuster la position des lectures. Cette valeur correspond à la moitié de la taille des fragments estimée.

--call-summits : Quand cette option est active, MACS ré-analyse le profil de chaque pic afin d'en déterminer le ou les sommets. Les sous-pics détectés auront, dans le fichier de sortie, les mêmes coordonnées que la région initiale dont ils sont issus mais des scores et position de sommet différents.

C. Résultats

MACS fournit en sortie une liste de pics, définis selon les critères imposés. Cette liste est générée dans plusieurs formats de fichiers, contenant différentes informations :

- **NAME_peaks.xls** : Fichier tabulé contenant des informations sur les pics détectés. Ces informations incluent : le nom du chromosome, les positions de début et de fin du pic, la longueur du pic, la position du sommet du pic, le nombre de lectures au sommet du pic, la valeur-p, le taux d'enrichissement dans ce pic et le FDR.

- **NAME_peaks.bed** : Fichier BED contenant la localisation des pics. Ce fichier peut être visualisé directement dans un explorateur de génome.

- **NAME_peaks.narrowPeak** : Fichier BED contenant la localisation des pics ainsi que la localisation des sommets des pics.

- **NAME_summits.bed** : Fichier BED contenant la localisation du sommet de chaque pic. Ce fichier peut être visualisé directement dans un explorateur de génome.

- **NAME_model.r** : Script R qui peut être utilisé pour générer une image PDF du modèle de shift en utilisant cette commande : R --vanilla < NAME_model.r.

Dans notre cas, le fichier qui nous intéresse est NAME_peaks.bed qui contient la liste de tous les pics significatifs au format BED et peut facilement être visualisé à l'aide d'un explorateur de génome (Figure 37).

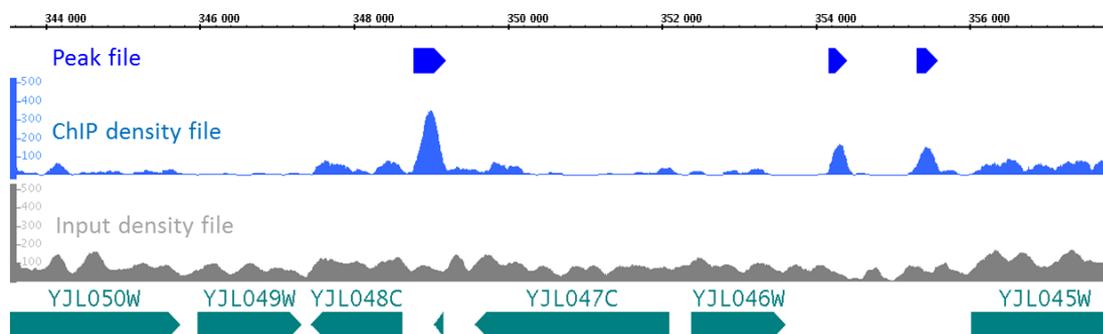


Figure 37 : Exemple de visualisation d'un fichier de pics issu de MACS2, superposé aux fichiers de densité des échantillons CHIP et Input.

D. Division des pics

Dans le cas d'un génome relativement petit comme celui de *S. cerevisiae*, l'analyse par MACS n'est pas assez fine pour distinguer les pics très proches. Dans ce cas, il peut être utile d'ajouter une étape de division des pics. Cette étape peut être réalisée à l'aide du programme PeakSplitter (<http://www.ebi.ac.uk/research/bertone/software>) développé par le groupe de Paul Bertone à l'EBI de Cambridge (Salmon-Divon et al., 2010). PeakSplitter utilise en entrée un fichier BED contenant la liste des pics significatifs triés par chromosome et coordonnée de début du pic et le fichier de densité correspondant au format WIG.

Usage :

```
PeakSplitter.jar -p Peaks.bed -w wig_for_IP-Prot.wig
-o Output-folder -u http://www.ensembl.org/das/Saccharomyces
_cerevisiae.EF4.reference -v 0.7 -c 20
```

Paramètres :

- p/--peakFile : Fichier contenant les pics au format BED.
- w/--wigFile : Fichier de densité au format WIG.
- o/--outDir : Dossier où seront extraits les fichiers de sortie.
- u/--url : URL DAS (« Distributed annotation system ») pour l'organisme étudié.
- c/--cutoff : Seuil de détection en hauteur. Seuls les sous-pics avec au moins ce nombre de lectures dans leur sommet seront conservés.

-v/--valley : Valeur indiquant la profondeur du creux requis entre deux pics pour les séparer. Par exemple une valeur de 0.5 indique que la profondeur du creux doit être au moins égale à la moitié de la hauteur du sommet du pic pour que les pics soient séparés.

PeakSplitter crée en sortie un fichier .subpeaks.bed contenant 5 colonnes : nom du chromosome, coordonnée de début du sous-pic, coordonnée de fin du sous-pic, nombre de lectures dans le pic et position du sommet du pic.

Les paramètres de MACS2 et PeakSplitter utilisés pour l'analyse de nos données sont indiqués en annexe (cf. Annexe VI. II).

6. Annotation des pics

Afin de pouvoir interpréter biologiquement les résultats de ChIP-seq, les pics identifiés dans les divers échantillons doivent être annotés. De nombreux logiciels existent qui permettent d'associer un pic au gène ou au TSS le plus proche. Cependant, le génome très condensé de *S. cerevisiae* rend souvent difficile l'association d'un pic à un gène en particulier. Ceci est notamment vrai pour les protéines liées dans les régions de gènes divergeants pour lesquels la résolution de nos données de ChIP-seq ne permet pas de distinguer 2 pics. De plus, l'annotation des TSS chez la levure reste à ce jour imprécise et incomplète, il n'est donc pas toujours possible de se baser sur ces données pour annoter les pics.

A. Annotation relative au nom de la région intergénique ou de l'ORF

Pour les analyses quantitatives ne nécessitant pas d'associer l'enrichissement à un gène précis, les pics ont simplement été annotés par rapport à la région intergénique dans laquelle ils se trouvent, selon la nomenclature suivante (cf. III de ce chapitre) :

- **Prom_Gene** : Dans le cas de gènes en tandem. Le gène est celui situé en aval de la région intergénique annotée.

- **Div_Gene1-Gene2** : Dans le cas de gènes divergeants. Le gène 1 correspond à celui situé en amont de la région intergénique annotée et le gène 2 celui situé en aval.

- **Conv_Gene1-Gene2** : Dans le cas de gènes convergeants. Le gène 1 correspond à celui situé en amont de la région intergénique annotée et le gène 2 celui situé en aval.

En ce qui concerne les pics de Pol II, ceux-ci ont été annotés de la même manière mais par rapport aux ORFs, après élimination des gènes « dubious » (cf. III de ce chapitre).

Pour associer chaque pic à une région (intergénique ou ORF), nous avons utilisé la commande « intersectBed » de BedTools :

```
intersectBed -wa -a Prom-Div_intergenic_regions.bed
-b Peaks.bed | sort -u > Peaks_Annot-Prom-div.bed
```

Paramètres :

- wa** : Pour chaque chevauchement, reportée l'entrée originale du fichier A.
- a** : Fichier A au format BED, à comparer au fichier B.
- b** : Fichier B au format BED.
- sort -u** : Permet de supprimer les doublons.

B. Annotation relative au gène le plus proche

Pour les données de Rad2, ou pour la comparaison qualitative des pics de différents échantillons, nous avons procédé à une annotation plus précise à l'aide du « package » de R ChIPpeakAnno (Zhu et al., 2010, <http://www.bioconductor.org>). L'annotation est effectuée en deux temps : (1) Annotation exhaustive de l'ensemble des pics par rapport au fichier d'annotation contenant les gènes codants et non codants de *S. cerevisiae* : commande « annotatePeakInBatch ». (2) Annotation des pics situés dans les régions de gènes divergeants par rapport au fichier d'annotation contenant les gènes codants uniquement : commande « peaksNearBDP ».

```
# Chargement des donnees et conversion au format RangedData :
library (ChIPpeakAnno)
Peaks = read.table("Peaks_IP-Prot.bed")
Peaks.ranged = BED2RangedData(Peaks)
Annot.codncod = read.table("coding_noncoding.bed")
Annot.codncod.ranged = BED2RangedData(Annot.codncod)
Annot.cod = read.table("coding.bed")
Annot.cod.ranged = BED2RangedData(Annot.cod)

# Annotation exhaustive de tous les pics :
annotated.peaks_All= annotatePeakInBatch(myPeakList = Peaks.ranged,
AnnotationData= Annot.codncod.ranged, multiple=TRUE, select="all",
output="both", PeakLocForDistance="middle",
FeatureLocForDistance = "TSS")

# Annotation des pics dans les regions de genes divergents :
BDP.Peaks= peaksNearBDP(myPeakList = Peaks.ranged, AnnotationData =
Annot.cod.ranged, MaxDistance = 600, PeakLocForDistance="middle",
FeatureLocForDistance = "TSS")
```

```
# Export des tables de donnees :
DF_annotated.peaks_All= as.data.frame(annotated.peaks_All)
write.table(DF_annotated.peaks_All,
"Annot_Peaks_codncod_All.txt", sep="\t", row.names=F, quote=F)
DF_BDP.peaks = as.data.frame(BDP.Peaks$peaksWithBDP)
write.table(DF_BDP.peaks, "BDP.Peaks_protcod.txt", sep="\t", row.names
=F, quote=F)
```

a. Commande « annotatePeaksInBatch

Cette commande permet d’obtenir la distance entre un pic et l’annotation la plus proche. Les fichiers d’annotation (« AnnotationData ») et de pics (« myPeakList ») donnés en entrée doivent être convertis au format RangedData avec la commande « BED2RangedData » pour pouvoir être utilisé dans ce programme.

Paramètres :

multiple : L’option « TRUE » exporte toutes les annotations pouvant correspondre à un même pic. L’option « FALSE » exporte une annotation maximum pour chaque pic.

select : L’option « all » exporte tous les pics correspondant à une annotation donnée. L’option « first » exporte seulement le 1^{er} pic, « last » seulement le dernier et « arbitrary » un pic au hasard.

output : L’option « nearestStart » exporte les annotations les plus proches après calcul de la distance entre le début du pic et le début de l’annotation. L’option « overlapping » exporte les annotations chevauchant le pic avec un « gap » maximum défini par l’option « maxgap ». L’option « both » exporte les deux annotations.

PeakLocForDistance : Spécifie la localisation sur le pic à partir de laquelle sera calculée la distance avec les annotations. L’option « middle » utilise le milieu du pic et l’option « start » le début du pic.

FeatureLocForDistance : Spécifie la localisation sur l’annotation à partir de laquelle sera calculée la distance avec les pics. L’option « middle » utilise le milieu de l’annotation, l’option « start » le début de l’annotation, l’option « TSS » utilise le début de l’annotation quand celle-ci est positionnée sur le brin « + » et la fin quand elle est positionnée sur le brin « - », l’option « geneEnd » utilise la fin de l’annotation quand celle-ci est positionnée sur le brin « + » et le début quand elle est positionnée sur le brin « - ».

Fichier de sortie :

```

I    9226  9377  152  0001  SUT001      0001 +   SUT001      9369
9601 overlapStart  -67   8   NearestStart
I    24826 24927 102  0002  YAL063C      0002 -   YAL063C      23999
27968 inside      3092 827  Overlapping
I    24826 24927 102  0002  YAL063C-A    0002 -   YAL063C-A    22394
22685 upstream  -2191 2141 NearestStart

```

Le fichier de sortie contient 15 colonnes donnant différents types d'informations :

- Coordonnées et nom du pic (en bleu ci-dessus) : chromosome, « start », « end », longueur, nom et brin.
- Nom du pic et de l'annotation associée (en rouge ci-dessus).
- Coordonnées et nom de l'annotation (en vert ci-dessus) : nom, « start », « end ».
- Informations sur l'association pic-annotation (en orange ci-dessus) : position de l'annotation par rapport au pic, distance entre le pic et l'annotation calculée selon les paramètres définis par l'utilisateur, distance la plus courte entre le pic et l'annotation, localisation utilisée pour le calcul de la distance la plus courte.

Les facteurs étudiés étant censés se lier sur la région promotrice des gènes codants, les pics se trouvant en position « downstream » ou « overlapEnd » sont éliminés. Cependant, pour les CUTs et SUTs on conserve tout car on ne sait pas exactement où démarre leur transcription.

```

awk '$6 !~ /^Y/ || $12 !~ /overlapEnd/' Annot_Peaks_codncod_All.txt
| awk '$6 !~ /^Y/ || $12 !~ /downstream/' >
Annot_Peaks_codncod_All_noDownstr_noOverlapEnd.txt

```

Paramètres :

awk '\$6 !~ /^Y/ || \$12 !~ /overlapEnd/' : Enlever les lignes qui contiennent un « Y » en début de colonne 6 (= gène codant une protéine) et « overlapEnd » en colonne 12.

awk '\$6 !~ /^Y/ || \$12 !~ /downstream/' : Enlever les lignes qui contiennent un « Y » en début de colonne 6 (= gène codant une protéine) et « downstream » en colonne 12.

Pour pouvoir être visualisé dans un explorateur de génome, le fichier de sortie est converti au format BED :

```

awk -F'\t' 'BEGIN {OFS="\t"} {print("chr"$1),$2,$3,$8,"0",$7}'
Annot_Peaks_codncod_All_noDownstr_noOverlapEnd.txt >
Annot_Peaks_codncod_All_noDownstr_noOverlapEnd.bed

```

b. Commande « peaksNearBDP »

Cette commande permet d'obtenir les pics près de promoteurs divergeants et leur pourcentage. Les fichiers d'annotation (« AnnotationData ») et de pics (« myPeakList ») donnés en entrée doivent être préalablement convertis au format RangedData avec la commande « BED2RangedData » pour pouvoir être utilisé dans ce programme.

Paramètres :

PeakLocForDistance : cf. II.6.B.a.

FeatureLocForDistance : cf. II.6.B. a.

MaxDistance : Spécifie le « gap » maximum autorisé entre un pic et l'annotation la plus proche.

Fichier de sortie :

```
I    58326 58777 452 0010 YAL044W-A 0010 + YAL044W-A 57517
57850 downstream 1035 476 NearestStart
I    60976 61377 402 0011 YAL042W 0011 + YAL042W 61315
62563 overlapStart -139 62 NearestStart
I    76076 76527 452 0015 YAL035W 0015 + YAL035W 76426
79435 overlapStart -124 101 NearestStart
```

Le fichier de sortie est du même format que précédemment. De même, les pics se trouvant en position « downstream » ou « overlapEnd » sont éliminés :

```
egrep -v 'downstream|overlapEnd' BDP.Peaks_protcod.txt >
BDP.Peaks_protcod_noDownstr_noOvlapEnd.txt
```

Pour pouvoir être visualisé dans un explorateur de génome, ce fichier est converti au format BED :

```
awk -F'\t' 'BEGIN {OFS="\t"} {print ("chr"$1), $2, $3, $8, "0", $7}'
BDP.Peaks_protcod_noDownstr_noOvlapEnd.txt >
BDP.Peaks_protcod_noDownstr_noOvlapEnd.bed
```

c. Concaténation des données

Pour obtenir le fichier d'annotation final, les résultats obtenus à l'issue des deux commandes « annotatePeaksInBatch » et « peaksNearBDB » doivent être concaténés :

```
cat Annot_Peaks_codncod_All_noDownstr_noOvlapEnd.bed
BDP_Peaks_protcod_noDownstr_noOvlapEnd.bed | sort -k7,7n | uniq >
Annot_Peaks_codncod_All_BDP_Peaks_protcod.bed
```

7. Comparaison qualitative des Pics et analyses GO

Les données de pics annotés peuvent être utilisées pour comparer qualitativement les enrichissements, entre différents échantillons. Les résultats peuvent alors être représentés sous forme de diagrammes de Venn. Ce type de représentation peut facilement être obtenu dans R à l'aide de différents « packages ». Dans notre cas, nous avons utilisé le script `plotVennDiag` écrit par Olivier Alibert qui permet de calculer l'intersection entre différentes listes de gènes puis affiche le résultat sous forme de diagramme de Venn (Figure 38).

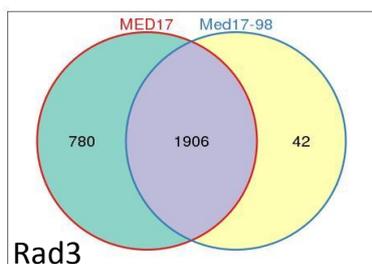


Figure 38 : Comparaison des pics de la protéine Rad3 entre la souche MED17 sauvage et le mutant *med17-98*.

Les listes de gènes contenus dans les différentes régions du diagramme (intersection par exemple) peuvent facilement être exportées et copiées sur le site SGD (<http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl>) pour effectuer une analyse d'ontologie qui permet de classer les gènes selon leur fonction moléculaire ou les processus biologiques dans lesquels ils sont impliqués.

8. Extraction des données de densité et du nombre de lectures

A. Densité par régions

Afin de pouvoir comparer les profils d'occupation des différentes protéines autour d'un point d'intérêt (TSS ou pic), il faut extraire les données de densité sur ces régions. Pour ce faire, nous avons utilisé le script « `get_tag_density` » écrit par Olivier Alibert de l'IRCM d'Evry.

Usage :

```
get_tag_density -s -f IP-Prot.bw TSS.bed > Tag-density_IP-Prot_TSS.tsv
```

Paramètres :

- s : Tiens compte du brin pour orienter les données de densité.
- f : Fichier de densité.
- TSS.bed : Fichier BED contenant les coordonnées des régions pour lesquelles on souhaite récupérer les données de densité, ici entre -800 et +800 autour des TSS.

Fichier de sortie :

```
chrI      170911    172511    YAR018C  0        -
442;444;447;450;451;451;451;451;451;451;451;451;451;451;451;
451;451;451;451;451;451;451;451;451;451;451;452;452;453;453;
453;453;454;454;454;456;457;458;458;458;460;460;461;461;461;
461;463;463;463;463;466;466;466;466;466;466;466;466;466;466;
466;466;466;466;466;466;466;466;466;466;466;466;466;466;
466;466;466;466;466;466;466;466;466;466;466;466;466;466;
466;466;466;466;466;466;466;466;469;469;468;469;469;469;
469;469;470;470;470;470;469;469;469;469;469;469;469;469;469;
468;467;467;467;467;464;462;460;460;456;450;445;441;435;431;
431;430;427;427;427;427;427;427;427;427;427;427;427;427;
425;419;419;413;411;408;408;408;408;407;407;407;406;406;405;...
449.98 56847 372 470
171301;171302;171303... 128486
```

chrI 170911 172511 YAR018C - : Informations extraites du fichiers BED indiquant les coordonnées de la région.

442;444;447... : Densité de la protéine par position sur l'ensemble de la région définie en entrée (ici entre -800 et +800 autour du TSS du gène YAR018C).

449.98 : moyenne des densités sur la région.

56847 : somme des densités sur la région.

372 470 : Densité minimale et maximale.

171301;171302;171303... : Positions génomiques où la densité est maximale.

128486 : Moyenne des positions de densité maximale.

Pour dessiner le profil de densité d'une protéine sur une région, il faut récupérer les données de densité pour chaque position (colonne 7) :

```
cut -f4,7 Tag-density_IP-Prot.tsv | sed 's/;/\t/g' > density-values-IP-Prot.tsv
```

Paramètres :

cut -f4,7 : Permet de récupérer les colonnes 4 (nom de la région) et 7 (densités par position) du fichier d'entrée.

sed 's/;/\t/g' : Remplace les « ; » par des tabulations dans le fichier de sortie.

Puis la représentation graphique est effectuée dans R (

Figure 39):

```
# Chargement des donnees :
MED15= read.table("density-values_IP-MED15_TSS.tsv", sep='\t')
RAD2= read.table("density-values_IP-RAD2_TSS.tsv", sep='\t')

# Calcul de la moyenne par position (=2x800 positions) :
MED15m= colMeans(MED15[,2:1601])
RAD2m= colMeans(RAD2[,2:1601])

# Positions de l'axe des abscisses :
x=c(-800:-1,1:800)

# Parametres graphiques et plot des valeurs de densite de chaque facteur :
par(cex.axis=1.5,cex.lab=1.5,cex.main=1.5,oma=c(0,2,0,0))
plot(x,MED15m,type='l',col='blue',xlab="TSS",ylab='',ylim=c(0,30),yaxt='n')
lines(x,RAD2m,col='red')
axis(2,las=2)
mtext(side=2,line=4,'Tag density',cex=1.5)

# Titre et legende :
title('TSS',font.main=2)
legend("topright", legend = c("MED15","RAD2"), col =
c("blue","red"),lwd=c(1.5),bty="o",cex=1)

# Sauvegarde en PDF :
dev.copy(pdf,"TSS_RAD2-MED15.pdf")
dev.off()
```

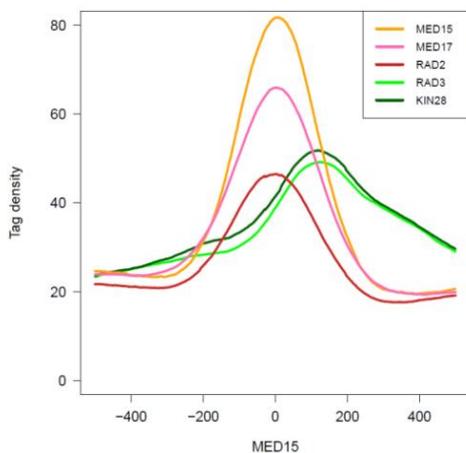


Figure 39 : Exemple de profil moyen de liaison des protéines Med15, Med17, Rad2, Rad3 et Kin28 sur une région comprise entre -500 et +500 autour des pics du Médiateur.

Ce type d'analyse peut également être effectué directement dans le logiciel SeqMiner (Ye et al., 2011, <http://sourceforge.net/projects/seqminer/>) qui prend en entrée une liste de fichiers de séquences au format BED, BAM ou SAM dont il analyse la distribution sur des intervalles de tailles définies autour d'une liste de coordonnées (TSS, Pics...) fournie au format BED. La distribution est alors représentée sous forme de courbe (similaire à ce qui est présenté ci-dessus) ou de « heatmap ». Un « heatMap » est une représentation graphique où les valeurs sont représentées par des couleurs (Figure 40). A la base, il s'agit d'une matrice à deux-dimensions où sont regroupées les valeurs. Plus une valeur est forte, plus la couleur utilisée pour l'illustrer sera foncée, et inversement, à une faible valeur sera associée une intensité de couleur plus faible.

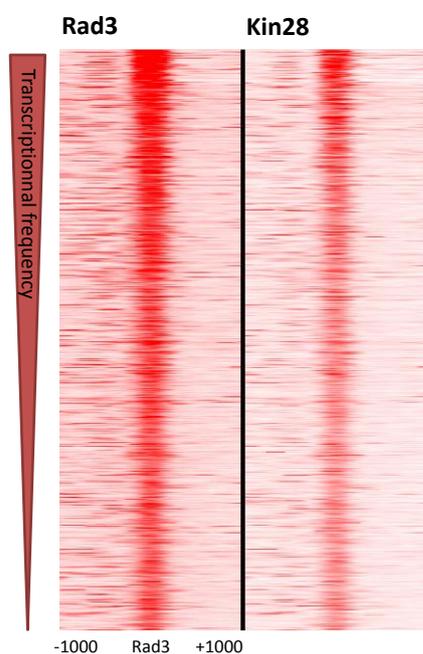


Figure 40 : Exemple de « heatmap » des protéines Rad3 et Kin28 sur une région comprise entre -1000 et +1000 autour du TSS.

Chaque ligne représente un intervalle de 2000 pb autour d'un TSS. Ces derniers sont classés selon la fréquence de transcription du gène qui leur est associé.

B. Nombre de lectures par régions

Pour pouvoir comparer le taux d'occupation d'une protéine entre les souches sauvages et mutantes, sur une région donnée, il faut calculer le nombre de lectures dans ces régions. Pour ce faire, nous avons utilisé la commande « coverageBed » de BedTools à partir du fichier BED contenant toutes les lectures, quel que soit leur taux de duplication.

Usage :

```
coverageBed -abam IP-Prot.bam -b POL2_peaks.bed > reads-on-POL2-peaks_IP-Prot.bed
```

Paramètres :

-abam : Fichier A au format BAM dans lequel les lectures vont être comptées.
-b : Fichier B au format BED, contenant les régions dans lesquels on veut compter les lectures.

Fichier de sortie :

```
chrI 31566 32940 YAL062W 0 + 2342 1351 1374 0.98
chrI 48563 51707 YAL051W 0 + 563 2031 3144 0.64
chrI 80709 81951 YAL034C 0 - 923 1177 1242 0.94
```

Les premières colonnes (en bleu) correspondent aux régions issues du fichier B. Puis coverageBed reporte :

- Le nombre de lecture dans A qui chevauchent la région issue de B (en rouge).
- Le nombre de positions de la région issue de B qui sont recouvertes par au moins une lecture de A.
- La longueur de la région issue de B.
- La fraction de la région issue de B qui est recouverte par au moins une lecture de A.

Pour comparer les taux d'occupation entre deux souches différentes, il faut récupérer les colonnes donnant le nombre de lecture pour ces deux échantillons après concaténation des 2 fichiers (colonnes 7 et 17) :

```
cat reads-on-POL2-peaks_IP-WT-Prot.bed reads-on-POL2-peaks_IP-
Mutant-Prot.bed | cut -f4,7,17 | sed -i
"1i\ Region\t WT\t Mutant" > reads-on-POL2-peaks_IPs-
WTvsMutant.tsv
```

Paramètres :

cut -f4,7,17 : Permet de récupérer les colonnes 4 (nom de la région), 7 (nombre de lectures dans le fichier WT) et 17 (nombre de lectures dans le fichier Mutant).
sed -i "...\t ...\t ..." : Rajoute une ligne d'en tête contenant les noms des 3 colonnes du fichier de sortie, séparés par des tabulations.

On obtient un fichier contenant 3 colonnes : nom de la région, valeur pour la souche sauvage et valeur pour la souche mutante.

Puis la représentation graphique est effectuée dans R (

Figure 41):

```
# Chargement des donnees :
table= read.table("reads-on-POL2-peaks_IPs-WTvsMutant.tsv",
header=T, row.names=1)

# Plot des reads dans chaque regions WT versus Mutant :
plot (table$WT,table$Mutant,pch=20,xlab="MED17-WT", ylab="med17-
mutant",main="Reads on POL II peaks",
xlim=c(0,40000),ylim=c(0,40000),col="blue",cex.main=1.5,
cex.lab=1.5,cex.axis=1.5)

# Droite de regression :
line= lm(table[,2]~table[,1])
abline (line, lwd=2, col="red")
abline (0,1,lwd=2,lty=2)

# Affichage des parametres de la droite de regression (pente,
ordonnee a l'origine et R²) :
summary (line)

# Legende :
legend ("topleft", legend =
c("y=x", "y=0.82x+136", "R²=0.83"), col=c("black", "red", "black"), cex=1
.5, lty=c(2,1,0), lwd=2, bg="white")

# Sauvegarde en PDF :
dev.copy(pdf, "Figure_Prot-WTvsMutant_Reads-on-Pol2-peaks.pdf")
dev.off()
```

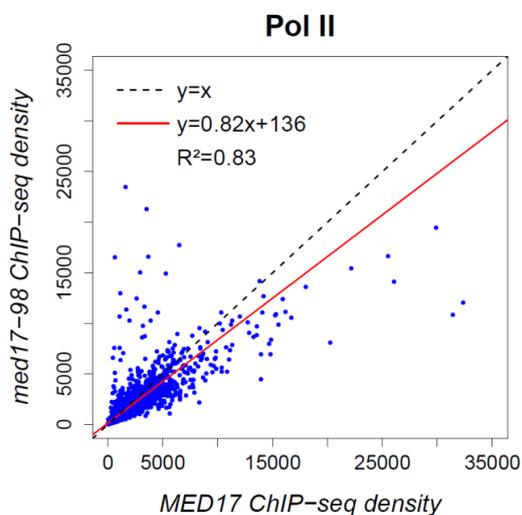


Figure 41: Exemple de comparaison des taux d'occupation de la Pol II entre une souche sauvage et une souche mutante, sur les régions liées par cette protéine.

III. BASES DE DONNEES D'ANNOTATION

Les bases de données utilisées pour réaliser l'annotation des analyses CHIP-seq de ce projet ont été téléchargées depuis les sites UCSC, Ensembl et SGD. Pour certaines annotations (TSS, CUT...) nous avons également utilisés des données de la littérature. Afin d'avoir un fichier d'annotation le plus exhaustif possible, les données provenant de ces différentes sources ont été compilées puis converties, le plus souvent au format BED.

Le tableau ci-après récapitule les tables utilisées dans ce projet, leur contenu et leur provenance.

saccharomyces_cerevisiae.gff	Sequence features of <i>S. c.</i> and related information such as GO annotation. http://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyces_cerevisiae.gff
Saccharomyces_cerevisiae.EF4.66.gtf	Gene sets for <i>S.c.</i> This file includes annotations of both coding and non-coding genes. ftp://ftp.ensembl.org/pub/release-75/gtf/saccharomyces_cerevisiae
SGD_features.tab	Information on current chromosomal features in SGD, including Dubious ORFs. Also contains coordinates of intron, exons, and other subfeatures that are located within a chromosomal feature. http://downloads.yeastgenome.org/curation/chromosomal_feature/SGD_features.tab
Chromosome_length.tab	Lengths of the chromosomes, with RefSeq identifiers. http://downloads.yeastgenome.org/curation/chromosomal_feature/chromosome_length.tab
NotFeature.fasta	DNA sequences which are not within the following types of features: ORF, ARS, CEN, rRNA, tRNA, snRNA, snoRNA, RNA genes, LTRs, telomeric elements and transposons. http://downloads.yeastgenome.org/sequence/S288C_reference/intergenic/NotFeature.fasta.gz
Mmc3.xls	Sequence and localization of TATA and TATA-like elements from Basehoar et al., 2004.
Rhee_SuppData1.xls	Localisation of TSS and TATA elements from Rhee and Pugh, 2012.
Nature07728-s2.xls	CUT and SUT coordinates from Xu et al., 2009. Also contains TSS and TES of coding genes.

Tableau 14 : Tables utilisées pour l'annotation des données de CHIP-seq.

IV. NORMALISATION DES DONNEES

1. Méthode

Le nombre total de lectures alignées n'est pas le même d'un échantillon à l'autre, ce qui peut amener à considérer un enrichissement sur les gènes de certains échantillons simplement parce qu'ils contiennent globalement plus de lectures que les autres. De plus, des biais peuvent subvenir lors des différents traitements subis par les échantillons lors de leur préparation. Ainsi, pour pouvoir effectuer une comparaison quantitative il faut donc rendre les échantillons comparables, en les remettant à la même échelle.

De nombreuses méthodes et outils existent pour normaliser les données de NGS. La plus simple consiste à diviser chaque échantillon par le nombre total de ses lectures (RPM « reads per million »). Une variante tient également compte de la longueur des gènes (RPKM « Reads per kilobase and per million »). Le script Polyphemus utilise, quant à lui, une normalisation par quantiles (Mendoza-Parra et al., 2011) et DESeq une normalisation par régression linéaire (Anders and Huber, 2010). Cependant, aucune de ces méthodes ne convient dans le cas où nous voulons vérifier si l'effet observé est global. En effet, si la protéine étudiée possède, chez une souche mutante, un défaut de liaison global, la quantité de chromatine récupérée dans cet échantillon sera moindre, et par conséquent le nombre de lectures obtenu plus faible que dans la souche sauvage. Si nous utilisons une des méthodes citée précédemment pour normaliser ces deux échantillons, les valeurs pour la souche sauvage vont être diminuées et celles pour la souche mutante augmentées, éliminant la différence de liaison globale de la protéine entre les deux souches.

Après avoir testé plusieurs méthodes de normalisation pour cette étude, nous avons finalement décidé de nous baser sur les valeurs de qPCR pour comparer nos données (Tableau 15). Pour ce faire, nous avons sélectionné une dizaine de régions (ORF ou promoteurs), que nous avons utilisées pour les analyses par qPCR. Nous avons ensuite calculé le rapport de quantification entre la souche mutante ($V_{qPCR\text{Mutant}}$) et la souche sauvage ($V_{qPCR\text{WT}}$) dans ces différentes régions : $R_{qPCR} = V_{qPCR\text{Mutant}} / V_{qPCR\text{WT}}$. En parallèle, nous avons regardé dans les données de ChIP-seq correspondantes, le nombre de lectures dans ces mêmes régions et calculé le rapport du nombre de lecture entre la souche mutante ($V_{seq\text{Mutant}}$) et la souche sauvage ($V_{seq\text{WT}}$) : $R_{seq} = V_{seq\text{Mutant}} / V_{seq\text{WT}}$. Nous avons alors pu calculer le coefficient à appliquer aux données de ChIP-seq pour que le rapport soit le même que pour les résultats de qPCR : $Q = R_{qPCR} / R_{seq}$. La médiane des coefficients obtenus pour les différentes régions analysée a ensuite été utilisée pour normaliser les données de ChIP-seq.

Les valeurs de normalisation obtenues pour nos données sont récapitulées en annexe (cf. Annexe VI. III.).

	WT seq	WT qpcr	Mutant seq	Mutant qpcr	Mut/WT seq	Mut/WT qpcr	qpcr/seq
ADH1_O12	9790	13	9596	9,63	0,98	0,74	0,76
PYK1_O12	7689	15	6063	9,13	0,79	0,61	0,77
PHO84_O56	3102	3,16	1235	1,06	0,40	0,34	0,84
							Mediane 0,77

Tableau 15 : Exemple de calcul du facteur de normalisation utilisé pour comparer les données de CHIP-seq des souches sauvages et mutantes.

2. Application

Ces facteurs de normalisation ont ensuite été appliqués aux souches mutantes sur les tables contenant le nombre de lectures par région, obtenues avec coverageBed :

```
awk -F'\t' 'BEGIN {OFS="\t"} {print $1,$2,($3*0.77)}'
reads-on-POL2-peaks_IPs-WTvsMutant.tsv >
reads-on-POL2-peaks_IPs-_norm-qPCR.tsv
```

Ils ont également servi à normaliser les fichiers de densité avant leur visualisation dans l'explorateur de génome. Cette normalisation a été effectuée sur les fichiers au format BEDGraph, plus facile à manipuler que les fichiers WIG.

```
awk -F'\t' 'BEGIN {OFS="\t"} {print $1,$2,$3($4*0.77)}'
NT_densityFile.bg > NT_densityFile_norm.bg
```

V. ELIMINATION DES ARTEFACTS

Lors de la visualisation de nos données dans un explorateur de génome, nous avons noté la présence d'artefacts dans les ORFs des gènes très exprimés. Ces artefacts sont également présent dans le contrôle « mock IP » (IP non spécifique avec anticorps anti-HA sur une souche non étiquetée HA que nous appellerons « NT » pour « non-tagged ») mais pas dans l'échantillon d'ADN total (« Input »). Nous avons donc conclu à un artefact de CHIP. Ces observations ont par la suite été confirmées par la publication de plusieurs papiers sur le sujet qui suggèrent une influence de la structure de la chromatine sur l'efficacité des IPs (Jeronimo and Robert, 2014; Park et al., 2013; Teytelman et al., 2009, 2013).

Bien que peu nombreux (~200 régions), ces artefacts se trouvent au niveau de gènes très exprimés, que nous n'avons pas voulu exclure de nos analyses. Afin de supprimer ces biais, nous avons donc décidé de soustraire les valeurs correspondant à l'échantillon NT de nos échantillons CHIP.

3. Données issues du « peakcalling »

La recherche de pics avec MACS2 est effectuée sur un fichier BAM ou BED. C'est fichiers ne contenant pas de données numériques, il n'est pas possible d'effectuer une soustraction du NT directement sur ces données. La solution aurait pu être d'effectuer le « peakcalling » par rapport au NT au lieu de l'échantillon d'ADN total, cependant le fichier NT contient très peu de lectures et n'est donc pas le contrôle le plus fiable pour ce genre d'analyse, laissant passer de nombreux pics non significatifs.

Nous avons donc décidé d'effectuer le « peakcalling » de manière classique par rapport à l'échantillon d'ADN total (fichier pics versus « Input ») et en parallèle par rapport au NT (fichier pics versus NT). Nous avons ainsi obtenu une première série de pics contenant les biais d'IP et une seconde série, ne les contenant pas. Nous avons ensuite identifié les régions artefactuelles en effectuant un « peakcalling » très stringeant sur l'échantillon NT versus « Input » à l'aide du package de R « bayespeak » (<http://www.bioconductor.org/packages/release/bioc/html/BayesPeak.html>). Puis, nous avons localisé ces régions artefactuelles dans le fichier de pics versus « Input » de notre échantillon d'intérêt et nous les avons éliminées. Nous avons ensuite récupéré ces mêmes régions dans le fichier de pics versus NT et nous les avons réintégré dans le fichier précédent. Ainsi notre fichier final est globalement un « peakcalling » effectué avec l'ADN total comme contrôle, sauf dans les régions artefactuelles où le « peakcalling » est effectué en prenant l'échantillon NT comme contrôle. Les étapes de réalisation sont détaillées ci-dessous :

```
# Identification des pics dans la souche NT avec BayesPeak :
bayespeak(job.size=1000000,bin.size=350,job.overlap=50)
summarize.peaks(threshold =0.9, method="lowerbound")

# Soustraction des pics artefactuels dans le fichier des pics IP vs Input :
intersectBed -a Peaks-IPvsINPUT.bed -b Peaks-NT.bed -wa -f 0.2 >
Artefact_IP

# Suppression de ces regions :
intersectBed -v -a Peaks-IPvsINPUT.bed -b Artefact_IP.bed -wa >
Peaks-IP_withoutArtefact.bed

# Récupération de ces regions dans le fichier des pics IP vs NT :
intersectBed -a Peaks-IPvsNT.bed -b Artefact_IP.bed -wa > Peaks-
IP_FromNT.bed

# Concaténation des pics vs Input en dehors des régions artefactuelles, et des pics vs NT dans ces mm régions :
cat Peaks-IP_withoutArtefact.bed Peaks-IP_FromNT.bed | sort -k1,1 -
k2,2n > Peaks_corrNT.bed
```

Paramètres :

-f : Fraction minimum de A requise comme chevauchement (Par exemple $-f 0.2 = 20\%$ de A).

-v : Reporter seulement les entrées de A qui ne chevauchent aucune entrée de B.

4. Fichiers de densité

Afin de pouvoir visualiser les fichiers de densité sans les artefacts observés précédemment, nous avons normalisé le fichier de densité du NT par rapport aux valeurs calculées par qPCR, puis nous avons soustrait ce fichier normalisé aux fichiers de densité des échantillons CHIP avec la commande « `getDifferentialBedGraph` » de l'utilitaire HOMER (<http://homer.salk.edu>) :

```
# Normalisation du fichier de densité NT au format BEDGraph :
awk -F'\t' 'BEGIN {OFS="\t"} {print $1,$2,$3($4*0.4)}'
NT_densityFile.bg > NT_densityFile_norm.bg

# Soustraction du fichier de densité NT normalisé au fichier de densité CHIP :
getDifferentialBedGraph.pl NT_densityFile_norm.bg
IP_densityFile.bg diff > IP_densityFile_subtrNT.bg
```

BIBLIOGRAPHIE

- Aalfs, J.D., and Kingston, R.E. (2000). What does “chromatin remodeling” mean? *Trends Biochem. Sci.* **25**, 548–555.
- Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protić, M., Hübscher, U., Egly, J.M., and Wood, R.D. (1995). Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* **80**, 859–868.
- Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* **17**, 103–112.
- Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405–409.
- Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393.
- Akoulitchev, S., Mäkelä, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIH kinase activity in transcription by RNA polymerase II. *Nature* **377**, 557–560.
- Akoulitchev, S., Chuikov, S., and Reinberg, D. (2000). TFIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**, 102–106.
- Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599–610.
- Amir-Zilberstein, L., Ainbinder, E., Toubé, L., Yamaguchi, Y., Handa, H., and Dikstein, R. (2007). Differential regulation of NF-kappaB by elongation factors is determined by core promoter type. *Mol. Cell. Biol.* **27**, 5246–5259.
- Andel, F., 3rd, Ladurner, A.G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-dimensional structure of the human TFIID-IIA-IIIB complex. *Science* **286**, 2153–2156.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106.
- Andrau, J.-C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C.P. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol. Cell* **22**, 179–192.
- Ansari, S.A., and Morse, R.H. (2013). Mechanisms of Mediator complex action in transcriptional activation. *Cell. Mol. Life Sci. CMLS* **70**, 2743–2756.
- Ansari, S.A., He, Q., and Morse, R.H. (2009). Mediator complex association with constitutively transcribed genes in yeast. *Proc. Natl. Acad. Sci. U.S.A* **106**, 16734–16739.
- Araújo, S.J., Tirorde, F., Coin, F., Pospiech, H., Syväoja, J.E., Stucki, M., Hübscher, U., Egly, J.M., and Wood, R.D. (2000). Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIH, and modulation by CAK. *Genes Dev.* **14**, 349–359.
- Araújo, S.J., Nigg, E.A., and Wood, R.D. (2001). Strong functional interactions of TFIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Mol. Cell. Biol.* **21**, 2281–2291.
- Archambault, J., and Friesen, J.D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.* **57**, 703–724.
- Archambault, J., Lacroute, F., Ruet, A., and Friesen, J.D. (1992). Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol. Cell. Biol.* **12**, 4142–4152.
- Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23**, 841–851.
- Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6964–6968.
- Armache, K.-J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134.
- Armstrong, J.A., Bieker, J.J., and Emerson, B.M. (1998). A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**, 93–104.
- Asada, S., Choi, Y., Yamada, M., Wang, S.-C., Hung, M.-C., Qin, J., and Uesugi, M. (2002). External control of Her2 expression and cancer cell growth by targeting a Ras-linked coactivator. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12747–12752.

- Asturias, F.J. (2004). RNA polymerase II structure, and organization of the preinitiation complex. *Curr. Opin. Struct. Biol.* *14*, 121–129.
- Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., and Kornberg, R.D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* *283*, 985–987.
- Atanassov, B.S., Evrard, Y.A., Multani, A.S., Zhang, Z., Tora, L., Devys, D., Chang, S., and Dent, S.Y.R. (2009). Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. *Mol. Cell* *35*, 352–364.
- Auble, D.T., Hansen, K.E., Mueller, C.G., Lane, W.S., Thorner, J., and Hahn, S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* *8*, 1920–1934.
- Bäckström, S., Elfving, N., Nilsson, R., Wingsle, G., and Björklund, S. (2007). Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. *Mol. Cell* *26*, 717–729.
- Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* *32*, 878–887.
- Baek, H.J., Malik, S., Qin, J., and Roeder, R.G. (2002). Requirement of TRAP/mediator for both activator-independent and activator-dependent transcription in conjunction with TFIID-associated TAF(II)s. *Mol. Cell. Biol.* *22*, 2842–2852.
- Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *J. Biol. Chem.* *281*, 15172–15181.
- Balamotis, M.A., Pennella, M.A., Stevens, J.L., Wasylyk, B., Belmont, A.S., and Berk, A.J. (2009). Complexity in transcription control at the activation domain-mediator interface. *Sci. Signal.* *2*, ra20.
- Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell* *116*, 699–709.
- Bastien, J., Adam-Stitah, S., Riedl, T., Egly, J.M., Chambon, P., and Rochette-Egly, C. (2000). TFIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J. Biol. Chem.* *275*, 21896–21904.
- Baumli, S., Hoepfner, S., and Cramer, P. (2005). A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J. Biol. Chem.* *280*, 18171–18178.
- Beaudenon, S.L., Huacani, M.R., Wang, G., McDonnell, D.P., and Huibregtse, J.M. (1999). Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *19*, 6972–6979.
- Belakavadi, M., and Fondell, J.D. (2010). Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation. *Mol. Cell. Biol.* *30*, 2437–2448.
- Bergink, S., Salomons, F.A., Hoogstraten, D., Groothuis, T.A.M., de Waard, H., Wu, J., Yuan, L., Citterio, E., Houtsmuller, A.B., Neefjes, J., et al. (2006). DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. *Genes Dev.* *20*, 1343–1352.
- Bernardes de Jesus, B.M., Bjørås, M., Coin, F., and Egly, J.M. (2008). Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Mol. Cell. Biol.* *28*, 7225–7235.
- Bernecky, C., Grob, P., Ebmeier, C.C., Nogales, E., and Taatjes, D.J. (2011). Molecular Architecture of the Human Mediator–RNA Polymerase II–TFIIF Assembly. *PLoS Biol.* *9*.
- Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. *Genome Biol.* *5*, R62.
- Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* *10*, 973–982.
- Berroteran, R.W., Ware, D.E., and Hampsey, M. (1994). The *sua8* suppressors of *Saccharomyces cerevisiae* encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to *sua7* (TFIIB) mutations. *Mol. Cell. Biol.* *14*, 226–237.
- Berti, L., Mittler, G., Przemek, G.K., Stelzer, G., Günzler, B., Amati, F., Conti, E., Dallapiccola, B., Hrabé de Angelis, M., Novelli, G., et al. (2001). Isolation and characterization of a novel gene from the DiGeorge chromosomal region that encodes for a mediator subunit. *Genomics* *74*, 320–332.
- Bertolaet, B.L., Clarke, D.J., Wolff, M., Watson, M.H., Henze, M., Divita, G., and Reed, S.I. (2001). UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nat. Struct. Biol.* *8*, 417–422.
- Bessho, T. (1999). Nucleotide excision repair 3' endonuclease XPG stimulates the activity of base excision repair enzyme thymine glycol DNA glycosylase. *Nucleic Acids Res.* *27*, 979–983.

- Bhatia, P.K., Verhage, R.A., Brouwer, J., and Friedberg, E.C. (1996). Molecular cloning and characterization of *Saccharomyces cerevisiae* RAD28, the yeast homolog of the human Cockayne syndrome A (CSA) gene. *J. Bacteriol.* *178*, 5977–5988.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. *Nature* *493*, 699–702.
- Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science* *280*, 298–301.
- Black, J.C., Choi, J.E., Lombardo, S.R., and Carey, M. (2006). A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Mol. Cell* *23*, 809–818.
- Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science* *281*, 60–63.
- Bleichenbacher, M., Tan, S., and Richmond, T.J. (2003). Novel Interactions Between the Components of Human and Yeast TFIIA/TBP/DNA Complexes. *J. Mol. Biol.* *332*, 783–793.
- Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* *11*, 1587–1598.
- Bohr, V.A., Smith, C.A., Okumoto, D.S., and Hanawalt, P.C. (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.
- Boiteux, S., and Jinks-Robertson, S. (2013). DNA Repair Mechanisms and the Bypass of DNA Damage in *Saccharomyces cerevisiae*. *Genetics* *193*, 1025–1064.
- Van den Boom, V., Citterio, E., Hoogstraten, D., Zotter, A., Egly, J.-M., van Cappellen, W.A., Hoeijmakers, J.H.J., Houtsmuller, A.B., and Vermeulen, W. (2004). DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *J. Cell Biol.* *166*, 27–36.
- Bosveld, F., van Hoek, S., and Sibon, O.C.M. (2008). Establishment of cell fate during early *Drosophila* embryogenesis requires transcriptional Mediator subunit dMED31. *Dev. Biol.* *313*, 802–813.
- Bouazoune, K., Mitterweger, A., Längst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. (2002). The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J.* *21*, 2430–2440.
- Boube, M., Faucher, C., Joulia, L., Cribbs, D.L., and Bourbon, H.M. (2000). *Drosophila* homologs of transcriptional mediator complex subunits are required for adult cell and segment identity specification. *Genes Dev.* *14*, 2906–2917.
- Boube, M., Joulia, L., Cribbs, D.L., and Bourbon, H.-M. (2002). Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* *110*, 143–151.
- Bourbon, H.-M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res.* *36*, 3993–4008.
- Bourbon, H.-M., Aguilera, A., Ansari, A.Z., Asturias, F.J., Berk, A.J., Bjorklund, S., Blackwell, T.K., Borggreffe, T., Carey, M., Carlson, M., et al. (2004). A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* *14*, 553–557.
- Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* *5*, 158–163.
- Boyer, T.G., Martin, M.E., Lees, E., Ricciardi, R.P., and Berk, A.J. (1999). Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* *399*, 276–279.
- Brower, C.S., Sato, S., Tomomori-Sato, C., Kamura, T., Pause, A., Stearman, R., Klausner, R.D., Malik, S., Lane, W.S., Sorokina, I., et al. (2002). Mammalian mediator subunit mMED8 is an Elongin BC-interacting protein that can assemble with Cul2 and Rbx1 to reconstitute a ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 10353–10358.
- Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., and Workman, J.L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* *292*, 2333–2337.
- Bryant, G.O., and Ptashne, M. (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. *Mol. Cell* *11*, 1301–1309.
- Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., et al. (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol. Cell* *44*, 942–953.

- Budd, M.E., and Campbell, J.L. (1995). DNA polymerases required for repair of UV-induced damage in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *15*, 2173–2179.
- Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* *334*, 37–42.
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* *56*, 549–561.
- Burke, T.W., and Kadonaga, J.T. (1996). *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* *10*, 711–724.
- Burton, Z.F., Killeen, M., Sopta, M., Ortolan, L.G., and Greenblatt, J. (1988). RAP30/74: a general initiation factor that binds to RNA polymerase II. *Mol. Cell. Biol.* *8*, 1602–1613.
- Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 6969–6973.
- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Å. *Science* *303*, 983–988.
- Butler, J.E.F., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* *16*, 2583–2592.
- Cai, G., Imasaki, T., Takagi, Y., and Asturias, F.J. (2009). Mediator structural conservation and implications for the regulation mechanism. *Struct. Lond. Engl.* *1993* *17*, 559–567.
- Cai, G., Imasaki, T., Yamada, K., Cardelli, F., Takagi, Y., and Asturias, F.J. (2010). Mediator head module structure and functional interactions. *Nat. Struct. Mol. Biol.* *17*, 273–279.
- Cai, G., Chaban, Y.L., Imasaki, T., Kovacs, J.A., Calero, G., Penczek, P.A., Takagi, Y., and Asturias, F.J. (2012). Interaction of the Mediator Head Module with RNA Polymerase II. *Struct. Lond. Engl.* *1993* *20*, 899–910.
- Cairns, B.R., Kim, Y.J., Sayre, M.H., Laurent, B.C., and Kornberg, R.D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 1950–1954.
- Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell* *87*, 1249–1260.
- Cantin, G.T., Stevens, J.L., and Berk, A.J. (2003). Activation domain-mediator interactions promote transcription preinitiation complex assembly on promoter DNA. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 12003–12008.
- Carcamo, J., Buckbinder, L., and Reinberg, D. (1991). The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 8052–8056.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* *309*, 1559–1563.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* *38*, 626–635.
- Carrera, I., Janody, F., Leeds, N., Duvéau, F., and Treisman, J.E. (2008). Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 6644–6649.
- Carter, R., and Drouin, G. (2010). The Increase in the Number of Subunits in Eukaryotic RNA Polymerase III Relative to RNA Polymerase II Is due to the Permanent Recruitment of General Transcription Factors. *Mol. Biol. Evol.* *27*, 1035–1043.
- Çevik, V., Kidd, B.N., Zhang, P., Hill, C., Kiddle, S., Denby, K.J., Holub, E.B., Cahill, D.M., Manners, J.M., Schenk, P.M., et al. (2012). MEDIATOR25 acts as an integrative hub for the regulation of jasmonate-responsive gene expression in *Arabidopsis*. *Plant Physiol.* *160*, 541–555.
- Chang, W.H., and Kornberg, R.D. (2000). Electron crystal structure of the transcription factor and DNA repair complex, core TFIIF. *Cell* *102*, 609–613.
- Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J.A. (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* *116*, 39–50.
- Chen, B.-S., and Hampsey, M. (2004). Functional interaction between TFIIF and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation. *Mol. Cell. Biol.* *24*, 3983–3991.
- Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* *14*, 696–703.

- Cheung, A.C.M., and Cramer, P. (2012). A Movie of RNA Polymerase II Transcription. *Cell* **149**, 1431–1437.
- Choder, M., and Young, R.A. (1993). A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol. Cell. Biol.* **13**, 6984–6991.
- Chua, G., Morris, Q.D., Sopko, R., Robinson, M.D., Ryan, O., Chan, E.T., Frey, B.J., Andrews, B.J., Boone, C., and Hughes, T.R. (2006). Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12045–12050.
- Chung, W.-H., Craighead, J.L., Chang, W.-H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., and Asturias, F.J. (2003). RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell* **12**, 1003–1013.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373.
- Citterio, E., Rademakers, S., van der Horst, G.T., van Gool, A.J., Hoeijmakers, J.H., and Vermeulen, W. (1998). Biochemical and biological characterization of wild-type and ATPase-deficient Cockayne syndrome B repair protein. *J. Biol. Chem.* **273**, 11844–11851.
- Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R.E., Hoeijmakers, J.H., and Vermeulen, W. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol. Cell. Biol.* **20**, 7643–7653.
- Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* **78**, 273–304.
- Clarkson, S.G. (2003). The XPG story. *Biochimie* **85**, 1113–1121.
- Cleaver, J.E. (2000). Common pathways for ultraviolet skin carcinogenesis in the repair and replication defective groups of xeroderma pigmentosum. *J. Dermatol. Sci.* **23**, 1–11.
- Cler, E., Papai, G., Schultz, P., and Davidson, I. (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell. Mol. Life Sci. CMLS* **66**, 2123–2134.
- Coin, F., Marinoni, J.C., Rodolfo, C., Fribourg, S., Pedrini, A.M., and Egly, J.M. (1998). Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat. Genet.* **20**, 184–188.
- Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W., and Egly, J.-M. (2004). Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity. *EMBO J.* **23**, 4835–4846.
- Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M., and Egly, J.-M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. *Mol. Cell* **21**, 215–226.
- Coin, F., Oksenysh, V., and Egly, J.-M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* **26**, 245–256.
- Coin, F., Oksenysh, V., Mocquet, V., Groh, S., Blattner, C., and Egly, J.M. (2008). Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol. Cell* **31**, 9–20.
- Cojocar, M., Jeronimo, C., Forget, D., Bouchard, A., Bergeron, D., Côte, P., Poirier, G.G., Greenblatt, J., and Coulombe, B. (2008). Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem. J.* **409**, 139–147.
- Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810.
- Compe, E., and Egly, J.-M. (2012). TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **13**, 343–354.
- Conaway, R.C., and Conaway, J.W. (1993). General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190.
- Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* **30**, 250–255.
- Constantinou, A., Gunz, D., Evans, E., Lalle, P., Bates, P.A., Wood, R.D., and Clarkson, S.G. (1999). Conserved residues of human XPG protein important for nuclease activity and function in nucleotide excision repair. *J. Biol. Chem.* **274**, 5637–5648.
- Corden, J.L. (1990). Tails of RNA polymerase II. *Trends Biochem. Sci.* **15**, 383–387.
- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* **209**, 1406–1414.

- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* 322, 1845–1848.
- Coulombe, B., Li, J., and Greenblatt, J. (1994). Topological localization of the human transcription factors IIA, IIB, TATA box-binding protein, and RNA polymerase II-associated protein 30 on a class II promoter. *J. Biol. Chem.* 269, 19962–19967.
- Craighead, J.L., Chang, W., and Asturias, F.J. (2002). Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Struct. Lond. Engl.* 1993 10, 1117–1125.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* 14, 218–226.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–1876.
- Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., et al. (2008). Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* 37, 337–352.
- Cui, J., Germer, K., Wu, T., Wang, J., Luo, J., Wang, S., Wang, Q., and Zhang, X. (2012). Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res.* 72, 5625–5634.
- Cui, X., Xu, D., Lv, C., Qu, F., He, J., Chen, M., Liu, Y., Gao, Y., Che, J., Yao, Y., et al. (2011). Suppression of MED19 expression by shRNA induces inhibition of cell proliferation and tumorigenesis in human prostate cancer cells. *BMB Rep.* 44, 547–552.
- D’Errico, M., Calcagnile, A., Canzona, F., Didona, B., Posteraro, P., Cavalieri, R., Corona, R., Vorechovsky, I., Nardo, T., Stefanini, M., et al. (2000). UV mutation signature in tumor suppressor genes involved in skin carcinogenesis in xeroderma pigmentosum patients. *Oncogene* 19, 463–467.
- D’Errico, M., Parlanti, E., Teson, M., Degan, P., Lemma, T., Calcagnile, A., Iavarone, I., Jaruga, P., Ropolo, M., Pedrini, A.M., et al. (2007). The role of CSA in the response to oxidative DNA damage in human cells. *Oncogene* 26, 4336–4343.
- Dang, W., Steffen, K.K., Perry, R., Dorsey, J.A., Johnson, F.B., Shilatifard, A., Kaeberlein, M., Kennedy, B.K., and Berger, S.L. (2009). Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* 459, 802–807.
- Daniel, J.A., Torok, M.S., Sun, Z.-W., Schieltz, D., Allis, C.D., Yates, J.R., 3rd, and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. *J. Biol. Chem.* 279, 1867–1871.
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5320–5325.
- Davis, C.A., and Ares, M., Jr (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3262–3267.
- Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* 10, 409–415.
- DeJong, J., and Roeder, R.G. (1993). A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA. *Genes Dev.* 7, 2220–2234.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* 295, 1306–1311.
- Delmas, V., Stokes, D.G., and Perry, R.P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc. Natl. Acad. Sci. U. S. A.* 90, 2414–2418.
- Deng, W., and Roberts, S.G.E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIIB. *Genes Dev.* 19, 2418–2423.
- DiGiovanna, J.J., and Kraemer, K.H. (2012). Shining a light on xeroderma pigmentosum. *J. Invest. Dermatol.* 132, 785–796.
- Van Dijk, E.L., Chen, C.L., d’ Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114–117.

- Ding, N., Zhou, H., Esteve, P.-O., Chin, H.G., Kim, S., Xu, X., Joseph, S.M., Friez, M.J., Schwartz, C.E., Pradhan, S., et al. (2008). Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol. Cell* *31*, 347–359.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* *485*, 376–380.
- Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *27*, 5575–5586.
- Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nat. Struct. Mol. Biol.* *17*, 194–201.
- Dorsett, D. (2011). Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.* *21*, 199–206.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* *16*, 1299–1309.
- Douki, T., Angelov, D., and Cadet, J. (2001). UV laser photolysis of DNA: effect of duplex stability on charge-transfer efficiency. *J. Am. Chem. Soc.* *123*, 11360–11366.
- Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M., and Coulombe, B. (2000). Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. *Mol. Cell. Biol.* *20*, 8168–8177.
- Doyen, C.-M., An, W., Angelov, D., Bondarenko, V., Mietton, F., Studitsky, V.M., Hamiche, A., Roeder, R.G., Bouvet, P., and Dimitrov, S. (2006). Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol. Cell. Biol.* *26*, 1156–1164.
- Dubaele, S., Proietti De Santis, L., Bienstock, R.J., Keriell, A., Stefanini, M., Van Houten, B., and Egly, J.-M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* *11*, 1635–1646.
- Dulk, B. den, Sun, S.M., de Ruijter, M., Brandsma, J.A., and Brouwer, J. (2006). Rad33, a new factor involved in nucleotide excision repair in *Saccharomyces cerevisiae*. *DNA Repair* *5*, 683–692.
- Dunand-Sauthier, I., Hohl, M., Thorel, F., Jaquier-Gubler, P., Clarkson, S.G., and Schärer, O.D. (2005). The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity. *J. Biol. Chem.* *280*, 7030–7037.
- Dvir, A., Conaway, R.C., and Conaway, J.W. (1997). A role for TFIIF in controlling the activity of early RNA polymerase II elongation complexes. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 9006–9010.
- Ebmeier, C.C., and Taatjes, D.J. (2010). Activator-Mediator binding regulates Mediator-cofactor interactions. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 11283–11288.
- Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* *266*, 71–75.
- Ehrensberger, A.H., and Kornberg, R.D. (2011). Isolation of an activator-dependent, promoter-specific chromatin remodeling factor. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 10115–10120.
- Eichner, J., Chen, H.-T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* *29*, 706–716.
- Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C.O., Koeck, P.J.B., Holmberg, S., Hebert, H., and Gustafsson, C.M. (2006). The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 15788–15793.
- Emmert, S., Slor, H., Busch, D.B., Batko, S., Albert, R.B., Coleman, D., Khan, S.G., Abu-Libdeh, B., DiGiovanna, J.J., Cunningham, B.B., et al. (2002). Relationship of neurologic degeneration to genotype in three xeroderma pigmentosum group G patients. *J. Invest. Dermatol.* *118*, 972–982.
- Enzlin, J.H., and Schärer, O.D. (2002). The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J.* *21*, 2045–2053.
- Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell* *31*, 337–346.
- Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J.* *16*, 6559–6573.
- Eyboulet, F., Cibot, C., Eychenne, T., Neil, H., Alibert, O., Werner, M., and Soutourina, J. (2013). Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment. *Genes Dev.* *27*, 2549–2562.

- Fagbemi, A.F., Orelli, B., and Schärer, O.D. (2011). Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair* 10, 722–729.
- Faghri, S., Tamura, D., Kraemer, K.H., and Digiovanna, J.J. (2008). Trichothiodystrophy: a systematic review of 112 published cases characterises a wide spectrum of clinical manifestations. *J. Med. Genet.* 45, 609–621.
- Fan, L., Arvai, A.S., Cooper, P.K., Iwai, S., Hanaoka, F., and Tainer, J.A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol. Cell* 22, 27–37.
- Farago, M., Nahari, T., Hammel, C., Cole, C.N., and Choder, M. (2003). Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol. Biol. Cell* 14, 2744–2755.
- Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* 7, 1728–1740.
- Feng, Z., Hu, W., Amin, S., and Tang, M. (2003). Mutational spectrum and genotoxicity of the major lipid peroxidation product, trans-4-hydroxy-2-nonenal, induced DNA adducts in nucleotide excision repair-proficient and -deficient human cells. *Biochemistry (Mosc.)* 42, 7848–7854.
- Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H., Vena, N., Ogino, S., et al. (2008). CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 455, 547–551.
- Fischer, E.S., Scrima, A., Böhm, K., Matsumoto, S., Lingaraju, G.M., Faty, M., Yasuda, T., Cavadini, S., Wakasugi, M., Hanaoka, F., et al. (2011). The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* 147, 1024–1039.
- Fish, R.N., and Kane, C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* 1577, 287–307.
- Fishburn, J., Mohibullah, N., and Hahn, S. (2005). Function of a eukaryotic transcription activator during the transcription cycle. *Mol. Cell* 18, 369–378.
- Fitch, M.E., Nakajima, S., Yasui, A., and Ford, J.M. (2003). In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *J. Biol. Chem.* 278, 46906–46910.
- Flanagan, J.F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* 438, 1181–1185.
- Flanagan, P.M., Kelleher, R.J., Sayre, M.H., Tschochner, H., and Kornberg, R.D. (1991). A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350, 436–438.
- Flores, O., Maldonado, E., and Reinberg, D. (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.* 264, 8913–8921.
- Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.* 267, 2786–2793.
- Fondell, J.D., Ge, H., and Roeder, R.G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8329–8333.
- Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* 147, 120–131.
- Forget, D., Langelier, M.-F., Therien, C., Trinh, V., and Coulombe, B. (2004). Photo-Cross-Linking of a Purified Preinitiation Complex Reveals Central Roles for the RNA Polymerase II Mobile Clamp and TFIIE in Initiation Mechanisms. *Mol. Cell. Biol.* 24, 1122–1131.
- Fousteri, M., Vermeulen, W., van Zeeland, A.A., and Mullenders, L.H.F. (2006). Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol. Cell* 23, 471–482.
- Frit, P., Kwon, K., Coin, F., Auriol, J., Dubaele, S., Salles, B., and Egly, J.M. (2002). Transcriptional activators stimulate DNA repair. *Mol. Cell* 10, 1391–1401.
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell. Biol.* 24, 787–795.
- Fujiwara, Y., Ichihashi, M., Kano, Y., Goto, K., and Shimizu, K. (1981). A new human photosensitive subject with a defect in the recovery of DNA synthesis after ultraviolet-light irradiation. *J. Invest. Dermatol.* 77, 256–263.

- Fullwood, M.J., Han, Y., Wei, C.-L., Ruan, X., and Ruan, Y. (2010). Chromatin interaction analysis using paired-end tag sequencing. *Curr. Protoc. Mol. Biol.* Ed. Frederick M Ausubel Al *Chapter 21*, Unit 21.15.1–25.
- Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLoS Genet* *8*, e1003036.
- Gaillard, H., Tous, C., Botet, J., González-Aguilera, C., Quintero, M.J., Viladevall, L., García-Rubio, M.L., Rodríguez-Gil, A., Marín, A., Ariño, J., et al. (2009). Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. *PLoS Genet.* *5*, e1000364.
- Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E., and Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell* *86*, 887–896.
- Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D., et al. (2013). HIF1A Employs CDK8-Mediator to Stimulate RNAPII Elongation in Response to Hypoxia. *Cell* *153*, 1327–1339.
- Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., and Weil, P.A. (2007). Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. *Mol. Cell. Biol.* *27*, 297–311.
- Gary, R., Ludwig, D.L., Cornelius, H.L., MacInnes, M.A., and Park, M.S. (1997). The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J. Biol. Chem.* *272*, 24522–24529.
- Ge, K., Guermah, M., Yuan, C.-X., Ito, M., Wallberg, A.E., Spiegelman, B.M., and Roeder, R.G. (2002). Transcription coactivator TRAP220 is required for PPAR gamma 2-stimulated adipogenesis. *Nature* *417*, 563–567.
- Georgaki, A., Strack, B., Podust, V., and Hübscher, U. (1992). DNA unwinding activity of replication protein A. *FEBS Lett.* *308*, 240–244.
- Gershenson, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinforma. Oxf. Engl.* *21*, 1295–1300.
- Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. *Science* *261*, 759–762.
- Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2012). Subunit architecture of general transcription factor TFIID. *Proc. Natl. Acad. Sci.* *109*, 1949–1954.
- Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G.J., Raams, A., Argentini, M., van der Spek, P.J., et al. (2004). A new, tenth subunit of TFIID is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat. Genet.* *36*, 714–719.
- Gillet, L.C.J., and Schärer, O.D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* *106*, 253–276.
- Gillette, T.G., Yu, S., Zhou, Z., Waters, R., Johnston, S.A., and Reed, S.H. (2006). Distinct functions of the ubiquitin-proteasome pathway influence nucleotide excision repair. *EMBO J.* *25*, 2529–2538.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996). Life with 6000 genes. *Science* *274*, 546, 563–567.
- Gong, F., Fahy, D., and Smerdon, M.J. (2006). Rad4-Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair. *Nat. Struct. Mol. Biol.* *13*, 902–907.
- Gonzalez, D., Hamidi, N., Del Sol, R., Benschop, J.J., Nancy, T., Li, C., Francis, L., Tzouros, M., Krijgsveld, J., Holstege, F.C.P., et al. (2014). Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 2500–2505.
- Goodrich, J.A., and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* *77*, 145–156.
- Van Gool, A.J., Verhage, R., Swagemakers, S.M., van de Putte, P., Brouwer, J., Troelstra, C., Bootsma, D., and Hoeijmakers, J.H. (1994). RAD26, the functional *S. cerevisiae* homolog of the Cockayne syndrome B gene ERCC6. *EMBO J.* *13*, 5361–5369.
- Van Gool, A.J., Citterio, E., Rademakers, S., van Os, R., Vermeulen, W., Constantinou, A., Egly, J.M., Bootsma, D., and Hoeijmakers, J.H. (1997). The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *EMBO J.* *16*, 5955–5965.
- Goppelt, A., Stelzer, G., Lottspeich, F., and Meisterernst, M. (1996). A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *EMBO J.* *15*, 3105–3116.

- Govind, C.K., Zhang, F., Qiu, H., Hofmeyer, K., and Hinnebusch, A.G. (2007). Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol. Cell* 25, 31–42.
- Green, C.M., and Almouzni, G. (2003). Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO J.* 22, 5163–5174.
- Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. *Struct. Lond. Engl.* 1993 14, 511–520.
- Grohmann, D., and Werner, F. (2011). Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase. *Res. Microbiol.* 162, 10–18.
- Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J., Saijo, M., Drapkin, R., Kisselev, A.F., Tanaka, K., and Nakatani, Y. (2003). The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 113, 357–367.
- Groisman, R., Kuraoka, I., Chevallerier, O., Gaye, N., Magnaldo, T., Tanaka, K., Kisselev, A.F., Harel-Bellan, A., and Nakatani, Y. (2006). CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev.* 20, 1429–1434.
- Grøntved, L., Madsen, M.S., Boergesen, M., Roeder, R.G., and Mandrup, S. (2010). MED14 tethers mediator to the N-terminal domain of peroxisome proliferator-activated receptor gamma and is required for full transcriptional activity and adipogenesis. *Mol. Cell. Biol.* 30, 2155–2169.
- Grünberg, S., Reich, C., Zeller, M.E., Bartlett, M.S., and Thomm, M. (2010). Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes. *Nucleic Acids Res.* 38, 1950–1963.
- Grünberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* 19, 788–796.
- Gu, W., Malik, S., Ito, M., Yuan, C.X., Fondell, J.D., Zhang, X., Martinez, E., Qin, J., and Roeder, R.G. (1999). A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* 3, 97–108.
- Guelman, S., Suganuma, T., Florens, L., Swanson, S.K., Kiesecker, C.L., Kusch, T., Anderson, S., Yates, J.R., 3rd, Washburn, M.P., Abmayr, S.M., et al. (2006). Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. *Mol. Cell. Biol.* 26, 871–882.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88.
- Guerrero-Santoro, J., Kapetanaki, M.G., Hsieh, C.L., Gorbachinsky, I., Levine, A.S., and Rapić-Otrin, V. (2008). The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. *Cancer Res.* 68, 5014–5022.
- Guglielmi, B., van Berkum, N.L., Klapholz, B., Bijma, T., Boube, M., Boschiero, C., Bourbon, H.-M., Holstege, F.C.P., and Werner, M. (2004). A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res.* 32, 5379–5391.
- Guglielmi, B., Soutourina, J., Esnault, C., and Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc. Natl. Acad. Sci.* 104, 16062–16067.
- Guo, R., Chen, J., Mitchell, D.L., and Johnson, D.G. (2011). GCN5 and E2F1 stimulate nucleotide excision repair by promoting H3K9 acetylation at sites of damage. *Nucleic Acids Res.* 39, 1390–1397.
- Guo, Y., Papachristoudis, G., Altshuler, R.C., Gerber, G.K., Jaakkola, T.S., Gifford, D.K., and Mahony, S. (2010). Discovering homotypic binding events at high spatial resolution. *Bioinforma. Oxf. Engl.* 26, 3028–3034.
- Guschin, D., and Wolffe, A.P. (1999). SWItched-on mobility. *Curr. Biol. CB* 9, R742–746.
- Guzder, S.N., Sung, P., Prakash, L., and Prakash, S. (1993). Yeast DNA-repair gene RAD14 encodes a zinc metalloprotein with affinity for ultraviolet-damaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5433–5437.
- Guzder, S.N., Sung, P., Bailly, V., Prakash, L., and Prakash, S. (1994). RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369, 578–581.
- Guzder, S.N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1995). Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIH. *J. Biol. Chem.* 270, 12973–12976.

- Guzder, S.N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996). RAD26, the yeast homolog of human Cockayne's syndrome group B gene, encodes a DNA-dependent ATPase. *J. Biol. Chem.* **271**, 18314–18317.
- Guzder, S.N., Sung, P., Prakash, L., and Prakash, S. (1997). Yeast Rad7-Rad16 complex, specific for the nucleotide excision repair of the nontranscribed DNA strand, is an ATP-dependent DNA damage sensor. *J. Biol. Chem.* **272**, 21665–21668.
- Guzder, S.N., Sung, P., Prakash, L., and Prakash, S. (1998). The DNA-dependent ATPase activity of yeast nucleotide excision repair factor 4 and its role in DNA damage recognition. *J. Biol. Chem.* **273**, 6292–6296.
- Guzder, S.N., Sung, P., Prakash, L., and Prakash, S. (1999). Synergistic interaction between yeast nucleotide excision repair factors NEF2 and NEF4 in the binding of ultraviolet-damaged DNA. *J. Biol. Chem.* **274**, 24257–24262.
- Guzder, S.N., Sommers, C.H., Prakash, L., and Prakash, S. (2006). Complex formation with damage recognition protein Rad14 is essential for *Saccharomyces cerevisiae* Rad1-Rad10 nuclease to perform its function in nucleotide excision repair in vivo. *Mol. Cell. Biol.* **26**, 1135–1141.
- Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L.U., Green, M., and Reinberg, D. (1993). Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* **7**, 1021–1032.
- Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**, 483–492.
- Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1993). Yeast excision repair gene RAD2 encodes a single-stranded DNA endonuclease. *Nature* **366**, 365–368.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403.
- Hahn, J.-S., and Thiele, D.J. (2004). Activation of the *Saccharomyces cerevisiae* heat shock transcription factor under glucose starvation conditions by Snf1 protein kinase. *J. Biol. Chem.* **279**, 5169–5176.
- Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* **97**, 833–842.
- Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503.
- Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**, 429–432.
- Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970.
- Hara, R., Mo, J., and Sancar, A. (2000). DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Mol. Cell. Biol.* **20**, 9173–9181.
- Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A.S., Conaway, R.C., Conaway, J.W., et al. (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20705–20710.
- Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369.
- Hashimoto, S., Boissel, S., Zarhrate, M., Rio, M., Munnich, A., Egly, J.-M., and Colleaux, L. (2011). MED23 Mutation Links Intellectual Disability to Dysregulation of Immediate Early Gene Expression. *Science* **333**, 1161–1163.
- He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. *Science* **322**, 1855–1857.
- He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486.
- He, Z., Henricksen, L.A., Wold, M.S., and Ingles, C.J. (1995). RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* **374**, 566–569.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.
- Hekmatpanah, D.S., and Young, R.A. (1991). Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. *Mol. Cell. Biol.* **11**, 5781–5791.

- Hengartner, C.J., Thompson, C.M., Zhang, J., Chao, D.M., Liao, S.M., Koleske, A.J., Okamura, S., and Young, R.A. (1995). Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev.* *9*, 897–910.
- Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* *2*, 43–53.
- Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* *21*, 133–153.
- Henning, K.A., Li, L., Iyer, N., McDaniel, L.D., Reagan, M.S., Legerski, R., Schultz, R.A., Stefanini, M., Lehmann, A.R., Mayne, L.V., et al. (1995). The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell* *82*, 555–564.
- Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. *Mol. Cell* *52*, 517–528.
- Henry, N.L., Campbell, A.M., Feaver, W.J., Poon, D., Weil, P.A., and Kornberg, R.D. (1994). TFIIF-TAF-RNA polymerase II connection. *Genes Dev.* *8*, 2868–2878.
- Hess, M.T., Gunz, D., Luneva, N., Geacintov, N.E., and Naegeli, H. (1997). Base pair conformation-dependent excision of benzo[a]pyrene diol epoxide-guanine adducts by human nucleotide excision repair enzymes. *Mol. Cell. Biol.* *17*, 7069–7076.
- Hey, T., Lipps, G., Sugawara, K., Iwai, S., Hanaoka, F., and Krauss, G. (2002). The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry (Mosc.)* *41*, 6583–6587.
- Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* *6*, 2288–2298.
- Hittelman, A.B., Burakov, D., Iñiguez-Lluhí, J.A., Freedman, L.P., and Garabedian, M.J. (1999). Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J.* *18*, 5380–5388.
- Hoeijmakers, J.H.J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* *411*, 366–374.
- Holstege, F.C., van der Vliet, P.C., and Timmers, H.T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* *15*, 1666–1677.
- Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* *16*, 7468–7480.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* *95*, 717–728.
- Hong, S.-K., Haldin, C.E., Lawson, N.D., Weinstein, B.M., Dawid, I.B., and Hukriede, N.A. (2005). The zebrafish *kohtalo/trap230* gene is required for the development of the brain, neural crest, and pronephric kidney. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 18473–18478.
- Horibata, K., Iwamoto, Y., Kuraoka, I., Jaspers, N.G.J., Kurimasa, A., Oshimura, M., Ichihashi, M., and Tanaka, K. (2004). Complete absence of Cockayne syndrome group B gene product gives rise to UV-sensitive syndrome but not Cockayne syndrome. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 15410–15415.
- Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A., and Roeder, R.G. (1989). Purification of a yeast TATA box-binding protein that exhibits human transcription factor IID activity. *Proc. Natl. Acad. Sci. U. S. A.* *86*, 4843–4847.
- Van der Horst, G.T., van Steeg, H., Berg, R.J., van Gool, A.J., de Wit, J., Weeda, G., Morreau, H., Beems, R.B., van Kreijl, C.F., de Gruijl, F.R., et al. (1997). Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. *Cell* *89*, 425–435.
- Hosfield, D.J., Daniels, D.S., Mol, C.D., Putnam, C.D., Parikh, S.S., and Tainer, J.A. (2001). DNA damage recognition and repair pathway coordination revealed by the structural biochemistry of DNA repair enzymes. *Prog. Nucleic Acid Res. Mol. Biol.* *68*, 315–347.
- Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* *334*, 683–686.
- Hsu, L.M., Vo, N.V., Kane, C.M., and Chamberlin, M.J. (2003). In vitro studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry (Mosc.)* *42*, 3777–3786.
- Hu, X., Malik, S., Negroiu, C.C., Hubbard, K., Velalar, C.N., Hampton, B., Grosu, D., Catalano, J., Roeder, R.G., and Gnatt, A. (2006). A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 9506–9511.

- Huang, Y., Li, W., Yao, X., Lin, Q.-J., Yin, J.-W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol. Cell* *45*, 459–469.
- Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M. (1998). Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nat. Struct. Biol.* *5*, 701–706.
- Imasaki, T., Calero, G., Cai, G., Tsai, K.-L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G.L., et al. (2011). Architecture of the Mediator head module. *Nature* *475*, 240–243.
- Imbalzano, A.N., Kwon, H., Green, M.R., and Kingston, R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* *370*, 481–485.
- Ito, M., Yuan, C.X., Okano, H.J., Darnell, R.B., and Roeder, R.G. (2000). Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. *Mol. Cell* *5*, 683–693.
- Ito, M., Okano, H.J., Darnell, R.B., and Roeder, R.G. (2002). The TRAP100 component of the TRAP/Mediator complex is essential in broad transcriptional events and development. *EMBO J.* *21*, 3464–3475.
- Ito, S., Kuraoka, I., Chymkowitch, P., Compe, E., Takedachi, A., Ishigami, C., Coin, F., Egly, J.-M., and Tanaka, K. (2007). XPG Stabilizes TFIIH, Allowing Transactivation of Nuclear Receptors: Implications for Cockayne Syndrome in XP-G/CS Patients. *Mol. Cell* *26*, 231–243.
- Iyer, N., Reagan, M.S., Wu, K.J., Canagarajah, B., and Friedberg, E.C. (1996). Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein. *Biochemistry (Mosc.)* *35*, 2157–2167.
- Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* *79*, 107–117.
- Jasiak, A.J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strässer, K., Martin, D.E., Söding, J., and Cramer, P. (2008). Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J. Biol. Chem.* *283*, 26423–26427.
- Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with Pervasive Transcription. *Mol. Cell* *52*, 473–484.
- Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.*
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* *10*, 161–172.
- Jiang, Y., Wang, X., Bao, S., Guo, R., Johnson, D.G., Shen, X., and Li, L. (2010). INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 17274–17279.
- Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 8538–8543.
- Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell* *29*, 232–242.
- Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIIF with Pol II. *Mol. Cell* *45*, 51–63.
- Johnson, A.D. (1995). The price of repression. *Cell* *81*, 655–658.
- Johnson, K.M., and Carey, M. (2003). Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. *Curr. Biol. CB* *13*, 772–777.
- Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev.* *16*, 1852–1863.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* *467*, 430–435.
- Kalogeraki, V.S., Tornaletti, S., Cooper, P.K., and Hanawalt, P.C. (2005). Comparative TFIIIS-mediated transcript cleavage by mammalian RNA polymerase II arrested at a lesion in different transcription systems. *DNA Repair* *4*, 1075–1087.

- Kamada, K., Roeder, R.G., and Burley, S.K. (2003). Molecular mechanism of recruitment of TFIIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 2296–2299.
- Kamiuchi, S., Saijo, M., Citterio, E., de Jager, M., Hoeijmakers, J.H.J., and Tanaka, K. (2002). Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 201–206.
- Kang, J.S., Kim, S.H., Hwang, M.S., Han, S.J., Lee, Y.C., and Kim, Y.J. (2001). The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.* *276*, 42003–42010.
- Kang, M.-S., Yu, S.-L., Kim, H.-Y., Gorospe, C.M., Choi, B.H., Lee, S.H., and Lee, S.-K. (2014). Yeast RAD2, a homolog of human XPG, plays a key role in the regulation of the cell cycle and actin dynamics. *Biol. Open* *3*, 29–41.
- Kang, X., Hu, Y., Li, Y., Guo, X., Jiang, X., Lai, L., Xia, B., and Jin, C. (2006). Structural, biochemical, and dynamic characterizations of the hRPB8 subunit of human RNA polymerases. *J. Biol. Chem.* *281*, 18216–18226.
- Kang, Y.K., Guermah, M., Yuan, C.-X., and Roeder, R.G. (2002). The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 2642–2647.
- Kapetanaki, M.G., Guerrero-Santoro, J., Bisi, D.C., Hsieh, C.L., Rapić-Otrin, V., and Levine, A.S. (2006). The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 2588–2593.
- Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* *301*, 1096–1099.
- Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Dutttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* *316*, 1484–1488.
- Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. *Science* *309*, 1564–1566.
- Kato, Y., Habas, R., Katsuyama, Y., Näär, A.M., and He, X. (2002). A component of the ARC/Mediator complex required for TGF beta/Nodal signalling. *Nature* *418*, 641–646.
- Kaufmann, J., and Smale, S.T. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* *8*, 821–829.
- Kaufmann, R., Straussberg, R., Mandel, H., Fattal-Valevski, A., Ben-Zeev, B., Naamati, A., Shaag, A., Zenvirt, S., Konen, O., Mimouni-Bloch, A., et al. (2010). Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am. J. Hum. Genet.* *87*, 667–670.
- Kelleher, R.J., Flanagan, P.M., and Kornberg, R.D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* *61*, 1209–1215.
- Kelman, Z. (1997). PCNA: structure, functions and interactions. *Oncogene* *14*, 629–640.
- Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res.* *12*, 656–664.
- Keriel, A., Stary, A., Sarasin, A., Rochette-Egly, C., and Egly, J.M. (2002). XPD mutations prevent TFIIH-dependent transactivation by nuclear receptors and phosphorylation of RARalpha. *Cell* *109*, 125–135.
- Kessler, K.J., Kaufmann, W.K., Reardon, J.T., Elston, T.C., and Sancar, A. (2007). A mathematical model for human nucleotide excision repair: damage recognition by random order assembly and kinetic proofreading. *J. Theor. Biol.* *249*, 361–375.
- Kim, B., Nesvizhskii, A.I., Rani, P.G., Hahn, S., Aebersold, R., and Ranish, J.A. (2007). The transcription elongation factor TFIIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl. Acad. Sci.* *104*, 16068–16073.
- Kim, J.K., Patel, D., and Choi, B.S. (1995). Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6-4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. *Photochem. Photobiol.* *62*, 44–50.
- Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature* *436*, 876–880.
- Kim, T.K., Lagrange, T., Wang, Y.H., Griffith, J.D., Reinberg, D., and Ebright, R.H. (1997). Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 12268–12273.

- Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599–608.
- Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* 275, 6530–6536.
- Klein, C., and Struhl, K. (1994). Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science* 266, 280–282.
- Klungland, A., Höss, M., Gunz, D., Constantinou, A., Clarkson, S.G., Doetsch, P.W., Bolton, P.H., Wood, R.D., and Lindahl, T. (1999). Base excision repair of oxidative DNA damage activated by XPG protein. *Mol. Cell* 3, 33–42.
- Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* 23, 439–451.
- Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol.* 2, E131.
- Koch, F., Jourquin, F., Ferrier, P., and Andrau, J.-C. (2008). Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.* 33, 265–273.
- Koh, S.S., Ansari, A.Z., Ptashne, M., and Young, R.A. (1998). An activator target in the RNA polymerase II holoenzyme. *Mol. Cell* 1, 895–904.
- Kohlhaw, G.B. (2003). Leucine biosynthesis in fungi: entering metabolism through the back door. *Microbiol. Mol. Biol. Rev. MMBR* 67, 1–15, table of contents.
- Kokubo, T., Swanson, M.J., Nishikawa, J.I., Hinnebusch, A.G., and Nakatani, Y. (1998). The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* 18, 1003–1012.
- Kolodziej, P., and Young, R.A. (1989). RNA polymerase II subunit RPB3 is an essential component of the mRNA transcription apparatus. *Mol. Cell. Biol.* 9, 5387–5394.
- Kolthur-Seetharam, U., Martianov, I., and Davidson, I. (2008). Specialization of the general transcriptional machinery in male germ cells. *Cell Cycle Georget. Tex* 7, 3493–3498.
- Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D., and Abrescia, N.G.A. (2009). Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol.* 7.
- Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* 30, 235–239.
- Koschubs, T., Seizl, M., Larivière, L., Kurth, F., Baumli, S., Martin, D.E., and Cramer, P. (2009). Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *EMBO J.* 28, 69–80.
- Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323–330.
- Koutelou, E., Hirsch, C.L., and Dent, S.Y.R. (2010). Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* 22, 374–382.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kowenz-Leutz, E., and Leutz, A. (1999). A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. *Mol. Cell* 4, 735–743.
- Krebs, A.R., Demmers, J., Karmodiya, K., Chang, N.-C., Chang, A.C., and Tora, L. (2010). ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. *EMBO Rep.* 11, 541–547.
- Krogan, N.J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V., et al. (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol. Cell* 12, 1565–1576.
- Kuehner, J.N., and Brow, D.A. (2006). Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* 281, 14119–14128.
- Kuhn, C.-D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell* 131, 1260–1272.
- Kuras, L., and Struhl, K. (1999). Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* 399, 609–613.

- Kuras, L., Borggreffe, T., and Kornberg, R.D. (2003). Association of the Mediator complex with enhancers of active genes. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 13887–13891.
- Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. *Annu. Rev. Genet.* *47*, 483–508.
- Kwon, J.Y., Park, J.M., Gim, B.S., Han, S.J., Lee, J., and Kim, Y.J. (1999). *Caenorhabditis elegans* mediator complexes are required for developmental-specific transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14990–14995.
- Laat, W.L. de, Jaspers, N.G.J., and Hoeijmakers, J.H.J. (1999). Molecular mechanism of nucleotide excision repair. *Genes Dev.* *13*, 768–785.
- De Laat, W.L., Appeldoorn, E., Sugawara, K., Weterings, E., Jaspers, N.G., and Hoeijmakers, J.H. (1998a). DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev.* *12*, 2598–2609.
- De Laat, W.L., Sijbers, A.M., Odijk, H., Jaspers, N.G., and Hoeijmakers, J.H. (1998b). Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res.* *26*, 4146–4152.
- Lacombe, T., Poh, S.L., Barbey, R., and Kuras, L. (2013). Mediator is an intrinsic component of the basal RNA polymerase II machinery in vivo. *Nucleic Acids Res.* gkt701.
- Lafrance-Vanasse, J., Arseneault, G., Cappadocia, L., Chen, H.-T., Legault, P., and Omichinski, J.G. (2012). Structural and functional characterization of interactions involving the Tfb1 subunit of TFIID and the NER factor Rad2. *Nucleic Acids Res.* *40*, 5739–5750.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebricht, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* *12*, 34–44.
- Lainé, J.-P., and Egly, J.-M. (2006). Initiation of DNA repair mediated by a stalled RNA polymerase II. *EMBO J.* *25*, 387–397.
- Lake, R.J., Geyko, A., Hemashettar, G., Zhao, Y., and Fan, H.-Y. (2010). UV-induced association of the CSB remodeling protein with chromatin requires ATP-dependent relief of N-terminal autorepression. *Mol. Cell* *37*, 235–246.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Längst, G., Bonte, E.J., Corona, D.F., and Becker, P.B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell* *97*, 843–852.
- Laprade, L., Rose, D., and Winston, F. (2007). Characterization of new Spt3 and TATA-binding protein mutants of *Saccharomyces cerevisiae*: Spt3 TBP allele-specific interactions and bypass of Spt8. *Genetics* *177*, 2007–2017.
- Larivière, L., Geiger, S., Hoepfner, S., Röther, S., Strässer, K., and Cramer, P. (2006). Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat. Struct. Mol. Biol.* *13*, 895–901.
- Larivière, L., Seizl, M., van Wageningen, S., Röther, S., van de Pasch, L., Feldmann, H., Strässer, K., Hahn, S., Holstege, F.C.P., and Cramer, P. (2008). Structure-system correlation identifies a gene regulatory Mediator submodule. *Genes Dev.* *22*, 872–877.
- Larivière, L., Plaschka, C., Seizl, M., Wenzek, L., Kurth, F., and Cramer, P. (2012a). Structure of the Mediator head module. *Nature* *492*, 448–451.
- Larivière, L., Seizl, M., and Cramer, P. (2012b). A structural perspective on Mediator function. *Curr. Opin. Cell Biol.*
- Larivière, L., Plaschka, C., Seizl, M., Petrotchenko, E.V., Wenzek, L., Borchers, C.H., and Cramer, P. (2013). Model of the Mediator middle module based on protein cross-linking. *Nucleic Acids Res.*
- Larochelle, S., Pandur, J., Fisher, R.P., Salz, H.K., and Suter, B. (1998). Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. *Genes Dev.* *12*, 370–381.
- Larschan, E., and Winston, F. (2005). The *Saccharomyces cerevisiae* Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription. *Mol. Cell Biol.* *25*, 114–123.
- Laurent, B.C., Treich, I., and Carlson, M. (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev.* *7*, 583–591.
- Leal, A., Huehne, K., Bauer, F., Sticht, H., Berger, P., Suter, U., Morera, B., Valle, G., Lupski, J.R., Ekici, A., et al. (2009). Identification of the variant Ala335Val of MED25 as responsible for CMT2B2: molecular data, functional studies of the SH3 recognition motif and correlation between wild-type MED25 and PMP22 RNA levels in CMT1A animal models. *Neurogenetics* *10*, 275–287.

- Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* *28*, 3290–3300.
- Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* *36*, 900–905.
- Lee, J.-H., Park, C.-J., Arunkumar, A.I., Chazin, W.J., and Choi, B.-S. (2003). NMR study on the interaction between RPA and DNA decamer containing cis-syn cyclobutane pyrimidine dimer in the presence of XPA: implication for damage verification and strand-specific dual incision in nucleotide excision repair. *Nucleic Acids Res.* *31*, 4747–4754.
- Lee, S.-K., Yu, S.-L., Prakash, L., and Prakash, S. (2001). Requirement for Yeast RAD26, a Homolog of the Human CSB Gene, in Elongation by RNA Polymerase II. *Mol. Cell. Biol.* *21*, 8651–8656.
- Lee, S.-K., Yu, S.-L., Prakash, L., and Prakash, S. (2002). Requirement of Yeast RAD2, a Homolog of Human XPG Gene, for Efficient RNA Polymerase II Transcription: Implications for Cockayne Syndrome. *Cell* *109*, 823–834.
- Lee, Y.C., Park, J.M., Min, S., Han, S.J., and Kim, Y.J. (1999). An activator binding module of yeast RNA polymerase II holoenzyme. *Mol. Cell. Biol.* *19*, 2967–2976.
- Lei, L., Ren, D., and Burton, Z.F. (1999). The RAP74 subunit of human transcription factor IIF has similar roles in initiation and elongation. *Mol. Cell. Biol.* *19*, 8372–8382.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* *25*, 1754–1760.
- Li, S., and Smerdon, M.J. (2002). Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *EMBO J.* *21*, 5921–5929.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* *128*, 707–719.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Li, L., Elledge, S.J., Peterson, C.A., Bales, E.S., and Legerski, R.J. (1994). Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 5012–5016.
- Li, L., Lu, X., Peterson, C.A., and Legerski, R.J. (1995). An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol. Cell. Biol.* *15*, 5396–5402.
- Li, X.Y., Virbasius, A., Zhu, X., and Green, M.R. (1999). Enhancement of TBP binding by activators and general transcription factors. *Nature* *399*, 605–609.
- Lieberman, P.M., and Berk, A.J. (1994). A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. *Genes Dev.* *8*, 995–1005.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289–293.
- Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* *18*, 1606–1617.
- Lim, J.-H., West, K.L., Rubinstein, Y., Bergel, M., Postnikov, Y.V., and Bustin, M. (2005). Chromosomal protein HMGN1 enhances the acetylation of lysine 14 in histone H3. *EMBO J.* *24*, 3038–3048.
- Lin, J.J., Lehmann, L.W., Bonora, G., Sridharan, R., Vashisht, A.A., Tran, N., Plath, K., Wohlschlegel, J.A., and Carey, M. (2011). Mediator coordinates PIC assembly with recruitment of CHD1. *Genes Dev.* *25*, 2198–2209.
- Lin, Y.C., Choi, W.S., and Gralla, J.D. (2005). TFIIF XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat. Struct. Mol. Biol.* *12*, 603–607.
- Linder, T., and Gustafsson, C.M. (2004). The Soh1/MED31 protein is an ancient component of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* Mediator. *J. Biol. Chem.* *279*, 49455–49459.
- Liu, Z., and Myers, L.C. (2012). Med5(Nut1) and med17(srb4) are direct targets of mediator histone h4 tail interactions. *PLoS One* *7*, e38416.
- Liu, W.-L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator-TFIID complexes. *Genes Dev.* *23*, 1510–1521.
- Liu, X., Vorontchikhina, M., Wang, Y.-L., Faiola, F., and Martinez, E. (2008). STAGA Recruits Mediator to the MYC Oncoprotein To Stimulate Transcription and Cell Proliferation. *Mol. Cell. Biol.* *28*, 108–121.
- Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA Polymerase II-TFIIB Complex and the Transcription Initiation Mechanism. *Science* *327*, 206–209.

- Liu, Y., Ranish, J.A., Aebersold, R., and Hahn, S. (2001). Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. *J. Biol. Chem.* **276**, 7169–7175.
- Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol. Cell. Biol.* **24**, 1721–1735.
- Lommel, L., Ortolan, T., Chen, L., Madura, K., and Sweder, K.S. (2002). Proteolysis of a nucleotide excision repair protein by the 26 S proteasome. *Curr. Genet.* **42**, 9–20.
- Loncle, N., Boube, M., Joulia, L., Boschiero, C., Werner, M., Cribbs, D.L., and Bourbon, H.-M. (2007). Distinct roles for Mediator Cdk8 module subunits in *Drosophila* development. *EMBO J.* **26**, 1045–1054.
- Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005). The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev.* **19**, 3004–3016.
- Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007). The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J. Cell Biol.* **178**, 1133–1143.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334.
- Luijsterburg, M.S., Goedhart, J., Moser, J., Kool, H., Geverts, B., Houtsmuller, A.B., Mullenders, L.H.F., Vermeulen, W., and van Driel, R. (2007). Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with UV-damaged DNA is independent of damage-recognition protein XPC. *J. Cell Sci.* **120**, 2706–2716.
- Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biol.* **1**, REVIEWS001.
- Lykke-Andersen, S., and Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**, 1177–1182.
- Lyndaker, A.M., and Alani, E. (2009). A tale of tails: insights into the coordination of 3' end processing during homologous recombination. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **31**, 315–321.
- Majello, B., and Napolitano, G. (2001). Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases. *Front. Biosci. J. Virtual Libr.* **6**, D1358–1368.
- Malagon, F., Tong, A.H., Shafer, B.K., and Strathern, J.N. (2004). Genetic Interactions of DST1 in *Saccharomyces cerevisiae* Suggest a Role of TFIIIS in the Initiation-Elongation Transition. *Genetics* **166**, 1215–1227.
- Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772.
- Malik, S., Gu, W., Wu, W., Qin, J., and Roeder, R.G. (2000). The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. *Mol. Cell* **5**, 753–760.
- Malik, S., Wallberg, A.E., Kang, Y.K., and Roeder, R.G. (2002). TRAP/SMCC/mediator-dependent transcriptional activation from DNA and chromatin templates by orphan nuclear receptor hepatocyte nuclear factor 4. *Mol. Cell. Biol.* **22**, 5626–5637.
- Malik, S., Chaurasia, P., Lahudkar, S., Durairaj, G., Shukla, A., and Bhaumik, S.R. (2010). Rad26p, a transcription-coupled repair factor, is recruited to the site of DNA lesion in an elongating RNA polymerase II-dependent manner in vivo. *Nucleic Acids Res.* **38**, 1461–1477.
- Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**, 571–574.
- Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature* **399**, 700–704.
- Matangkasombut, O., Auty, R., and Buratowski, S. (2004). Structure and function of the TFIID complex. *Adv. Protein Chem.* **67**, 67–92.
- Mathieu, N., Kaczmarek, N., and Naegeli, H. (2010). Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17545–17550.
- Matsuda, T., Saijo, M., Kuraoka, I., Kobayashi, T., Nakatsu, Y., Nagai, A., Enjoji, T., Masutani, C., Sugawara, K., and Hanaoka, F. (1995). DNA repair protein XPA binds replication protein A (RPA). *J. Biol. Chem.* **270**, 4152–4157.
- Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIF: a model for promoter clearance. *Genes Dev.* **8**, 515–524.

- Le May, N., Mota-Fernandes, D., Vélez-Cruz, R., Iltis, I., Biard, D., and Egly, J.M. (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.
- Le May, N., Fradin, D., Iltis, I., Bougnères, P., and Egly, J.-M. (2012). XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell* **47**, 622–632.
- Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**, 1723–1725.
- Mayne, L.V., and Lehmann, A.R. (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 Res.* **42**, 1473–1478.
- McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306–3318.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361.
- McKay, B.C., Chen, F., Clarke, S.T., Wiggin, H.E., Harley, L.M., and Ljungman, M. (2001). UV light-induced degradation of RNA polymerase II is dependent on the Cockayne's syndrome A and B proteins but not p53 or MLH1. *Mutat. Res.* **485**, 93–105.
- McKune, K., Moore, P.A., Hull, M.W., and Woychik, N.A. (1995). Six human RNA polymerase subunits functionally substitute for their yeast counterparts. *Mol. Cell. Biol.* **15**, 6895–6900.
- Medler, S., Al Husini, N., Raghunayakula, S., Mukundan, B., Aldea, A., and Ansari, A. (2011). Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. *J. Biol. Chem.* **286**, 33709–33718.
- Mei Kwei, J.S., Kuraoka, I., Horibata, K., Ubukata, M., Kobatake, E., Iwai, S., Handa, H., and Tanaka, K. (2004). Blockage of RNA polymerase II at a cyclobutane pyrimidine dimer and 6-4 photoproduct. *Biochem. Biophys. Res. Commun.* **320**, 1133–1138.
- Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**, 223–226.
- Mellon, I., and Hanawalt, P.C. (1989). Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342**, 95–98.
- Mencía, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., and Struhl, K. (2002). Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol. Cell* **9**, 823–833.
- Menck, C.F., and Munford, V. (2014). DNA repair diseases: What do they tell us about cancer and aging? *Genet. Mol. Biol.* **37**, 220–233.
- Mendoza-Parra, M.A., Sankar, M., Walia, M., and Gronemeyer, H. (2011). POLYPHEMUS: R package for comparative analysis of RNA polymerase II ChIP-seq profiles by non-linear normalization. *Nucleic Acids Res.*
- Meyer, K.D., Lin, S., Bernecky, C., Gao, Y., and Taatjes, D.J. (2010). p53 activates transcription by directing structural shifts within Mediator. *Nat. Struct. Mol. Biol.* **17**, 753–760.
- Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat. Struct. Mol. Biol.* **13**, 603–610.
- Min, J.-H., and Pavletich, N.P. (2007). Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575.
- Mittler, G., Kremmer, E., Timmers, H.T., and Meisterernst, M. (2001). Novel critical role of a human Mediator complex for basal RNA polymerase II transcription. *EMBO Rep.* **2**, 808–813.
- Miyao, T., Barnett, J.D., and Woychik, N.A. (2001). Deletion of the RNA polymerase subunit RPB4 acts as a global, not stress-specific, shut-off switch for RNA polymerase II transcription at high temperatures. *J. Biol. Chem.* **276**, 46408–46413.
- Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348.
- Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270.

- Mocquet, V., Kropachev, K., Kolbanovskiy, M., Kolbanovskiy, A., Tapias, A., Cai, Y., Broyde, S., Geacintov, N.E., and Egly, J.-M. (2007). The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. *EMBO J.* 26, 2923–2932.
- Mocquet, V., Lainé, J.P., Riedl, T., Yajin, Z., Lee, M.Y., and Egly, J.M. (2008). Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J.* 27, 155–167.
- Mohibullah, N., and Hahn, S. (2008). Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* 22, 2994–3006.
- Moreland, R.J., Tirode, F., Yan, Q., Conaway, J.W., Egly, J.M., and Conaway, R.C. (1999). A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *J. Biol. Chem.* 274, 22127–22130.
- Morettini, S., Tribus, M., Zeilner, A., Sebald, J., Campo-Fernandez, B., Scheran, G., Wörle, H., Podhraski, V., Fyodorov, D.V., and Lusser, A. (2011). The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization. *Nucleic Acids Res.* 39, 3103–3115.
- Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8253–8257.
- Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H.F., and Fousteri, M.I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol. Cell* 27, 311–323.
- Mu, D., Park, C.H., Matsunaga, T., Hsu, D.S., Reardon, J.T., and Sancar, A. (1995). Reconstitution of human DNA repair excision nuclease in a highly defined system. *J. Biol. Chem.* 270, 2415–2418.
- Mu, D., Hsu, D.S., and Sancar, A. (1996). Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.* 271, 8285–8294.
- Muftuoglu, M., Selzer, R., Tuo, J., Brosh, R.M., Jr, and Bohr, V.A. (2002). Phenotypic consequences of mutations in the conserved motifs of the putative helicase domain of the human Cockayne syndrome group B gene. *Gene* 283, 27–40.
- Mukundan, B., and Ansari, A. (2011). Novel Role for Mediator Complex Subunit Srb5/Med18 in Termination of Transcription. *J. Biol. Chem.* 286, 37053–37057.
- Van Mullem, V., Wery, M., Werner, M., Vandenhoute, J., and Thuriaux, P. (2002). The Rpb9 subunit of RNA polymerase II binds transcription factor TFIIE and interferes with the SAGA and elongator histone acetyltransferases. *J. Biol. Chem.* 277, 10220–10225.
- Mullenders, L.H., Vrieling, H., Venema, J., and van Zeeland, A.A. (1991). Hierarchies of DNA repair in mammalian cells: biological consequences. *Mutat. Res.* 250, 223–228.
- Muncke, N., Jung, C., Rüdiger, H., Ulmer, H., Roeth, R., Hubert, A., Goldmuntz, E., Driscoll, D., Goodship, J., Schön, K., et al. (2003). Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries). *Circulation* 108, 2843–2850.
- Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., et al. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342, 1238724.
- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nat. Genet.* 39, 1507–1511.
- Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. *Annu. Rev. Biochem.* 69, 729–749.
- Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* 12, 45–54.
- Näär, A.M., Beurang, P.A., Zhou, S., Abraham, S., Solomon, W., and Tjian, R. (1999). Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* 398, 828–832.
- Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. *Annu. Rev. Biochem.* 70, 475–501.
- Näär, A.M., Taatjes, D.J., Zhai, W., Nogales, E., and Tjian, R. (2002). Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes Dev.* 16, 1339–1344.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349.

- Nakamura, Y., Yamamoto, K., He, X., Otsuki, B., Kim, Y., Murao, H., Soeda, T., Tsumaki, N., Deng, J.M., Zhang, Z., et al. (2011). Wwp2 is essential for palatogenesis mediated by the interaction between Sox9 and mediator subunit 25. *Nat. Commun.* 2, 251.
- Nakatsu, Y., Asahina, H., Citterio, E., Rademakers, S., Vermeulen, W., Kamiuchi, S., Yeo, J.P., Khaw, M.C., Saijo, M., Kodo, N., et al. (2000). XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J. Biol. Chem.* 275, 34931–34937.
- Nardo, T., Oneda, R., Spivak, G., Vaz, B., Mortier, L., Thomas, P., Orioli, D., Laugel, V., Stary, A., Hanawalt, P.C., et al. (2009). A UV-sensitive syndrome patient with a specific CSA mutation reveals separable roles for CSA in response to UV and oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6209–6214.
- Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., et al. (2003). Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol. Cell. Biol.* 23, 1863–1873.
- Nechaev, S., Fargo, D.C., Santos, G. dos, Liu, L., Gao, Y., and Adelman, K. (2010). Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in *Drosophila*. *Science* 327, 335–338.
- Neil, H., Malabat, C., d' Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042.
- Ng, J.M.Y., Vermeulen, W., van der Horst, G.T.J., Bergink, S., Sugasawa, K., Vrieling, H., and Hoeijmakers, J.H.J. (2003). A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev.* 17, 1630–1645.
- Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.-R., and Lis, J.T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol. Cell* 13, 55–65.
- Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G., and Burley, S.K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4862–4867.
- Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005). Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol. Cell. Biol.* 25, 5664–5674.
- Nock, A., Ascano, J.M., Barrero, M.J., and Malik, S. (2012). Mediator-regulated transcription through the +1 nucleosome. *Mol. Cell* 48, 837–848.
- Nonet, M.L., and Young, R.A. (1989). Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* 123, 715–724.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Oda, Y., Ishikawa, M.H., Hawker, N.P., Yun, Q.-C., and Bikle, D.D. (2007). Differential role of two VDR coactivators, DRIP205 and SRC-3, in keratinocyte proliferation and differentiation. *J. Steroid Biochem. Mol. Biol.* 103, 776–780.
- Oda, Y., Chalkley, R.J., Burlingame, A.L., and Bikle, D.D. (2010). The transcriptional coactivator DRIP/mediator complex is involved in vitamin D receptor function and regulates keratinocyte proliferation and differentiation. *J. Invest. Dermatol.* 130, 2377–2388.
- Oelgeschläger, T., Tao, Y., Kang, Y.K., and Roeder, R.G. (1998). Transcription activation via enhanced preinitiation complex assembly in a human cell-free system lacking TAFII. *Mol. Cell* 1, 925–931.
- Ogi, T., and Lehmann, A.R. (2006). The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair. *Nat. Cell Biol.* 8, 640–642.
- Ogi, T., Limsirichaiikul, S., Overmeer, R.M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N.G., et al. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol. Cell* 37, 714–727.
- Oh, K.-S., Khan, S.G., Jaspers, N.G.J., Raams, A., Ueda, T., Lehmann, A., Friedmann, P.S., Emmert, S., Gratchev, A., Lachlan, K., et al. (2006). Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Hum. Mutat.* 27, 1092–1103.
- Ohkuma, Y., and Roeder, R.G. (1994). Regulation of TFIIF ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* 368, 160–163.
- Ohkuma, Y., Hashimoto, S., Wang, C.K., Horikoshi, M., and Roeder, R.G. (1995). Analysis of the role of TFIIE in basal transcription and TFIIF-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. *Mol. Cell. Biol.* 15, 4856–4866.

- Oksenyich, V., Bernardes de Jesus, B., Zhovmer, A., Egly, J.-M., and Coin, F. (2009). Molecular insights into the recruitment of TFIH to sites of DNA damage. *EMBO J.* *28*, 2971–2980.
- Orelli, B., McClendon, T.B., Tsodikov, O.V., Ellenberger, T., Niedernhofer, L.J., and Schärer, O.D. (2010). The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *J. Biol. Chem.* *285*, 3705–3712.
- Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. *Cell* *108*, 439–451.
- Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* *3*, 379–387.
- Ottaviani, A., Gilson, E., and Magdinier, F. (2008). Telomeric position effect: from the yeast paradigm to human pathologies? *Biochimie* *90*, 93–107.
- Overmeer, R.M., Moser, J., Volker, M., Kool, H., Tomkinson, A.E., van Zeeland, A.A., Mullenders, L.H.F., and Foustari, M. (2011). Replication protein A safeguards genome integrity by controlling NER incision events. *J. Cell Biol.* *192*, 401–415.
- Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J.* *19*, 6141–6149.
- Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* *61*, 1053–1095.
- Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The Role of the Transcription Bubble and TFIIB in Promoter Clearance by RNA Polymerase II. *Mol. Cell* *19*, 101–110.
- Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., and Schultz, P. (2009). Mapping the Initiator Binding Taf2 Subunit in the Structure of Hydrated Yeast TFIID. *Structure* *17*, 363–373.
- Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* *465*, 956–960.
- Pardee, T.S., Bangur, C.S., and Ponticelli, A.S. (1998). The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J. Biol. Chem.* *273*, 17859–17864.
- Park, C.H., and Sancar, A. (1994). Formation of a ternary complex by human XPA, ERCC1, and ERCC4(XPF) excision repair proteins. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 5017–5021.
- Park, C.-J., and Choi, B.-S. (2006). The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway. *FEBS J.* *273*, 1600–1608.
- Park, D., Lee, Y., Bhupindersingh, G., and Iyer, V.R. (2013). Widespread Misinterpretable ChIP-seq Bias in Yeast. *PLoS ONE* *8*, e83506.
- Park, J.M., Kim, H.S., Han, S.J., Hwang, M.S., Lee, Y.C., and Kim, Y.J. (2000). In vivo requirement of activator-specific binding targets of mediator. *Mol. Cell Biol.* *20*, 8709–8719.
- Park, J.M., Werner, J., Kim, J.M., Lis, J.T., and Kim, Y.J. (2001). Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Mol. Cell* *8*, 9–19.
- Patrick, S.M., and Turchi, J.J. (2002). Xeroderma pigmentosum complementation group A protein (XPA) modulates RPA-DNA interactions via enhanced complex stability and inhibition of strand separation activity. *J. Biol. Chem.* *277*, 16096–16101.
- Pei, Y., Schwer, B., and Shuman, S. (2003). Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J. Biol. Chem.* *278*, 7180–7188.
- Peng, J., and Zhou, J.-Q. (2012). The tail-module of yeast Mediator complex is required for telomere heterochromatin maintenance. *Nucleic Acids Res.* *40*, 581–593.
- Van de Peppel, J., Kettelarij, N., van Bakel, H., Kockelkorn, T.T.J.P., van Leenen, D., and Holstege, F.C.P. (2005). Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* *19*, 511–522.
- Perdiz, D., Grof, P., Mezzina, M., Nikaido, O., Moustacchi, E., and Sage, E. (2000). Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. Possible role of Dewar photoproducts in solar mutagenesis. *J. Biol. Chem.* *275*, 26732–26742.
- Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* *23*, 297–305.
- Peterson, C.L. (2000). ATP-dependent chromatin remodeling: going mobile. *FEBS Lett.* *476*, 68–72.

- Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. (2002). The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 14670–14675.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* *20*, 2922–2936.
- Pines, A., Vrouwe, M.G., Martejijn, J.A., Typas, D., Luijsterburg, M.S., Cansoy, M., Hensbergen, P., Deelder, A., de Groot, A., Matsumoto, S., et al. (2012). PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. *J. Cell Biol.* *199*, 235–249.
- Pinto, I., Wu, W.H., Na, J.G., and Hampsey, M. (1994). Characterization of *sua7* mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J. Biol. Chem.* *269*, 30569–30573.
- Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* *122*, 517–527.
- Politi, A., Moné, M.J., Houtsmuller, A.B., Hoogstraten, D., Vermeulen, W., Heinrich, R., and van Driel, R. (2005). Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies. *Mol. Cell* *19*, 679–690.
- Polo, S.E., Roche, D., and Almouzni, G. (2006). New histone incorporation marks sites of UV repair in human cells. *Cell* *127*, 481–493.
- Porrúa, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* *20*, 884–891.
- Powell, W., and Reines, D. (1996). Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. *J. Biol. Chem.* *271*, 6866–6873.
- Prakash, S., and Prakash, L. (2000). Nucleotide excision repair in yeast. *Mutat. Res.* *451*, 13–24.
- Prather, D.M., Larschan, E., and Winston, F. (2005). Evidence that the Elongation Factor TFIIS Plays a Role in Transcription Initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *25*, 2650–2659.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* *322*, 1851–1854.
- Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* *14*, 105–110.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* *108*, 501–512.
- Qiu, H., Park, E., Prakash, L., and Prakash, S. (1993). The *Saccharomyces cerevisiae* DNA repair gene RAD25 is required for transcription by RNA polymerase II. *Genes Dev.* *7*, 2161–2171.
- Qiu, H., Hu, C., Zhang, F., Hwang, G.J., Swanson, M.J., Boonchird, C., and Hinnebusch, A.G. (2005). Interdependent Recruitment of SAGA and Srb Mediator by Transcriptional Activator Gcn4p. *Mol. Cell. Biol.* *25*, 3461–3474.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* *26*, 841–842.
- Rachez, C., and Freedman, L.P. (2001). Mediator complexes and transcription. *Curr. Opin. Cell Biol.* *13*, 274–280.
- Rachez, C., Suldan, Z., Ward, J., Chang, C.P., Burakov, D., Erdjument-Bromage, H., Tempst, P., and Freedman, L.P. (1998). A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes Dev.* *12*, 1787–1800.
- Rachez, C., Lemon, B.D., Suldan, Z., Bromleigh, V., Gamble, M., Näär, A.M., Erdjument-Bromage, H., Tempst, P., and Freedman, L.P. (1999). Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* *398*, 824–828.
- Rademakers, S., Volker, M., Hoogstraten, D., Nigg, A.L., Moné, M.J., Van Zeeland, A.A., Hoeijmakers, J.H.J., Houtsmuller, A.B., and Vermeulen, W. (2003). Xeroderma pigmentosum group A protein loads as a separate factor onto DNA lesions. *Mol. Cell. Biol.* *23*, 5755–5767.
- Rana, R., Surapureddi, S., Kam, W., Ferguson, S., and Goldstein, J.A. (2011). Med25 is required for RNA polymerase II recruitment to specific promoters, thus regulating xenobiotic and lipid metabolism in human liver. *Mol. Cell. Biol.* *31*, 466–481.

- Rani, P.G., Ranish, J.A., and Hahn, S. (2004). RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Mol. Cell. Biol.* **24**, 1709–1720.
- Ranish, J.A., and Hahn, S. (1991). The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J. Biol. Chem.* **266**, 19320–19327.
- Ranish, J.A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev.* **13**, 49–63.
- Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X., Eng, J., and Aebersold, R. (2004). Identification of TFB5, a new component of general transcription and DNA repair IIF. *Nat. Genet.* **36**, 707–713.
- Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J. Biol. Chem.* **287**, 23549–23561.
- Rau, M.J., Fischer, S., and Neumann, C.J. (2006). Zebrafish Trap230/Med12 is required as a coactivator for Sox9-dependent neural crest, cartilage and ear development. *Dev. Biol.* **296**, 83–93.
- Ray, A., Mir, S.N., Wani, G., Zhao, Q., Battu, A., Zhu, Q., Wang, Q.-E., and Wani, A.A. (2009). Human SNF5/INI1, a component of the human SWI/SNF chromatin remodeling complex, promotes nucleotide excision repair by influencing ATM recruitment and downstream H2AX phosphorylation. *Mol. Cell. Biol.* **29**, 6206–6219.
- Reardon, J.T., and Sancar, A. (2002). Molecular anatomy of the human excision nuclease assembled at sites of DNA damage. *Mol. Cell. Biol.* **22**, 5938–5945.
- Reardon, J.T., and Sancar, A. (2003). Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes Dev.* **17**, 2539–2551.
- Rechkunova, N.I., Maltseva, E.A., and Lavrik, O.I. (2008). Nucleotide excision repair in higher eukaryotes: Mechanism of primary damage recognition. *Mol. Biol.* **42**, 20–26.
- Reeves, W.M., and Hahn, S. (2003). Activator-independent functions of the yeast mediator sin4 complex in preinitiation complex formation and transcription reinitiation. *Mol. Cell. Biol.* **23**, 349–358.
- Reeves, W.M., and Hahn, S. (2005). Targets of the Gal4 transcription activator in functional transcription complexes. *Mol. Cell. Biol.* **25**, 9092–9102.
- Reid, J., and Svejstrup, J.Q. (2004). DNA damage-induced Def1-RNA polymerase II interaction and Def1 requirement for polymerase ubiquitylation in vitro. *J. Biol. Chem.* **279**, 29875–29878.
- Reinberg, D., Orphanides, G., Ebright, R., Akoulitchev, S., Carcamo, J., Cho, H., Cortes, P., Drapkin, R., Flores, O., Ha, I., et al. (1998). The RNA polymerase II general transcription factors: past, present, and future. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 83–103.
- Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301.
- Risheg, H., Graham, J.M., Clark, R.D., Rogers, R.C., Opitz, J.M., Moeschler, J.B., Peiffer, A.P., May, M., Joseph, S.M., Jones, J.R., et al. (2007). A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat. Genet.* **39**, 451–453.
- Risley, M.D., Clowes, C., Yu, M., Mitchell, K., and Hentges, K.E. (2010). The Mediator complex protein Med31 is required for embryonic growth and cell proliferation during mammalian development. *Dev. Biol.* **342**, 146–156.
- Robert, F., Forget, D., Li, J., Greenblatt, J., and Coulombe, B. (1996). Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. *J. Biol. Chem.* **271**, 8517–8520.
- Robinson, P.J.J., Bushnell, D.A., Trnka, M.J., Burlingame, A.L., and Kornberg, R.D. (2012). Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.*
- Rocha, P.P., Scholze, M., Bleiss, W., and Schrewe, H. (2010). Med12 is essential for early mouse development and for canonical Wnt and Wnt/PCP signaling. *Dev. Camb. Engl.* **137**, 2723–2731.
- Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J.M., and Chambon, P. (1997). Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIF and phosphorylation by CDK7. *Cell* **90**, 97–107.
- Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell. Biol.* **28**, 1883–1891.

- Ryu, S., Zhou, S., Ladurner, A.G., and Tjian, R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* 397, 446–450.
- Sadhale, P.P., and Woychik, N.A. (1994). C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Mol. Cell. Biol.* 14, 6164–6170.
- Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493, 437–440.
- Sakurai, H., and Fukasawa, T. (1999). Activator-specific requirement for the general transcription factor IIE in yeast. *Biochem. Biophys. Res. Commun.* 261, 734–739.
- Sakurai, H., and Fukasawa, T. (2003). Artificial recruitment of certain Mediator components affects requirement of basal transcription factor IIE. *Genes Cells Devoted Mol. Cell. Mech.* 8, 41–50.
- Sakurai, H., Kim, Y.J., Ohishi, T., Kornberg, R.D., and Fukasawa, T. (1996). The yeast GAL11 protein binds to the transcription factor IIE through GAL11 regions essential for its in vivo function. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9488–9492.
- Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., and Tansey, W.P. (2001). Regulation of transcriptional activation domain function by ubiquitin. *Science* 293, 1651–1653.
- Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* 11, 415.
- Sanders, S.L., and Weil, P.A. (2000). Identification of two novel TAF subunits of the yeast *Saccharomyces cerevisiae* TFIID complex. *J. Biol. Chem.* 275, 13895–13900.
- Sanders, S.L., Garbett, K.A., and Weil, P.A. (2002). Molecular characterization of *Saccharomyces cerevisiae* TFIID. *Mol. Cell. Biol.* 22, 6000–6013.
- Sarkar, S., Kiely, R., and McHugh, P.J. (2010). The Ino80 chromatin-remodeling complex restores chromatin structure during UV DNA damage repair. *J. Cell Biol.* 191, 1061–1068.
- Sarker, A.H., Tsutakawa, S.E., Kostek, S., Ng, C., Shin, D.S., Peris, M., Campeau, E., Tainer, J.A., Nogales, E., and Cooper, P.K. (2005). Recognition of RNA Polymerase II and Transcription Bubbles by XPG, CSB, and TFIIH: Insights for Transcription-Coupled Repair and Cockayne Syndrome. *Mol. Cell* 20, 187–198.
- Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybilov, B., Swanson, S.K., Banks, C.A., Jin, J., Cai, Y., and Washburn, M.P. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol. Cell* 14, 685–691.
- Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P., and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260, 58–63.
- Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J.H., and Egly, J.M. (1994). The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *EMBO J.* 13, 2388–2392.
- Scharer, O.D. (2008). XPG: its products and biological roles. *Adv. Exp. Med. Biol.* 637, 83–92.
- Scherly, D., Nospikel, T., Corlet, J., Ucla, C., Bairoch, A., and Clarkson, S.G. (1993). Complementation of the DNA repair defect in xeroderma pigmentosum group G cells by a human cDNA related to yeast RAD2. *Nature* 363, 182–185.
- Schiano, C., Casamassimi, A., Rienzo, M., de Nigris, F., Sommese, L., and Napoli, C. (2014). Involvement of Mediator complex in malignancy. *Biochim. Biophys. Acta BBA - Rev. Cancer* 1845, 66–83.
- Schmickel, R.D., Chu, E.H., Trosko, J.E., and Chang, C.C. (1977). Cockayne syndrome: a cellular sensitivity to ultraviolet light. *Pediatrics* 60, 135–139.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* 132, 887–898.
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J.M. (2000). Molecular structure of human TFIIH. *Cell* 102, 599–607.
- Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155, 1075–1087.
- Schulz, D., Pirkl, N., Lehmann, E., and Cramer, P. (2014). Rpb4 functions mainly in mRNA synthesis by RNA polymerase II. *J. Biol. Chem.*
- Schwartz, C.E., Tarpey, P.S., Lubs, H.A., Verloes, A., May, M.M., Risheg, H., Friez, M.J., Futreal, P.A., Edkins, S., Teague, J., et al. (2007). The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. *J. Med. Genet.* 44, 472–477.

- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* 322, 1849–1851.
- Seizl, M., Larivière, L., Pfaffeneder, T., Wenzek, L., and Cramer, P. (2011). Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic Acids Res.*
- Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* 18, 735–748.
- Selby, C.P., and Sancar, A. (1994). Mechanisms of transcription-repair coupling and mutation frequency decline. *Microbiol. Rev.* 58, 317–329.
- Selzer, R.R., Nyaga, S., Tuo, J., May, A., Muftuoglu, M., Christiansen, M., Citterio, E., Brosh, R.M., Jr, and Bohr, V.A. (2002). Differential requirement for the ATPase domain of the Cockayne syndrome group B gene in the processing of UV-induced DNA damage and 8-oxoguanine lesions in human cells. *Nucleic Acids Res.* 30, 782–793.
- Serizawa, H., Conaway, J.W., and Conaway, R.C. (1994). An oligomeric form of the large subunit of transcription factor (TF) IIE activates phosphorylation of the RNA polymerase II carboxyl-terminal domain by TFIIF. *J. Biol. Chem.* 269, 20750–20756.
- Sermwittayawong, D., and Tan, S. (2006). SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. *EMBO J.* 25, 3791–3800.
- Sexton, T., Kurukuti, S., Mitchell, J.A., Umlauf, D., Nagano, T., and Fraser, P. (2012). Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat. Protoc.* 7, 1335–1350.
- Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* 406, 541–544.
- Shin, C.H., Chung, W.-S., Hong, S.-K., Ober, E.A., Verkade, H., Field, H.A., Huisken, J., and Stainier, D.Y.R. (2008). Multiple roles for Med12 in vertebrate endoderm development. *Dev. Biol.* 317, 467–479.
- Shiyonov, P., Nag, A., and Raychaudhuri, P. (1999). Cullin 4A Associates with the UV-damaged DNA-binding Protein DDB. *J. Biol. Chem.* 274, 35309–35312.
- Shopland, L.S., Hirayoshi, K., Fernandes, M., and Lis, J.T. (1995). HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. *Genes Dev.* 9, 2756–2769.
- Shpakovski, G.V., Acker, J., Wintzerith, M., Lacroix, J.F., Thuriaux, P., and Vigneron, M. (1995). Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between *Homo sapiens* and *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 4702–4710.
- Shuck, S.C., Short, E.A., and Turchi, J.J. (2008). Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.* 18, 64–72.
- Siaut, M., Zarus, C., Levivier, E., Ferri, M.-L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. (2003). An Rpb4/Rpb7-Like Complex in Yeast RNA Polymerase III Contains the Orthologue of Mammalian CGRP-RCP. *Mol. Cell. Biol.* 23, 195–205.
- Sijbers, A.M., de Laat, W.L., Ariza, R.R., Biggerstaff, M., Wei, Y.F., Moggs, J.G., Carter, K.C., Shell, B.K., Evans, E., de Jong, M.C., et al. (1996). Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 86, 811–822.
- Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr. Opin. Cell Biol.* 21, 344–351.
- Sil, A.K., Alam, S., Xin, P., Ma, L., Morgan, M., Lebo, C.M., Woods, M.P., and Hopper, J.E. (1999). The Gal3p-Gal80p-Gal4p transcription switch of yeast: Gal3p destabilizes the Gal80p-Gal4p complex in response to galactose and ATP. *Mol. Cell. Biol.* 19, 7828–7840.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38, 1348–1354.
- Sims, R.J., 3rd, Rojas, L.A., Beck, D., Bonasio, R., Schüller, R., Drury, W.J., 3rd, Eick, D., and Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* 332, 99–103.
- Singh, B.N., and Hampsey, M. (2007). A transcription-independent role for TFIIB in gene looping. *Mol. Cell* 27, 806–816.

- Singh, M.V., Bland, C.E., and Weil, P.A. (2004). Molecular and genetic characterization of a Taf1p domain essential for yeast TFIID assembly. *Mol. Cell. Biol.* *24*, 4929–4942.
- Smale, S.T., and Baltimore, D. (1989). The “initiator” as a transcription control element. *Cell* *57*, 103–113.
- Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* *72*, 449–479.
- Smale, S.T., Schmidt, M.C., Berk, A.J., and Baltimore, D. (1990). Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. U. S. A.* *87*, 4509–4513.
- Smerdon, M.J., and Lieberman, M.W. (1978). Nucleosome rearrangement in human chromatin during UV-induced DNA-repair synthesis. *Proc. Natl. Acad. Sci. U. S. A.* *75*, 4238–4241.
- Smith, R.L., and Johnson, A.D. (2000). Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem. Sci.* *25*, 325–330.
- Somesh, B.P., Reid, J., Liu, W.-F., Sjøgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* *121*, 913–923.
- Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* *21*, 319–330.
- Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* *331*, 1451–1454.
- Spaeth, J.M., Kim, N.H., and Boyer, T.G. (2011). Mediator and human disease. *Semin. Cell Dev. Biol.*
- Spahr, H., Khorosjutina, O., Baraznenok, V., Linder, T., Samuelsen, C.O., Hermand, D., Mäkela, T.P., Holmberg, S., and Gustafsson, C.M. (2003). Mediator influences *Schizosaccharomyces pombe* RNA polymerase II-dependent transcription in vitro. *J. Biol. Chem.* *278*, 51301–51306.
- Spåhr, H., Calero, G., Bushnell, D.A., and Kornberg, R.D. (2009). *Schizosaccharomyces pombe* RNA polymerase II at 3.6-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 9185–9190.
- Staresincic, L., Fagbemi, A.F., Enzlin, J.H., Gourdin, A.M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S.G., Vermeulen, W., and Scharer, O.D. (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* *28*, 1111–1120.
- Van Steeg, H., and Kraemer, K.H. (1999). Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer. *Mol. Med. Today* *5*, 86–94.
- Stefanini, M., Botta, E., Lanzafame, M., and Orioli, D. (2010). Trichothiodystrophy: from basic mechanisms to clinical implications. *DNA Repair* *9*, 2–10.
- Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell* *24*, 735–746.
- Stiller, J.W., and Hall, B.D. (2002). Evolution of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci.* *99*, 6091–6096.
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., et al. (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* *306*, 655–660.
- Struhl, K. (1984). Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. *Proc. Natl. Acad. Sci. U. S. A.* *81*, 7865–7869.
- Struhl, K. (1995). Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* *29*, 651–674.
- Stumpf, M., Waskow, C., Krötschel, M., van Essen, D., Rodriguez, P., Zhang, X., Guyot, B., Roeder, R.G., and Borggrefe, T. (2006). The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 18504–18509.
- Sugasawa, K., and Hanaoka, F. (2007). Sensing of DNA damage by XPC/Rad4: one protein for many lesions. *Nat. Struct. Mol. Biol.* *14*, 887–888.
- Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell* *2*, 223–232.
- Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.* *15*, 507–521.

- Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K., et al. (2005). UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* *121*, 387–400.
- Sun, Z.W., and Hampsey, M. (1995). Identification of the gene (*SSU71/TFG1*) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 3127–3131.
- Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W., and Reinberg, D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* *2*, 213–222.
- Sung, P., Higgins, D., Prakash, L., and Prakash, S. (1988). Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP. *EMBO J.* *7*, 3263–3269.
- Sung, P., Guzder, S.N., Prakash, L., and Prakash, S. (1996). Reconstitution of TFIIF and requirement of its DNA helicase subunits, Rad3 and Rad25, in the incision step of nucleotide excision repair. *J. Biol. Chem.* *271*, 10821–10826.
- Svejstrup, J.Q. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim. Biophys. Acta* *1677*, 64–73.
- Svejstrup, J.Q., Wang, Z., Feaver, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1995). Different forms of TFIIF for transcription and DNA repair: holo-TFIIF and a nucleotide excision repairosome. *Cell* *80*, 21–28.
- Swamy, K.B.S., Cho, C.-Y., Chiang, S., Tsai, Z.T.-Y., and Tsai, H.-K. (2009). Impact of DNA-binding position variants on yeast gene expression. *Nucleic Acids Res.* *37*, 6991–7001.
- Taatjes, D.J., Näär, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* *295*, 1058–1062.
- Takagi, Y., and Kornberg, R.D. (2006). Mediator as a general transcription factor. *J. Biol. Chem.* *281*, 80–89.
- Takagi, Y., Calero, G., Komori, H., Brown, J.A., Ehrensberger, A.H., Hudmon, A., Asturias, F., and Kornberg, R.D. (2006). Head module control of mediator interactions. *Mol. Cell* *23*, 355–364.
- Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A.S., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. *Cell* *146*, 92–104.
- Tan, Q., Linask, K.L., Ebright, R.H., and Woychik, N.A. (2000). Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. *Genes Dev.* *14*, 339–348.
- Tang, J.Y., Hwang, B.J., Ford, J.M., Hanawalt, P.C., and Chu, G. (2000). Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol. Cell* *5*, 737–744.
- Tantin, D., Kansal, A., and Carey, M. (1997). Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes. *Mol. Cell. Biol.* *17*, 6803–6814.
- Tapias, A., Auriol, J., Forget, D., Enzlin, J.H., Schärer, O.D., Coin, F., Coulombe, B., and Egly, J.-M. (2004). Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J. Biol. Chem.* *279*, 19074–19083.
- Teng, Y., Liu, H., Gill, H.W., Yu, Y., Waters, R., and Reed, S.H. (2008). *Saccharomyces cerevisiae* Rad16 mediates ultraviolet-dependent histone H3 acetylation required for efficient global genome nucleotide-excision repair. *EMBO Rep.* *9*, 97–102.
- Terriente-Félix, A., López-Varea, A., and de Celis, J.F. (2010). Identification of genes affecting wing patterning through a loss-of-function mutagenesis screen and characterization of med15 function during wing development. *Genetics* *185*, 671–684.
- Teytelman, L., Özyayın, B., Zill, O., Lefrançois, P., Snyder, M., Rine, J., and Eisen, M.B. (2009). Impact of Chromatin Structures on DNA Processing for Genomic Analyses. *PLoS One* *4*.
- Teytelman, L., Thurtle, D.M., Rine, J., and Oudenaarden, A. van (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl. Acad. Sci.* *110*, 18602–18607.
- Thakur, J.K., Arthanari, H., Yang, F., Chau, K.H., Wagner, G., and Näär, A.M. (2009). Mediator subunit Gal11p/MED15 is required for fatty acid-dependent gene activation by yeast transcription factor Oaf1p. *J. Biol. Chem.* *284*, 4422–4428.
- Theisen, J.W.M., Lim, C.Y., and Kadonaga, J.T. (2010). Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol. Cell. Biol.* *30*, 3471–3479.

- Theron, T., Fousteri, M.I., Volker, M., Harries, L.W., Botta, E., Stefanini, M., Fujimoto, M., Andressoo, J.-O., Mitchell, J., Jaspers, N.G.J., 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.
- Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* *23*, 853–864.
- Thompson, C.M., and Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 4587–4590.
- Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* *73*, 1361–1375.
- Thorel, F., Constantinou, A., Dunand-Sauthier, I., Nospikel, T., Lalle, P., Raams, A., Jaspers, N.G.J., Vermeulen, W., Shivji, M.K.K., Wood, R.D., et al. (2004). Definition of a short region of XPG necessary for TFIIH interaction and stable recruitment to sites of UV damage. *Mol. Cell. Biol.* *24*, 10670–10680.
- Tietjen, J.R., Zhang, D.W., Rodríguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S., et al. (2010). Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.* *17*, 1154–1161.
- Timmers, H.T.M., and Tora, L. (2005). SAGA unveiled. *Trends Biochem. Sci.* *30*, 7–10.
- Tirode, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* *3*, 87–95.
- Topalidou, I., and Thireos, G. (2003). Gcn4 occupancy of open reading frame regions results in the recruitment of chromatin-modifying complexes but not the mediator complex. *EMBO Rep.* *4*, 872–876.
- Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* *16*, 673–675.
- Trego, K.S., and Turchi, J.J. (2006). Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination. *Biochemistry (Mosc.)* *45*, 1961–1969.
- Tremeau-Bravard, A., Riedl, T., Egly, J.-M., and Dahmus, M.E. (2004). Fate of RNA polymerase II stalled at a cisplatin lesion. *J. Biol. Chem.* *279*, 7751–7759.
- Treuner, K., Ramsperger, U., and Knippers, R. (1996). Replication Protein A Induces the Unwinding of Long Double-stranded DNA Regions. *J. Mol. Biol.* *259*, 104–112.
- Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* *5*, 190–196.
- Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J.H. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* *71*, 939–953.
- Tsai, K.-L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2014). Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex. *Cell*.
- Tsodikov, O.V., Enzlin, J.H., Schärer, O.D., and Ellenberger, T. (2005). Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 11236–11241.
- Tudor, M., Murray, P.J., Onufryk, C., Jaenisch, R., and Young, R.A. (1999). Ubiquitous expression and embryonic requirement for RNA polymerase II coactivator subunit Srb7 in mice. *Genes Dev.* *13*, 2365–2368.
- Ura, K., Araki, M., Saeki, H., Masutani, C., Ito, T., Iwai, S., Mizukoshi, T., Kaneda, Y., and Hanaoka, F. (2001). ATP-dependent chromatin remodeling facilitates nucleotide excision repair of UV-induced DNA lesions in synthetic dinucleosomes. *EMBO J.* *20*, 2004–2014.
- Urahama, N., Ito, M., Sada, A., Yakushijin, K., Yamamoto, K., Okamura, A., Minagawa, K., Hato, A., Chihara, K., Roeder, R.G., et al. (2005). The role of transcriptional coactivator TRAP220 in myelomonocytic differentiation. *Genes Cells Devoted Mol. Cell. Mech.* *10*, 1127–1137.
- Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992). Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell* *69*, 871–881.
- Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among Distant Regulatory Elements at the β -Globin Locus Requires GATA-1 and FOG-1. *Mol. Cell* *17*, 453–462.

- Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* *474*, 516–520.
- Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A., and Mullenders, L.H. (1991). Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* *11*, 4128–4134.
- Verger, A., Baert, J.-L., Verreman, K., Dewitte, F., Ferreira, E., Lens, Z., de Launoit, Y., Villeret, V., and Monté, D. (2013). The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res.* *41*, 4847–4859.
- Verhage, R., Zeeman, A.M., de Groot, N., Gleig, F., Bang, D.D., van de Putte, P., and Brouwer, J. (1994). The RAD7 and RAD16 genes, which are essential for pyrimidine dimer removal from the silent mating type loci, are also required for repair of the nontranscribed strand of an active gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *14*, 6135–6142.
- Verrijzer, C.P., Chen, J.L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* *81*, 1115–1125.
- Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. (2000). ATP-Dependent Chromatin-Remodeling Complexes. *Mol. Cell. Biol.* *20*, 1899–1910.
- Vijayvargia, R., May, M.S., and Fondell, J.D. (2007). A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. *Cancer Res.* *67*, 4034–4041.
- Volker, M., Moné, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H.J., van Driel, R., van Zeeland, A.A., and Mullenders, L.H.F. (2001). Sequential Assembly of the Nucleotide Excision Repair Factors In Vivo. *Mol. Cell* *8*, 213–224.
- Van Vuuren, A.J., Appeldoorn, E., Odijk, H., Yasui, A., Jaspers, N.G., Bootsma, D., and Hoeijmakers, J.H. (1993). Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO J.* *12*, 3693–3701.
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998a). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* *12*, 343–356.
- Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998b). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* *17*, 7395–7403.
- Wakasugi, M., and Sancar, A. (1999). Order of assembly of human DNA repair excision nuclease. *J. Biol. Chem.* *274*, 18759–18768.
- Wakasugi, M., Reardon, J.T., and Sancar, A. (1997). The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J. Biol. Chem.* *272*, 16030–16034.
- Wal, M., and Pugh, B.F. (2012). Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq. *Methods Enzymol.* *513*, 233–250.
- Wallberg, A.E., Flinn, E.M., Gustafsson, J.A., and Wright, A.P. (2000). Recruitment of chromatin remodelling factors during gene activation via the glucocorticoid receptor N-terminal domain. *Biochem. Soc. Trans.* *28*, 410–414.
- Walmacq, C., Kireeva, M.L., Irvin, J., Nedialkov, Y., Lubkowska, L., Malagon, F., Strathern, J.N., and Kashlev, M. (2009). Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J. Biol. Chem.* *284*, 19601–19612.
- Walmsley, R.W., Chan, C.S.M., Tye, B.-K., and Petes, T.D. (1984). Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* *310*, 157–160.
- Wang, Z., and Roeder, R.G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev.* *11*, 1315–1326.
- Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Mol. Cell* *17*, 683–694.
- Wang, H., Zhai, L., Xu, J., Joo, H.-Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y., and Zhang, Y. (2006a). Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol. Cell* *22*, 383–394.
- Wang, Q., Sharma, D., Ren, Y., and Fondell, J.D. (2002). A coregulatory role for the TRAP-mediator complex in androgen receptor-mediated gene expression. *J. Biol. Chem.* *277*, 42852–42858.
- Wang, W., Huang, L., Huang, Y., Yin, J., Berk, A.J., Friedman, J.M., and Wang, G. (2009). Mediator MED23 links insulin signaling to the adipogenesis transcription cascade. *Dev. Cell* *16*, 764–771.

- Wang, W., Yao, X., Huang, Y., Hu, X., Liu, R., Hou, D., Chen, R., and Wang, G. (2013). Mediator MED23 regulates basal transcription in vivo via an interaction with P-TEFb. *Transcription* 4, 39–51.
- Wang, X., Yang, N., Uno, E., Roeder, R.G., and Guo, S. (2006b). A subunit of the mediator complex regulates vertebrate neuronal development. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17284–17289.
- Wang, X., Sun, Q., Ding, Z., Ji, J., Wang, J., Kong, X., Yang, J., and Cai, G. (2014). Redefining the modular organization of the core Mediator complex. *Cell Res.*
- Wang, Y., Fairley, J.A., and Roberts, S.G.E. (2010). Phosphorylation of TFIIIB links transcription initiation and termination. *Curr. Biol. CB* 20, 548–553.
- Warfield, L., Ranish, J.A., and Hahn, S. (2004). Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes Dev.* 18, 1022–1034.
- Watanabe, T., Hayashi, K., Tanaka, A., Furumoto, T., Hanaoka, F., and Ohkuma, Y. (2003). The Carboxy Terminus of the Small Subunit of TFIIIE Regulates the Transition from Transcription Initiation to Elongation by RNA Polymerase II. *Mol. Cell. Biol.* 23, 2914–2926.
- Weber, C.A., Salazar, E.P., Stewart, S.A., and Thompson, L.H. (1990). ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. *EMBO J.* 9, 1437–1447.
- Webster, N., Jin, J.R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell* 52, 169–178.
- Weeda, G., Ma, L.B., van Ham, R.C., van der Eb, A.J., and Hoeijmakers, J.H. (1991). Structure and expression of the human XPBC/ERCC-3 gene involved in DNA repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Nucleic Acids Res.* 19, 6301–6308.
- De Weerd-Kastelein, E.A., Keijzer, W., and Bootsma, D. (1972). Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nature. New Biol.* 238, 80–83.
- Wei, C.-L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. (2006). A global map of p53 transcription-factor binding sites in the human genome. *Cell* 124, 207–219.
- Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol.* 17, 2973–2984.
- Wellinger, R.J., Wolf, A.J., and Zakian, V.A. (1993). *Saccharomyces telomeres acquire single-strand TG1–3 tails late in S phase.* *Cell* 72, 51–60.
- Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98.
- Werner, M., Thuriaux, P., and Soutourina, J. (2009). Structure-function analysis of RNA polymerases I and III. *Curr. Opin. Struct. Biol.* 19, 740–745.
- West, M.L., and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140, 1223–1233.
- Westerling, T., Kuuluvainen, E., and Mäkelä, T.P. (2007). Cdk8 is essential for preimplantation mouse development. *Mol. Cell. Biol.* 27, 6177–6182.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* 153, 307–319.
- Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim. Biophys. Acta* 1829, 151–157.
- Wind, M., and Reines, D. (2000). Transcription elongation factor SII. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 22, 327–336.
- Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol. Cell.*
- Wood, R.D., Robins, P., and Lindahl, T. (1988). Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. *Cell* 53, 97–106.
- Woudstra, E.C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* 415, 929–933.

- Woychik, N.A., and Young, R.A. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol. Cell. Biol.* *9*, 2854–2859.
- Woychik, N.A., and Young, R.A. (1990). RNA polymerase II subunit RPB10 is essential for yeast cell viability. *J. Biol. Chem.* *265*, 17816–17819.
- Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. (1990). Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev.* *4*, 313–323.
- Wu, C.-C., Herzog, F., Jennebach, S., Lin, Y.-C., Pai, C.-Y., Aebersold, R., Cramer, P., and Chen, H.-T. (2012). RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 19232–19237.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* *121*, 725–737.
- Xie, Z., Liu, S., Zhang, Y., and Wang, Z. (2004). Roles of Rad23 protein in yeast nucleotide excision repair. *Nucleic Acids Res.* *32*, 5981–5990.
- Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* *457*, 1033–1037.
- Xue, W., Wang, J., Shen, Z., and Zhu, H. (2004). Enrichment of transcriptional regulatory sites in non-coding genomic region. *Bioinforma. Oxf. Engl.* *20*, 569–575.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* *97*, 41–51.
- Yan, M., and Gralla, J.D. (1997). Multiple ATP-dependent steps in RNA polymerase II promoter melting and initiation. *EMBO J.* *16*, 7457–7467.
- Yan, Q., Moreland, R.J., Conaway, J.W., and Conaway, R.C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* *274*, 35668–35675.
- Yang, S., Macol, C., Iyer, L., Tjian, R., van den Heuvel, S., Hart, A.C., Wagner, G., Näär, A.M., Yang, F., Vought, B.W., et al. (2006). An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* *442*, 700–704.
- Ye, T., Krebs, A.R., Choukallah, M.-A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* *39*, e35.
- Yin, J.-W., and Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. *Dev. Camb. Engl.* *141*, 977–987.
- Yin, J., Liang, Y., Park, J.Y., Chen, D., Yao, X., Xiao, Q., Liu, Z., Jiang, B., Fu, Y., Bao, M., et al. (2012). Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation. *Genes Dev.* *26*, 2192–2205.
- Yokoi, M., Masutani, C., Maekawa, T., Sugawara, K., Ohkuma, Y., and Hanaoka, F. (2000). The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIF to damaged DNA. *J. Biol. Chem.* *275*, 9870–9875.
- Yoon, S., Qiu, H., Swanson, M.J., and Hinnebusch, A.G. (2003). Recruitment of SWI/SNF by Gcn4p does not require Snf2p or Gcn5p but depends strongly on SWI/SNF integrity, SRB mediator, and SAGA. *Mol. Cell. Biol.* *23*, 8829–8845.
- You, J.-S., Wang, M., and Lee, S.-H. (2003). Biochemical analysis of the damage recognition process in nucleotide excision repair. *J. Biol. Chem.* *278*, 7476–7485.
- Yu, S., Owen-Hughes, T., Friedberg, E.C., Waters, R., and Reed, S.H. (2004). The yeast Rad7/Rad16/Abf1 complex generates superhelical torsion in DNA that is required for nucleotide excision repair. *DNA Repair* *3*, 277–287.
- Yu, S., Smirnova, J.B., Friedberg, E.C., Stillman, B., Akiyama, M., Owen-Hughes, T., Waters, R., and Reed, S.H. (2009). ABF1-binding sites promote efficient global genome nucleotide excision repair. *J. Biol. Chem.* *284*, 966–973.
- Yu, S., Teng, Y., Waters, R., and Reed, S.H. (2011). How chromatin is remodelled during DNA repair of UV-induced DNA damage in *Saccharomyces cerevisiae*. *PLoS Genet.* *7*, e1002124.
- Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* *309*, 626–630.

- Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* *408*, 225–229.
- Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 6975–6980.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* *39*, 1512–1516.
- Zhang, D.-Y., Carson, D.J., and Ma, J. (2002). The role of TFIIB-RNA polymerase II interaction in start site selection in yeast cells. *Nucleic Acids Res.* *30*, 3078–3085.
- Zhang, M., Wang, X.J., Chen, X., Bowman, M.E., Luo, Y., Noel, J.P., Ellington, A.D., Etkorn, F.A., and Zhang, Y. (2012). Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chem. Biol.* *7*, 1462–1470.
- Zhao, Q., Wang, Q.-E., Ray, A., Wani, G., Han, C., Milum, K., and Wani, A.A. (2009). Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *J. Biol. Chem.* *284*, 30424–30432.
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* *38*, 1341–1347.
- Zhou, Y., Kou, H., and Wang, Z. (2007). Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast. *Nucleic Acids Res.* *35*, 861–871.
- Zhu, C., Byers, K.J.R.P., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* *19*, 556–566.
- Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* *11*, 237.
- Zhu, X., Wirén, M., Sinha, I., Rasmussen, N.N., Linder, T., Holmberg, S., Ekwall, K., and Gustafsson, C.M. (2006). Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol. Cell* *22*, 169–178.
- Zhu, X., Zhang, Y., Bjornsdottir, G., Liu, Z., Quan, A., Costanzo, M., Davila Lopez, M., Westholm, J.O., Ronne, H., Boone, C., et al. (2011a). Histone modifications influence mediator interactions with chromatin. *Nucleic Acids Res.* *39*, 8342–8354.
- Zhu, X., Liu, B., Carlsten, J.O.P., Beve, J., Nystrom, T., Myers, L.C., and Gustafsson, C.M. (2011b). Mediator Influences Telomeric Silencing and Cellular Life Span. *Mol. Cell. Biol.* *31*, 2413–2421.
- Ziegler, L.M., Khapersky, D.A., Ammerman, M.L., and Ponticelli, A.S. (2003). Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF. *J. Biol. Chem.* *278*, 48950–48956.
- Zotter, A., Luijsterburg, M.S., Warmerdam, D.O., Ibrahim, S., Nigg, A., van Cappellen, W.A., Hoeijmakers, J.H.J., van Driel, R., Vermeulen, W., and Houtsmuller, A.B. (2006). Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. *Mol. Cell. Biol.* *26*, 8868–8879.
- Zou, J., Guo, Y., Guettouche, T., Smith, D.F., and Voellmy, R. (1998). Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* *94*, 471–480.
- Zurita, M., and Merino, C. (2003). The transcriptional complexity of the TFIIH complex. *Trends Genet. TIG* *19*, 578–584.

ANNEXES

ANNEXE I : PROTOCOLE DE CHIP

I. CULTURE ET PONTAGE AU FORMALDEHYDE

- Ensemencer 100mL d'YPD à 30°C dans un erlen de 500mL avec une préculture pour être le lendemain à $DO_{600} = 0.6$.
- Ajouter sous sorbonne 3 ml de **Formaldéhyde 37%**. Laisser agir 10 min en agitant toutes les 2 min.
- Arrêter le pontage avec 20 ml de **Glycine 2,5 M** pendant 5 min.
- Centrifuger 3 min à 4000 rpm (Centrifuge 5810, Eppendorf) dans des tubes Falcons de 50 ml.
- Reprendre les cellules dans 100 ml de **Tris 20 mM pH8 à 4°C** et centrifuger idem.
- Reprendre les cellules dans **5 ml de FA-SDS + PMSF**. Toujours travailler dans la glace.
- Transférer dans des tubes Falcon de 14 ml, centrifuger idem.
- A cette étape on peut congeler dans l'azote puis conserver les cellules à -80°C.

II. BROyage ET PREPARATION DE LA CHROMATINE

- Reprendre les cellules avec **1 ml de FA-SDS + PMSF** et transvaser par 500 µl dans 2 tubes Eppendorf de 2 ml. Ajouter 0,75 ml de billes de verre (425-600 µm, G8772-500G Sigma) par tube.
- Vortexer 30 min en chambre froide dans le Vibrax (puissance maximale), retourner les tubes toutes les 10 min.
- Percer un tube Greiner de 12 ml avec une aiguille (0,7x30 mm ou 0,5x15mm) portée au rouge.
- Placer le tube dans un Falcon de 50 ml à l'aide d'un anneau de midiprep.
- Transvaser le broyat avec les billes dans le tube percé. Laver chaque tube avec 2 fois 1 ml de **FA-SDS + PMSF** (4 ml par extrait). Transvaser et centrifuger le montage à 2000 rpm (Centrifuge 5810, Eppendorf) pendant 1 min.
- Resuspendre doucement le culot avec une pipette de 5 ml et transférer dans un tube Corex stérile de 15 mL.

- Centrifuger à 12 000 rpm (JA-17, Beckman) pendant 20 min 4°C avec des adaptateurs. Aspirer le surnageant à la trompe à vide. La chromatine cross-linkée fait un anneau transparent autour d'un culot opaque.
- Transférer le culot dans un tube Eppendorf de 2 ml avec le bout fermé d'une pipette Pasteur. Ajouter dans le Corex 0,8 ml de **FA-SDS + PMSF** et resuspendre avec la pipette Pasteur, transvaser dans le tube Eppendorf et laver le Corex avec 0,8 ml de **FA-SDS + PMSF**. Homogénéiser au maximum dans le tube Eppendorf avec la pipette Pasteur puis mettre sur roue rotative à 4°C pendant 1h minimum.
- Centrifuger à 12 000 rpm (JA-17, Beckman) pendant 20 min à 4°C.
- Préparer les pipettes Pasteur.
- Aspirer le surnageant, ajouter 1,6 ml de **FA-SDS + PMSF** et remettre en suspension avec le bout fermé d'une pipette Pasteur.

III. SONICATION DE LA CHROMATINE

- La sonde doit atteindre le niveau de 0,5 ml dans un tube Eppendorf de 2 ml sans toucher la paroi. Les tubes sont dans un mélange eau + glace + NaCl pendant la sonication.
- Réglage de l'appareil (Vibra-Cell™, Sonics®) : pulse, 60% du cycle actif, puissance 4. Soniquer 3 fois 40 secondes avec un intervalle de 20 secondes (taille moyenne des fragments 400 pb). Pour le ChIP-seq, ajouter 6 cycles de 30 secondes au Biorupteur (Bioruptor® UCD-200, Diagenode), puissance moyenne (pour une taille moyenne des fragments de 200 pb) en refroidissant quelques minutes avec de la glace au bout de 3 cycles.
- Transférer dans un tube Greiner de 12 ml et ajouter 1,2ml de **FA-SDS + PMSF**.
- Centrifuger à 10 000 rpm (JA-17, Beckman) 30 min à 4°C. La chromatine est maintenant soluble, reprendre le surnageant.
- Aliquoter par 650 µl et congeler dans l'azote. Conserver 100 µl pour contrôler la taille.

IV. CONTROLE DE LA TAILLE DES FRAGMENTS D'ADN

- Prendre un aliquot de 100 µl de chromatine, ajouter 25 µl de **Tampon Pronase 5X** et 6,25 µl de **Pronase 20mg/ml**, incuber 1h à 37°C (facultatif) puis au moins 6h à 65°C dans une étuve ce qui permet de digérer les protéines et de réverser le pontage.
- Ajouter 3,3 µl de **RNase 1 mg/ml** à partir de la solution RNase A (EN0531 Fermentas) à 10 mg/ml diluée 10x dans H₂O, et incuber 1h à 37°C dans une étuve.

- Purifier l'ADN sur colonne (PCR purification kit, Qiagen) en suivant le protocole. Charger en 2 fois sur la colonne (2 x 750 µL). Avant élution mettre 15 min en étuve à 37°C. Eluer avec 2 x 25 µl d'H₂O stérile.
- Analyser la taille des fragments par électrophorèse sur gel d'agarose sur 12 µl + 2,7 µl de [Tampon d'échantillon Xylène-Cyanol](#).

V. IMMUNOPRECIPITATION

1. Préparation des billes couplées aux anticorps

- Prélever 50 µl de la solution de stock de [billes IgG anti-souris](#) (Dynabeads[®] Pan mouse IgG, Invitrogen)
- Laver 3 x avec 500 µl de [PBS-BSA 0,1%](#) à 4°C.
- Reprendre dans 100 µl de [PBS-BSA 0,1%](#).
- Ajouter 1 µl d'anticorps 12CA5 (anti-HA) ou 5 µl de 8WG16 (anti-Pol II).
- Incuber à 1300rpm dans un agitateur (Thermomixer[®], Eppendorf) pendant 30 min à 30°C.
- Laver 2 x avec 500 µl de [PBS-BSA 0,1%](#).
- Laver 1 x avec 500 µl de [PBS-BSA 0,1%](#) sous agitation à 1300 rpm (Thermomixer[®], Eppendorf) 10 min à 30°C.
- Laver de nouveau rapidement avec 500 µl de [PBS-BSA 0,1%](#).

2. Réaction d'immunoprécipitation

- Décongeler un tube de chromatine et centrifuger à 12000 rpm (Centrifuge 4415R, Eppendorf) pendant 15 min à 4°C.
- Ajouter aux billes 500 µl de la chromatine avec 50 µl de [PBS-BSA 10 mg/ml](#).
- Incuber à 1300 rpm (Thermomixer[®], Eppendorf) pendant 2h à 21°C.
- Reprendre avec 500 µl de [FA-SDS](#), changer de tube.
- Laver 2 x avec 1 ml de [FA-SDS + NaCl](#) (concentration finale de 500 mM).
- Laver 1 x avec 1 ml de [FA-SDS + NaCl](#) sous agitation à 1300 rpm (Thermomixer[®], Eppendorf) 10 min à 21°C.
- Laver avec 500 µl de [Tampon IP](#).

- Laver avec 500 μ l de **Tampon TE**.
- Eluer à 600 rpm (Thermomixer[®], Eppendorf) pendant 20 min à 65°C avec 125 μ l de **Tampon pronase 1X**. Refroidir à température ambiante.
- Reprendre le surnageant, ajouter 6,25 μ l de **Pronase 20 mg/ml** (Roche).
- Incuber 1h à 37°C (facultatif) puis au moins 6h dans une étuve à 65°C.
- Sur les 100 μ l de chromatine restant non immunoprécipitée (Input), rajouter 25 μ l de **Tampon Pronase 5X** et 6,25 μ l de **Pronase**. (Roche) Incuber 1h à 37°C (facultatif) puis au moins 6h dans une étuve à 65°C.
- Ajouter 3,3 μ l de **RNase 1 mg/ml** aux échantillons IP et Input et incuber 1h à 37°C.
- Purifier l'ADN sur colonne Qiagen (PCR purification kit) en suivant le protocole.

VI. qPCR

- Préparer les dilutions de l'ADN suivantes :
 - Standard : T1 1/40 (5 μ l dans 195 μ l d'H₂O) ; T2 1/200 (1 μ l dans 199 μ l d'H₂O) ; T3 1/1000 (40 μ l de T2 dans 160 μ l d'H₂O) ; T4 1/5000 (40 μ l de T3 160 μ l d'H₂O).
 - Echantillons Inputs : 1/200 (1 μ l dans 199 μ l d'H₂O).
 - Echantillons IP : 1/40 (5 μ l dans 195 μ l d'H₂O).
- Réaction de PCR : 12,5 μ l de **mix Mesagreen qPCR** (SYBR[®] Eurogentec) + 1 μ l d'oligos + 6,5 μ l d'eau + 5 μ l d'ADN (IP ou Input).

VII. TAMPONS

- **Tris 20mM pH 8 (Trizma HCl)** : 20 mL de Tris 1 M + 980 mL d'H₂O.
- **Glycine 2,5 M** : 187,5 g de Glycine + H₂O qsp 1L. Solution finale filtrée et stérilisée.
- **FA-SDS + NaCl** : FA-SDS + 7 mL de NaCl pour 100mL.
- **Tampon Pronase 5X** : solutions stocks stériles.
 - Tris pH7,5 125 mM (Trizma HCl) 250 µl de 1 M
 - EDTA 25 mM 100 µl de 0,5 M
 - SDS 2,5% 500 µl de 10%
 - H₂O qsp 2 mL
- **FA-SDS** : solutions stocks stériles. Tampon final filtré (sans le PMSF).
 - Hepes KOH pH 7,5 50 mM 25 mL de 1 M
 - NaCl 150 mM 15 mL de 5 M
 - EDTA 1 mM 1 mL de 0,5 M
 - Triton 100X 1% 5 mL de 100%
 - Déoxycholate de Na 0,1% 5 mL de 10%
 - SDS 0,1% 5 mL de 10%
 - H₂O qsp 500 mL
 + PMSF 1 mM à ajouter extemporanément ($T_{1/2}=30$ min en solution aqueuse) : 1%.
- **Tampon IP** : solutions stocks stériles. Tampon final filtré.
 - Tris 10 mM pH8 0,5 mL de 1 M
 - LiCl 0,25 M 12,5 mL de 1 M
 - EDTA 1 mM 0,1 mL de 0,5 M
 - NP40 0,5% (Igepal) 0,25 mL de 100%
 - Déoxycholate de Na 0,5% 2,5 mL de 10%
 - H₂O stérile qsp 50ml
- **TE** : 0,5 mL de Tris 1M pH8 + 0,1 mL d'EDTA 0,5M + H₂O qsp 50 mL. Solution finale filtrée.
- **Tampon d'échantillon** :
 - Xylène cyanol 12,5 mg
 - Glycérol 1 mL
 - EDTA 0,5M 40 µl
 - H₂O qsp 2 mL
- **PBS-BSA 0,1%** : 0,1 g de BSA, 10 mL de PBS 10X, H₂O qsp 100 mL. Filtrer et garder à 4°C.
- **PBS-BSA 10mg/ml** : 0,1 g de BSA, 1 mL de PBS 10X, H₂O qsp 10 mL. Filtrer et garder à 4°C.

ANNEXE II : PROTOCOLE DE DOSAGE AU PICOGREEN

I. PREPARATION DE LA GAMME ETALON

- A partir de marqueur 1kb Plus (Ready-load™ 1kb Plus, Invitrogen), préparer la gamme étalon selon le protocole du kit Quant-iT™ PicoGreen® dsDNA Reagent (Invitrogen) dans du **TE 1X pH8**.
- Préparer la gamme en ng/μl : 0,1 ; 0,2 ; 0,3 ; 0,4 ; 0,5 ; 0,6 ; 0,8 ; 1 ; 2 ; 4 ; 6. Validée, puis stockée à -20°C.

II. PREPARATION DES ECHANTILLONS A DOSER

- Mettre dans des tubes Eppendorf 1μl de chaque échantillon à doser. Faire de même avec les échantillons de la gamme.
- Décongeler le tube de PicoGreen à l'abri de la lumière.
- Préparer une dilution de la **solution de PicoGreen** avec le ratio 10μl de solution mère pour 2ml de **TE 1X pH8** (prévoir 200μl par dosage).
- Rajouter 200μl de **solution de PicoGreen diluée** à chaque tube contenant l'échantillon à doser ou un échantillon de la gamme, et agiter.
- Transférer chaque solution sur plaque noire FluoroNunc F96 (VWR), et entourer de papier d'aluminium

III. MESURES

- La mesure de la fluorescence est effectuée avec le lecteur de plaques Wallac EnVision et les filtres : excitation 485 et émission 535.
- Choisir le protocole 485/535 puis sélectionner les puits à mesurer et sauver.
- Appuyer sur « LOAD » sur la machine Wallac : le portoir pour la plaque sort, poser la plaque sans le couvercle, et appuyer de nouveau sur « LOAD » pour charger la plaque.
- Lancer la mesure (« START ») puis aller dans : « results » et rechercher le fichier avec la date adéquate, sélectionner les valeurs du tableau et copier/coller dans Excel.

IV. TAMPONS

- **TE 1X** : 0,5 mL de Tris 1M pH8 + 0,1 mL d'EDTA 0,5M + H₂O qsp 50 mL. Solution finale filtrée.

ANNEXE III : PROTOCOLE DE CO-IP

I. CULTURE

- Ensemencer 50mL d'YPD à 30°C dans un Erlen de 250mL avec une préculture pour être le lendemain à $DO_{600} = 0.6-0.8$
- Centrifuger à 4000 rpm (Centrifuge 5810, Eppendorf) pendant 2min puis laver avec 2 x 50 mL d'H₂O.
- Reprendre le culot avec 10 mL de **Tampon WB-**, centrifuger à nouveau.
- Reprendre dans 1 mL du même tampon.
- Transférer dans des tubes Eppendorf de 2 mL et centrifuger à nouveau.
- Congélation à -80°C possible.

II. LYSE DES CELLULES

- Resuspendre dans 0.25 mL de **Tampon WB + DTT + PMSF + Complete (= Tp WB+)**.
- Ajouter 200 µL de billes de verres (425-600 µm, G8772-500G Sigma) et mettre sur un agitateur (VXR basic Vibrax®) à puissance maximum pendant au moins 30 min en retournant les tubes toutes les 10min.
- Centrifuger 2 x 15 min à 15000 rpm à 4°C et récupérer le surnageant.
- Conserver à -80°C et garder 2 µL pour le dosage des protéines.

III. DOSAGE DE BRADFORD

- Diluer au 1/10 les échantillons à doser (2 µL dans 18 µL d'H₂O).
- Dans une cuve de spectrométrie mettre 2 µL de solution d'échantillon diluée et compléter à 100 µL avec H₂O.
- Préparer la gamme étalon à partir d'une **solution de BSA diluée au 1/10** (1 µg/ µL). Dans 10 cuves mettre 0 ; 0,5 ; 1 ; 1,5 ; 2 ; 2,5 ; 3 ; 4 ; 5 ; 6 µL de la solution de BSA diluée et compléter à 100 µL avec H₂O.
- Diluer le **réactif de Bradford** au 1/5 et mettre 900 µL de réactif par cuve.
- Effectuer le dosage à 595 nm.

IV. IMMUNOPRECIPITATION

1. Préparation des billes couplées aux anticorps

- Prélever 50µl de la solution de stock de billes IgG anti-souris (Dynabeads® Pan mouse IgG, Invitrogen)
- Laver 3 x avec 500µl de PBS-BSA 0,1% à 4°C.
- Laver une fois supplémentaire dans 100 µl de PBS-BSA 0,1% sous agitation à 1300rpm pendant 30 min à 4°C.
- Reprendre dans 100 µl de PBS-BSA 0,1%.
- Ajouter 1µl d'anticorps anti-HA (12CA5) ou 5 µl d'anticorps anti-CTD Rpb1 Pol II (8WG16).
- Incuber à 1300 rpm dans un agitateur (Thermomixer®, Eppendorf) pendant 1h à 4°C.
- Laver 3 x avec 500 µl de PBS BSA 0,1 % avec incubation 5 min à 4°C sous agitation.
- Faire deux lavages rapides supplémentaires avec 500 µl de Tampon WB+.

(Plusieurs IP : 1 ml pour 2 IP, regrouper en un seul tube et distribuer pour chaque IP).

- Mettre à centrifuger les extraits bruts de protéines pendant 15min à 12000 rpm (Centrifuge 4415R, Eppendorf) à 4°C.

2. Réaction d'immunoprécipitation

- Ajouter les extraits bruts contenant 1,5 mg de protéines et ajuster avec le Tampon WB+ pour avoir le même volume dans chaque tube.
- Incuber sous agitation à 1300 rpm (Thermomixer®, Eppendorf) pendant 3h à 4°C. Au bout de 1h30 rajouter 1 µl de PMSF par tube.
- Laver 4 x avec 1ml de Tampon WB+ avec incubation 5min à 4°C sous agitation. Changer de tube après le 1^{er} lavage.
- Reprendre dans 40 µl de Bleu de charge 1X (+ 10% DTT ou 4% β-mercapto-éthanol).
- Eluer les protéines à 95°C pendant 10min afin de les détacher des billes et conserver le surnageant.

V. WESTERN BLOT

1. Préparation du gel d'électrophorèse

- Préparer les gels de « stacking » et « resolving » selon les proportions indiquées dans le tableau ci-dessous. Plus le poids moléculaire de la protéine à étudier est petit plus le pourcentage d'acrylamide doit être élevé pour permettre une séparation optimale.

Composants	Stacking 5% (pour 2mL)	Resolving 8% (pour 5mL)	Resolving 10% (pour 5mL)	Resolving 12% (pour 5mL)
H ₂ O	1,48 mL	2,63	2,3	2,1
Acrylamide mix 40%	0,250 mL	0,970	1,275	1,5
Tris 1M pH6,8	0,250 mL	/	/	/
Tris 1M pH8,8	/	1,3	1,3	1,3
SDS	0,020 mL	0,050	0,050	0,050
APS*	0,020 mL	0,050	0,050	0,050
TEMED*	0,002 mL	0,003	0,002	0,002

* ajoutés extemporanément

2. Préparation des échantillons

- Mettre à bouillir les extraits bruts 5 min à 85°C afin de les dénaturer.
- Bien laver les puits du gel avec du **Tampon TGS 1 X** avant de déposer.
- Dépôt des échantillons (sur 2 gels séparés si besoin) :
- Pour les extraits bruts (Input) : 30 µg d'extrait + 5 µL de **Bleu de charge 1X** (+ 10% DTT ou 4% β-mercapto-éthanol).
- Pour les IP : 10 µL.

3. Electrophorèse

- Sur chaque gel déposer 3 µL de marqueur de poids moléculaire (Page Ruler™ Plus Prestained Protein Ladder, Fermentas®).
- Faire migrer ~45 min à 160 V (A max, W max) dans le **Tampon d'électrophorèse TGS 1X**.

4. Transfert semi-sec sur membrane

- Retirer le gel de stacking et déposer le reste du gel dans du **Tampon de transfert** puis sur membrane de nitrocellulose 0,2 µm (Protran BA83 Wattman™, GE Healthcare) préalablement imprégnée du même tampon.

- Mouiller 3 à 5 papiers Wattman™ 3 mm (GE Healthcare) dans le **Tampon de transfert** et les poser sur l'appareil de transfert (Fastblot B43, Labgene), ajouter la membrane de nitrocellulose recouverte du gel d'acrylamide puis à nouveau 3 à 5 papiers Wattman préalablement imprégnés de **Tampon de transfert**.
- Régler l'appareil à 120 mA pour un gel de 6x10cm (20V, 8W max) et laisser transférer pendant 1h à 1h30.
- Colorer la membrane au **rouge Ponceau** pour vérifier l'efficacité du transfert puis bien laver la membrane avec H₂Omq.

5. Saturation et exposition aux anticorps

- Incuber la membrane dans une **solution à 5 % de lait dans du Tampon TBST** (50mL) pendant 1h à TA sur agitateur ou ON à 4°C.
- Effectuer un lavage rapide au **TBST** puis ajouter le 1^{er} anticorps dans une **solution à 2% de lait dans du Tampon TBST** (10mL) (réutilisable 2-3x).

- 12CA5 (souris) : 2 µL/10 mL	- αSrb4 (lapin) : 1 µL/10 mL
- 8WG16 (souris) : 5 µL/10 mL	- αMed6 (lapin) : 5 µL/10 mL
- 9E10 (souris) : 4 µL/10mL	- αMed20 (lapin) : 1 µL/10 mL
- αMed14 (lapin) : 3,3 µL/10 mL	- αMed17 (lapin) : 2 µL/10 mL
- Incuber 1 à 2h sous agitation.
- Faire 3 lavages de 10 min au **TBST**.
- Ajouter l'anticorps secondaire (anti-souris ou anti-lapin) dans une **solution à 2% de lait dans du Tampon TBST** (10mL) : 1µL/10mL.
- Incuber 30-45 min sous agitation.
- Faire 3 lavages de 10 min au **TBST**.

6. Détection

- Déposer la membrane sur un plastique et ajouter dessus la **solution ECL Prime** (Amersham ECL Prime™ Western Blotting Detection Reagents, GE Healthcare®).
- Incuber 5min dans le noir.
- Insérer la membrane dans une pochette plastique et éliminer l'excès de liquide puis déposer la pochette dans une cassette.

- Révéler avec un film photographique ou un appareil de visualisation des gels de type EGel® Imager (Life Technologies).

VI. TAMPONS

- **Tampon WB** : Tampon final filtré (sans PMSF, DTT, Complete) et conservé à 4°C.
 - Glycérol 10 mL de 100%
 - Hepes KOH pH 7,5 2,5 ml de 1M
 - NaCl 1 ml de 5M
 - EDTA 50 µl de 0,5M
 - NP40 25 µL (ajouté après filtration)
 - H₂O qsp 50 mL
 - Ajouter extemporanément : PMSF 100 mM ($T_{1/2}$ =30 min en solution aqueuse) : 500 µL + DTT 1M : 50µL + Complete 50X (= 1pastille dans 2mL) : 1mL.
- **PBS-BSA 0,1%** : 0,1 g de BSA + 10 mL de PBS 10X + H₂O qsp 100 mL. Tampon final filtré et conservé à 4°C.
- **Bleu de charge 5X** :
 - Tris HCl 250 mM pH6,8 2,5 ml de 1M
 - SDS 10% 1 g
 - Bleu de bromophénol 1 pte de spatule
 - Glycérol 50 % 5 ml
 - H₂O qsp 10 mL
 - Ajouter extemporanément 4 % de β-Mercapto-Ethanol ou 10% de DTT.
- **TGS 5X** : 30 g de TrisBase + 144 g de Glycine + 5 g de SDS + H₂O qsp 1L.
- **Tampon de transfert** : 2,5 g de Glycine + 5,8 g de Tris + 3,7 mL de SDS 10% + 200 mL d'EtOH + H₂O qsp 1L.
- **TBST** : 100mL de TBS 10X + 0,5 mL de Tween + H₂O qsp 1L.

ANNEXE IV : PROTOCOLE D'EXTRACTION DES ARN TOTAUX DE LEVURE

I. CULTURE

- Ensemencer 10 mL d'YPD à 30°C avec une préculture pour être le lendemain à $DO_{600} = 0.5$.
- Centrifuger à 4000 rpm (Centrifuge 5810, Eppendorf) pendant 2 min puis laver avec 1 mL d'**H₂O DEPC**
- Centrifuger idem et congeler dans l'azote puis conserver à -80°C.

II. EXTRACTION DES ARNS

- Préparer 3 séries de tubes : 1,5 mL vides + 2 mL contenant 400 µL de **PCI à 4°C** (UltraPure™ Phenol:Chloroform:Isoamyl Alcohol, Invitrogen) sous la hotte + 2 mL contenant 500 µL de **Phénol acide à 4°C** (UltraPure™ Phenol Water, Invitrogen) sous la hotte.
- Reprendre les culots de cellules dans 500 µL de **Tampon AE**.
- Ajouter 50 µL de **SDS 10%** et vortexer.
- Transférer dans les tubes contenant le Phénol acide et vortexer vigoureusement.
- Incuber à 65°C pendant 8min dans le thermomixeur à 1300rpm.
- Congeler les cellules dans un mélange Ethanol + Carboglace (< 5min).
- Centrifuger à 14000 rpm (Sigma-202MK) pendant 10 min à TA.
- Prélever 400 µL de la phase aqueuse (phase supérieure) et les transférer dans les tubes contenant le PCI sous la hotte et vortexer vigoureusement.
- Centrifuger à 15000 rpm (Sigma-202MK) pendant 5min à 4°C.
- Prélever 200-300µL de la phase aqueuse.

III. PRECIPITATION DES ARNS

- Ajouter 1/10 volume d'**Acétate de Na 3M** + 2,5 volumes d'**éthanol absolu** sous la hotte.
- Incuber au moins 10 min à -80°C.

- Centrifuger à 14000 rpm (Sigma-202MK) pendant 20 min à 4°C. Eliminer le surnageant et rincer le culot avec 1 mL d'éthanol 80% à -20°C.
- Centrifuger idem 10 min et éliminer le surnageant.
- Sécher le culot au Speed-vac ~3 min puis reprendre dans 20 µL d'H₂O DEPC.

IV. ESTIMATION DE LA CONCENTRATION D'ARN ET DE SA PURETE

- Dans une microcuve mettre 1 µL d'ARN dans 100 µL d'H₂O DEPC.
- Mesurer la DO₂₆₀ (concentration) et le rapport DO_{260/280} (pureté : doit être proche de 2).

V. VISUALISATION SUR GEL

- Diluer au 1/10 les ARNs pour pouvoir déposer 1 µg d'ARN + H₂O DEPC qsp 10 µL + 2 µL de Bleu de charge 1X.
- Migrer sur gel d'agarose 1% dans du TAE 1X.
- Marqueur : 10 µL de 1kb Plus (Ready-load™ 1kb Plus, Invitrogen).

VI. TAMPONS

- **Tampon AE :**
 - Acétate de Na 0,5mL de 3M
 - EDTA 0,6mL de 0.5M
 - Eau DEPC qsp 30mL
- **H₂O DEPC :** 500 mL d'H₂Omq + 500 µL de DEPC. Laisser reposer 30min à TA puis autoclaver.
- **Ethanol 80% :** 40 mL d'EtOH absolu + H₂O DEPC qsp 50 mL. Conserver à -20°C.
- **Bleu de charge 5X :** Tampon final filtré.
 - Glycérol 5mL
 - EDTA 200 µL
 - Xylène cyanol 1 pte de spatule
 - Bleu de bromophénol 1 pte de spatule

ANNEXE V : PROTOCOLE DE RT-PCR

I. REACTION DE RT

- La RT est réalisée avec le kit iScript™ cDNA Synthesis (Bio-Rad) qui contient un tube de « mix » contenant un mélange d'oligo(dT) et d'amorces aléatoires et un tube de « RT » qui contient l'enzyme.

- Par réaction de 10µL préparer :

- 2 µL de 5X iScript reaction mix + 0.5 µL iScript RT (mix des deux)

- 0.5 µg d'ARN

- H₂O nuclease free qsp 10µL.

- Pour chaque série, toujours faire une réaction sans mix, ni RT (= RT-).

- Mettre dans l'appareil de PCR avec le programme :

- 5min à 25°C

- 30min à 42°C

- 5min à 85°C

- ∞ à 4°C

- A la fin de la réaction de RT, transférer les 10 µL dans un Ependorf et ajouter 90 µL d'H₂O (dilution 1/10).

II. REACTION DE PCR

- Par réaction de 25µL préparer :

- 12.5µL de mix Mesagreen qPCR (SYBR® Eurogentec) + 1 µL d'oligos à 10 µM + 6,5 µL d'H₂O.

- 5 µL d'ADN (RT) (dilution au 1/10 pour tous les gènes sauf 25S rRNA dilué au 1/100).

- Réaliser une courbe standard : T1 1/40 (5µL dans 195µL), T2 1/400 (20 µL de T1 dans 180 µL d'H₂O), T3 1/4000, T4 1/40000, T5 1/400000. Utiliser les 4 derniers points de la courbe pour 25S rRNA et les 4 premiers pour tous les autres gènes.

ANNEXE VI : PROTOCOLE DE TRANSFORMATION DE LEVURE CLASSIQUE PAR LIAC

I. PRECULTURE

- Ensemencer 10 mL d'YPD à 30°C pour 5 transformations, pour être le lendemain à $DO_{600} = 0.7$.
- Vérifier que l'on a bien des cellules bourgeonnantes et pas de contamination.

II. LAVAGES

- Veiller à travailler stérilement tout au long des étapes qui suivent.
- Centrifuger 1 min à 4000 rpm (Centrifuge 5810, Eppendorf).
- Jeter le surnageant et laver le culot avec 10 mL d'[H₂O stérile](#).
- Centrifuger 1 min à 4000 rpm (Centrifuge 5810, Eppendorf).
- Jeter le surnageant et laver le culot avec 500 µL de [Tampon LiAc](#).
- Centrifuger 1 min à 4000 rpm (Centrifuge 5424, Eppendorf).
- Jeter le surnageant et resuspendre le culot dans 500 µL de [Tampon LiAc](#).

III. TRANSFORMATION

- Dans un eppendorf stérile mettre : 1 µL d'ADN + 5µL d' « [ADN carrier](#) » (YeastMaker™ Carrier DNA, Clontech) préalablement bouilli 5min à 95°C puis mis dans la glace.
- Toujours faire un contrôle négatif sans ADN.
- Puis ajouter stérilement : 100 µL de levures + 700µL de [LiAc PEG](#).
- Bien resuspendre à la pipette
- Incuber 1h à 30°C puis 15 min à 42°C, au bain-marie.
- Centrifuger 5 min à 6000 rpm et resuspendre le culot dans 300 µL d'[H₂O stérile](#).
- Etaler entre 10 et 100 µL sur milieu spécifique (par exemple CAU) et placer à 30°C pendant 3 jours.

IV. TAMPONS

- **Tampon LiAc :**

- LiAc 1 mL de 1 M
- TE 1 mL de 0,1 M
- H₂O qsp 10 mL

- **Tampon LiAc PEG :**

- LiAc 1 mL de 1 M
- TE 1 mL de 0,1 M
- PEG qsp 10 mL

ANNEXE VII : PARAMETRES UTILISES POUR LE TRAITEMENTS DES DONNEES DE CHIP-SEQ

I. PARAMETRES D'ALIGNEMENT (BOWTIE)

Nom de fichier	Trim 3'	Trim5'	Longueur de lecture	Nbr. de lectures alignées
INPUT-WT	0	0	40	2 869 572
INPUT-98	0	0	40	3 015 982
INPUT-444	0	0	40	4 048 455
INPUT-670	0	0	40	3 467 920
INPUT-504	0	0	50	15 222 648
POL2-WT#IP	0	0	40	7 869 899
POL2-98#IP	0	0	40	8 699 675
POL2-444#IP	0	0	40	10 395 906
POL2-670#IP	0	0	40	7 788 546
POL2-WT#IP	0	0	50	18 706 153
POL2-504#IP	0	0	50	16 173 935
TBP-WT#IP	0	0	40	2 844 255
TBP-98#IP	0	0	40	2 839 579
TBP-444#IP	0	0	40	3 208 885
TBP-670#IP	0	0	40	3 644 017
TBP-WT#IP	0	0	50	9 858 797
TBP-504#IP	0	0	50	11 744 791
RAD3-WT#IP	0	0	40	4 954 980
RAD3-98#IP	0	0	40	3 372 139
RAD3-670#IP	0	0	40	3 948 003
RAD3-WT#IP	0	0	50	10 641 851
RAD3-504#IP	0	0	50	7 998 733
RAD3-444#IP	0	0	50	11 272 674
KIN28-WT#IP	19	1	20	2 869 572
KIN28-98#IP	19	1	20	3 015 982
KIN28-444#IP	19	1	20	4 048 455
KIN28-670#IP	19	1	20	3 467 920
KIN28-WT#IP	29	1	20	5 003 742
KIN28-504#IP	29	1	20	3 246 545
MED15-WT#IP	19	1	20	3 525 193
MED15-98#IP	19	1	20	9 220 210
MED15-444#IP	19	1	20	3 037 533
MED15-670#IP	19	1	20	3 733 705
MED15-WT#IP	29	1	20	6 287 453
MED15-504#IP	29	1	20	3 513 743
RAD2-WT#IP	19	1	20	4 907 177
NT-WT#IP	0	0	40	1 676 789
IgG-WT#NAc	29	1	20	2 577 634

II. PARAMETRES DE DETECTION DES PICS (MACS2 ET PEAKSPLITTER)

1. Série 1 (MED17, med17-98, med17-444, med17-670)

Protéine	1- Paramètres MACS2			2- Paramètres PeakSplitter		3- NT correction
	mfold	bandwidth	qvalue	valley (V)	cutoff (C)	
Pol II	2,100	300	0.05	0.5	20	N
Kin28	2,100	300	0.05	0.5	80	O
Rad3	2,100	300	0.05	0.5	80	O
TBP	2,100	100	0.05	0.6	50	O
Rad2	2,100	300	0.01	0.6	50	O
Med15	2,100	300	0.01	0.7	80	O

2. Série 2 (MED17, med17-504)

Protéine	1- Paramètres MACS2			2- Paramètres PeakSplitter		3- NT correction
	mfold	bandwidth	qvalue	valley (V)	cutoff (C)	
Pol II	2,100	300	0.05	0.5	20	N
Kin28	2,100	300	0.01	0.5	80	O
Rad3	2,100	300	0.01	0.5	80	O
TBP	2,100	100	0.01	0.6	50	O
Med15	2,101	300	0.01	0.7	80	O

III. VALEURS DE NORMALISATION DU NOMBRE DE LECTURES

Protéine	<i>med17-98</i>	<i>med17-444</i>	<i>med17-670</i>	<i>med17-504</i>
Pol II	0.83	0.38	0.70	0.88
Kin28	1.26	0.28	0.88	1.42
Rad3	1.25	0.85	0.59	1.09
TBP	1.24	0.62	0.89	0.82
Med15	0.31	0.72	0.73	1.96

Résumé

Chez les eucaryotes, la synthèse des ARN messagers (ARNm) est un processus hautement régulé en réponse à la fixation d'activateurs spécifiques sur des régions régulatrices. Cette étape permet le recrutement de co-activateurs, des facteurs généraux de la transcription (GTFs) et de l'ARN polymérase II (Pol II) pour former le complexe de préinitiation (PIC). Le Médiateur est un complexe co-activateur essentiel à ce processus et bien qu'il ait fait l'objet de nombreuses études ces dernières années, sa complexité a empêché de parvenir à une compréhension détaillée de son mécanisme de fonctionnement *in vivo*. Au cours de ma thèse, je me suis intéressée à la sous-unité Med17 qui joue un rôle central au sein du module de tête du Médiateur et interagit directement avec la Pol II. Nous avons construit une collection de mutants thermosensibles de cette sous-unité chez la levure *Saccharomyces cerevisiae*, que nous avons ensuite caractérisés par différentes approches de biologie moléculaire et génomique fonctionnelle. Nos analyses par CHIP-seq montrent que le Médiateur influence indépendamment le recrutement et/ou la stabilisation de la TBP ainsi que des modules cœur et kinase de TFIID sur le génome. Ces résultats indiquent que, contrairement à la séquence d'assemblage linéaire observée *in vitro*, l'assemblage du PIC *in vivo* est un processus à plusieurs étapes non-séquentielles et que le Médiateur est important pour orchestrer l'arrivée des différents composants du PIC. Par ailleurs, nous avons mis en évidence un contact direct entre le Médiateur et Rad2/XPG, une endonucléase qui intervient dans la réparation de l'ADN. Une analyse à l'échelle du génome a révélé que cette protéine est présente sur les gènes de classe II, en absence de stress génotoxique et que sa localisation génomique corrèle avec celle du Médiateur. Nous avons ainsi démontré que le Médiateur est important pour le recrutement de Rad2, suggérant un nouveau rôle pour ce complexe dans la réparation de l'ADN, en plus de son rôle de co-activateur dans la transcription par la Pol II.

Abstract

In eukaryotes, the synthesis of messenger RNA (mRNA) is highly regulated in response to the binding of specific activators to regulatory regions. This step allows the recruitment of coactivators, general transcription factors (GTFs) and RNA polymerase II (Pol II) to form the preinitiation complex (PIC). Mediator is a co-activator complex essential to this process and although it has been studied intensively during the last few years, its complexity has precluded a detailed understanding of the molecular mechanisms of its function *in vivo*. During my PhD, I focused on the Med17 subunit which plays a central role within the Mediator head module and interacts directly with Pol II. We obtained a large collection of temperature-sensitive mutants of this subunit in the yeast *Saccharomyces cerevisiae*, and then characterized these mutants by different molecular biology and functional genomics approaches. Our CHIP-seq analyses show that Mediator influences independently the recruitment and/or the stabilization of TBP as well as TFIID core and kinase modules on the genome. These results indicate that, unlike a linear sequence observed *in vitro*, *in vivo* the PIC assembly is a non-sequential multistep process and that Mediator is important to orchestrate the recruitment of different PIC components. Furthermore, we identified a direct contact between Mediator and Rad2/XPG, an endonuclease involved in DNA repair. A genome-wide analysis reveals that this protein is present on class II genes in the absence of genotoxic stress, and that its genomic localization correlates with that of Mediator. We thus demonstrated that Mediator is important for Rad2 recruitment, suggesting a new role for this complex in DNA repair, in addition to its co-activator role in Pol II transcription.