

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE**

**UNIVERSITE SAAD DAHLEB BLIDA 1
FACULTE DES SCIENCES DE LA NATURE ET DE VIES**

Département de Biologie Physiologie cellulaire



Mémoire

En vue de l'obtention du diplôme
Master 2 en Science de la nature et de la vie
Filière : Sciences Biologiques
Option : Génétique & Physiologie

Thème

**Identification des gènes bidirectionnels et leur
caractéristique intrinsèque chez la souris.**

Présenté et soutenu par : Karima RABEHI

Les nombres de jury :

Mr GEUTARNI Djamel	Professeur	USDB	Président
Mm Abdul Hussein Maria Stella	MCA	USDB	Examinatrice
Mr. MOHAMED SAID R	MAA	USDB	Promoteur
Mr BELHOCINE Mohamed	MAB	PHD	Co-promoteur

Remerciements

Est venu le moment d'écrire les remerciements.... Ca y est, c'est la fin d'un long chapitre.

Nous remercions «**ALLAH**» le tout puissant de nous avoir donné le courage et la Volonté de mener à terme ce présent travail.

Je remercie tous ceux qui m'ont aidé et soutenu dans l'élaboration de ce travail, plus particulièrement:

- J'aimerais commencer par remercier profondément mon promoteur *Monsieur Remdhane MOHAMED SAID* et mon Co-promoteur *Monsieur Mohamed BELHOCINE*, pour leur soutien et leurs encouragements, qui m'ont aidé à surmonter les difficultés et les doutes qui sont apparus tout au long de cette préparation. Leurs commentaires, suggestions et critiques ont permis d'améliorer bien l'aspects de ce travail.

- je remercie vivement les membres du jury, qui ont accepté d'évaluer mon travail. Je remercie **Monsieur GEUTARNI, D.** professeur à l'université SAAD DAHLEB BLIDA, qui a bien voulu présider le jury. J'éprouve un grand respect pour son travail, son parcours, ainsi que pour ses grandes qualités humaines. Je prends aussi un réel plaisir à remercier chaleureusement Madame **Abdul Hussein Maria Stella Maitre** Maitre assistante au département de biologie d'avoir accepté d'examiner de ce travail.

- Mes très sincères remerciements à mes parents :

À *Mon très cher père « Salim »*, je le remercie pour ses efforts et son cœur qu'il a pour faire vivre la famille et pour son aide précieuse.

À *Ma très chère maman « Nadia »*, pour toutes ses prières qu'elle n'a sans cesse réitérer à mon endroit, pour toutes les peines que je lui ai fais endurées. Je lui en serais éternellement reconnaissant.

Pour tous les sacrifices consentis à mon égard. Vous avez fait de moi ce que je suis : « une femme ». Je ne vous oublierai jamais. Que dieu leur procure bonne santé et longue vie.

À mes très chère sœurs «*Meriem et soumia*» à Mon unique frère «*Abed allatif*» pour leurs soutiens et les bons moments passés ensemble, un grand merci à mes grands-mères aussi «*Mama Hada et Mamie Lucienne*» qui m'ont aidé a leur façon. Sans oublier à dédier mon cher neveu «*Adem*» Je t'aime tant et Avec tout mon amour, je te souhaite un avenir souriant.

Merci enfin à tous les membres de ma famille pour leur soutien durant ces années et pour leur affection.

Sans oublier tout les professeurs que ce soit du primaire, du moyen, du secondaire ou de l'enseignement supérieur.

Merci a tout ...

Résumé

Plusieurs études ont montré que les promoteurs des gènes codants pour des protéines peuvent initier la transcription dans les deux directions. La bioinformatique nous a aidé tout au long de notre étude à mieux identifier cette transcription de bidirectionnalités et leurs caractéristiques intrinsèques grâce à l'utilisation des résultats d'une nouvelle technique à haut débit Chromatin Immunoprecipitation-Sequencing (ChIP-Seq) réalisée auparavant par le laboratoire de Technologie Avancée pour le Génome et la Clinique (TAGC) au Marseille. Dans ce manuscrit nous démontrons que cette activité de bidirectionnalité s'exprime préférentiellement dans les cellules des lymphocytes T murins au stade double négatif (DN).

Les résultats des profils moyens obtenus ont montré l'existence d'une classe de gènes bidirectionnels codant_Non codant présentant des profils d'enrichissements en polymérase II (Pol II) au tour du site d'initiation de la transcription (TSS) qui mène aussi à la production des ARNm dans les deux sens .

Grâce à des expériences basées sur la capture d'interactions chromosomiques de CHIP_Seq visualisées par l'outil IGB (Integrated Genome Browser), nous avons eu la possibilité de trouvé un lien entre l'enrichissement en Pol II suivi par une marque épigénétique des histones (H3k4me3) la triméthylation de l'histone3 lysine4 sur la chromatine intervient dans la régulation de la transcription.

En outre, les résultats issus de l'expérience de ChIP-Seq suggèrent que cette classe des gènes riche en séquence GC répété dans leurs promoteurs ce qui confirme que cette polymérase est fonctionnelle et reproductrice dans les deux sens. Donc elle partage les mêmes caractéristiques intrinsèques des classes des gènes unidirectionnels. Ainsi l'analyse fonctionnelle de cette classe des gènes de bidirectionnalités montre un score élevé, leur participation à des fonctions spécialement transcriptionnelles est prouvée d'après l'analyse fonctionnelle par DAVID (*Database for Annotation, Visualization and Integrated Discovery*)

Mot clés : transcription bidirectionnelles, CHIP-seq, TSS , Pol II.

Abstract

Several studies have shown that promoters of genes encoding proteins can initiate transcription in both directions. Bioinformatics helped us throughout our study to better identify the bidirectionality transcription and their intrinsic characteristics by the use of results of a new broadband technology Chromatin ImmunoPrecipitation-Sequencing (ChIP-Seq) fulfilled by the laboratory of Advanced technology for the Genome and the Clinic (TAGC)in Marseille –France . In this manuscript we show that this bidirectional activity is preferentially expressed in T cells murine in double negative stage (DN).

The results of the mean profiles obtained, showed the existence of bidirectional genes class encoding-noncoding having enrichments profiles of polymerase II (Pol II) around the initiation site of transcription (TSS), which also leads to the production of the mRNA in both directions.

Due to experiences based on the capture of chromosomal interactions of CHIP-Seq visualized by the tool IGB (Integrated Genome Browser), we have been able to find a link between the enrichment Pol II followed by an epigenetic histone marks (H3K4me3) ,the trimethylation of histone H3 lysine 4 of chromatin involved in the regulation of transcription.

Furthermore, the results from the experiment of CHIP-Seq suggest that this class of genes rich of GC- repeated sequence in their promoters which confirms that this polymerase is functional and reproductive in both directions. So it shares the same intrinsic characteristics of the classes of unidirectional genes. Functional analysis of bidirectional genes class shows a high score, their participation in specially transcriptional functions is proved from functional analysis by DAVID (*Database for Annotation, Visualization and Integrated Discovery*).

Keywords: bidirectional transcription, CHIP-seq, TSS, Pol II.

الملخص

وقد أظهرت العديد من الدراسات أن رموز الجينات التي ترمز إلى بروتينات يمكنها بدء الاستنساخ في كلتا الاتجاهين. ساعدتنا المعلوماتية الحيوية لل bioinformatique طوال دراستنا في تحديد بشكل أفضل هذا الاستنساخ ثنائي الاتجاه وكذا خصائصه الجوهرية وذلك من خلال استخدام لتنتائج تقنية جديدة لل *CHIP-Seq* التي أنجزت من قبل في مخبر للتكنولوجيا المتقدم (*TAGC*) في مرسيليا . في هذه المذكرة علينا أن نظهر هذا النشاط ثنائي الاتجاه الذي يظهر بشكل واضح في مرحلة DN من مراحل تحول الخلايا المناعية التائية للفئران .

وقد أظهرت نتائج *Profils moyens* المتحصل عليها من وجود فئة من الجينات ثنائية الاتجاه الغير دالة على وجود ثرية ببيانات *Profils d'enrichissements en Pol II* (Pol II) على جانبي موقع البداية للاستنساخ (TSS) . مما يؤدي أيضا الى انتاج الحمض النووي الريبي الرسول ARNm في كلتا الاتجاهين .

من خلال تجاربه التي تستند على قبض التفاعلات الكروموزومات لل *CHIP-Seq* وعرضها بواسطة أداة (Integrated Genome Browser) IGB , لدينا فرصة لاجاد رابط بين التخصيب بال *كان* (Pol II) تاليها للهيستونات (*H3K4me3*) التي تتدخل في تنظيم الاستنساخ .

وبالإضافة الى , تشير النتائج المتحصل عليها من تجربة ال *CHIP-Seq* أن هذا النوع من الجينات غني بتسلسل من القطع ذلك

المتكررة لل GC في قطع بداياتها .

وبالتالي هي تشترك في نفس الخصائص الداخلية لفئة الجينات أحادية الاتجاه . بالإضافة الى التحليل الوظيفية لهذه الفئة من الجينات ثنائية الاتجاه التي تظهر بدرجة عالية , و مشاركتها خصوصا في وظائف الاستنساخ التي تظهر بوضوح بعد

الاستعمال للأداة (Database for Annotation, DAVID Visualization and Integrated Discovery) .

الكلمات المفتاحية :

بوليميراز , POI II , *CHIP-Seq* , TSS , الاستنساخ ثنائي الاتجاه

Sommaire

Introduction

Chapitre I : Synthèse bibliographique

I. Chromatine	3
I.1. Historique de découvertes majeures sur la chromatine	3
I.2. Structure de la chromatine	3
I.3. Domaines chromatines	4
I.4. Modification de la chromatine	6
II. La transcription des ARNm	11
II.1. Structure de la polymérase II	12
II.2. Formation du complexe de pré initiation	13
II.3. Initiation et pause	14
II.4. Reprise de l'élongation	15
II.5. Terminaison.....	17
II.6. Régulation de la transcription basale	17
III. Transcription bidirectionnelle	18

Chapitre II : Matériel et Méthodes

II. 1. Matériel	21
II.1.1. La bioinformatique	21
II.1.2.les outils bioinformatiques	21
II .1.2. 1. Logiciel R	21
II .1.2.2. Chromatine Immuno-Précipitation – séquençage (CHIP-Seq)	22
II .1.2.3. Visualisation des résultats par deux outils bioinformatiques.....	23
II .2. Méthode	25
II.2.1. Model de travail	25
II.2.2. Identification des gènes bidirectionnelle	26
II.2.2.1. Récupération du génome de la souris mm9	26
II.2.2.2. Trier les gènes non chevauchant	27
II.2.2.3. Isoler les coordonnées des promoteurs	28
II.2.2.4. Expérience de CHIP-Seq sur Pol II dans les lymphocytes-DN.....	29

II.2.2.5. Analyser l'enrichissement en Pol II autour de chaque TSS.....	29
II.2.2 6. Réalisation des Profils moyens d'enrichissement en Pol II.....	31
II .2.3. Analyse des trois classes des gènes	31
II .2.3.1. Analyse graphique IGB ' <i>Integrated Genome Browser</i> '	31
II.2.3.2. Analyse d'annotation fonctionnelle par David	32
II.2.4. Les Caractéristiques intrinsèques des gènes bidirectionnels C-NC	32
II .2.4.1. Récupérer la séquence FASTA des trois classes des gènes.....	32
II.2.4.2. Analyse GC Skew	33
II.2.5. Traitement de Pol II par le KM05283	33

Chapitre III : Résultat

III. Résultat	34
----------------------------	-----------

Chapitre IV : Discussion

Iv. Discussion.....	41
Conclusion	45
Références bibliographiques.....	51
Annexes.....	60

Liste des figures

Figure 1: Les différents niveaux de condensation de la chromatine	4
Figure 2: Noyau d'une cellule à l'interphase du cycle cellulaire. Image obtenue par cryo-microscopie Indique Les différents domaines chromatiniennes sur la figure.....	5
Figure 3: Représentation en fausses couleurs des territoires chromosomiques d'un fibroblaste humain	6
Figure 4: Les principales modifications post-traductionnelles des histones sur leur queue N-Terminale	7
Figure 5: Alignement des séquences de différentes variantes d'histones.....	10
Figure 6: Représentation schématique du « core Promoter » de l'ARN polymérase de type II	12
Figure 7: Représentation schématique de la formation du complexe de pré-initiation (PIC)	13
Figure 8: Mécanisme de maintien et de sortie de pause	15
Figure 9: Etapes de la transcription par l'ARN polymérase II	17
Figure 10: Représentation schématique des régions régulatrices permettant la modulation transcriptionnelle de l'expression des gènes	18
Figure 11 : Organisation génomique d'une paire de gènes bidirectionnels	19
Figure 12 : Principe générale de l'immuno précipitation.....	23
Figure 13 : La fenêtre principale de l'IGB.....	23
Figure 14: Présentation d'interface des données génomique avec IGB.....	24
Figure 15: l'interface La partie principale de la fenêtre du navigateur est la fenêtre de visualisation.....	24
Figure 16: DAVID outil de visualisation fonctionnelle.	25
Figure 17: Différenciation des lymphocytes T.....	25
Figure 18: les bons paramètres indiqués dans le tableau du site web UCSC.....	27
Figure 19: Cliquer sur BED, pour sauvegarder le fichier « RefSeq.txt » qui porte la liste de tous les gènes de la souris avec leur coordonnées.	27
Figure 20: liste de toutes les gènes issu du site UCSC avec ces coordonnées en « brin, début, fin, nom » sous forme d'un fichier ' RefSeq' porte enivrant 34989 gènes.....	28
Figure 21: TSS site d'initiation de la transcription des promoteurs sens et antisens.....	28
Figure 22: Liste des gènes avec son TSS calculé.	29
Figure 23: Fichier « WIG » qui contient l'enrichissement en Pol II.....	29
Figure 24: Présentation des deux parties d'enrichissement en PolII « A » et « B » autour de TSS.....	29
Figure 25: Fichier d'enrichissement au tour du TSS.....	30
Figure 26: Sélections des listes selon le ratio d'enrichissement calculé.....	30
Figure 27- Profils moyens d'enrichissement des classes des gènes distingué en fonction du score.	31
Figure 28: Schémas comment récupère FASTA séquence de UCSC.....	32
Figure 29: Présente le format FASTA de nos gènes.....	32
Figure 30: Marque épigénétique de Pol II par la phosphorylation en ser5P au début de la transcription et en ser2P au cours de l'élongation).....	33

Figure 31: Liste des gènes récupérer du site UCSC avec son ratio d'enrichissement des deux cotés A et B.	35
Figure 32 : – Profil d'enrichissement en Pol II des gènes du part et d'autre du site d'initiation du TSS.	36
Figure 33 - Diagramme représente le pourcentage de la distribution des gènes dans le génome de la souris.	36
Figure 34 : Visualisation et représentation des pics d'enrichissement des gènes en Pol II par IGB	37
Figure 35: Représentation en barre le Score d'enrichissement des gènes Unidirectionnelle.....	38
Figure36 : Représentation en barre le Score d'enrichissement des gènes coding_coding.....	38
Figure 37- Représentation en barre le Score d'enrichissement des gènes coding_Non coding....	39
Figure 38 : Profil du GC skew des trois classes des gènes.....	40
Figure 39 : Exemple d'un gène unidirectionnel avant et après le traitement par Km par IIGB.....	40
Figure 40 : Exemple d'un gène bidirectionnel avant et après le traitement par Km par IGB.....	41
Figure 41 : Positionnement de Pol II fonctionnelle sur les promoteurs unidirectionnels et bidirectionnels.....	44
Tableau 1- Paramètre à sélectionné dans le tableau des donnée.....	26

Abréviations

Ac : Anticorps

ADN : Acide DésoxyriboNucléique

ADNc : Acide Désoxyribo-Nucléique complémentaire

ARN : Acide Ribo-Nucléique

ARNm: ARN messenger

ARN pol II: ARN polymérase de type II

ARNt: ARN de transfert

CBP: CREB-binding protein

C/EBP: CCAAT/Enhancer Binding Protein

ChIP-Seq : Chromatin ImmunoPrecipitation-Sequencing

CIML : Centre d'Immunologie Marseille Luminy

CPE: Core Promoter Elements

CTD: C-terminal Domain

C_C : Codant-codant

C_NC : codant Non codant

DAVID: Database for Annotation, Visualization and Integrated Discovery.

DBD : DNA Binding Domain .

DN : Double Négatif.

DP : Double Positif.

dNTP: Désoxyribonucléotide triphosphate

DP : Double Positive

DPE: Downstream Promoter Element

DRB : Dichloro-1-β-D-Ribofuranosylbenzimidazole

DRE: Downstream Regulatory Element

DSIF: DRB Sensitivity-Inducing Factor

FT : Facteur de Transcription

GC skew : Le biais en GC

HAT: Histone Acetyl Transferase

HDAC: Histone Deacetylase

IGB: Integrated Genome Browser.

IGV: Integrated Genome viewer.

Inr: Initiator

NCBI : National Center for Biotechnology Information

ORF: Open Reading Frame

PCR: Polymerase Chain Reaction

Pol : Polymérase

NELF: Négative Elongation Factor

RNAi: RNA interference

RNP: Ribonucléoprotéine

rNTP : ribonucléotide Triphosphate

RE: Regulatory Element

Seq: séquençage.

SP : Simple Positif.

SnRNP: Small Nuclear Ribonucleoprotei

TAG : Codon terminateur de la transcription

TAGC : Technologie Avancée pour le Génome et la Clinique

TBP: TATA-Binding Protein

TF: Transcription Factor

TFII: Transcription Factor for RNA pol II

TSS: Transcription Start Site

UBD: ubiquitin binding domain

UCSC: Genome browser (University California Santa Cruz - USA).

U3 : Unique 3' séquence

U5 : Unique 5' séquence

Introduction

Dans cette introduction je vais présenter la transcription en accord avec les dernières données disponibles sur la transcription bidirectionnelle, pour offrir un cadre adéquat lors de la présentation de ce mémoire. Ce manuscrit comporte trois chapitres principaux.

Le premier chapitre s'agit d'une présentation du contexte générale de la transcription. Pour mieux comprendre la première partie de ce chapitre on a défini la chromatine et détaillé sa structure et sa dynamique. La seconde partie expose les mécanismes de la transcription par l'ARN polymérase II qui est responsable de la synthèse des ARN messagers, ce qui constitue la première étape de régulation de l'expression des gènes. Ensuite vient la partie présentant la transcription bidirectionnelle dont la fonction n'est pas connue. Cette transcription peut être initiée par des promoteurs des gènes codant ou non pour des protéines. Notre principal objectif est d'identifier les promoteurs bidirectionnels et leurs caractéristiques partagés avec des promoteurs classiques. Suite à ces identifications nous pourrions déterminer des gènes intervenant dans ce processus de bidirectionnalité.

Afin de répondre à ces questions nous avons analysé les données de séquençage à haut débit issues des expériences de ChIP-Seq réalisées auparavant par le laboratoire de TAGC à Marseille, et les confronter aux données de transcriptome murin produites par le laboratoire ou téléchargées sur les serveurs publics, et l'ensemble d'analyses utilisé dans cette étude seront décrits dans le second chapitre de matériel et méthodes.

Il s'agit de faire comprendre un nouveau domaine appelé : Bioinformatique qui a pour but de développer les outils et les méthodes pour pouvoir décrypter le génome et permettre à leur tour de développer un nouveau concept en biologie.

Après la présentation des principaux résultats obtenus dans le troisième chapitre ou nous sommes arrivé à identifier l'intervention de certains gènes dans ce processus de bidirectionnalité et déterminer leurs caractéristiques intrinsèques par rapport à celle des gènes unidirectionnelles. Ces derniers résultats seront remis dans les prochaines parties de discussion et conclusion.

Chapitre I

Synthèse Bibliographique

I. Chromatine

La chromatine (du grec khroma: couleur et sôma: corps) est une substance présente dans le noyau des cellules et qui contient une énorme quantité d'ADN emballé de manière à s'intégrer dans le noyau. Sa structure varie tout au long du cycle cellulaire (Maillet, 1981).

I.1. Historique des découvertes majeures sur la chromatine

Depuis la découverte historique des acides nucléiques, la compréhension des niveaux d'organisation de la chromatine n'a jamais cessé d'évoluer.

L'acide désoxyribonucléique (ADN) génomique eucaryote se localise dans le noyau de la cellule. La structure organisée et compacte de l'ADN dans le noyau fut observée, bien avant de connaître sa fonction, vers 1880 par Walther Flemming. Il la nomma chromatine à cause de la couleur dense observée au microscope dans le noyau de la cellule (Olins et al., 2003). Ce n'est qu'à partir de la deuxième moitié du XXème siècle que de nombreuses études ont déterminé de façon plus précise que l'ADN était porteur de l'information génétique (Avery et al., 1944) et qu'il existe une structure tridimensionnelle bien précise à la molécule d'ADN élucidée par Watson et Crick 1953 . En 1974, Kornberg et Olins ont visualisé de façon plus détaillée la structure compacte de l'ADN démontrant ainsi que des protéines nommées histones s'alliaient à l'ADN. Dans ce cas précis, il s'agit de la chromatine. L'enroulement de l'ADN autour des histones a donné la sous unité chromatinienne: le nucléosome (Kornberg, 1974). L'ensemble de ces recherches a permis d'affiner la compréhension de la chromatine tant sur sa structure et sa composition que son organisation et ses fonctions.

I.2 Structure de la chromatine

La structure de la chromatine est étroitement liée à sa fonction et permet ainsi la condensation de l'ADN selon plusieurs niveaux, tel qu'illustré à la figure 1. L'unité de base de la chromatine est le nucléosome.

Bien que la chromatine et les nucléosomes soient connus depuis un certain nombre d'années, Ce n'est que récemment que des informations sur leurs modes de régulation et sur leurs Fonctions biologiques ont été obtenues. La chromatine fait actuellement l'objet de nombreuses recherches régulièrement, des découvertes importantes apportent de nouvelles données sur sa dynamique et son rôle biologique (Noll et Kornberg, 1977).

➤ Nucléosome

L'unité élémentaire de la chromatine, comprenant une particule cœur et une région internucléosomique. La particule cœur est composée de 146 paires de bases d'ADN enroulées autour d'un octamère protéique comprenant un tétramère H3-H4 associé à deux dimères H2A-H2B. (Luger, 2003).

La région internucléosomique étant pour sa part caractérisée par la présence de l'histone H1 (histone « lien »). C'est donc la chromatine, et non l'ADN seul, qui est impliquée dans tous les événements moléculaires faisant intervenir le matériel génétique, à savoir la réplication, la transcription, la réparation et la recombinaison. (Bird, 2002).

Le complexe d'ADN et de protéines (les histones) constituent le nucléosome, qui est l'unité fondamentale de la chromatine. (Olins et Olins, 1974).

➤ Différents niveaux de compaction de la chromatine

L'enroulement de l'ADN autour de l'octamère d'histones pour former le premier niveau de compaction de la chromatine qui est le nucléofilament de 11nm, une structure dite en « collier de perles » constituée d'un enchaînement régulier de nucléosomes.

Le nucléofilament subit un second niveau de repliement grâce à la mise en place d'une cinquième histone : H1 Elle se fixe sur l'ADN qui relie 2 nucléosomes et contribue à enrouler d'avantage l'ADN pour former une fibre de chromatine de 30 nm de diamètre appelé solénoïde. Chaque tour d'enroulement de la fibre comporte 6 nucléosomes (Shao, 1999).

➤ Structures plus condensées

La fibre de 30 nm peut également se replier et former des structures dont la topologie est encore moins bien décrite. Le solénoïde forme des boucles de 300nm et des tours de longueur variable grâce à l'intervention de protéines non histones pour atteindre un niveau de compaction maximum et aboutir à la forme du chromosome métaphasique. (Felsenfeld et Groudine, 2003). Au cours de la mitose et de la méiose, il existe transitoirement un niveau extrême de compaction des chromosomes. Ils sont tellement condensés qu'on peut étudier leur morphologie au microscope optique (Wood, Severson et al., 2010).

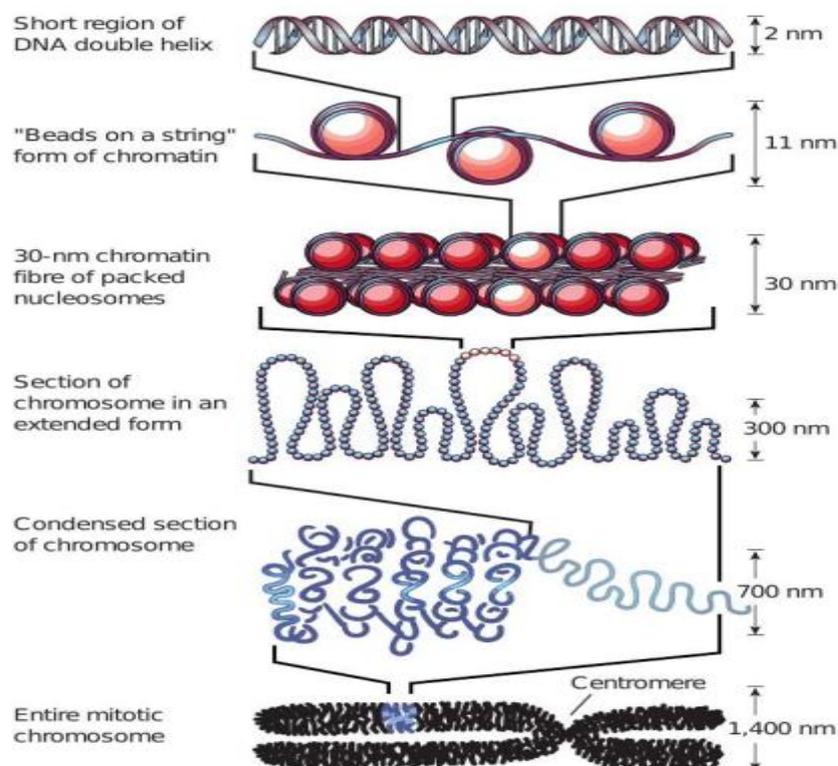


Figure 1 - Les différents niveaux de compaction de la chromatine (Felsenfeld et Groudine 2003).

I.3 Domaines chromatines

Les différents niveaux de compaction permettent la définition de domaines chromatiniens en interphase : l'hétérochromatine et l'euchromatine, que l'on peut distinguer dans le noyau après incorporation d'intercalant à l'ADN ou en microscopie électronique (Bierne et al., 2009) (Figure 2).

➤ L'euchromatine :

L'euchromatine est la chromatine décondensée riche en gène actif, que l'on retrouve répartie à l'intérieur du nucléosome, l'euchromatine est hyperacétylée et hypométhylée comparativement à l'hétérochromatine (Sarg et al., 2004). Les portions d'ADN nucléaire constituant l'euchromatine contiennent des histones acétylées et riches en méthylation sur la lysine 4 de l'histone H3. (Strahl et Allis, 2000).

➤ L'hétérochromatine :

Elle correspond à une structure condensée qui ne change pas de degré de compaction au cours du cycle cellulaire. Pas ou peu d'expression de gènes. Elle se situe principalement en périphérie du noyau et du nucléole. (Sarg et al., 2004). L'hétérochromatine est caractérisée par une hypoacétylation des histones et une hyperméthylation des lysines 9 et 27 de l'histone H3 et de la lysine 20 de l'histone H4 (Kouzarides, 2007).

On distingue deux types de séquences hétérochromatiniennes :

❖ L'hétérochromatine constitutive : est définie comme la chromatine qui reste condensée dans tous les types cellulaires. Elle est composée de séquences répétées et est située au niveau des centromères et des télomères des chromosomes.

❖ L'hétérochromatine facultative : est définie comme de l'euchromatine qui est mise en silence à certains stades du développement ou dans certains types cellulaires (Grigoriev *et al.*, 2006). Elle peut s'étendre sur un chromosome entier (X inactif dans les cellules mammifères femelles), une grande région génomique (locus des gènes HOX.), ou être localisée sur des régions régulatrices de gènes (gènes soumis à l'empreinte) (Trojer et Reinberg, 2007).

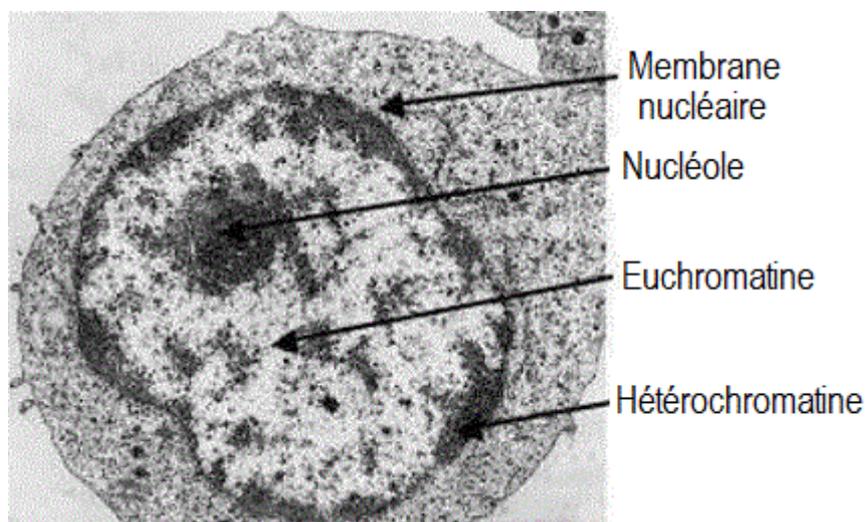


Figure 2 - Noyau d'une cellule à l'interphase du cycle cellulaire. Image obtenue par microscopie électronique montrant la définition histologique des deux formes de chromatine (Jackson, 1997).

➤ Organisation nucléaire de la chromatine

Les fibres de chromatine d'un chromosome interphasique donné ne sont pas dispersées dans le noyau comme dans un plat de spaghetti, mais sont concentrées dans un territoire spécifique qui n'empiète guère sur les territoires des autres chromosomes. (Karp ,2010).

Les territoires chromosomiques ne sont pas disposés de façon aléatoire mais sont organisés en fonction de leur taille et de leur contenu en gènes. Ainsi, les petits chromosomes sont préférentiellement localisés au cœur du noyau tandis que les plus grands sont situés près de la membrane nucléaire. Pour une taille similaire, un chromosome plus riche en gènes est situé plus près du centre qu'un chromosome pauvre en gènes (Cremer *et al.* 1982; Cremer et Cremer 2010) (Figure .3).

En outre, des gènes localisés sur des chromosomes différents mais participant a un même processus peuvent se réunir dans le noyau et y être transcrits simultanément. Ceci est illustré par les gènes codant L'ARN ribosomique, qui sont localisés sur plusieurs chromosomes différents, mais sont capables de converger aux confins du nucléole (Karp ,2010).

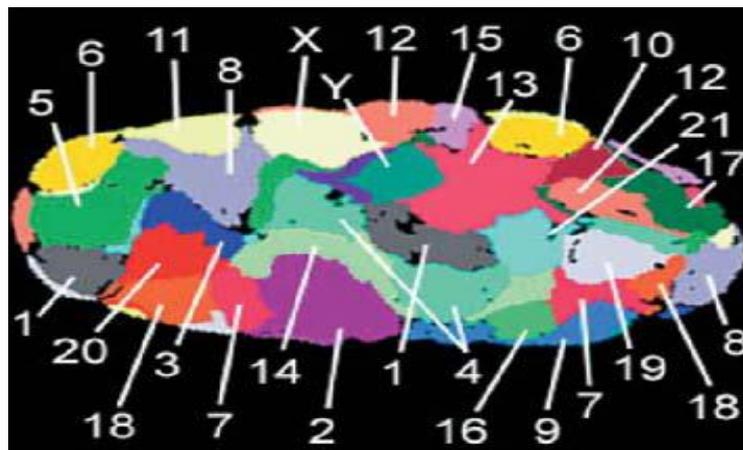


Figure –3 Représentation en fausses couleurs des territoires chromosomiques d'un fibroblaste humain (Bolzer *et al.* 2005). Chaque région colorée représente un chromosome.

Cette organisation nucléaire semble importante pour la régulation des gènes. En effet, quelques exemples montrent que la mise en silence ou l'activation d'un gène est accompagnée d'une relocalisation de ce gène dans le noyau (Dundr *et al.*, 2007; Chuang *et al.*, 2006). De plus, en déplaçant artificiellement un gène vers une autre région du noyau (comme la Périphérie nucléaire) on observe un changement dans son expression (Finlan et Bickmore 2008; Reddy *et al.* 2008), ce qui suggère que la position d'un gène dans le noyau participe à sa régulation et n'est pas seulement une conséquence du niveau d'expression de ce gène.

I.4 Modifications de la chromatine

On sait que la modification locale de la structure de la chromatine peut conduire soit à une activation soit au contraire à une répression de l'expression génique (Orlando et al., 1997).

Dès les premières étapes de l'assemblage de la chromatine, la particule élémentaire peut être soumise à des variations au niveau de l'ADN (notamment par méthylation), ou au niveau des histones qui peuvent soit présenter différentes modifications post-traductionnelles, soit exister sous forme de variantes (Campos et Reinberg, 2009).

Ces variations peuvent avoir des effets intrinsèques, extrinsèques ou dépendants d'un facteur effecteur. On parle d'un effet intrinsèque quand la modification altère la structure même du nucléosome. C'est notamment le cas des variantes d'histones qui modifient la structure du cœur du nucléosome. Un effet extrinsèque décrit un effet direct sur les contacts internucléosomiques et la structure de la chromatine.

Enfin, les effets dépendants d'un effecteur sont les variations dépendantes de la liaison de facteurs spécifiques non-histones modifiant la chromatine. Ce sont ces effets qui sont à la base de l'hypothèse du code des histones.

➤ Les modifications des queues des histones

Les queues N-Terminales des histones qui sont situées à l'extérieur de la particule nucléosomale, sont assez courtes (moins de 40 acides aminés) et peuvent donc être la cible d'enzymes qui les modifient post-traductionnellement (Luger et al., 1997). Les histones sont sujettes à une grande diversité de modifications covalentes telles que : les acétylations, méthylations, phosphorylations, ubiquitinations (Figure 4).

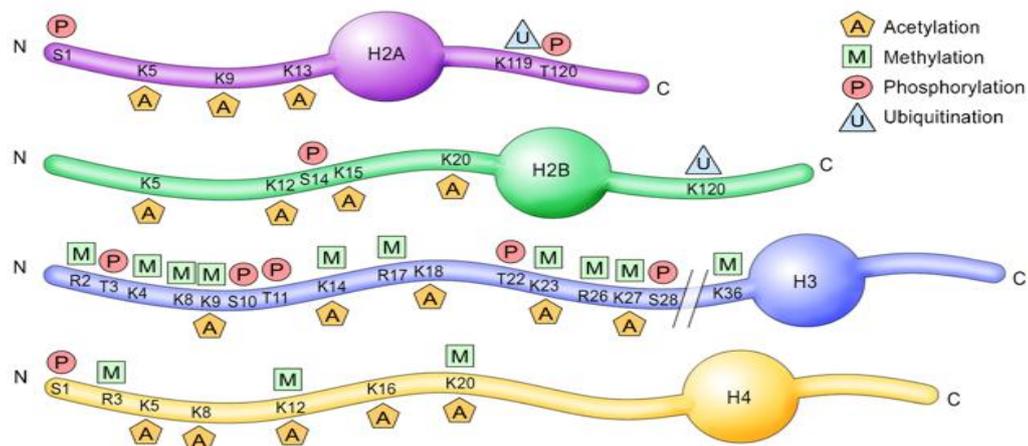


Figure 4– Les principales modifications post-traductionnelles des histones sur leur queue N-Terminale (Graff et Mansuy, 2008).

La modification la mieux étudiée jusqu'à présent est l'acétylation des résidus lysines. Cette modification correspond au transfert d'un groupement acétyle sur un résidu lysine, ce qui suggère une structure moins compacte. Elle neutralise la charge électrostatique des histones et elle change leurs propriétés structurales et/ou modifie les interactions de leurs extrémités (Allfrey, 1964).

L'état d'acétylation résulte d'un équilibre entre deux activités antagonistes : l'activité histone-acétyltransférase (HAT) et l'activité histone-désacétylase (HDAC) (Suka *et al.* 2002). Un exemple bien étudié est l'acétylation de la lysine K16 de l'histone H4 qui modifie directement les contacts inter-nucléosomiques ce qui perturbe la structure chromatinienne. En effet, les résidus 16 à 20 de l'histone H4 interagissent avec deux zones acides des hélices. C terminales de l'histone H2A du nucléosome adjacent (Zhou *et al.*, 2007).

Cette modification est liée à l'activation transcriptionnelle (Grunstein, 1997).

Elle joue un rôle dans le recrutement des complexes de remodelage et de facteurs d'initiation de la transcription (Agalioti, Chen *et al.*, 2002) et est reconnue par des protéines possédant des bromodomains, comme certains régulateurs transcriptionnels (Strahl et Allis, 2000).

La méthylation correspond à l'ajout d'un ou plusieurs groupements méthyle sur un résidu lysine ou arginine par des enzymes appelées Histone Méthyl Transférases (HMT). Cette modification est portée par les queues N-Terminales des histones H3 et H4 uniquement. Elle est également liée à la régulation transcriptionnelle et peut marquer des zones de chromatine active comme des zones de chromatine silencieuse, suivant la position de s résidus méthyles.

D'une manière générale, ces modifications comme (Lysine : H3K27me, H3K9me, H4K20me1, H3K4me1/2/3 ou H3K36me3, Arginine : H3R17me (1/2)) qui peuvent être liées à l'activation transcriptionnelle, tandis que celle (Lysine : H3K9me1/2/3, H4K20me3 ou H3K27me2/3, Arginine : H3R2me2) a la répression (Barski *et al.*, 2007).

Les groupements méthyles portés par les queues d'histones peuvent interagir avec des protéines possédant des chromodomains. Ces protéines peuvent être des activateurs ou des répresseurs transcriptionnels (Bannister et Kouzarides ., 2005).

La méthylation au niveau de la lysine 9 de l'histone H3 crée un site d'interaction de haute affinité pour la protéine HP1 (Heterochromatin Protein1), une protéine impliquée dans la formation de l'hétérochromatine (Eissenberg JC, Elgin SC, 2000). C'est par intermédiaire de son chromodomaine que HP1 reconu l'histone 3 méthylé, qui représente certainement le site de liaison de HP1 sur l'hétérochromatine (Bannister, 2001).

Les régions transcrites présentent généralement des groupements méthyles sur H3K4 ainsi que la marque H3K36me3. Cependant, l'étude de la répartition de ces marques à l'échelle du génome montre qu'il existe de nombreuses exceptions (par exemple, H3K9me3 peut être associé à des gènes actifs ou H3K4me3 à des gènes mis en silence (Barski *et al.*, 2007).

La méthylation des histones sur les arginines est caractérisée quant à elle par l'ajout de un ou 2 groupements méthyle, la diméthylation pouvant être symétrique ou asymétrique.

On sait que pour les arginines la nature de la diméthylation apporte également une information épigénétique supplémentaire et donc un nouveau niveau de complexification.

Elle est mise en place par des enzymes spécifiques de résidu, les PRMT (proteins arginine methyltransferase), qui sont classées en deux groupes : les PRMT de type I catalysent la diméthylation asymétrique est impliquée dans l'activation transcriptionnels tandis que les PRMT de type II sont responsables de la diméthylation symétrique des arginines associées à la répression transrationnelle (Wysocka, Allis et al., 2006).

La phosphorylation est également une modification importante qui peut se trouver sur les sérines, tyrosines et thréonines .Le niveau de phosphorylation des histones est contrôlées par l'action antagoniste de kinases et de phosphatases.

La majorité des travaux effectués sur cette modification ont été réalisés sur la phosphorylation de la sérine 10 de l'histone H3 (H3S10), cette marque intervient notamment pour activer rapidement certains gènes en réponse à des signaux extérieurs à la cellule (Prigent et Dimitrov, 2003).

L'ubiquitine est une petite protéine de 76 acides aminés, Un de ses rôles consiste à marquer une protéine pour la dégradation. Dans le cas qui nous intéresse ici, c'est seulement une mono-ubiquitination non déprédative d'une lysine 119 et 120 respectivement chez les mammifères. La fonction de l'ubiquitination passe en général par sa reconnaissance par des protéines effectrices contenant un domaine UBD (ubiquitin binding domain).La relation entre ubiquitination des histones et état transcriptionnels dépend du résidu modifié. En effet, l'ubiquitination de H2A (H2AK119ub) sur le corps des gènes est corrélée à la répression de la transcription, tandis que l'ubiquitination de H2B (H2BK120ub) au niveau des promoteurs et des régions codantes est associée à une chromatine transcriptionnellement active (Weake et Workman, 2008).

Strahl et Allis [2000] ont proposé que ces modifications post traductionnelles des queues des histones pouvaient constituer un "code" compréhensible par la cellule. Ceci suppose l'existence de protéines capables d'établir le code ("writer"), de lire ce code ("reader"), et de l'effacer au besoin ("eraser"). Pour la plupart des modifications post traductionnelles d'histones, des complexes protéines sont connus dans les 3 catégories (Kouzarides, 2007; Yun et al, 2011).

Il existe encore d'autres types de modifications covalentes d'histones : la glycosylation, l'ADP-ribosylation, la biotinylation, la sumoylation. Ces modifications pourraient également participer à la régulation de l'expression des gènes.

➤ **Les variantes d'histones :**

Outre les enzymes de modifications des histones et les complexes de remodelage de la chromatine ATP-dépendants, il existe une autre voie permettant de moduler la structure et les propriétés fonctionnelles de la chromatine : l'incorporation de variantes d'histones.

La séquence des variantes d'histones peut différer de celle des histones canoniques au niveau de la queue N-terminale, du domaine globulaire ou seulement de quelques acides aminés.

Ces divergences entre les histones canoniques et leurs variantes permettent, quand ces derniers sont intégrés dans les nucléosomes, d'acquiescer une structure chromatiniene spécifique. L'incorporation de variantes d'histone au sein d'un locus précis soit la mise en place de domaines chromatiniens spécifiques ou bien participe à la régulation de processus tel que la transcription ou la réparation (Campos et Reinberg, 2009).

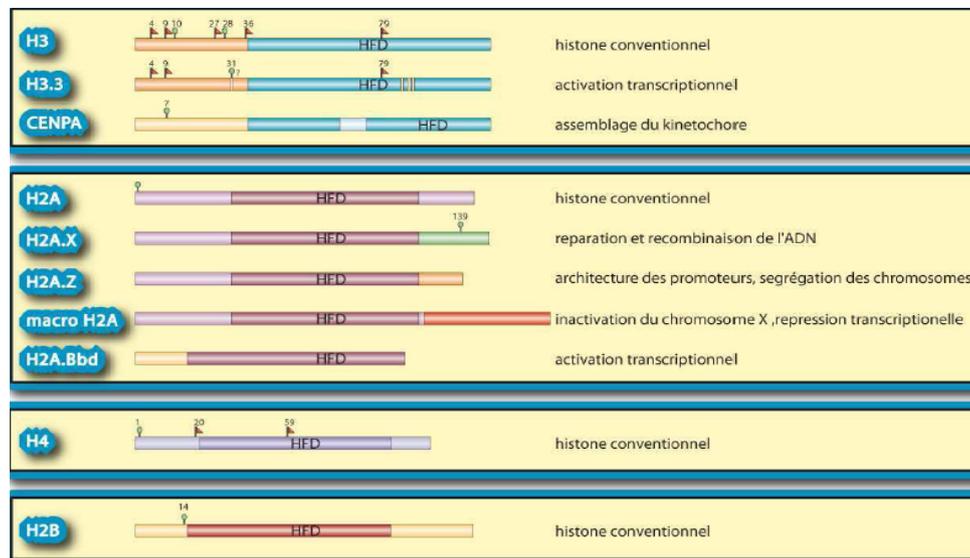


Figure 5 - Alignement des séquences de différentes variantes d'histones. HFD correspond à l' « histone fold ». Les marqueurs sont les zones de modifications post-traductionnelles. (Sarma et Reinberg (2005)).

Selon les espèces, il existe plusieurs isoformes des histones H3 (Figure 5), l'histone H3.1 correspond à la forme canonique d'H3. H3.2 et H3.3 qui divergent peu entre elles et CENPA constituant les trois variantes. Le quatrième variant CENPA (Centromere ProteinA) est associé aux régions centromériques d'H3. Tandis que le variant d'histone H3.3 est incorporé au cours de la transcription (Schwartz et Ahmad 2005; Wirbelauer, Bell et al., 2005; Daury, Chailleux et al., 2006). Le variant H3.3 porte par ailleurs les marques épigénétique de la chromatine active (méthylation des lysines 4 et acétylation des lysines 9, 14, 18 et 23) ce qui conforte l'idée de son rôle d'activateur de la transcription.

Les principales variantes d'histone H2A sont au nombre de quatre (H2AZ, H2AX, macroH2A et H2A.Bbd) (Figure 5), un variant de l'histone H2B, spécifique des tissus testiculaires et des spermatozoïdes (TSH2B).

Le variant H2A.X est impliquée dans le maintien de l'intégrité du génome en réponse aux lésions double-brins de l'ADN et dans le contrôle du développement de tumeurs (Halicka et al., 2005; Bassing et al., 2003). Quant à H2A.Z, ce variant pourrait jouer un rôle en modifiant la structure de la chromatine afin de réguler la transcription de façon positive ou négative (Leachet al., 2000). En 2004, une étude a montrée que l'histone H2AZ stabilise l'octamère d'histone dans le nucléosome (Park et al., 2004).

Chez les vertébrés, le variant Macro-H2A participe à l'inactivation d'une des deux copies du chromosome X de la femelle sous la forme du corpuscule de Bar (Costanzi et Pehrson,

1998; Mermoud, Costanzi et al., 1999). Il semble donc jouer un rôle de répression de la transcription lié à la compensation de dose. À l'inverse, H2ABBD (pour Barr Body déficient) est un nouveau variant d'histone (Chadwick et Willard, 2001) et pourrait jouer un rôle dans l'activation de la transcription comme cela semble être le cas pour le variant H2AZ.

Plusieurs fonctions ont été attribuées à ce dernier variant ; il pourrait jouer le rôle de frontière en empêchant l'invasion des régions euchromatiniennes par l'hétérochromatine (Raisner, Hartley et al., 2005), mais pourrait également avoir une fonction de répression. Ce variant, qu'on retrouve chez les mammifères associé aux régions péri-centriques, aurait également une fonction dans la ségrégation des chromosomes mitotiques (Rangasamy, Greaves et al., 2004).

De façon générale, deux classes d'histones peuvent alors être distinguées : les histones canoniques, synthétisées et assemblées en nucléosome lors de la réplication (H3.1, H4, H2A, H2B) et les variantes d'histones, synthétisées et déposés sur l'ADN indépendamment de la réplication (H3.3, CENPA, H2AZ, H2AX, macro-H2A) (Sakai *et al.*, 2009).

➤ La méthylation de l'ADN

La méthylation de l'ADN est le mécanisme épigénétique le mieux caractérisé, associé surtout à la répression transcriptionnelles

Il résulte de réactions chimiques réversibles catalysées par les ADN méthyltransférases (DNMTs, DNA méthyltransférase), dans lesquelles un groupement méthyle (-CH₃) est ajouté sur le carbone 5 d'une cytosine dans un dinucléotide CpG. Les dinucléotides CpG sont regroupés dans le génome en îlots CpG et souvent associés aux promoteurs des gènes (Hermann et al., 2003).

L'hypométhylation est un processus croissant et affecte principalement les régions intergéniques (Wilson et al., 2007). Paradoxalement, cette hypométhylation globale est souvent accompagnée d'une hyperméthylation locale et spécifique de certains promoteurs de gènes particulièrement au sein des îlots CpG. Elle entraîne la répression d'une grande variété de gènes aux fonctions anti-prolifératives appartenant à une nouvelle génération de gènes suppresseurs de tumeurs (Esteller, 2006).

II. La transcription des ARN messagers chez les eucaryotes

Dans les cellules eucaryotes, la transcription des ARN messagers est réalisée par l'ARN polymérase de type II (ARN Pol II) ; cette transcription s'initie au niveau de régions dites promotrices. L'activité RNA polymérase est régulée par un certain nombre de cofacteurs, de façon positive ou négative. De façon générale, on peut distinguer d'un côté les domaines composants le promoteur et de l'autre les domaines permettant le recrutement de ces facteurs de transcription régulateurs. Les règles de fonctionnement de ces promoteurs eucaryotes sont parfaitement établies (Myers et al., 1986).

Le « core promoter » est composé de quatre éléments minimum (Figure 6), permettant le recrutement de l'ARN polymérase au niveau du promoteur et la formation du PIC (complexe de pré-initiation) (Butler et Kadonaga, 2002) : son élément central est la TATA box, mais il

comprend également les domaines BRE (TFIIB Recognition Element), Inr (initiator) et DPE (Downstream Promoter Element) ; les promoteurs sont également des régions riches en G-C.

Le PIC est une structure multi-protéique composée de l'ARN polymérase et des facteurs de transcription TFII (Transcription Factor for RNA Pol II) ; ceux-ci sont des complexes comprenant un grand nombre de sous-unités (Figure 6).

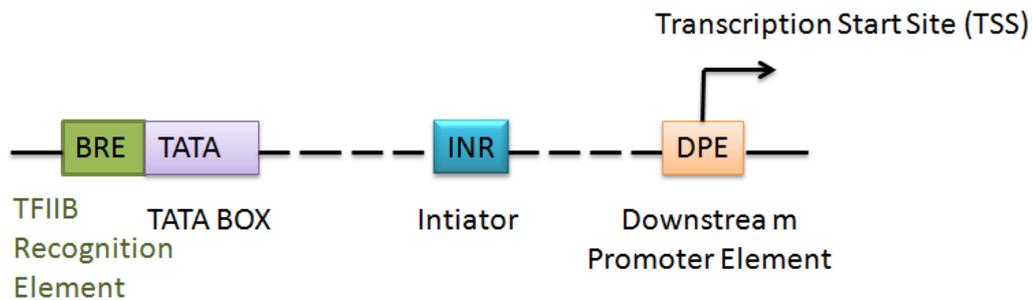


Figure 6 - Représentation schématique du « core promoter » de l'ARN polymérase de type II. La position des différents éléments est exprimée par rapport au TSS ; les séquences théoriques sont également présentées : BRE -37 à -32, TATA -31 à -26, Inr -2 à +4, DPE +28 à +32.

II.1. Structure de l'ARN polymérase II

L'ARN polymérase II est une enzyme complexe constituée de 12 sous-unités (Rpb1 à Rpb12) chez les mammifères. Les deux plus grosses sous-unités, Rpb1 et Rpb2, portent le site actif, les autres permettent l'interaction avec l'ADN, la stabilisation, des interactions avec des facteurs transcriptionnels ou une forte processivité par exemple une enzyme processive est une enzyme capable de catalyser la même réaction de façon répétée le long d'un polymère sans se détacher de la chaîne (Alberts et al., 2011). À l'extrémité carboxy-terminale de la plus grande sous-unité Rpb1 se trouve une structure spécifique remarquable, qui n'est retrouvée dans aucune autre polymérase, appelée CTD (Carboxy Terminal Domain), et qui est constituée d'une répétition du motif Tyr-Ser-Pro-Thr-Ser-Pro-Ser (52 fois chez les mammifères). Des études de délétion ont pu montrer que cette structure était indispensable et jouait un rôle important dans la transition entre l'initiation et l'élongation.

L'ARN polymérase II ne se fixe pas directement sur l'ADN, mais par l'intermédiaire de plusieurs facteurs. Les premiers à intervenir sont les facteurs généraux de transcription (GTF, *General Transcription Factors*). Le mécanisme décrit dans la Figure 7 est séquentiel et correspond à ce qui doit être réalisé *in vitro* pour obtenir une transcription. Il n'est pas certain qu'il en soit ainsi *in vivo* et il est possible que la cellule utilise des complexes préformés.

II .2. Formation du complexe de pré initiation

La première étape est la formation du complexe de pré initiation (PIC) au niveau du promoteur. Le PIC est le complexe minimal nécessaire pour l'initiation de la transcription *in vitro*. Il est composé de l'ARN polymérase II ainsi que des facteurs généraux de transcription TFIIB, TFIID, TFIIIE, TFIIF et TFIIH .TFIIA. Les facteurs généraux de transcription interviennent dans la reconnaissance du promoteur, le recrutement de l'ARN polymérase II et l'ouverture de la bulle de transcription (Figure 7).

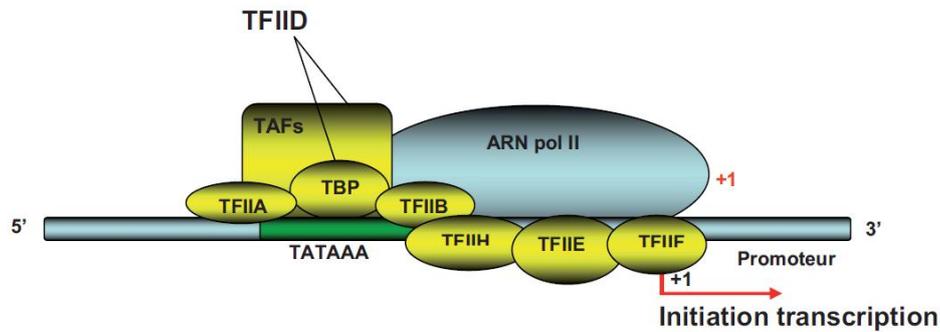


Figure 7- Représentation schématique de la formation du complexe de pré-initiation (PIC).

D'après (Kaplan et Delpech, 2007). Le premier acteur est le facteur TFIID (*transcription factor IID*) qui a une structure particulièrement complexe. Sa composition est variable en fonction des tissus et de l'état cellulaire. On y retrouve toujours une sous-unité appelée TBP (*TATA Binding Protein*) et une dizaine de polypeptides appelés TAF (*TBP- Associated Factors*). TBP, qui a une forme qui ressemble à une selle, se fixe sur une séquence riche en T et en A, la boîte TATA (*TATA box1*), située 25 à 30 bases en amont du premier nucléotide qui sera transcrit. TBP interagit avec la boîte TATA au niveau du petit sillon de l'ADN et induit une torsion d'environ 80 ° de la molécule d'ADN, ce qui a pour effet de légèrement dissocier les bases complémentaires de la double hélice à ce niveau. La fixation de TBP est suivie de la fixation des différents TAF qui constituent TFIID. Dans un second temps, le facteur TFIIA (*transcription factor IIA*) se fixe à TBP et stabilise son interaction avec l'ADN. Il semble que sa fixation puisse induire l'expulsion de facteurs inhibiteurs de la transcription. Le troisième acteur est une protéine monomérique appelée TFIIB (*transcription factor IIB*). Le premier tiers de la protéine correspond à un domaine fixant le zinc, dont le rôle est inconnu. Les deux tiers restants interagissent à la fois avec les protéines fixées et à la fois avec l'ADN. Une extrémité se fixe en amont de la boîte TATA (au niveau du grand sillon de l'ADN), l'autre en aval (au niveau du petit sillon de l'ADN). Le quatrième acteur est le facteur TFIIF (*transcription factor IIF*) qui est une protéine constituée de quatre sous-unités deux à deux identiques. Il stabilise le complexe créé précédemment, induit des torsions de l'ADN qui facilitent sa fusion au niveau du promoteur. Il possède aussi une grande affinité pour l'ARN polymérase II ce qui permet son association au complexe et vraisemblablement prévient sa fixation à des sites non spécifiques. L'acteur suivant est le facteur TFIIIE (*transcription factor IIE*). Les activités de ce facteur sont multiples. Il sépare les deux brins de l'ADN au niveau du promoteur à l'aide d'ATP (adénosine triphosphate). Sa forte affinité pour l'ADN simple brin permet aux deux brins de l'ADN du promoteur de rester bien séparés. Enfin il recrute le sixième et dernier facteur, TFIIH (*transcription factor IIH*), protéine très complexe qui

possède 9 sous-unités. Six d'entre elles (XPB, XPD3, p34, p44, p52 et p62) constituent le «cœur» (*core*) de la protéine. La sous-unité XPD fait le pont entre ce « cœur » et les trois autres sous-unités (Cdk7, cycline H et Mat1) qui constituent le complexe CTD kinase, appelé CAK (*Cyclin-Activating Kinase*), dont l'action est, une fois activée par TFIIE, de phosphoryler le domaine CTD de l'ARN polymérase II, étape indispensable à la transition initiation/élongation. La taille du complexe ainsi formé est proche de celle d'un ribosome et il est possible de l'observer au microscope électronique lorsqu'il est attaché à l'ADN.

Lorsque la polymérase se fixe au complexe d'initiation, le complexe CAK (*Cyclin-Activating Kinase*) phosphoryle le domaine CTD de l'ARN polymérase sur les sérines 2 et 5 de chacune des 52 répétitions qui constituent le domaine. Cette phosphorylation rend cette partie de la polymérase très acide, ce qui induit la polymérase à quitter le site d'initiation et à passer sous la forme d'un complexe d'élongation stable. Des facteurs agissent aussi sur la transition entre initiation et élongation, en permettant notamment à la polymérase de quitter le site d'initiation de la transcription (comme P-TEFb, *Positive Transcription Elongation Factor b*). Dans le même temps de nombreux facteurs qui composaient le complexe d'initiation sont libérés et remplacés par des facteurs d'élongation. Une fois le site d'initiation de la transcription libéré, une nouvelle molécule de polymérase peut amorcer un nouveau transcrit. Le nombre de molécules de polymérase engagées dans la transcription d'un gène est proportionnel au taux de transcription de ce gène.

Ce PIC minimal suffit pour la transcription *in vitro* mais non *in vivo* où la polymérase doit pouvoir interagir avec les activateurs et initier la transcription sur une matrice chromatinienne. Ces étapes nécessitent l'intervention de différents Co-activateurs tels que les complexes de remodelage de la chromatine ou le Médiateur.

II.3. Initiation et pause

L'initiation de la transcription corrèle avec la phosphorylation de la sérine 5 du CTD de la polymérase (Komarnitsky *et al.*, 2000). Cette marque est déposée par la kinase Cdk7 (Cyclin-dépendante kinase 7) appartenant au complexe TFIIF. Il est proposé que la kinase Cdk9 (Cyclin-dépendante Kinase 9) du complexe P-TEFb (Positive Transcription Elongation Factor b) soit également impliquée dans la phosphorylation de ce résidu. Cette modification est essentielle pour le recrutement des facteurs nécessaires à la mise en place de la coiffe des ARN messagers (Fabrega *et al.*, 2003). Avant l'élongation, plusieurs cycles d'initiation abortive ont lieu, pendant lesquels l'ARN polymérase II synthétise de courts ARN de quelques nucléotides qui sont relargués (Hsu, 2009). Des progrès dans la détermination de la structure du PIC ont permis de comprendre le mécanisme de production de ces petits ARN. En effet, il a été montré que le domaine N-terminal du facteur TFIIB forme une boucle à l'intérieur du site actif de la polymérase II à l'endroit normalement occupé par l'hybride ARN-ADN. Tant que l'ARN ne dépasse pas 9 bases, cette boucle ne gêne pas (elle pourrait même faciliter l'incorporation des premières bases). Par contre, quand l'ARN dépasse 9 bases, il y a compétition entre la boucle et l'ARN qui amène soit la victoire de l'ARN, l'éjection de TFIIB et le passage en élongation productive « Promoter escape », soit la victoire de TFIIB et l'avortement de l'initiation qui conduit aux petits ARN observés (Boeger

et al., 2005). La fuite du promoteur « Promoter escape » décrit le moment où l'ARN polymérase II rompt ses contacts avec les séquences promotrices et une partie des facteurs liés au promoteur et entre en phase d'élongation (Saunders *et al.*, 2006). Pour certains gènes il a été montré que la polymérase s'arrête peu de temps après son départ du promoteur.

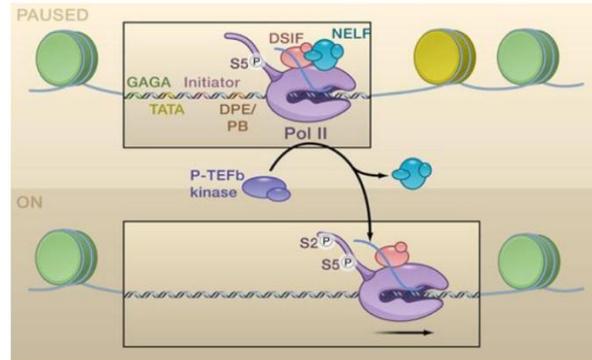


Figure 8- Mécanisme de maintien et de sortie de pause (Levine, 2011).

Schéma représentant un promoteur type possédant une ARN Pol II en pause (en haut). Le promoteur contient la majorité des séquences pour recruter et maintenir l'ARN Pol II en pause, à savoir les séquences GAGA, TATA, Initiator et DPE/PB. L'ARN Pol II phosphorylée sur la Ser2 de son domaine CTD après avoir échappé au promoteur, est maintenue en pause par le complexe répresseur NELF/DSIF juste après la séquence DPE/PB. L'activation du facteur d'élongation P-TEFb entraîne la libération de NELF, la phosphorylation de l'ARN Pol II sur la Ser5 et son entrée en élongation (en bas).

Cette étape est un point important de régulation de l'élongation de la transcription par l'ARN polymérase II *in vivo*. La pause pourrait à la fois servir de point de contrôle pour la vérification de la préparation de l'ARN polymérase II à son entrée en élongation productive, et permettre une induction rapide de l'expression des gènes. En effet, beaucoup des gènes régulés par la pause appartiennent à la catégorie des gènes répondant à un stimulus environnemental ou développemental (Muse *et al.*, 2007). Cependant, la déplétion de NELF supprime la pause de la Pol II mais ne permet pas d'atteindre les niveaux d'expression observés après induction, ce qui suggère la nécessité d'un activateur supplémentaire dont la fonction n'est pas seulement de faire repartir la polymérase arrêtée. Il est également possible que la déplétion de NELF diminue la quantité de Pol II prête à entrer en élongation. Une telle diminution globale du recrutement de la Pol II a été observée et permet d'expliquer le comportement de centaines de gènes dont l'expression est diminuée après déplétion de NELF (Gilchrist *et al.* 2008). Cette étude montre une augmentation de la densité des nucléosomes près du promoteur en l'absence de NELF. Ainsi, le rôle de la polymérase en pause pourrait être de maintenir la chromatine dans un état ouvert permettant un haut niveau de transcription. En effet, beaucoup de gènes parmi les plus hautement transcrits présentent un enrichissement de NELF et de la Pol II en pause (Lee *et al.*, 2008).

II.4. Reprise de L'élongation

L'entrée en élongation productive coïncide avec le recrutement du complexe P-TEFb, dont la kinase Cdk9 est responsable de la phosphorylation de DSIF, NELF et de la sérine 2 du CTD de l'ARN polymérase II. La phosphorylation de la sérine 2 de la Pol II par P-TEFb semble être l'étape clé du relargage de la pause (Saunders *et al.*, 2006).

En effet, c'est la forme hyperphosphorylée de la polymérase où le CTD est phosphorylé sur les sérines 2 et 5 qui est impliquée dans l'élongation. Le CTD subit plusieurs cycles de phosphorylation/déphosphorylation au cours du processus de la transcription. Cependant, l'utilisation du flavopiridol, un inhibiteur spécifique de P-TEFb (Chao et Price, 2001) soulève de nouvelles questions. En effet, après traitement avec cette drogue, la polymérase reprend effectivement l'élongation mais s'arrête de nouveau après avoir transcrit environ 100 nucléotides. Quand la Pol II reprend l'élongation, NELF se dissocie alors que DSIF reste associé et devient un activateur de l'élongation (Yamada *et al.*, 2006).

Lors de la transition entre l'initiation et l'élongation, les facteurs généraux de transcription, à l'exception de TFIIF, sont progressivement relâchés et la polymérase peut alors s'associer aux facteurs d'élongation. Ces derniers préviennent les pauses et les arrêts de la polymérase durant la synthèse de l'ARNm. Au niveau du complexe d'élongation, un hybride ADN-ARN de 8 à 9 nucléotides se forme au sein de la bulle qui reste ouverte entre -9 et +2 (le site +1 correspondant à la position du dernier rNTP (ribonucléotide Triphosphate) ajouté dans le complexe en élongation. Les sous-unités Rpb1 et Rpb2 forment le centre catalytique de l'enzyme (Conaway *et al.*, 2000).

La vérification de la fidélité de la transcription est indispensable pour une expression correcte des gènes. Celle-ci se fait tout au long de l'élongation. Une boucle de Rpb1 située près du site actif interagit avec le nucléotide entrant. Si celui-ci est complémentaire avec la base de l'ADN qui se trouve dans le site actif, la conformation obtenue favorise l'attaque nucléophile et la formation de la liaison phosphodiester entre l'ARN déjà formé et le nouveau nucléotide. Si le nucléotide entrant n'est pas complémentaire, il se forme un mésappariement dans le site actif et la vitesse de formation de la liaison est nettement réduite. À ce moment, soit le nucléotide ressort du site actif, ce qui permet d'en tester un autre, soit la liaison se forme malgré tout. La conformation anormale de l'hybride ADN/ARN ralentit le processus d'ajout du nucléotide suivant et le transcrit devient sensible à l'activité endonucléase intrinsèque de la polymérase qui peut revenir en arrière sur quelques nucléotides et couper le transcrit puis recommencer à ajouter des nucléotides (Sydow et Cramer, 2009). Cette activité endonucléase importante pour la fidélité de la transcription est stimulée par TFIIS.

L'élongation est couplée à la biogenèse de l'ARNm qui comporte la mise en place de la coiffe, l'épissage, le clivage du transcrit et sa polyadénylation. La transcription affecte, et est affectée par, les autres processus métaboliques de l'ADN que sont la réparation, la recombinaison et la réplication.

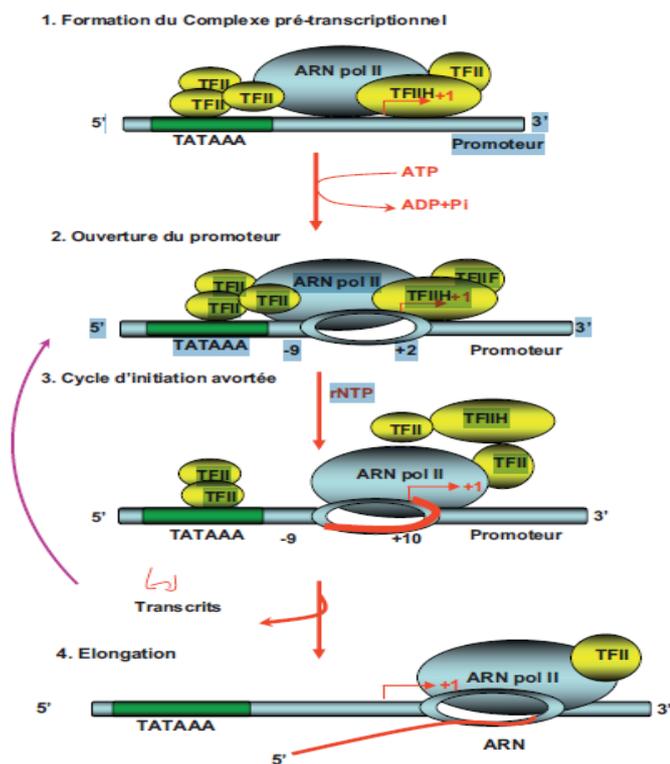


Figure 9- Etapes de la transcription par l'ARN polymérase II. (Langelier, 2001).

II.5. Terminaison

Au fur et à mesure de la transcription, le CTD de la polymérase est phosphorylé sur la sérine 2 d'un nombre croissant de répétitions de l'heptapeptide Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Cette marque est importante pour le recrutement de la machinerie de maturation des ARN en 3' des gènes. Pcf11 (Protein 1 of Cleavage and polyadenylation Factor 1) est un facteur important dans cet étape. Il interagit avec le CTD et l'ARN et permet la dissociation du complexe d'élongation (Zhang et Gilmour 2006). Comme la phosphorylation, la déphosphorylation du CTD joue un rôle important dans le processus transcriptionnels en permettant à l'ARN Pol II de revenir à sa forme hypophosphorylée (la seule capable d'intégrer le complexe de pré-initiation) et d'entamer un nouveau cycle de transcription. La protéine phosphatase Fcp1 (TFIIF interacting CTD phosphatase) est responsable de la déphosphorylation du CTD (Cho *et al.*, 1999). Fcp1 pourrait jouer un rôle dans le recyclage de la polymérase bien qu'il ne soit pas connu si la déphosphorylation du CTD est une cause ou une conséquence de l'arrêt de la transcription.

II.6. Régulation de la transcription

La régulation de la transcription est un des mécanismes fondamentaux qui contrôlent la quantité de protéines produites par la cellule. Cette régulation diffère selon les conditions environnementales et le stade de développement de l'organisme. Un vaste panel de protéines parmi lesquelles les ARN polymérases, les histones, les modificateurs d'histones, les facteurs de transcription et les cofacteurs, est impliqué dans le maintien de la précision et de la spécificité du processus de régulation.

Les gènes des eucaryotes possèdent des séquences régulatrices présentes à proximité du site d'initiation de la transcription (ou TSS pour « Transcription Start Site ») constituant le promoteur proximal. Celui-ci est le lieu de formation du complexe de préinitiation de la transcription. La modulation de la transcription de l'ADN par la Pol II est effectuée par des facteurs de transcription spécifiques. Leur liaison non-covalente au niveau de l'ADN s'effectue au niveau de sites spécifiques, appelés sites de fixation de facteurs de transcription (TFBS, « Transcription Factor Binding Site »), afin d'activer ou d'inhiber l'expression d'un gène donné.

Ces régions régulatrices sont de plusieurs types : (1) les promoteurs lorsqu'ils sont situés à proximité du site d'initiation de la transcription de la région codante, (2) les enhanceurs quand ils se localisent à distance du site d'initiation de la transcription et qu'ils potentialisent l'action du promoteur, (3) des silencers lorsqu'ils se situent comme les enhanceurs à distance du gène mais qu'ils le répriment, et (4) les insulateurs qui correspondent à une séquence régulatrice affectant l'interaction entre l'enhancer et le promoteur (Geyer, 2002). Ces régions régulatrices sont particulièrement conservées lors de l'évolution, car elles sont composées de courtes séquences de 6 à 15 paires de bases (pb), appelées éléments régulateurs (RE, « Regulatory Element ») permettant le recrutement spécifique des facteurs de transcription sur l'ADN (Figure10).

Les facteurs de transcription n'agissent pas de manière indépendante, mais forment des complexes avec d'autres facteurs de transcription et des cofacteurs protéiques, comme c'est également le cas pour les facteurs de transcription généraux (Ravasi et al., 2010). Ces facteurs de transcription se lient à leurs sites de liaison spécifiques souvent regroupés en modules cis-régulateurs.

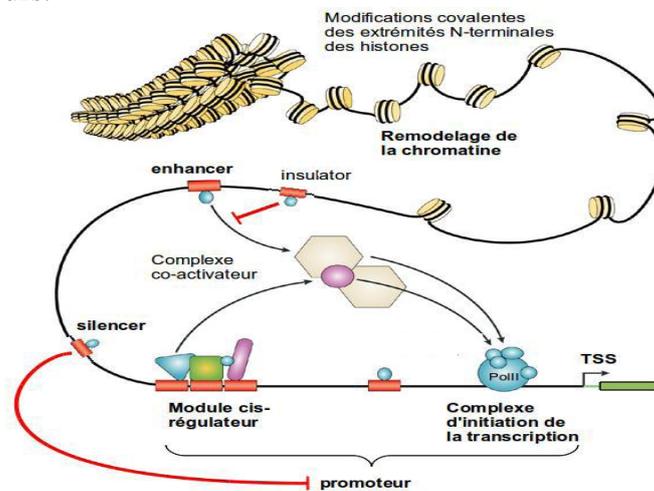


Figure 10 – Représentation schématique des régions régulatrices permettant la modulation transcriptionnelle de l'expression des gènes (Wasserman et Sandelin. 2004).

III. Transcription bidirectionnelle

Ces dernières années, la compréhension du fonctionnement des promoteurs eucaryotes a beaucoup évolué. Le séquençage haut-débit et l'évolution des outils d'analyse ont permis de mettre en évidence l'existence, dans le génome humain notamment, de paires de gènes arrangés en « tête-à-tête », sur chacun des deux brins d'ADN (Figure 11), et partageant donc une région promotrice en 5' (Adachi et Lieber, 2002).

Cette structure génomique particulière, bidirectionnelle, était connue depuis de nombreuses années. Dans la plupart des cas, elle avait été relevée pour des paires de gènes dont l'expression était coordonnée; ceci a par exemple été démontré dans le cas des gènes des histones, dont la stœchiométrie d'expression est primordiale (Albig et al, 1997 ; Maxson et al, 1983), ou encore dans le cas de gènes coexprimés dans la même voie de signalisation (Momota et al, 1998 ; Schmidt et al, 1993). Des études plus récentes, réalisées sur l'ensemble du génome humain, ont permis de démontrer que cet arrangement bidirectionnel ne concernait pas moins de 10% de nos gènes (Trinklein et al, 2004) ; ce constat est généralisable aux autres génomes eucaryotes (Neil et al, 2009 ; Wang et al, 2009).

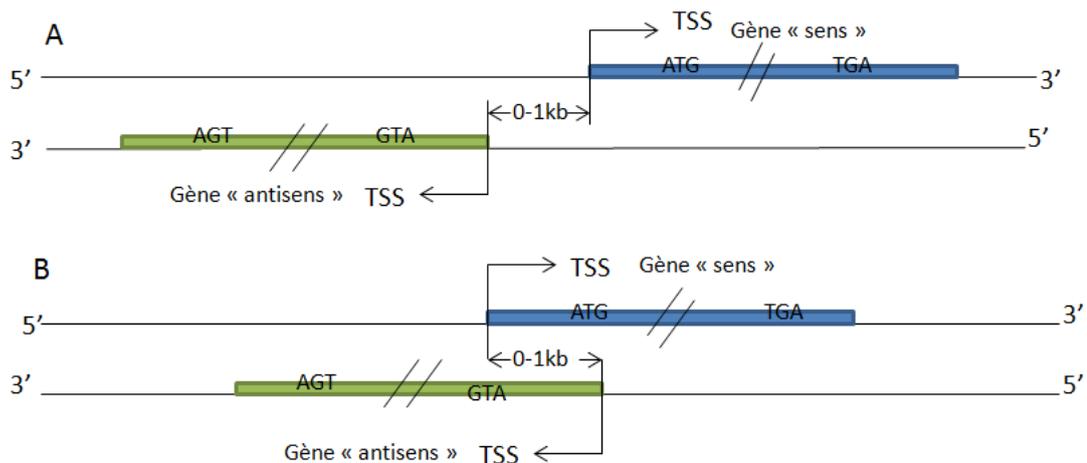


Figure 11 – Organisation génomique d'une paire de gènes bidirectionnels.

Les sites d'initiation de la transcription (TSS) des deux gènes « sens » et « antisens » sont séparés de moins de 1kb. Les extrémités 5' des gènes bidirectionnels peuvent être non-chevauchantes (A) ou chevauchantes (B). ATG : codon initiateur ; TGA :codon terminateur.

Les promoteurs bidirectionnels sont donc définis comme les séquences d'ADN permettant la transcription simultanée de deux gènes d'orientations opposées ; les analyses à grande échelle des génomes ont montré que la distance séparant les sites d'initiation de la transcription (TSS) de chacun des deux gènes était inférieure à 1kb (Adachi et Lieber, 2002 ; Trinklein et al, 2004).

Les transcrits initiés de part et d'autre du promoteur peuvent être codants ou non-codants. Des études ont d'ores et déjà pu mettre en évidence des promoteurs bidirectionnels permettant l'expression en « sens » et en « antisens » de transcrits codants pour des protéines (les dénominations sens et antisens étant dans ce cas purement arbitraires), c'est par exemple le cas du promoteur régulant l'expression des gènes de p53 et WDR79 chez l'humain et la souris (Polson et al, 2011). Dans de nombreux autres cas, la transcription bidirectionnelle permet d'initier deux transcrits, l'un codant et donc défini comme le transcrit « sens », l'autre non-codant ; ces derniers transcrits, appelés « *divergents* », sont retrouvés chez tous les eucaryotes (Li et al, 2006 ; Neil et al, 2009 ; Pickrell et al, 2010).

Chapitre II

Matériel Et

Méthode

II.1. Matériels

II.1.1. La bioinformatique

La bioinformatique est la discipline de l'analyse de l'information biologique, qui s'intéresse principalement au contenu en des séquences des macromolécules (acides nucléiques et protéines) et leur structure tridimensionnelle. C'est une branche théorique de la biologie, largement antérieure à la « révolution génomique » des années 1990.

Les bio-informaticiens sont à la conjonction entre biologie et informatique, mais peuvent aussi utiliser des outils statistiques, mathématique (exemple : logiciel R).

La bioinformatique n'est pas une simple application des concepts et des outils de l'informatique traditionnelle aux données biologiques. Elle recouvre un ensemble de techniques très spécifiques, intimement liées aux objets d'étude de la biologie moléculaire et de la génomique. Plus récemment, l'introduction de techniques expérimentales massivement parallèles, produisant une masse de donnée. (exemple : les puces à ADN, RAN-seq,CHIP-seq..) .

Enfin, la bioinformatique est indissociable de l'existence de grandes bases de données internationales publiques de la mise en place de nombreux serveurs tel que :

- GenBank , NCBI (National Cancer for Biology Information),USA .
- EMBL, EBI, (European Molecular Biology Organisation), Européenne
- DDBJ, DNA Data Bank of Japan

Les plus populaires sont :

- Ensembl (European Bioinformatics Institute / Wellcome Trust Sanger Institute)
- NCBI (National Cancer for Biology Information)
- UCSC (University of California Santa Cruz)

II.1.2.les outils bioinformatiques

II .1.2. 1 Logiciel R :

R est une implémentation d'un langage de programmation interactif appelé S ,Ross Ihaka et Robert Gentleman proposent un langage pour l'analyse de données et les graphiques qu'ils nomment R.C'est un langage de programmation, il peut être largement et facilement enrichi grâce à des extensions, appelées les "packages".

C'est un logiciel de type Open Source (Logiciel Libre) qui se développe grâce au bon-vouloir de certains contributeurs, soucieux de l'intérêt général.

Le système R fournit un environnement intégrant un grand nombre de fonctionnalités statistiques et graphiques qui en font un outil particulièrement adapté au traitement et à l'analyse des données.

R permet de tracer toute sorte de graphiques (histogramme, camembert,...), de calculer différents paramètres (moyenne, l'écart type,...) et de réaliser les tests statistiques.

R « connaît » les formules pour effectuer ces tests statistiques, en revanche c'est à vous de lui dire quel test utiliser ! De plus, R est très utilisé pour la bioinformatique vous pourrez écrire vos propres fonctions et packages pour personnaliser les graphiques, ou utiliser ceux créés par d'autres et mis à disposition dans la bibliothèque R"

II .1.2.2. Chromatine Immuno-Précipitation – séquençage (CHIP-Seq)

Comme son nom l'indique, la méthode CHIP-Seq se fait en deux parties distinctes. D'un côté l'immuno-précipitation de la chromatine et de l'autre côté le séquençage.

- **ChIP :**

La méthode d'immuno-précipitation de chromatine est une méthode permettant la sélection de fragments d'ADN possédant une interaction avec une protéine d'intérêt.

Dans un premier temps, la protéine d'intérêt est fixée sur son site de liaison grâce à, l'utilisation de formaldéhyde. L'ADN est ensuite extrait et découpé en courts brins de 200 à 400 paires de bases, généralement par sonication. Un anticorps spécifique à la protéine étudiée est ensuite incorporé afin de former un complexe ADN-protéine-anticorps.

L'immuno-précipitation de ces complexes permet ensuite de récupérer les brins d'ADN associés à la protéine après élimination du surnageant (i.e. ADN non associé à la protéine d'intérêt). Pour finir, la protéine d'intérêt est séparée de l'ADN par chauffage par exemple. Les fragments d'ADN obtenus sont ainsi ceux où la protéine s'est fixée.

- **Séquençage :**

La collection de fragments ainsi obtenus est ensuite introduite dans un séquenceur à haut débit (exemple : illumina). Ce dernier va alors séquencer une ou les deux extrémités des brins d'ADN sur une courte distance, souvent entre 35 et 75 paires de bases. Les séquenceurs actuels permettent de séquencer un génome entier en une seule fois, générant ainsi des millions de séquences. Il est nécessaire ensuite d'utiliser des moyens informatiques afin d'aligner ces séquences sur un génome de référence. Apparaissent alors des zones fortement séquencées correspondant à une des extrémités des séquences d'ADN collectées.

Les sites de liaison des facteurs de transcription, Poly A, ou des modifications de chromatine sont les zones situées entre deux de ces régions. (Eloi, Mercier .2011).

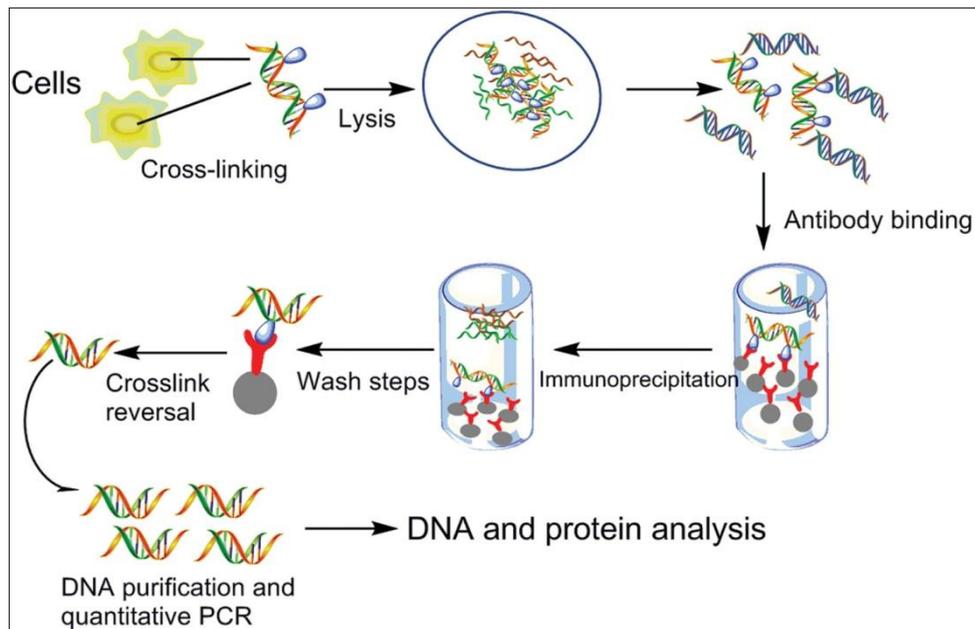


Figure 12 -Principe générale de l'immuno précipitation (Shah, 2009; Vinckevicius et Chakravarti, 2012).

II .1.2.3. Deux outils bioinformatiques ont été utilisés pour visualises nos résultats

A- IGB *Integrated Genome Browser* :

Pour la visualisation graphique en pic des données génomique : Un outil permettant de visualiser et d'explorer des données génomiques et d'annotation. Il est utile notamment pour la visualisation de données de CHIP-Seq. Il peut ouvrir des fichiers wig.....

Afin de commencer d'analysez vous devez acquérir plusieurs compétences sur IGB :

- Ouvrir le programme
- Utiliser l'interface
- Charger des fichiers génomiques et visualiser les données qu'ils contiennent
- Croiser des informations
- Sauvegarder une session
- Exporter des images

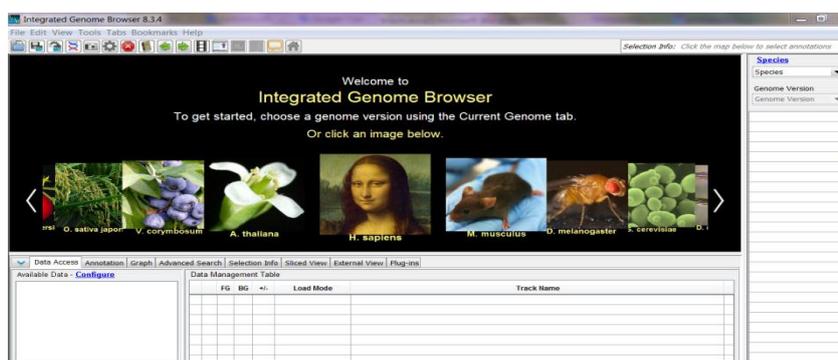


Figure 13- La fenêtre principale de l'IGB.

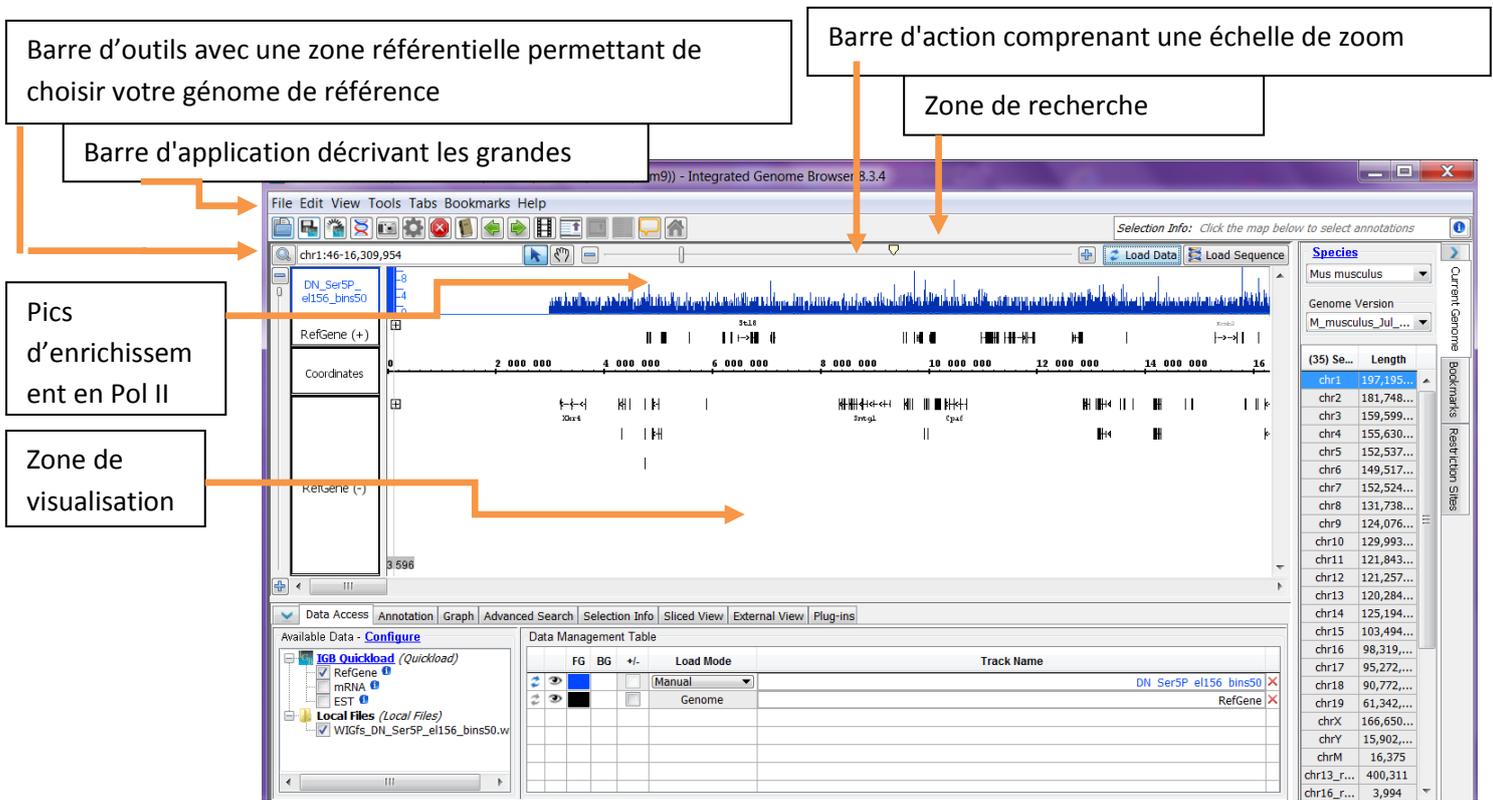


Figure 14- Présentation d'interface des données génomique avec IGB .

Lorsque des données d'annotation est chargé dans IGB, le navigateur ressemble à ceci :

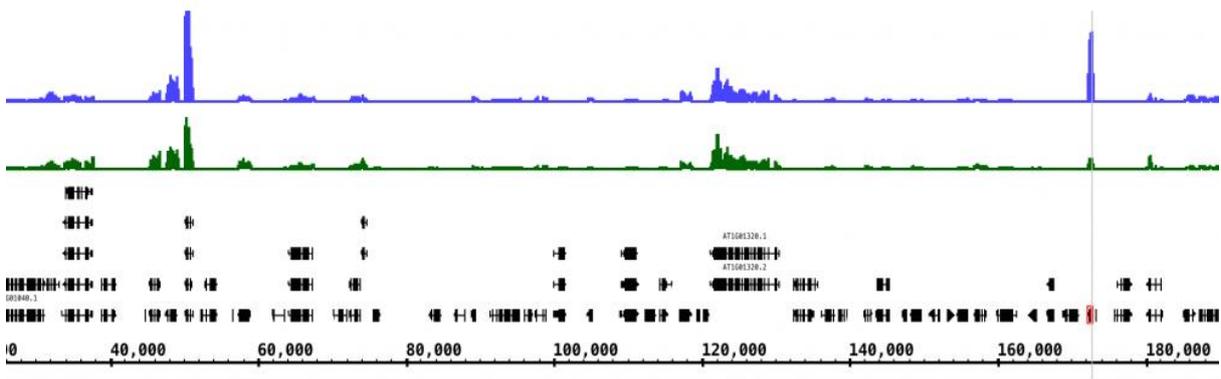


Figure 15- l'interface La partie principale de la fenêtre du navigateur est la fenêtre de visualisation.

Il contient Plusieurs pistes d'annotation ou niveaux. Ces pistes sont comparables à IGB pistes d'annotation dans les autres navigateurs de génome. Typiquement, un coordonne piste.

On verra quelle que s'exemples ci-dessus sur nos gènes des différentes classe (**voir chapitre résultat partie III.3.I**).

B - L'outil DAVID (Database for Annotation, Visualization and Integrated Discovery):

Est l'un des programmes d'annotation fonctionnelle les plus populaires auprès des biologistes et s'appuie sur une annotation fonctionnelle des gènes particulièrement les plus riches. En effet, cette annotation intègre plusieurs milliers de termes fonctionnels provenant de 14 bases de données différentes (NCBI, PIR, SWISS-PROT, GO, OMIM, PubMed, KEGG,

BIOCARTA, AffyMetrix, TIGR, Pfam, BIND, MINT, DIP...)(da Huang W .2009) .



Figure 16- DAVID outil de visualisation fonctionnelle.

II .2. Méthodes

II.2.1 Model de travail : Cellules T de la Souris

Le modèle spontané d'étude de la transcription bidirectionnelle similaire à l'homme c'est la souris qui constitue un tremplin pour l'étude des gènes.

Les principales analyses utilisées pour ce travail sont fournies au labo TAGC à Marseille. Sur les précurseurs des lymphocytes T (thymocytes) en stade double négatif CD4⁻/CD8⁻ (ou DN). La maturation intra-thymique est marquée par une succession de stades de différenciation. Les thymocytes immatures CD4⁻ / CD8⁻ (DN), se différencient en thymocytes CD4⁺ / CD8⁺, désignés double positifs (DP) puis en thymocytes matures CD4⁺ ou CD8⁺, désignés simple positif (SP).

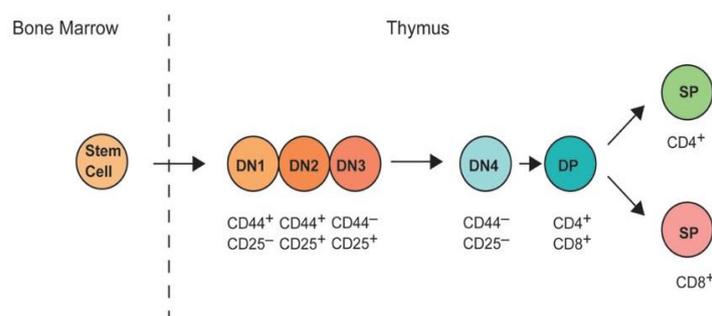


Figure 17- Différenciation des lymphocytes T (Koch et Radtke, 2011).

➤ Notre travail a comme objectif d'identifier les promoteurs bidirectionnels responsables de la transcription opposée au cours de cette différenciation des cellules T.

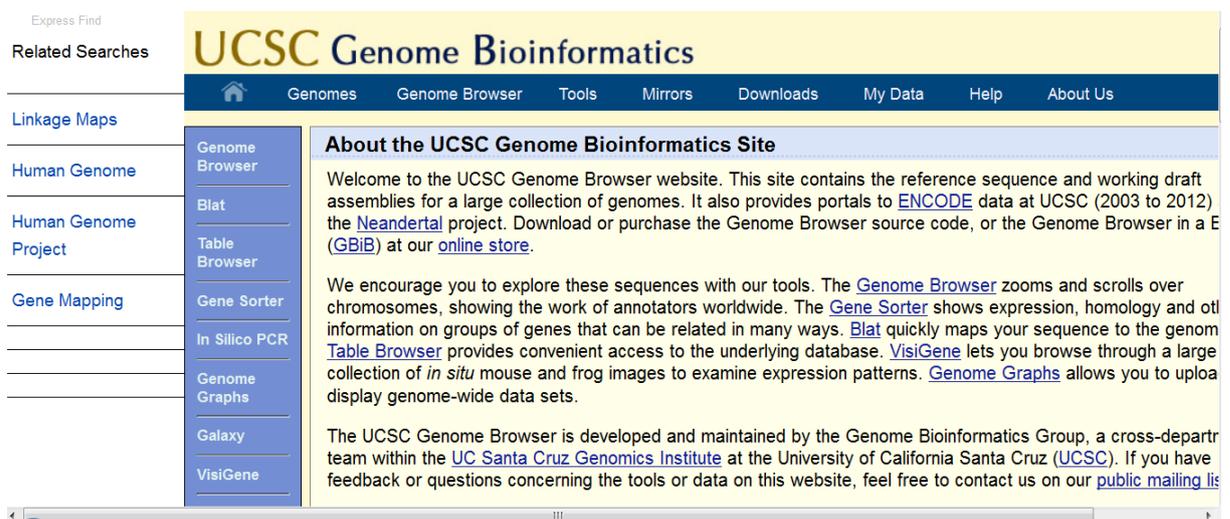
II.2.2. Identification des gènes bidirectionnelle

II.2.2.1. Récupération du génome de la souris mm9

Pour cela nous avons récupéré le génome de la version mm9 de la souris à partir du site web généraliste UCSC, qui propose une synthèse des informations connues sur les génomes séquencés et de leurs annotations. Quand Ensembl est plutôt orienté génomique fonctionnelle, UCSC est plus orienté génomique structurale. Ces 2 sites utilisent les données d'autres sites comme EBI, Uniprot... L'accès aux données est plus ou moins facilité.

A travers le Navigateur d'UCSC Genome Browser (University California Santa Cruz - USA) <http://genome.ucsc.edu/>.

- Cette page doit s'afficher dans le volet central :



Nous avons obtenu le génome entier de la souris, après avoir Indiqué les bons paramètres dans le table browser du site web UCSC, afin de récupérer tout les gènes de la souris.

Tableau 1- Paramètre à sélectionné dans le tableau des donnée.

• Clade	mammifère
• Genome	souris
• assemblage	mm9 version juillet 200
• Group	Les gènes et les prédictions de gènes
• Piste	les gènes RefSeq
• Tableau	refGene
• Région	génome
• Format de sortie	tous les champs de la table sélectionnée
• Fichier de sortie	RefSeq.tx

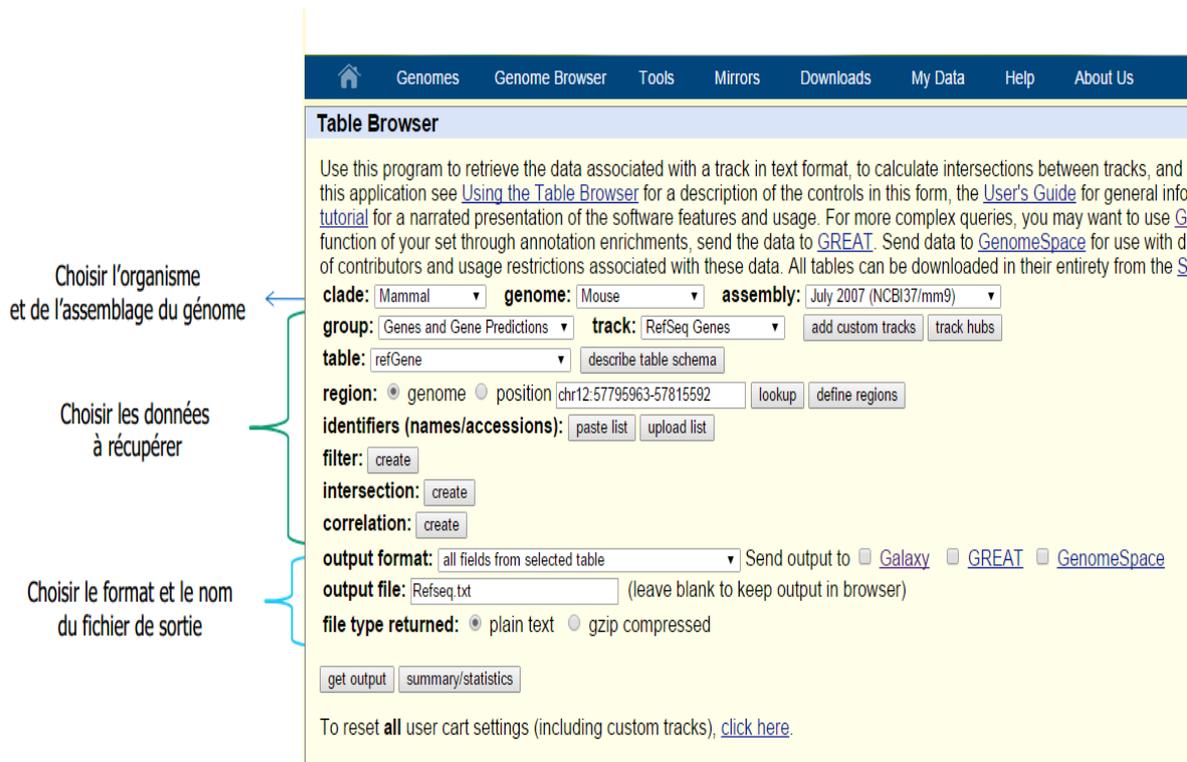


Figure 18- les bons paramètres indiqués dans le tableau du site web UCSC.

Par une simple clique sur Get output pour avoir le récapitulatif de la demande.

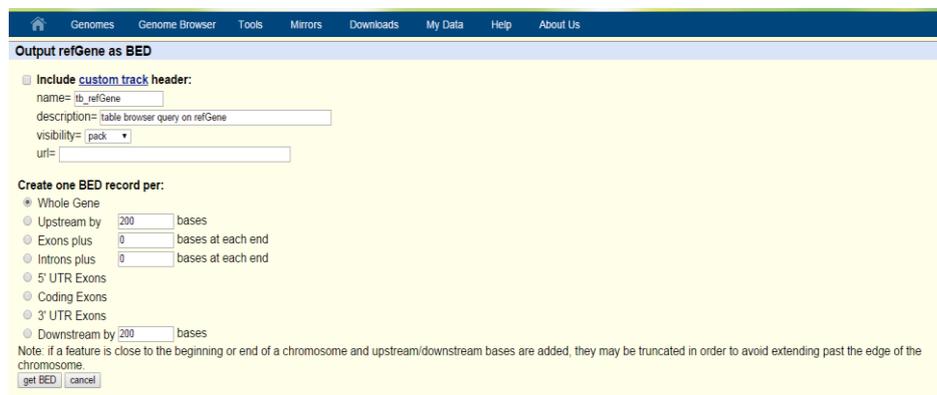


Figure 19- Cliquer sur get BED, pour sauvegarder le fichier « RefSeq.txt » qui porte la liste de tous les gènes de la souris avec leur coordonnées.

La liste de nos gènes obtenus d'après le site web UCSC va être traitée selon les étapes suivantes :

II.2.2.2. Trier les gènes non chevauchant

À partir de la base de données UCSC (mm9, RefSeq). Dans le cas où plusieurs TSS provenant de plusieurs isoformes d'un même gène étaient éloignés de moins de 100 pb, un seul représentant de la transcription a été sélectionné (qui génère un filtre et produit une liste avec 8000 gènes) à l'aide d'un script sur R (**voir Annexe script**)

	A	B	C	D	E	F
1	name	chrom	strand	txStart	txEnd	name2
2	NM_0011950	chr1	+	134212701	134230065	Nuak2
3	NM_028778	chr1	+	134212701	134230065	Nuak2
4	NM_0012903	chr1	-	8349819	9289958	Sntg1
5	NM_0012903	chr1	-	8349819	9289958	Sntg1
6	NM_175370	chr1	-	58714963	58752833	Als2cr12
7	NM_027671	chr1	-	8349819	9289958	Sntg1
8	NM_008922	chr1	-	33510652	33726639	Prim2
9	NM_0012903	chr1	-	8349819	9289958	Sntg1
10	NM_175642	chr1	-	25124320	25886552	Adgrb3
11	NM_178884	chr1	-	75482399	75503027	Obsl1
12	NM_199021	chr1	-	125228714	125942136	Dpp10
13	NM_198680	chr1	-	109050537	109057691	Serpinb3b
14	NM_0011133	chr1	+	167718811	167792243	Cd247
15	NM_130890	chr1	+	184495137	184562483	Capn8
16	NM_011465	chr1	+	176102906	176178580	Spta1
17	NM_031162	chr1	+	167718811	167797567	Cd247
18	NM_0011133	chr1	+	167718811	167801284	Cd247
19	NM_0011133	chr1	+	167718811	167801284	Cd247

Figure 20 - liste de toutes les gènes issu du site UCSC avec ces coordonnées en « brin, début, fin, nom » sous forme d'un fichier ' RefSeq' porte enivrnt 34989 gènes.

II.2.2.3. Isoler les coordonnées des promoteurs

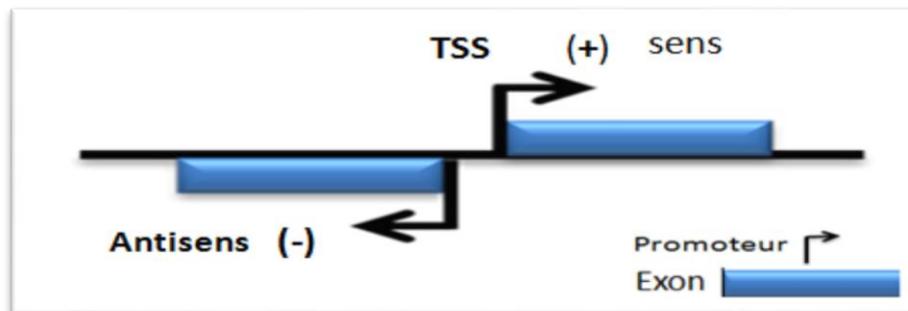


Figure 21 -TSS site d'initiation de la transcription des promoteurs sens et anti sens.

Cette étape a pour but isoler le site du début de la transcription TSS (start (+),end (-)) pour tous les transcrits codants. On applique une procédure de sélection grâce à une petite manipulation via Excel ou Logiciel R par exemple la formule suivante sur Excel : TSS =SI (C2="+";D2;E2).

On aura un fichier RefSeq.txt avec tout les gènes et leur cordonné en (Sens du brin, TSS, début, fin, nom du gène)

Le fichier « wig » porte l'enrichissement du signal en Pol II en glissant dans une fenêtre de taille fixe d'environ 50 nucléotides sur le génome complet.

Ce fichier « wig » a été traité grâce à un script sur logiciel R (Annexe I) la stratégie de ce script est qu'on récupère le fichier des coordonnées des TSS (RefSeq.txt) et puis il va récupérer dans le 2eme fichier Wig les valeurs d'enrichissement en Pol II et cela dans les deux régions de part et d'autre du TSS selon les coordonnées -1 et -4 kb et aussi +1 et +4kb. En fin l'outil R fournira en sortie un fichier avec ces valeurs associer à chaque gènes voir Figure N°25.

	A	B	C	D	E	F	G	H
1	chr	start	end	name	name2	strand	A	B
2	chr1	134212701	134230065	NM_0011950	Nuak2	+	37.44	6.24
3	chr1	134212701	134230065	NM_028778	Nuak2	+	37.44	6.24
4	chr1	8349819	9289958	NM_0012903	Sntg1	-	24.6	21.84
5	chr1	8349819	9289958	NM_0012903	Sntg1	-	24.6	21.84
6	chr1	58714963	58752833	NM_175370	Als2cr12	-	11.6	3.12
7	chr1	8349819	9289958	NM_027671	Sntg1	-	24.6	21.84
8	chr1	33510652	33726639	NM_008922	Prim2	-	34.32	18.72
9	chr1	8349819	9289958	NM_0012903	Sntg1	-	24.6	21.84
10	chr1	25124320	25886552	NM_175642	Adgrb3	-	15.6	15.6
11	chr1	75482399	75503027	NM_178884	Obsl1	-	46.92	30.24
12	chr1	125228714	125942136	NM_199021	Dpp10	-	6.24	31.2
13	chr1	109050537	109057691	NM_198680	Serpnb3b	-	0	0
14	chr1	167718811	167792243	NM_0011133	Cd247	+	14.48	15.6
15	chr1	184495137	184562483	NM_130890	Capn8	+	29.1	21.84
16	chr1	176102906	176178580	NM_011465	Spta1	+	12.48	0
17	chr1	167718811	167797567	NM_031162	Cd247	+	14.48	15.6
18	chr1	167718811	167801284	NM_0011133	Cd247	+	14.48	15.6
19	chr1	167718811	167801284	NM_0011133	Cd247	+	14.48	15.6

Figure 25- Fichier d'enrichissement au tour du TSS.

-Des procédures supplémentaire on été effectué sur ce dernier tableau (N°25) à l'aide d'un script sur R (Annexe script R) ou une simple manipulation via Excel.

- R Calcule la sum d'enrichissement autours du TSS par cette formule $sum=A+B$, pour but de récupérer que les endroits enrichis en Pol II.
- Pour qualifier et quantifier l'enrichissement en Pol II entre les deux parties A et B, un ratio est calculé $=ABS (LOG ((G2+1)/ (H2+1);2))$.

	A	B	C	D	E	F	G	H	I	J
1	chr	start	end	TSS	name2	strand	A	B	Sum	ratio
2	chr17	39979941	39985774	39979941	Rn45s	+	467.5	0	467.5	8.87190524
3	chr4	102658983	102678199	102658983	Tctex1d1	+	125.12	0	125.12	6.97865326
4	chr7	19888044	19895394	19895394	Fosb	-	0	106.02	106.02	6.74173662
5	chr9	103014403	103087600	103014403	Rab6b	+	0	98.74	98.74	6.6401003
6	chr2	181313184	181319820	181313184	Uckl1os	+	93.6	0	93.6	6.56376828
7	chr7	7250711	7253562	7253562	Clcn4-2	-	0	73.44	73.44	6.21800615
8	chr16	92058580	92112472	92058580	Mrps6	+	71.76	0	71.76	6.18507364
9	chr16	92058566	92087718	92058566	Slc5a3	+	71.76	0	71.76	6.18507364
10	chr5	109983727	109988012	109988012	Crlf2	-	0	67.94	67.94	6.10726939
11	chr5	24256819	24263661	24263661	Crygn	-	0	66.14	66.14	6.06910063
12	chr1	74438182	74443859	74438182	Ctdsp1	+	63.96	0	63.96	6.02147973
13	chr5	135487991	135489027	135489027	Abhd11os	-	0.26	76.78	77.04	5.9479036
14	chr8	75096001	75111183	75096001	Tmem38a	+	0	60.7	60.7	5.94719858
15	chr9	79446793	79566529	79566529	Col12a1	-	0	59.28	59.28	5.91360751
16	chr7	108570204	108661343	108570204	Pde2a	+	0	58.86	58.86	5.90352037
17	chr11	88825980	88828231	88828231	Gm15698	-	58.54	0	58.54	5.89578732
18	chr12	86814850	86818219	86814850	Fos	+	57.16	0	57.16	5.86195536
19	chr13	93864787	93868923	93864787	A630019I02Ri	+	0	56.16	56.16	5.83693401

Figure 26- Sélections des listes selon le ratio d'enrichissement calculé.

II.2.2.6. Réalisation des Profils moyens d'enrichissement de la Pol II.

Nous avons tout d'abord généré d'autres fichiers valeurs à partir de fichier WIG, à l'aide d'un script R (Annexe Script R). La particularité de ces fichiers est qu'il retranscrit un score d'enrichissement pour chaque nucléotide ou groupe de nucléotides, qui vont servir à calculer les moyennes d'enrichissement et les traduits en des profils plus explicatifs. (Annexes III)

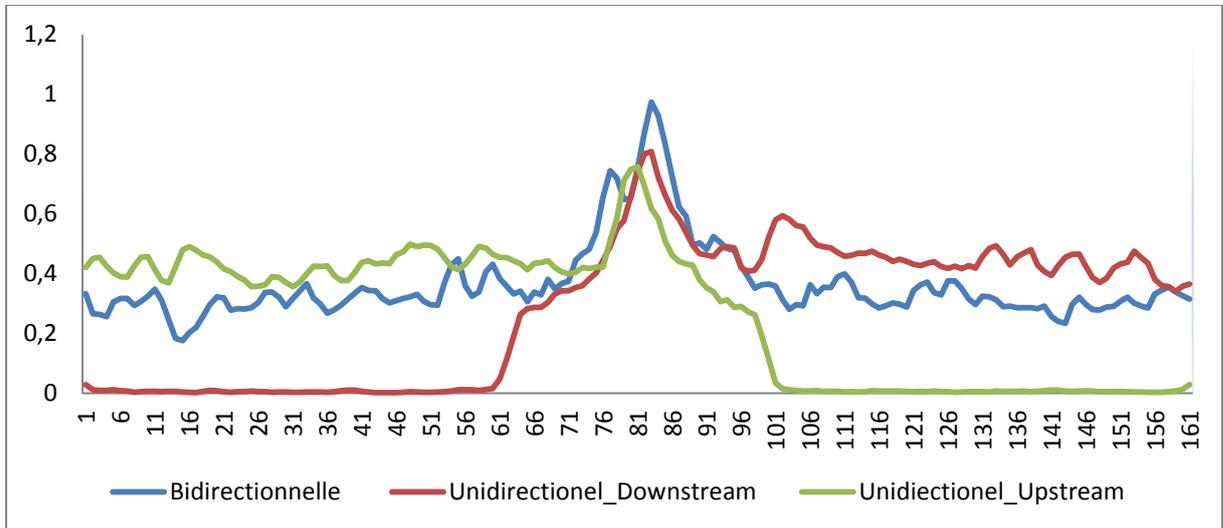


Figure 27 - Profils moyens d'enrichissement des classes des gènes distingués en fonction du score.

La Classe Bidirectionnelle va être divisée en deux listes, en se basant sur les listes d'annotation d'UCSC, une contient les gènes pairs Codant-codant connus, et donc par déduction l'autre liste va contenir les paires codant-non codant qui sont inconnus dans les bases de données et qu'il reste à étudier.

Ainsi, en fin on dispose des différentes classes :

- Les gènes unidirectionnels (connus)
- Les gènes pairs codant-codant (connus)
- Les gènes pairs codant-non codant (inconnus)

D'après cette étape on doit étudier et identifier la classe des paires des gènes codant non codant par rapport aux autres.

II .2.3. Analyse des trois classes des gènes

Deux outils bioinformatique ont été utilisés pour analyser nos classes des gènes.

II .2.3.1. Analyse graphique IGB '*Integrated Genome Browser* ':

Pour visualiser les sites d'enrichissement en Pol II de nos gènes (voir chapitre III résultat III.1.2), on doit charger nos fichiers génomiques nécessaires afin d'arriver de les visualiser leur données qui résulte d'un croisement des informations.

II.2.3.2. Analyse d'annotation fonctionnelle par David

Cette étape consiste en effet en une caractérisation fonctionnelle de chaque groupe de gènes afin d'identifier une ou plusieurs fonctions biologiques qui seraient associées de façon spécifique à ces gènes. Cet outil prend nos listes des gènes celle de Bidirectionnelle coding _Non coding ; coding _coding et celle d'unidirectionnelle (Annexe IV). Et exploité les annotations fonctionnelles disponibles sur ces gènes dans des bases de données publiques afin de trouver des fonctions communes et suffisamment spécifiques à ces gènes.

-On prend les 10 premières fonctions de chaque groupe des gènes afin de calculer le score ($S = -\log_{10}(p\text{-value})$) et tracer leurs graphes (voir chapitre résultat III.2.2).

II.2.4. Les Caractéristiques intrinsèques des gènes bidirectionnels C-NC

II .2.4.1. Récupérer la séquence FASTA des trois classes des gènes:

D'après Interface web UCSC : on récupère les séquences promotrices en amont et en aval du site d'initiation TSS pour chaque gène des trois listes.

The screenshot shows the UCSC Table Browser interface. At the top, it says "Table Browser" and provides a brief description of the tool's purpose. Below this, there are several dropdown menus and input fields for configuring the search:

- clade:** Mammal
- genome:** Mouse
- assembly:** July 2007 (NCBI37/mm9)
- group:** Genes and Gene Predictions
- track:** UCSC Genes
- table:** knownGene
- region:** genome (selected), position: chr12:57795963-57815592
- identifiers (names/accessions):** paste list, upload list
- filter:** create
- interrelation:** create
- correlation:** create
- output format:** sequence
- output file:** bidirectionel.fasta
- file type returned:** plain text (selected), gzip compressed

 There are also buttons for "get output" and "summary/statistics". At the bottom, a link is provided to "reset all user cart settings".

Figure 28 - Schémas comment récupère FASTA séquence de UCSC.

```

1 >mm9_ct_UserTrack_3545_Casp8 range=chr11:58851218-58853217 5'pad=0 3'pad=0 strand=+ repeatMasking=none
2 ATGGAACTGGGGGCACTGTGGGGATGGGAGCAGAGGATCAGGTGGGGCGA
3 GCATGGAGGGAGAGATATGGCAGAGAAAGACTGGAAATGTGGGGCTAGAAA
4 ACCTATCAGGGGACCCCTGGCTAAGACTCCAGTAAGAGGGGTTGGAG
5 CCGTCTGGTGTAGCCAGGCAAGCCTCCAGTGGAGAGACTGGGACACCAA
6 CTCAGACCCCAAACTCTGAGCTACCATTTATCCTGCCTGCAAGATGTGC
7 TGGAGAAAAGATGGCAGAGACTGGGGAGTCCCTCCACCAATCACTAGT
8 CCTCTTGAAGACTGTGCCACAAGAGGGAGGGCCACTCTGACACTGAATT
9 CATGGAGGTCAAGAAACAAGAGGCTGGATGCCCAAGACAGCTAAGATAGAA
10 CTCGACATGGACAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
11 TCTAATCATATCTCTGCCATCTGTAGACAGAAATCTAACATATGTCGCC
12 ATCAGAGAGCTACAGCCACACATAAGCCAGACTGGGGAAATCCTACA
13 GGATATGGGGAGGAAGGATGTAGGAATCAGAGCTCAAGGGCACCACAAG
14
15
16 >mm9_ct_UserTrack_3545_Csrx1 range=chr11:137624774-137626773 5'pad=0 3'pad=0 strand=+ repeatMasking=none
17 ACATAGTTTCCCTGGCAGATCTCTCACTCCGTAGGTCCTCAAGGAAAGTG
18 TGTGCTCTCTGCTGGAGCTGTGTGTACCCATAGTGAATAATGGCTGGG
19 CTGGAGACAGGCTCAGTGGTAAAGAGCCACAGACTGGCTCTCCAGATGCC
20 CTTCTCTGGTGTCTCAAGACAGCCGACAAATGTATCCACATACATGGAA
21 AAATTAAAAGAGAGGAAAGAAAGAAAGTGGGCCCTTCCCTTAAAGTCCCA
22 GGGGATCAATTCACCTTAGGATGGATTTCTGGAGACAGCCCTCTGGG
23 GACTGGGAGCAGTGGGCCCACTAACACATATGTTAATCACTCCAGAGG
24 CATTTCTCAATGGCTGAGACTGGTTTTCATCTGTAGTGTAGCCAG
25 AGGAAGATCCACTCAGAGCCGTTAGAAATGATCTCTTGTGGGTGGCAGGGA
26 GATGGGAAAGCAGAGGACAGAGAGCCCTGGAGGAGGGCAGCTAGGCTA
27 ATGCTCCCGGGGTATAGATATCTCATCCCCATCAGACTCTGGTGGCA
28 CATAGCCCTTTACTGATGAACCTAAAGCCAGCTGTCTGGCAGAAAGTAAG
29
30 >mm9_ct_UserTrack_3545_Pcmd1 range=chr11:7078001-7080000 5'pad=0 3'pad=0 strand=+ repeatMasking=none
31 CGAACCCCAAGACTGTATGTGATGCCCATCTGTCCAGGGGCTCTGTG
32 GTCTAGAGAGAGACCAAGATAGGACTAAGGAATCCGGCCACAGCCCTC
33 APTCAAGGGCTAAAGGCTCAACCTAAGCTCACTCTGACTGGAGATAA
34 AGGTGAGTGAAGTCAAGCCATCGCATAAAGTTAAGTTTCTCCAGTATT
35 CAAGTCAATGGAGCCCTCCATTTACCTACCTCCCTGTTCTAAAACACTTACA
36 CAGCTACAGGCTAAAGCTCAACCTAAGCTCAAGACTCTGTCTTTG
37 CTGATAGGACTGTGATAATATATGATGATGATGATGATGATGATGATGATG
38 CAAGTCCAGGCTGATGATAATATATGATGATGATGATGATGATGATGATG
39 GTACAAATTACAGCATTGGGAAAGCAGTGTCTTTTATGTGCTGGAAAC
40 AAACAAACAAACAAAATAACAGTCCCAAGTAGGAATCGACATGTTTAC
41 TCGGAGGACTACAGATGTAGATGCTGCTCCCTCTGACTGACCTCGCCA
42 CTCTCAACAAAGCACTCTCTCCCGACCCGACACAGCCAGGGGCTCTCGGC
43 GTGGACCAAGAGGCTCAGCTGCCCTGCACAGGGGGCCAGCCATGCAC
44 CCGCCAGGCTCAAGGGTCCGGGCTCCAGAGGACTCCCTCCAGTCA
45 CCGGATACAGAGACCCCGGAGAGGCTGGAGATGAAGTGGCCCTCCTAGAT

```

Figure 29 – Présente le format FASTA de nos gènes

Afin d'évaluer si les *promoteurs* de nos listes *présentent* des *caractéristiques* spécifiques de la chromatine, nous avons fait l'analyse CG Skew pour les gènes unidirectionnels et bidirectionnels.

II.2.4.2. Analyse GC Skew :

Le GC skew est prouvé pour être utile comme un indicateur et présentateur d'enrichissement en CG par cette formule $GC\ Skew = (G - C) / (G + C)$.

Pour tester le lien entre les îlots GpC et la bidirectionnalité de nos trois classe, nous avons calculé les profils skew GC on glissant par une fenêtre de 50 nucléotides pour chacun des trois classes des gènes, générer et représenter a l'aide d'un script sur R .

II.2.5. Traitement de Pol II par le KM05283 (Maybridge, Cornwall, UK) :

-Dans le but de savoir s'il ya une relation entre la Pol II et la marque épigénétique des histones dans la régulation de la transcription des lymphocytes T DN, Ils ont réalisé une analyse afin de savoir l'effet de l'inhibition de l'élongation de Pol II sur la marque épigénétique H3k4me3 on utilise le KM05283, c'est un produit qu'on l'ajoute à la cellule afin d'inhiber et enlever tous les Pol II (Ser2 phosphorylé) fonctionnelles (qui sont au cours de l'élongation de la transcription).

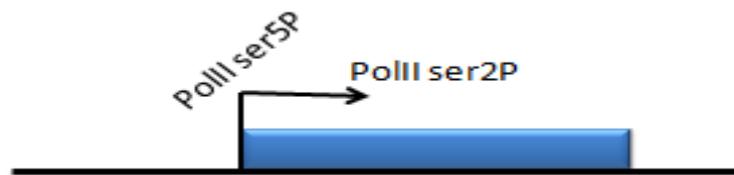


Figure 30- Marque épigénétique de Pol II par la phosphorylation en ser5P au début de la transcription et en ser2P au cours de l'élongation)

L'analyse des données de CHIP-Seq généré au laboratoire TAGC de l'enrichissement en marque épigénétique H3K4me3 avant et après le traitement avec le Contrôleur DMSO et l'inhibiteur de l'élongation KM.

Pour ça nous avons comparé les régions environnantes de trois ensembles de promoteurs affichant des Caractéristiques communes de la chromatine des gènes unidirectionnelles et bidirectionnels.

Chapitre

III

Résultats

Dans ce chapitre on présente nos résultats brutes obtenus par des études réalisées sur le génome de la souris (mm9, UCSC).

III.1. Identification des gènes bidirectionnelle :

D'après le site web UCSC on a récupéré le génome de la souris quand réalise sur tableau transférer sur Excel avec toutes ces coordonnées en (Chr, début, fin, nom, somme, ratio), nécessaire aux prochaines analyses.

	A	B	C	D	E	F	G	H	I	J
1	chr	start	end	TSS	name2	strand	A	B	Sum	ratio
2	chr17	39979941	39985774	39979941	Rn45s	+	467.5	0	467.5	8.87190524
3	chr4	102658983	102678199	102658983	Tctex1d1	+	125.12	0	125.12	6.97865326
4	chr7	19888044	19895394	19895394	Fosb	-	0	106.02	106.02	6.74173662
5	chr9	103014403	103087600	103014403	Rab6b	+	0	98.74	98.74	6.6401003
6	chr2	181313184	181319820	181313184	Uckl1os	+	93.6	0	93.6	6.56376828
7	chr7	7250711	7253562	7253562	Clcn4-2	-	0	73.44	73.44	6.21800615
8	chr16	92058580	92112472	92058580	Mrps6	+	71.76	0	71.76	6.18507364
9	chr16	92058566	92087718	92058566	Slc5a3	+	71.76	0	71.76	6.18507364
10	chr5	109983727	109988012	109988012	Crf2	-	0	67.94	67.94	6.10726939
11	chr5	24256819	24263661	24263661	Crygn	-	0	66.14	66.14	6.06910063
12	chr1	74438182	74443859	74438182	Ctdsp1	+	63.96	0	63.96	6.02147973
13	chr5	135487991	135489027	135489027	Abhd11os	-	0.26	76.78	77.04	5.9479036
14	chr8	75096001	75111183	75096001	Tmem38a	+	0	60.7	60.7	5.94719858
15	chr9	79446793	79566529	79566529	Col12a1	-	0	59.28	59.28	5.91360751
16	chr7	108570204	108661343	108570204	Pde2a	+	0	58.86	58.86	5.90352037
17	chr11	88825980	88828231	88828231	Gm15698	-	58.54	0	58.54	5.89578732
18	chr12	86814850	86818219	86814850	Fos	+	57.16	0	57.16	5.86195536
19	chr13	93864787	93868923	93864787	A630019I02Ri	+	0	56.16	56.16	5.83693401

Figure 31–Liste des gènes récupéré du site UCSC avec son ratio d'enrichissement des deux cotés A et B.

On a trié les gènes en fonction de la somme d'enrichissement en Pol II, afin de définir une liste des gènes codant qui portent environ (8000 gènes).

III.1.1. Analyse des données de CHIP_Seq :

L'utilisation de l'immunoprécipitation de la Chromatine (ChIP) Pour étudier plus précisément la présence d'un certain site d'accrochage en Pol II sur les régions promotrices autour de TSS, et ainsi confirmer des prédictions théoriques.

Les résultats de l'expérience de CHIP_Seq fournis par le labo TAGC avec le ratio calculé nous présentons trois profils d'enrichissement en Pol II de part et d'autre du TSS (-1-4 et +1+4).

- On distingue par ce ratio analysé trois classes des gènes :
 - Classe1 : Unidirectionnelle Downstream avec $R < 6$. (Annexe III)
 - Classe2 : Unidirectionnelle Upstream avec $R > 6$. (Annexe III)
 - Classe3 : Bidirectionnelle avec $R \sim 0$. (Annexe III)

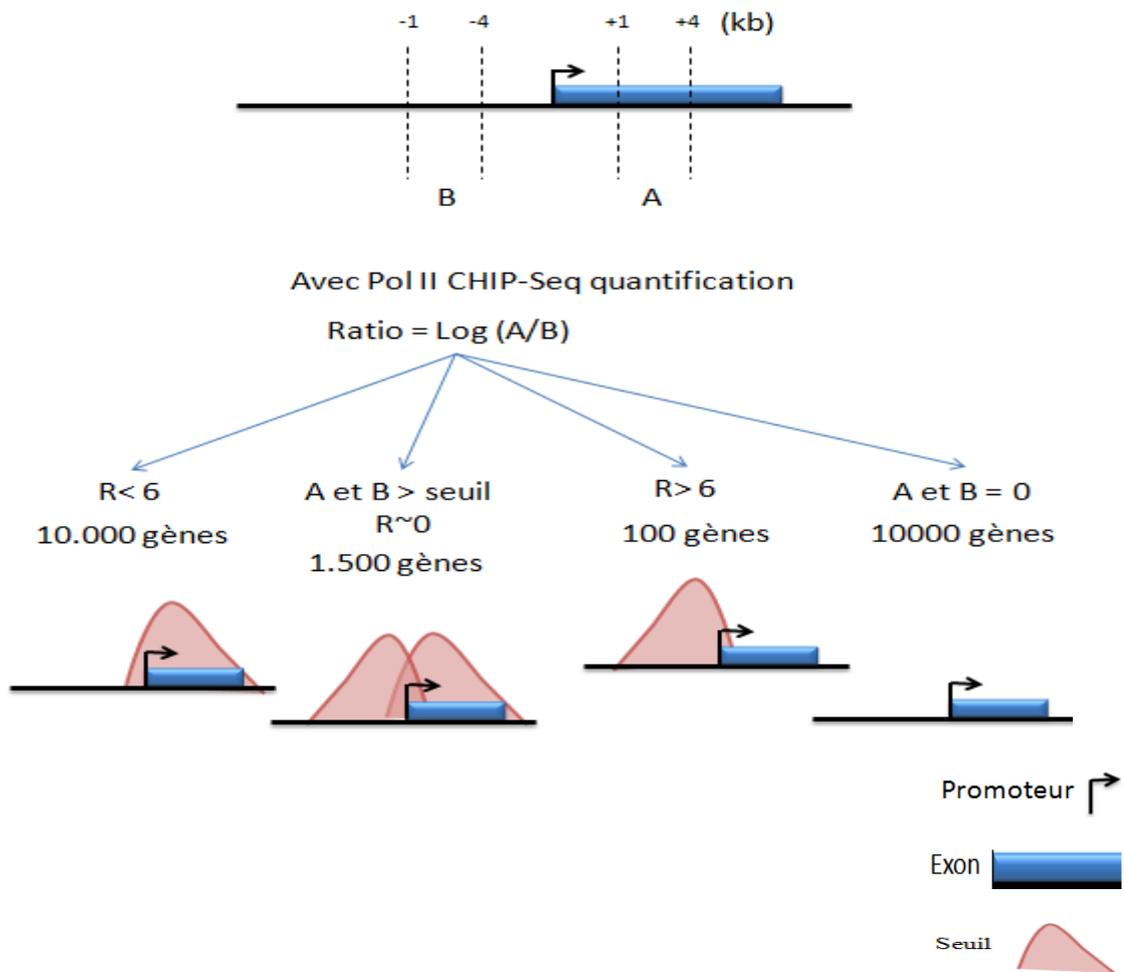


Figure 32 – Profil d’enrichissement en Pol II des gènes du part et d’autre du site d’initiation du TSS.

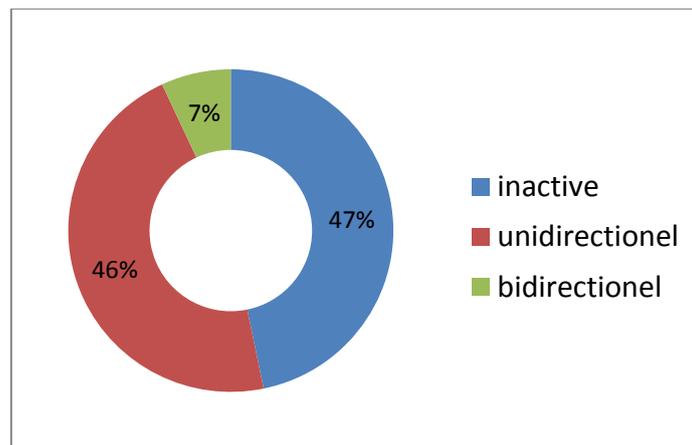


Figure 33 - Diagramme représente le pourcentage de la distribution des gènes dans le génome de la souris.

-R traduit les résultats d’enrichissement en diagramme significative, qui nous présente le pourcentage d’expression des trois classes des gènes.

- on trouve un pourcentage proche entre 46% et 47% pour les gènes dont leurs transcriptions unidirectionnelles en rouge, et celles qui ne sont pas encore identifier noté en bleu, tandis que 7% présente les gènes opposés de bidirectionnalité notée en vert.

III.1.2. Profils moyens d'enrichissement de la Pol II :

Les profils moyens acquiers d'après les table valeurs générés par R et les liste d'annotation sur le site UCSC nous permettra de divisé la classe de bidirectionnalité en deux sous classes :

- Les gènes unidirectionnels (upstream ou downstream)(connus)
 - Les gènes pairs codant-codant (connus)
 - Les gènes pairs codant-non codant (inconnus)
- } Classe de bidirectionnalité

III.2. Analyse des trois classes des gènes par

II.2.1. IGB_I 'Integrated Genome Browser ':

A fin de visualiser les pics d'enrichissements des gènes on a pris au hasard un gène unidirectionnelle connus et l'autre d'une classe bidirectionnelle inconnu pour la visualiser sur un navigateur génomique telqu' IGB.

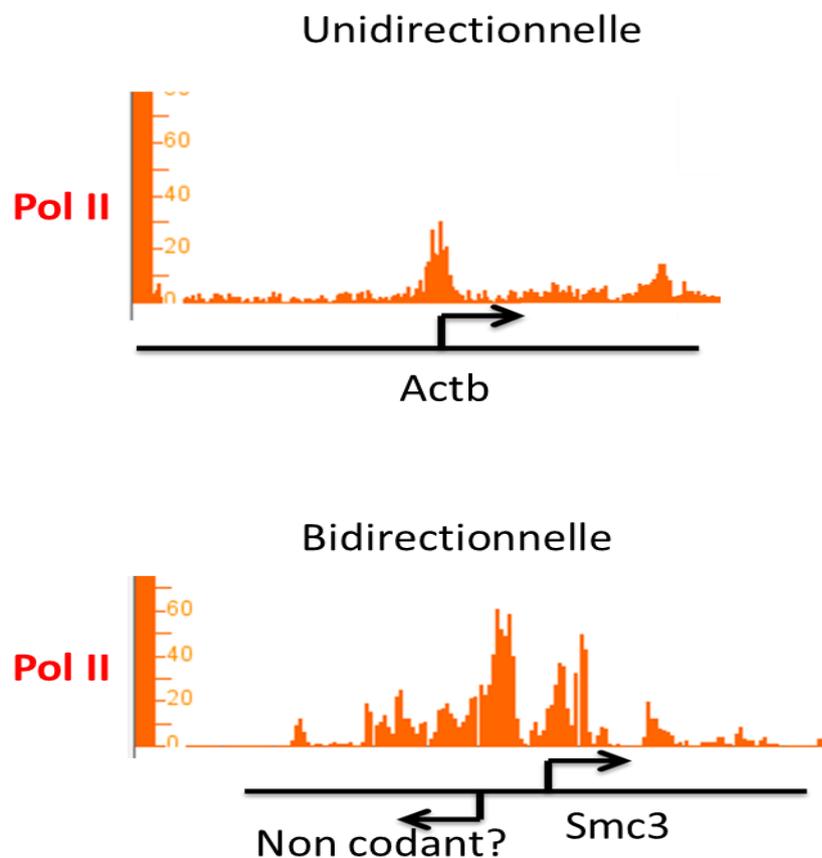


Figure 34: Visualisation et représentation des pics d'enrichissement des gènes en Pol II par IGB.

Les profils présentent clairement les pics d'enrichissement en Pol II du TSS. De sorte qu'il est présenté par un seul côté pour les gènes unidirectionnels l'exemple : Actb et dans les deux sens du TSS pour celle de bidirectionnalité exemple : Smc3.

III.2.2. DAVID pour l'analyse fonctionnelle des gènes :

Nous avons utilisé cet outil pour trouver les différentes fonctions biologiques dont nos gènes sont intégrés, ces fonctions sont présentées en score calculé d'après leur P-value (Annexe IV). Pour clarifier ces résultats, nous avons représenté les meilleurs motifs obtenus avec DAVID sur différentes classes des gènes d'intérêt.

Figure 35- Représentation en barre le Score d'enrichissement des gènes Unidirectionnelle

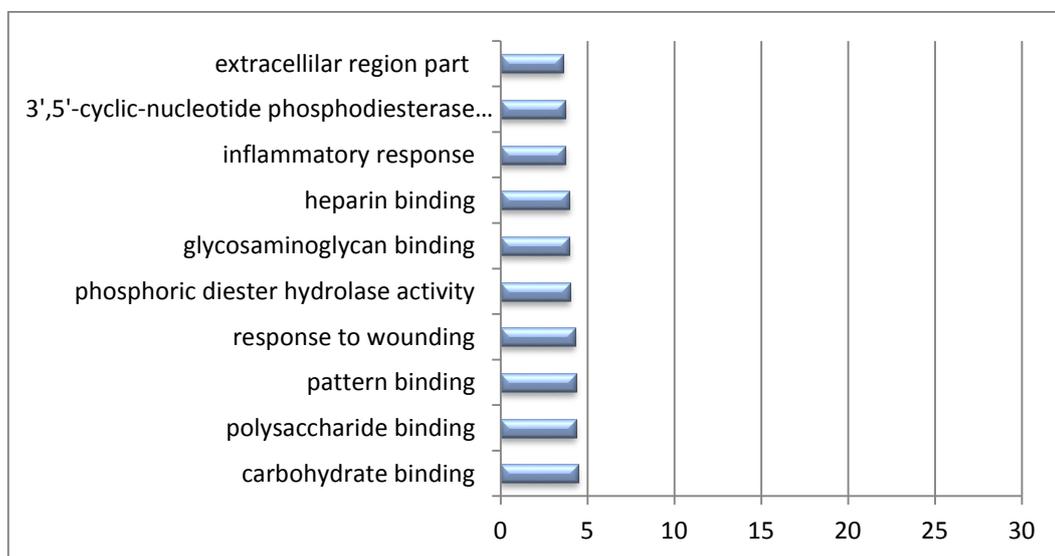


Figure 36- Représentation en barre le Score d'enrichissement des gènes coding_coding

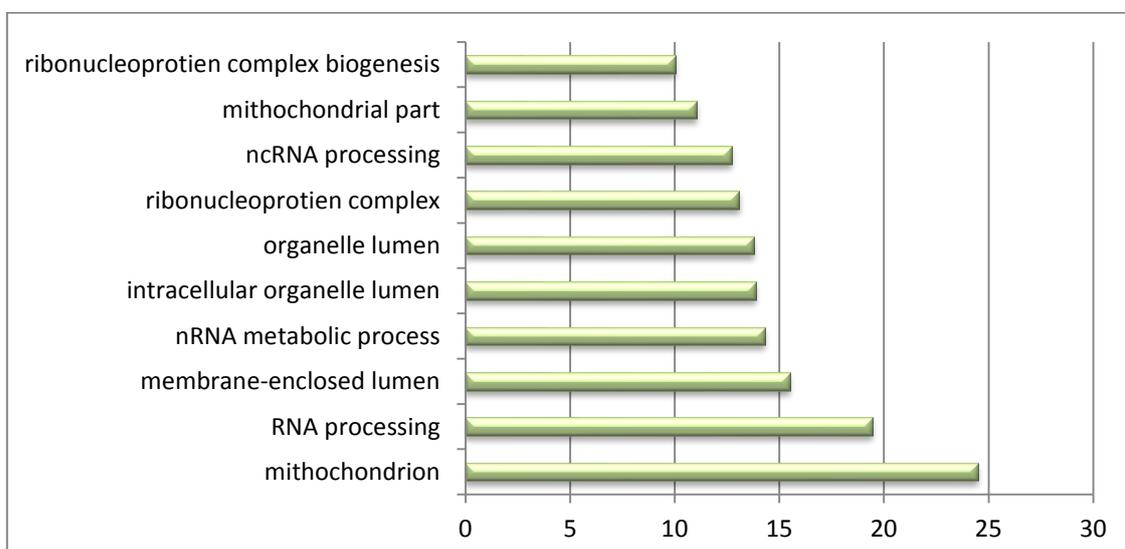
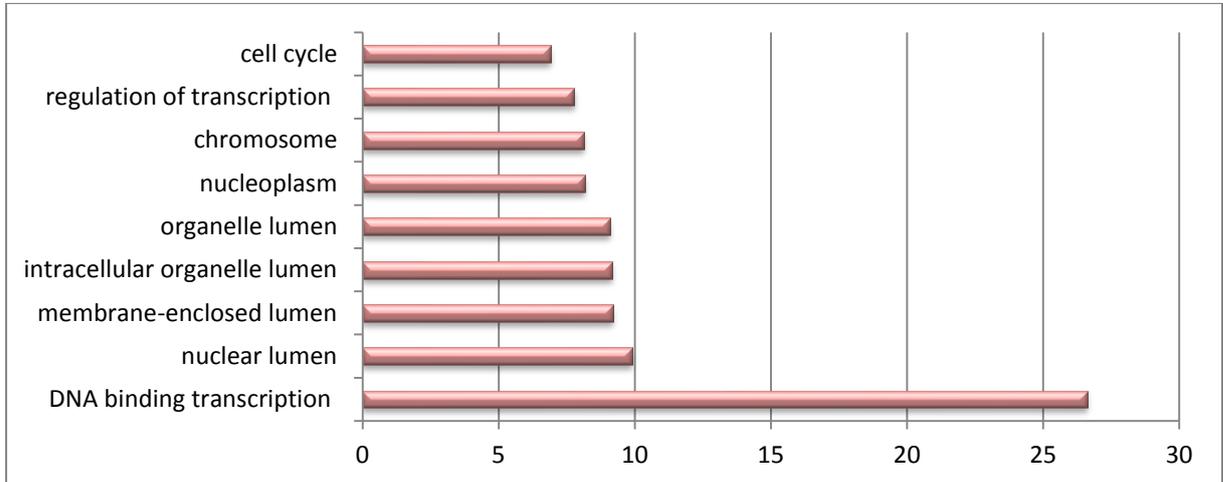


Figure 37- Représentation en barre le Score d'enrichissement des gènes coding_Non coding



-Nous avons trouvé le score des gènes codant non codant est très élevés pour les voies impliqués dans la transcription (voir Annexe IV.1).

➤ L'analyse de ces 2 dernier graphes montres que les 2 classes celle de bidirectionnel C-C et l'autre d'unidirectionnel, leurs gènes sont impliqué en des fonctions différentes, par contre les gènes de bidirectionnel Codant Non codant sont impliqué spécialement dans les fonctions transcriptionnelles.

III.3. Les Caractéristiques intrinsèques des gènes bidirectionnels Codant Non Codant

La visualisation du taux de GC skew des gènes unidirectionnels montrent un profil positif des GC immédiatement en aval du TSS, par contre les gènes codant-codant et les gènes codant-Non codant 2 profil inversés, une positif et l'autre négative respectivement en amont et en aval de la TSS.

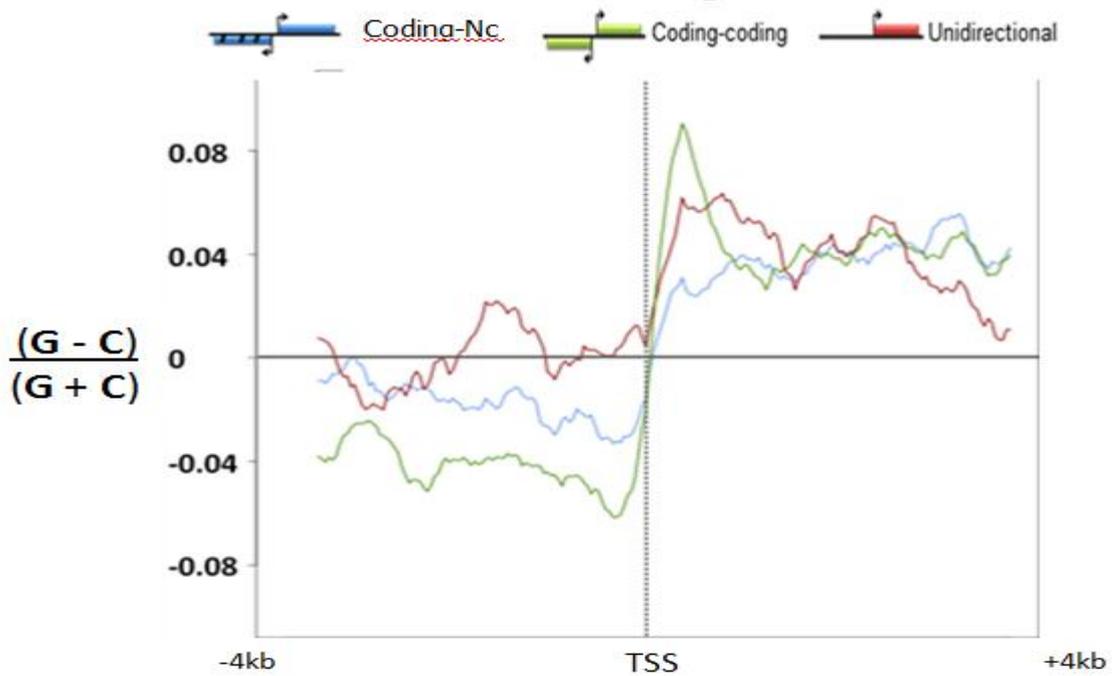


Figure 38 – Profil du GC skew des trois classes des gènes.

- Les promoteurs bidirectionnels des gènes codant Non codant partagent une caractéristique intrinsèque liée à un profil inclinaison en GC, symétrique à celle des promoteurs des gènes codant-codant.

III.4. Analyse des résultats de la technique de KM

Le résultat de CHIP-Seq visualisé par IGB montre clairement l'effet de KM sur l'inhibition de Pol II ser2 qui mène par la suite à la disparition de la marque épigénétique H3K4me3, pour les deux liste des gènes unidirectionnels et bidirectionnels (C-NC).

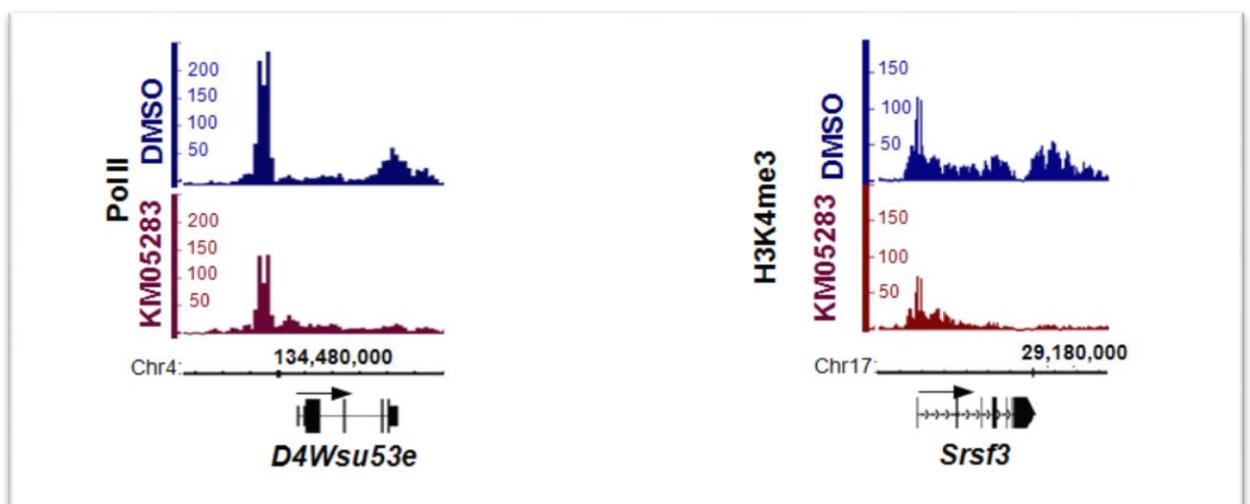


Figure 39 - Exemple d'un gène unidirectionnel avant et après le traitement par Km visualisé sur IGB.

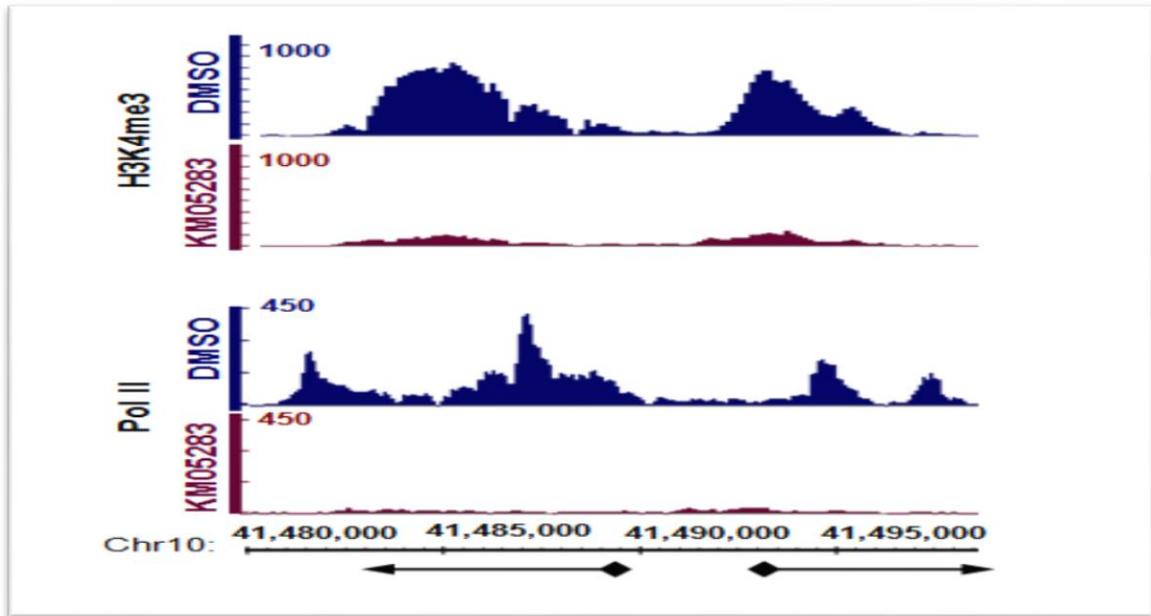


Figure 40 -Exemple d'un gène bidirectionnel avant et après le traitement par Km visualisé par IGB.

- Donc on conclut que la régulation de la transcription par la Pol II est associée à la marque épigénétique tri-méthylation de l'histone 3 en lysine 4.

Chapitre IV

Discussion

IV- Discussion

Les gènes d'intérêts que nous avons utilisées dans cette étude ont été obtenues par un certains nombres de centre généralistes offrent des ressources dédiées aux promoteurs des gènes .Parmi eux le NCBI et l'UCSC sont les plus utilisés, principalement par des expériences biologiques importantes tel que la technique de la Chromatine Immuno-Précipitation – séquençage (CHIP-Seq). Ils permettent également la visualisation de la conservation de séquences génomiques de plusieurs génomes. Ceci nous permettent de réalises plusieurs analyses génomiques comparative au différentes classes des gènes.

Dans notre étude nous avons choisir les précurseurs des cellules immunitaires T de la souris, génomes de référence mm9 sur UCSC, permettant d'étudier une nouvelle sorte de la transcription bidirectionnelle ainsi leurs promoteur qui est une région inter génique de moins de 1Kpb séparant deux gènes en une orientation inverse. Il est aujourd'hui communément admis que cette transcription bidirectionnelle à partir d'un site d'initiation (TSS) mais dans le sens opposé au gène, tend à être remplacée par transcription divergente (Quinlan AR ,2010).

L'organisation de notre données du génome RefSeq (mm9 USCS) en collaboration avec des données brutes « wig » issues de l'expérience de la technique CHIP-Seq généré par Mr BELHOCINE au laboratoire TAGC à Marseille, résulte le ratio d'enrichissement en Pol II de part et d'autre du site d'initiation de la transcription TSS (-1 et -4 kb et aussi +1 et +4kb). Par conséquence ce ratio divise notre liste des gènes en différentes classes, celle des gènes unidirectionnelle downstream, upstream et celle des gènes bidirectionnelle codant_codant connus sur le site UCSC , Bien que la liste des paires des gènes bidirectionnelle codant_ Non codant n'est pas assez clair . Ce qui ouvre un champ vaste à des recherches approfondies sur cette classe.

Afin de mieux visualiser les voies dont elle est impliquée cette transcription des gènes bidirectionnelle codant_Non codant on a utilisé l'outil DAVID, Le but de l'analyse d'annotation fonctionnelle par DAVID est d'identifier les classes des gènes différentielle ment exprimés. Les paires des gènes codant_Non codant montrent avec un score très élevé leur intégration à des fonctions spécialement transrationnelles (Figure 37) trouvées aussi par da Huang et al ,2009).

Nous avons fait des études sur les promoteurs des gènes bidirectionnelle codant_Non codant, ont révélés le partage des mêmes caractéristiques intrinsèques liée à un profil inclinaison en GC skew, symétrique au celle des promoteurs des gènes codant-codant. Cette caractéristique était connu depuis les années 90 mais le séquençage haut débit a révélé qu'il était commun et qu'il avait un impact sur la régulation du transcrit sens. L'enrichissement des promoteurs des gènes bidirectionnelles codant_codant et codant_Non codant par ces séquences répétées en GC, confirme l'enrichissement et le positionnement de Pol II autour du site TSS (-1 -4 / +1 +4), apparue par des profils inclinaisons exprimer en taux élevé en GC skew (Figure 38).

Les résultats de la technique de CHIP-Seq sur l'enrichissement en Polymérase II et en marque épigénétique H3K4me3 montrent une similarité étroite entre les deux listes des gènes bidirectionnels et la liste des gènes unidirectionnels.

On a observé le positionnement et l'enrichissement en Polymérase II phosphorylé en serine 2 suivis par l'apparition d'une marque épigénétique tri méthylation de lysine4 d'histone 3 sur les gènes unidirectionnels et bidirectionnelles codant_Non codant. D'autre part le traitement de ces gènes par le KM05283 (Maybridge, Cornwall, UK) inhibiteur de Pol II ser2, abouti à la disparition de cette marque épigénétique H3k4me3 (Figure 39, 40). Alors que les résultats de Mr BELHOCINE et al (2012) faites avec la technique de CHIP-Seq sur d'autres types de tissu au laboratoire TAGC confirment nos résultats (Cyrille Lepoivre, Mohamed Belhocine et al ,2013).

Enfin, cette étude nous a permis de confirmer que la Pol II de la transcription bidirectionnelle (C_NC) est fonctionnelle et productif dans les deux sens, qui conduit a la production des ARNm du part et d'autre du site d'initiation TSS. (ce qui est trouvé par Cyrille Lepoivre, Mohamed Belhocine et al ,2013).

De manière intéressante, le sens de la transcription d'un gène est souvent corrélé à l'architecture de son promoteur.

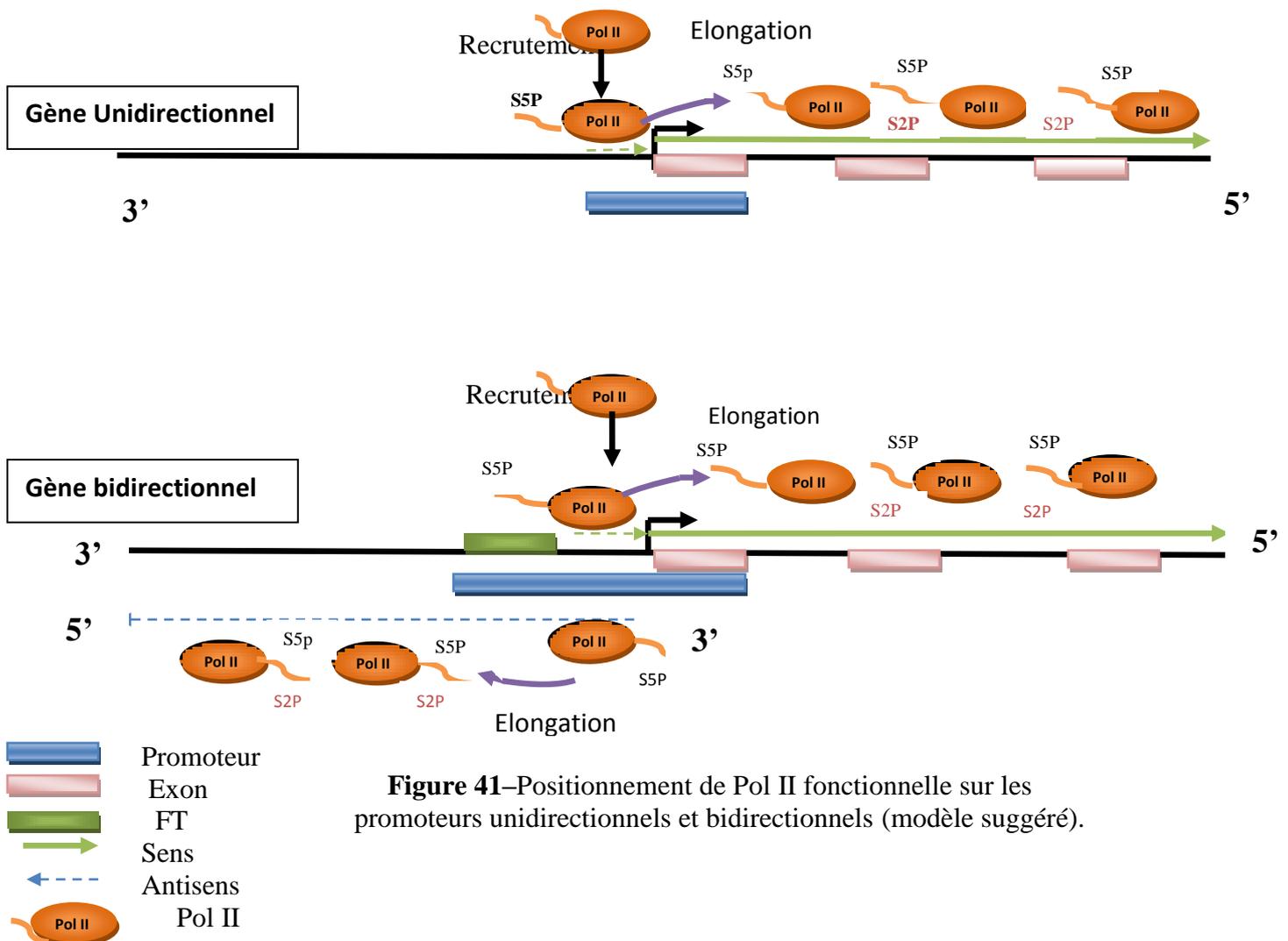


Figure 41–Positionnement de Pol II fonctionnelle sur les promoteurs unidirectionnels et bidirectionnels (modèle suggéré).

Conclusion

Notre étude est basée sur l'identification d'une classe des gènes bidirectionnelle codant non codant, dans les lymphocytes T au cours de sa différenciation du stade DN au stade DP.

D'après les résultats obtenus par DAVID ils confirment l'implication de cette classe des gènes bidirectionnelle codant non codant à des fonctions spécialement transcriptionnelles.

La Pol II de la transcription bidirectionnelle est fonctionnelle et productive car elle peut se déplacer facilement dans les deux sens des gènes bidirectionnelles (Codant Codant / Codant-Non Codant) confirmé par son enrichissement en GC Skew et en tri-méthylation de son Histone 3 lysine 4.

Les gènes qui codent pour les FT semblent présenter une configuration séquentielle spéciale riche en GC skew dans les deux sens du promoteur qui offre la possibilité à la Pol II de faire une transcription bidirectionnelle et ainsi produire des transcrits d'ARN qui ne codent pas pour une protéine mais restent des ARN fonctionnelles.

On suggère que les ARN de la transcription anti-sens jouent un rôle dans la régulation EN-CIS de l'activité transcriptionnelles ou post traductionnelle des FT.

On note un article qui a été publié récemment dans le journal Molecular Cell par une équipe Américaine qui confirme cette hypothèse.

Références

Bibliographiques

Références bibliographiques

- **Adachi, N., M. R. Lieber. 2002.** Bidirectional gene organization: a common architectural feature of the human genome. *Cell* 109:807-809.
- **Agalioti T, Chen G, Thanos D. 2002.** Deciphering the transcriptional histone acetylation code for a human gene. *Cell*. 111 (3): 381-392.
- **Alberts B, Johnson A, Lewis J .2011.** Biologie moléculaire de la cellule, cinquième. ed. Médecine Sciences Publications, Paris.
- **Allfrey, V. G., Faulkner, R., Mirsky, A. E. 1964.** Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* . 51, 786-794.
- **Avery, O. T., Macleod, C. M., McCarty, M. 1944.** Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 79, 137–158.
- **Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas Jo .2001.** Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromodomain. *Nature*. 410 : 120-4.
- **Bannister, A. J. and T. Kouzarides .2005.** Reversing histone methylation. *Nature* 436 (7054): 1103-1106.
- **Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. Zhao, K. 2007.** High-resolution profiling of histone methylations in the human genome. *Cell*. 129, 823–837.
- **Bierne H, Tham TN, Batsche E, Dumay A, Leguillou M, Kernéis-Golsteyn S, Regnault B, Seeler JS, Muchardt C, Feunteun J, Cossart P. 2009.** Human BAH1D1 promotes heterochromatic gene silencing. *Proc Natl Acad Sci U S A*. Aug .18;106(33):13826-31.
- **Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M. R. , Cremer, T. 2005 .** Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 3,e157.
- **Bird A. 2002.** DNA methylation patterns and epigenetic memory. *Genes Dev.* 16: 6-21.

- **Boeger, H., Bushnell, D. A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J. S., Westover, K. D. Kornberg, R. D. 2005.** Structural basis of eukaryotic gene transcription. *FEBS Lett.* 579, 899–903, [DOI :10.1016/j.febslet.2004.11.027] [PubMed :15680971].

- **Butler, J. E. F., J. T. Kadonaga .2002.** The rna polymerase ii core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16:2583-2592.

- **Chadwick, B. P. H. F. Willard .2001.** A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* ,152(2): 375-84.

- **Chao, S. H. Price, D. H. 2001 .**Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J. Biol. Chem.* 276, 31793–31799, [DOI:10.1074/jbc.M102306200][PubMed :11431468].

- **Chuang, C. H., Carpenter, A. E., Fuchsova, B., Johnson, T., de Lanerolle, P. Belmont, A. S. 2006.** Long-range directional movement of an interphase chromosome site.

- **Campos, E.I., Reinberg, D. 2009.** Histones: annotating chromatin. *Annu.Rev. Genet.*43, 559–599.

- **Conaway, J.W., Shilatifard, A., Dvir, A., and Conaway, R.C. 2000. Control of elongation by RNA polymerase II.** *Trends Biochem Sci* 25(8): 375-380.

- **Costanzi, C. and J. R. Pehrson .1998.** Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* .393(6685): 599-601.

- **Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L. Kirsch-Volders, M. 1982.** Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum. Genet.* 62, 201–209, .

- **Cremer, T. Cremer, M. 2010.** Chromosome territories. *Cold Spring Harb Perspect Biol* 2, a003889 .

- **Cyrille Lepoivre, Mohamed Belhocine, Aur_elie Bergon, Aur_elen Gri_on, Miriam Yammine, et al .2013.** Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics, BioMed Central* .14 (1), pp.914.

- **da Huang W, Sherman BT, Lempicki RA .2009.** Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* . 4:44–57.

- **Daury, L., C. Chailleux, J. Bonvallet, D. Trouche . 2006.** Histone H3.3 deposition at E2F-regulated genes is linked to transcription. *EMBO Rep.* 7(1): 66-71.

- **Dundr, M., Ospina, J. K., Sung, M. H., John, S., Upender, M., Ried, T., Hager, G. L. & Matera, A. G. 2007.** Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J. Cell Biol.* 179, 1095–1103.

- **Eissenberg JC, Elgin SC. The HP1 pro-teïn family : getting a grip on chromatin.** *Curr Opin Genet Dev* .2000. 10 : 204-10.

- **Eloi, Mercier .2011. mémoire : «Développement d’outils pour l’analyse de données de ChIP seq et l’identification des facteurs de transcription » .**Soutenue le 01/12/2011.

- **Esteller M .Epigenetics provides a new generation of oncogenes and tumour-suppressor genes.** *Br J Cancer* .2006.94:179-183.

- **Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. 2003.** Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *MolCell* .11, 1549–1561, .

- **Felsenfeld, G. & Groudine, M. 2003.** Controlling the double helix. *Nature* .421, 448–453.

- **Finlan, L. E. & Bickmore, W. A. 2008 .**Porin new light onto chromatin and nuclear organization. *Genome Biol.* 9, 222 .

- **Geyer, P. K. & Clark, I. 2002.** Protecting against promiscuity: the regulatory role of insulators. *Cell. Mol. Life Sci.* 59, 2112–2127 .

- **Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K., Collins, J. B., Li, L., Gilmour, D. S. & Adelman, K. 2008 .**NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* 22, 1921–1933.

- **Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F.2012.** R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*, 45:814–825.

- **Grigoriev, S. A., Bulyanko, Y. A. & Popova, E. Y. 2006.** The end adjusts the means: heterochromatin remodelling during terminal cell differentiation *Chromosome Res.* 14, 53– 69 .

- **Grunstein, M. 1997.** Histone acetylation in chromatin structure and transcription. *Nature*. 389(6649): 349-352.

- **Heitz, E. 1928 .** Das Heterochromatin der Moose. *Jahrb Wiss Botanik* .69, 762-818.

- **Hermann A, Schmitt S, and Jeltsch A .2003.** The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J Biol Chem*, 278, 31717-31721.

- **Hsu, L. M. 2009 .** Monitoring abortive initiation. *Methods* .47, 25–36 .

- **Jackson, D. A. 1997 .** Chromatin domains and nuclear compartments: establishing sites of gene expression in eukaryotic nuclei. *Mol Biol* .24:209-20.

- **Kaplan J-C, Delpech M .2007.** Biologie moléculaire et médecine, troisième. ed, De la biologie à la clinique. Médecine Sciences Flammarion. Paris.

- **Karp, G. 2010.** Cell and Molecular Biology Concepts and Experiments .Hoboken : John Wiley & Son, Inc.

- **Koch, U., and Radtke, F. 2011 .** Mechanisms of T cell development and transformation. *Annu. Rev. Cell Dev. Biol.* 27, 539–562. Pubmed Abstract , Pubmed Full Text , CrossRef Full Text.

- **Komarnitsky, P., Cho, E. J. & Buratowski, S. 2000.** Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* 14,2452–2460.

- **Kornberg, R. D. and Y. Lorch .1974.** Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*.98 (3): 285-94.

- **Kouzarides, T. 2007.** Chromatin modifications and their function. *Cell*.128,693–705.

- **Langelier, M. F., D. Forget, A. Rojas, Y. Porlier, Z. F. Burton, and B. Coulombe. 2001.** Structural and functional interactions of transcription factor (TF) IIA with TFIIIE and TFIIIF in transcription initiation by RNA polymerase II. *J. Biol. Chem.* 276:38652-38657

- **Leach, T. J., Mazzeo, M., Chotkowski, H. L., Madigan, J. P., Wotring, M. G., and Glaser, R. L. 2000.** Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster*. *J Biol Chem* .275, 23267-23272.

- **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*. *Molecular and cellular biology* 28, 3290-3300.

- **Levine, M. 2011.** Paused RNA polymerase II as a developmental checkpoint. *Cell*. 145,502-11.

- **Lorch, Y., Maier-Davis, B., and Kornberg, R.D. 2006.** Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A* .103, 3090-3093.

- **Li, L., X. Wang, V. Stolc, X. Li, D. Zhang, N. Su, W. Tongprasit, S. Li, Z. Cheng, J. Wang, and X. W. Deng. 2006.** Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* .38:124-129.

- **Lorch, Y., Maier-Davis, B., and Kornberg, R.D. 2006.** Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A* .103, 3090-3093.

- **Luger, K., A. W. Mader, et al. 2003.** Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*.389(6648): 251-260.

- **Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. 1997.** Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389, 251–260.

- **Lorch, Y., Maier-Davis, B., and Kornberg, R.D. 2006.** Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A* .103, 3090-3093.

- **Maillet, M., Biologie cellulaire, Masson, paris, 1981.** *Annales de l'Institut Pasteur / Virologie*, Volume 132, Issue 1, January–March 1981, Pages 119-120.

- **Mermoud, J. E., C. Costanzi, et al. 1999.** "Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation." *J Cell Biol* 147(7):1399-408.

- **Momota, R., M. Sugimoto, T. Oohashi, K. Kigasawa, H. Yoshioka, and Y. Ninomiya. 1998.** Two genes, col4a3 and col4a4 coding for the human alpha3(iv) and alpha4(iv) collagen chains are arranged head-to-head on chromosome 2q36. *FEBS Lett*. 424:11-16.

- **Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J. & Adelman, K. 2007.** RNA polymerase is poised for activation across the genome. *Nat. Genet.* 39, 1507–1511.

- **Myers, R. M., K. Tilly, and T. Maniatis.1986.** Fine structure genetic analysis of a beta-globin promoter.*Science*232:613-618.

- **Neil, H., C. Malabat, Y. d'Aubenton-Carafa, Z. Xu,L. M. Steinmetz, and A. Jacquier. 2009** .Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.*Nature* .457:1038-1042.

- **Noll, M., and Kornberg, R.D. 1977.** Action of micrococcal nuclease on chromatin and the location of histone H1. *Journal of Molecular Biology.* 109, 393–404.

- **Olins, A. L. & Olins, D. E. 1974.**Spheroid chromatin units (v bodies). *Science* .183, 330 332,[PubMed :4128918].

- **Olins, D. E. & Olins, A. L . 2003.** Chromatin history: our view from the bridge. *Nat. Rev. Mol.Cell Biol.* 4, 809–814.

- **Park, Y. J., Dyer, P. N., Tremethick, D. J., and Luger, K. 2004.** A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. *J Biol Chem* .279, 24274-24282.

- **Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J.Veyrieras, M. Stephens, Y. Gilad, and J. K. Pritchard.2010.** Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature.* 1;464(7289):768-72.
- **Polson, A., E. Durrett, and D. Reisman.2011.** A bidirectional promoter reporter vector for the analysis of the p53/wdr79 dual regulatory element.*Plasmid* .66:169-179.

- **Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, 26:841–842.

- **Raisner, R. M., P. D. Hartley, et al. 2005.** Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin.*Cell* .123(2): 233-48.

- **Rangasamy, D., I. Greaves, et al. 2004.** RNA interference demonstrates a novel role for H2A.Z in chromosome segregation.*Nat Struct Mol Biol* .11(7): 650-5.

- **Ravasi T1, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, et al.** An atlas of combinatorial transcriptional regulation in mouse and man. **2010**.140(5):744-52. doi: 10.1016/j.cell.2010.01.044

- **Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. 2008.** Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*. 452, 243–247.

- **Sakai, A., Schwartz, B. E., Goldstein, S. & Ahmad, K. 2009.** Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. *Curr. Biol*. 19, 1816–1820.

- **Sarg B, Helliger W, Talasz H, Koutzamani E, Lindner HH. 2004.** Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. *The Journal of biological chemistry*. 279(51):53458-53464.

- **Sarma, K. et D. Reinberg .2005.** **Histone variants meet their match.** *Nat Rev Mol Cell Biol* 6(2): 139-49.

- **Saunders, A., Core, L. J. & Lis, J. T. 2006 .**Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol*. 7, 557–567.

- **Schmidt, C., G. Fischer, H. Kadner, E. Genersch, K. Kühn, and E. Pöschl.1993.** Differential effects of dna-binding proteins on bidirectional transcription from the common promoter region of human collagen type iv genes col4a1 and col4a2.*Biochim Biophys Acta* .1174:1-10.
- **Schwartz, B. E. and K. Ahmad .2005.** Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* .19(7): 804-14.

- **Seila, A. C., J. M. Calabrese, S. S. Levine, G. W.Yeo, P. B. Rahl, R. A. Flynn, R. A. Young, and P. A. Sharp.2008.** Divergent transcription from active promoters.*Science*. 322:1849-1851.

- **Seila, A. C., L. J. Core, J. T. Lis, and P. A. Sharp.2009.** Divergent transcription: a new feature of active promoters.*Cell Cycle* .8:2557-2564.

- **Shah, A. 2009 .** Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD™ system. *Nat. Methods* 6 .Article.

- **Shao, Z. 1999.** **Probing Nanometer Structures with Atomic Force Microscopy.** *News Physiol Sci*, 14, 142-149.

- **Strahl, B. D. and C. D. Allis 2000.** The language of covalent histone modifications. *Nature* 403(6765): 41-45.

- **Suka, N., Luo, K. & Grunstein, M. 2002.** Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat. Genet.* 32, 378–383,[DOI:10.1038/ng1017] .

- **Sydow, J. F. & Cramer, P. 2009.** RNA polymerase fidelity and transcriptional proofreading. *Curr. Opin. Struct. Biol.* 19, 732–739 .

- **Trinklein, N. D., S. F. Aldred, S. J. Hartman, D.I. Schroeder, R. P. Otilar, and R. M. Myers .2004.** An abundance of bidirectional promoters in the human genome. *Genome Res.* 14:62-66.

- **Trojer P, Reinberg D .2007.** Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* 28(1): 1-13.

- **Vinckevicius, A., and Chakravarti, D. 2012.** Chromatin immunoprecipitation: advancing analysis of nuclear hormone signaling. *J. Mol. Endocrinol.* 49, R113–R123. doi: 10.1530/JME-12-0016 .

- **Wang, Q., L. Wan, D. Li, L. Zhu, M. Qian, and M. Deng.2009.** Searching for bidirectional promoters in arabidopsis thaliana. *BMC Bioinformatics* 10 Suppl. 1:S29.

- **Weake VM, Workman JL .2008.** Histone ubiquitination: triggering gene activity. *Mol Cell* 29(6):653-663 .

- **Wilson AS, Power BE, Molloy PL.** DNA hypomethylation and human diseases. *Biochim Biophys Acta* .2007.1775:138-162.

- **Wirbelauer, C., O. Bell, et al. 2005.** Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev.* 19(15): 1761-6.

- **Wood, A. J., Severson, A. F. & Meyer, B. J. 2010.** Condensin and cohesin complexity :the expanding repertoire of functions. *Nat. Rev. Genet.* 11, 391–404, .

- **Wysocka,J.,C.D.Allis,et al .2006.** Histone arginine methylation and its dynamic regulation. *FrontBiosci* .11:344-55.

- **Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. & Handa, H. 2006.** PTEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell.* 21, 227–237, .

- **Yun, M., Wu, J., Workman, J. L., and Li, B. 2011.** Readers of histone modifications. *Cell Res*, 21(4):564–578.

- **Zhang, Z. & Gilmour, D. S. 2006.** Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell* .21, 65–74.

- **Zhou, J., Fan, J. Y., Rangasamy, D. & Tremethick, D. J. 2007.** The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat. Struct. Mol. Biol.* 14, 1070–1076.

- **Zhengdong D. Zhang, Joel Rozowsky, Michael Snyder, Joseph Chang, and Mark Gerstein .2008.** Modeling ChIP Sequencing In Silico with Applications. *Biol. Aug*; 4(8): e1000158.

Annexes

Script R

```
#####Récupérer les données de fichier "wig"

### FUCTION read WIG

readWIG = function(fileName){

  df <- read.table(fileName,sep="^",header=F,quote="");
  charData <- as.character(df$V1);

#####Récupérer les données suivants pour chaque chromosome.

  indexTrack <- grep("chr",charData);

  lines <- charData[indexTrack]; indexTrack <- c(indexTrack,length(charData)+2);

#####Récupérer les nomes des chromosomes et tout ce qui est entre les
parenthèses

  chromNames <- gsub("fixedStep chrom=chr(\\w+) start=\\d+
step=\\d+", "\\1",lines,perl=T);

  wigList <- list();

  num<- 0;

  for(i in 1:length(chromNames)){

    num <- num + 1;

    startOfChromi <- (indexTrack[i]+1)

    endOfChromi <- (indexTrack[i+1]-2);

##### Met les données de chaque chromosome dans une liste « RefSeq.txt »

    wigList[[chromNames[i]]] <- as.numeric(charData[startOfChromi:endOfChromi]);
    print(paste(num," chromosome",chromNames[i], " :",startOfChromi," ->
",endOfChromi,sep=""));

  }

  return(wigList);

}
```

récupère le fichier des coordonnées des TSS (RefSeq.txt) puis le 2eme fichier Wig

Avec leur valeurs d'enrichissement en Pol II et cela dans les deux régions de part et d'autre du TSS .

```
###load WIG

setwd("C:/Users/karima/Desktop/projet/DATA")

file = "WIGfs_DN_Ser5P_el156_bins50.wig"

df = readWIG(file)

###load PEAKS

setwd("C:/Users/karima/Desktop/projet")

peaks <- read.delim("RefSeq.txt", header = TRUE, stringsAsFactors = FALSE)

peaks <- peaks[-(grep("_random", peaks$chrom)),]

### get values and max cordonies

resultList=list();

for ( i in 1:nrow(peaks))

{
  chr= substr(peaks$chrom[i] , 4 , 5)
  resultList[[i]] = sum(df[[chr]][(round(((peaks$TSS[i]+1000))/50,0)) :
    (round(((peaks$TSS[i]+4000))/50,0)))]
}

resultList2=list();

for ( i in 1:nrow(peaks))

{
  chr= substr(peaks$chrom[i] , 4 , 5)
  resultList2[[i]] = sum(df[[chr]][(round(((peaks$TSS[i]-1000))/50,0)) :
(round(((peaks$TSS[i]-4000))/50,0)))]
}

resultMatrix=do.call(rbind, resultList)
```

```

resultMatrix2=do.call(rbind, resultList2)

values=cbind(peaks$chrom,peaks$txStart,peaks$txEnd,peaks$TSS,peaks$name2,
peaks$strand ,resultMatrix,resultMatrix2)

colnames(values) <- c("chr","start","end","TSS","name2","strand","A","B")

values <- as.data.frame(values)

values <- unique(values)

values$ratio <-
abs(log(((as.numeric(as.character(values$A))+1)/(as.numeric(as.character(values$B))+1)
),2))

values$Sum <- as.numeric(as.character(values$A))+
as.numeric(as.character(values$B))

values <- values[values$Sum !=0 ,]

values <- values[order(values$ratio, decreasing = TRUE ),]

values1 <- values[(values$ratio) > 2,]

values1 <- values1[(as.numeric(as.character(values1$A)) <
as.numeric(as.character(values1$B))),]

values2 <- values[(values$ratio) > 2,]

values2 <- values2[(as.numeric(as.character(values2$A)) >
as.numeric(as.character(values2$B))),]

values3 <- values[values$ratio < 0.3 ,]

##### Généré trios listes des gènes valeurs a partir de fichier wig avec leur score
d'enrichissement pour chaque nucléotide ou un groupe nc .

setwd("C:/Users/karima/Desktop/projet/Resultats")

values1=values1[-nrow(values1),]

values2=values2[-nrow(values2),]

values3=values3[-nrow(values3),]

values1=values1[1:1000,]

```

```

values2=values2[1:1000,]
values3=values3[1:1000,]

setwd("C:/Users/karima/Desktop/projet/Resultats")

###

write.table(values, file = "All_Valuers_PoIII_RefSeq.txt", sep = "\t", row.names =
FALSE, col.names = TRUE)

write.table(values1, file = "Unidirectional_Upstream_PoIII.txt", sep = "\t",
row.names = FALSE, col.names = TRUE)

write.table(values2, file = "Unidirectional_Downstream_PoIII.txt", sep = "\t",
row.names = FALSE, col.names = TRUE)

write.table(values3, file = "Bidirectional_PoIII.txt", sep = "\t", row.names = FALSE,
col.names = TRUE)

#####

values1$TSS <- as.numeric(as.character(values1$TSS))
values2$TSS <- as.numeric(as.character(values2$TSS))
values3$TSS <- as.numeric(as.character(values3$TSS))

#####

### get values 1
resultList=list();
for ( i in 1:nrow(values1))
{
chr= substr(values1$chr[i] , 4 , 5)
resultList[[i]] = df[[chr]][(round(((values1$TSS[i]-4000))/50,0)) :
(round(((values1$TSS[i]+4000))/50,0))]
}

resultMatrix=do.call(rbind, resultList)

values=cbind(values1$chr,values1$txStart,values1$txEnd
,values1$TSS,values1$name2, values1$strand ,resultMatrix)

```

```
write.table(values, file = "Unidirectional_Upstream_Valuers_PoIII.txt.txt", sep =
"\t", row.names = FALSE, col.names = TRUE)
```

```
#####
```

```
### get values 2
```

```
values2$TSS <- as.numeric(values2$TSS)
```

```
resultList=list();
```

```
for ( i in 1:nrow(values2))
```

```
{
```

```
chr= substr(values2$chr[i] , 4 , 5)
```

```
resultList[[i]] = df[[chr]][(round(((values2$TSS[i]-4000))/50,0)) :
(round(((values2$TSS[i]+4000))/50,0))]
```

```
}
```

```
resultMatrix=do.call(rbind, resultList)
```

```
values=cbind(values2$chr,values2$txStart,values2$txEnd
,values2$TSS,values2$name2, values2$strand ,resultMatrix)
```

```
write.table(values, file = "Unidirectional_Downstream_Valuers_PoIII.txt.txt", sep =
"\t", row.names = FALSE, col.names = TRUE)
```

```
#####
```

```
### get values 3
```

```
values3$TSS <- as.numeric(values3$TSS)
```

```
resultList=list();
```

```
for ( i in 1:nrow(values3))
```

```
{
```

```
chr= substr(values3$chr[i] , 4 , 5)
```

```
resultList[[i]] = df[[chr]][(round(((values3$TSS[i]-4000))/50,0)) :
(round(((values3$TSS[i]+4000))/50,0))]
```

```
}
```

```
resultMatrix=do.call(rbind, resultList)
```

```
values=cbind(values3$chr,values3$Start,values3$txEnd
,values3$TSS,values3$name2, values3$strand ,resultMatrix)
```

```
write.table(values, file = "Bidirectional_Valuers_PoII.txt.txt", sep = "\t", row.names
= FALSE, col.names = TRUE)
```

```
#####
```

Annexe II

Listes des Profils moyens :

Valeurs d'enrichissement des gènes bidirectionnels

Début

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	chr	start	end	TSS	name2	strand	V7	V8	V9	V10	V11	V12	V13
2	chr17	33016165	33028548	33028548	Zfp870	-	0	0	0	0	0	0	0
3	chr1	169302514	169323928	169323928	Mgst3	-	0	0	0	0	0	0	0
4	chr1	20991477	21069306	21069306	Tram2	-	0	0	0	0	2.46	3	
5	chr6	145123391	145159490	145159490	Casc1	-	0	0	0	0	0	0	0
6	chr4	88248483	88249280	88249280	Ifna12	-	0	0	0	0	0	0	0
7	chr9	50413344	50425569	50425569	AU019823	-	0	0	0	0	0	0	0
8	chr13	24104352	24372659	24372659	Lrrc16a	-	1	1	1	0.08	0	0	0
9	chr1	9932105	9934199	9934199	Snhg6	-	0	0	0	0	1.46	2	
10	chr12	41041632	41172377	41172377	Zfp277	-	0.46	1	1	0.66	0.58	2	
11	chr4	154528183	154596644	154596644	Ski	-	0	0	0	0	0	0	0
12	chr13	99612035	99696339	99696339	Tnpo1	-	0	0	0	0	0	0	0
13	chr4	95766008	95807843	95807843	Cyp2j12	-	0	0	0	0	0	0	0
14	chr2	14916757	14976499	14976499	Nsun6	-	0	0	0	0	0	0	0
15	chr1	57445087	57463518	57463518	Tyw5	-	0	0	0	0	0	0	0
16	chr15	43081585	43114282	43114282	Eif3e	-	1.62	1	1	0.14	0	0	0
17	chr2	164215128	164230318	164230318	Matn4	-	0	0	0	0	0	0	0
18	chr4	129103008	129118062	129118062	Zbtb8b	-	2	1.78	0.34	0	0	0	0
19	chr2	94081520	94081610	94081610	Mir129-2	-	0	0	0	0	0	0	0

Fin du tableau

	A	B	C	D	E	F	G	H	I	J	K	L	
987	chr19	6418737	6533217	6418737	Nrxn2	+	0	0	0	0	0	0	
988	chr13	22648069	22648978	22648069	Vmn1r204	+	0	0	0	0	0	0	
989	chr14	30781565	30788260	30781565	Selk	+	1.22	0	0	0	0	0	
990	chr6	132818025	132819027	132818025	Tas2r110	+	0	0	0	0	0	0	
991	chr13	74075285	74098343	74075285	Brd9	+	0	0	0	0	0	1.02	
992	chr7	134355121	134357801	134355121	Mylpf	+	2.94	0	0	0.98	1	1.04	
993	chr6	121195923	121220935	121195923	Usp18	+	1	1	0.16	0	0	0	
994	chr13	13876573	13920162	13876573	Gng4	+	0	0	0	0	0	0	
995	chr14	67852128	68049789	67852128	Ebf2	+	3	2.6	0	0	0	0	
996	chr2	32485019	32490578	32485019	Ak1	+	3.2	2.1	0.44	0	0	0	
997	chr11	68715067	68718036	68715067	Rpl26	+	2	1.06	0	0	0	0	
998	chr5	136592560	136598817	136592560	Polr2j	+	1	1	0	0	0	0	
999	chr9	63928567	63935852	63928567	Scarletttr	+	0	0	0	0	0	0	
1000	chr11	87912647	88007642	87912647	Cuedc1	+	0	0	0	0	0.96	1	
1001	chr8	3515383	3544267	3515383	Pnpla6	+	0	0	0	0	0	0	
1002	moyenne						0.33305466	0.26572347	0.26334405	0.25646302	0.305209	0.31710611	0.3
1003													
1004													
1005													

Valeurs d'enrichissement des gènes Downstream

	A	B	C	D	E	F	G	H	I	J	K	L	V
1	chr	strat	end	TSS	name2	strand	V7	V8	V9	V10	V11	V12	
2	chr17	39979941	14616	39979941	Rn45s	+	0	0	0	0	0	0	0
3	chr4	102658983	472	102658983	Tctex1d1	+	0	0	0	0	0	0	0
4	chr2	181313184	9367	181313184	Uckl1os	+	0	0	0	0	0	0	0
5	chr16	92058580	25030	92058580	Mrps6	+	0	0	0	0	0	0	0
6	chr16	92058566	25026	92058566	Slc5a3	+	0	0	0	0	0	0	0
7	chr1	74438182	21230	74438182	Ctdsp1	+	0	0	0	0	0	0	0
8	chr11	88825980	24321	88828231	Gm15698	-	0	0	0	0	0	0	0
9	chr12	86814850	23673	86814850	Fos	+	0	0	0	0	0	0	0
10	chr10	92623620	25139	92623620	Cdk17	+	0	0	0	0	0	0	0
11	chr3	96381794	25807	96398104	Polr3gl	-	0	0	0	0	0	0	0
12	chr1	58768334	18412	58768334	Cflar	+	0	0	0	0	0	0	0
13	chr8	122638051	4295	122638051	6430548M08f	+	0	0	0	0	0	0	0
14	chr8	122638051	4292	122638051	6430548M08f	+	0	0	0	0	0	0	0
15	chr18	74427941	21228	74427941	Mbd1	+	0	0	0	0	0	0	0
16	chr18	49682942	16396	49682942	1700044K03R	+	0	0	0	0	0	0	0
17	chr12	29279037	11847	29308155	Colec11	-	0	0	0	0	0	0	0
18	chr5	33678220	12899	33678220	Maea	+	0	0	0	0	0	0	0
19	chr1	136302566	6355	136302566	Cyb5r1	+	0	0	0	0	0	0	0

Valeurs d'enrichissement des gènes Upstream

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
2	chr7	19888044	19895394	19895394	Fosb	-	0.08	0	0	0	0	0	0
3	chr9	103014403	103087600	103014403	Rab6b	+	2	1.82	0	0	0	0	0
4	chr7	7250711	7253562	7253562	Clcn4-2	-	2	2	1.84	0	0	0	0
5	chr5	109983727	109988012	109988012	Crif2	-	0	0	0	0	0.22	1	
6	chr5	24256819	24263661	24263661	Crygn	-	2	1.74	0	0	0	0	0
7	chr5	135487991	135489027	135489027	Abhd11os	-	14.22	13.28	6.88	6.04	8.02	8.94	
8	chr8	75096001	75111183	75096001	Tmem38a	+	1	0.86	0	0	0.42	1.96	
9	chr9	79446793	79566529	79566529	Col12a1	-	0	0	0	0	0.76	1.6	
10	chr7	108570204	108661343	108570204	Pde2a	+	11.36	10.74	2.8	2.36	2	1	
11	chr13	93864787	93868923	93864787	A630019I02Ri	+	0	0	0.06	1.98	2	2	
12	chr6	29150436	29166364	29166364	Impdh1	-	0	0	1	3.4	4	3.24	
13	chr6	114812107	114871770	114871770	Vgll4	-	2	2	1.64	0	0.46	1	
14	chr16	4480227	4523053	4523053	Srl	-	0	0	0	0	0.54	2	
15	chr1	80497642	80755128	80755128	Dock10	-	0	0	0.92	1	1	0.2	
16	chr17	32643406	32665839	32643406	Cyp4f17	+	0	0	0	1.68	3.48	4.52	
17	chr14	56297917	56300654	56300654	Tinf2	-	0	0	0	0	0	0	
18	chr9	107877897	107886387	107877897	Uba7	+	0.04	0	0.06	1.84	2	2.84	
19	chr7	108988501	108993845	108993845	Folr2	-	0	0	0	0	0	0	

Annexe III

Liste des gènes Unidirectionnelle Downstream

	A	B	C	D	E	F	G	H	I	J
1	chr	start	end	TSS	name2	strand	A	B	ratio	Sum
2	chr17	39979941	39985774	39979941	Rn45s	+	467.5	0	8.87190524	467.5
3	chr4	102658983	102678199	102658983	Tctex1d1	+	125.12	0	6.97865326	125.12
4	chr2	181313184	181319820	181313184	Uckl1os	+	93.6	0	6.56376828	93.6
5	chr16	92058580	92112472	92058580	Mrps6	+	71.76	0	6.18507364	71.76
6	chr16	92058566	92087718	92058566	Slc5a3	+	71.76	0	6.18507364	71.76
7	chr1	74438182	74443859	74438182	Ctdsp1	+	63.96	0	6.02147973	63.96
8	chr11	88825980	88828231	88828231	Gm15698	-	58.54	0	5.89578732	58.54
9	chr12	86814850	86818219	86814850	Fos	+	57.16	0	5.86195536	57.16
10	chr10	92623620	92704087	92623620	Cdk17	+	54.72	0	5.80012335	54.72
11	chr3	96381794	96398104	96398104	Polr3gl	-	53.44	0	5.76659516	53.44
12	chr1	58768334	58816053	58768334	Cflar	+	52.54	0	5.74254523	52.54
13	chr8	122638051	122689207	122638051	6430548M08f	+	49.92	0	5.67016051	49.92
14	chr8	122638051	122678474	122638051	6430548M08f	+	49.92	0	5.67016051	49.92
15	chr18	74427941	74442338	74427941	Mbd1	+	49.92	0	5.67016051	49.92
16	chr18	49682942	49684094	49682942	1700044K03R	+	48.54	0	5.63052196	48.54
17	chr12	29279037	29308155	29308155	Colec11	-	48.04	0	5.61588707	48.04
18	chr5	33678220	33715943	33678220	Maea	+	47.7	0	5.60584987	47.7
19	chr1	136302566	136308315	136302566	Cyb5r1	+	47.58	0	5.60229058	47.58

-Liste des gènes Unidirectionnelle Upstream

chr	start	end	TSS	name2	strand	A	B	ratio	Sum
chr7	19888044	19895394	19895394	Fosb	-	0	106.02	6.74173662	106.02
chr9	103014403	103087600	103014403	Rab6b	+	0	98.74	6.6401003	98.74
chr7	7250711	7253562	7253562	Clcn4-2	-	0	73.44	6.21800615	73.44
chr5	109983727	109988012	109988012	Crif2	-	0	67.94	6.10726939	67.94
chr5	24256819	24263661	24263661	Crygn	-	0	66.14	6.06910063	66.14
chr5	135487991	135489027	135489027	Abhd11os	-	0.26	76.78	5.9479036	77.04
chr8	75096001	75111183	75096001	Tmem38a	+	0	60.7	5.94719858	60.7
chr9	79446793	79566529	79566529	Col12a1	-	0	59.28	5.91360751	59.28
chr7	108570204	108661343	108570204	Pde2a	+	0	58.86	5.90352037	58.86
chr13	93864787	93868923	93864787	A630019I02Ri	+	0	56.16	5.83693401	56.16
chr6	29150436	29166364	29166364	Impdh1	-	0	54.54	5.79545527	54.54
chr6	114812107	114871770	114871770	Vgll4	-	0	54.04	5.78240856	54.04
chr16	4480227	4523053	4523053	Srl	-	0	53.04	5.75595577	53.04
chr1	80497642	80755128	80755128	Dock10	-	0	51.26	5.70763522	51.26
chr17	32643406	32665839	32643406	Cyp4f17	+	0	49.92	5.67016051	49.92
chr14	56297917	56300654	56300654	Tinf2	-	0	48.5	5.62935662	48.5
chr9	107877897	107886387	107877897	Uba7	+	0.24	59.32	5.6042244	59.56
chr7	108988501	108993845	108993845	Folr2	-	0	46.8	5.57893871	46.8

-Liste des gènes bidirectionnelle

	A	B	C	D	E	F	G	H	I	J
1	chr	start	end	TSS	name2	strand	A	B	ratio	Sum
2	chr1	169238800	169264109	169238800	Tmco1	+	15.02	18.72	0.2997854	33.74
3	chr5	138012032	138029309	138012032	Tfr2	+	15.02	18.72	0.2997854	33.74
4	chr14	52636539	52639462	52636539	Tmem253	+	65.52	53.04	0.29976049	118.56
5	chr5	147196581	147282701	147196581	Wasf3	+	4.98	6.36	0.29956028	11.34
6	chr17	33016165	33028548	33028548	Zfp870	-	17.56	21.84	0.29936594	39.4
7	chr9	31078985	31082970	31078985	Gm7244	+	21.84	17.56	0.29936594	39.4
8	chr1	169302514	169323928	169323928	Mgst3	-	11.32	14.16	0.2992675	25.48
9	chr1	20991477	21069306	21069306	Tram2	-	30.94	24.96	0.29907393	55.9
10	chr15	59480208	59488654	59480208	Trib1	+	24.96	30.94	0.29907393	55.9
11	chr14	84843369	84936867	84843369	Pcdh17	+	20.1	24.96	0.29904738	45.06
12	chr6	145123391	145159490	145159490	Casc1	-	10.86	8.64	0.29899896	19.5
13	chr9	20970157	21017692	20970157	Pde4a	+	59.28	48	0.29889767	107.28
14	chr8	108225546	108227394	108225546	Pard6a	+	19.42	15.6	0.29879962	35.02
15	chr19	7492114	7500394	7492114	2700081O15F+	+	30.86	24.9	0.29879417	55.76
16	chr2	102546594	102630941	102546594	Slc1a2	+	15.58	12.48	0.29862351	28.06
17	chr2	102546594	102618263	102546594	Slc1a2	+	15.58	12.48	0.29862351	28.06
18	chr14	77556569	78121570	77556569	Enox1	+	20.72	16.66	0.29853876	37.38
19	chr11	70783701	70791435	70783701	Rpain	+	18.72	15.04	0.29798541	33.76

Annexe IV :

IV.1. Voies d'enrichissement des gènes codant_Non codant en PolII

288 chart records							Download File	
Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_FAT	DNA binding	RT		133	18.2	3.8E-15	2.1E-12
<input type="checkbox"/>	GOTERM_BP_FAT	transcription	RT		125	17.1	6.3E-13	1.2E-9
<input type="checkbox"/>	GOTERM_CC_FAT	nuclear lumen	RT		67	9.2	1.3E-10	4.7E-8
<input type="checkbox"/>	GOTERM_CC_FAT	membrane-enclosed lumen	RT		79	10.8	6.3E-10	1.1E-7
<input type="checkbox"/>	GOTERM_CC_FAT	intracellular organelle lumen	RT		77	10.5	7.1E-10	8.2E-8
<input type="checkbox"/>	GOTERM_CC_FAT	organelle lumen	RT		77	10.5	8.0E-10	7.0E-8
<input type="checkbox"/>	GOTERM_CC_FAT	nucleoplasm	RT		49	6.7	6.8E-9	4.8E-7
<input type="checkbox"/>	GOTERM_CC_FAT	chromosome	RT		37	5.1	7.6E-9	4.4E-7
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of transcription	RT		132	18.1	1.7E-8	1.7E-5
<input type="checkbox"/>	GOTERM_BP_FAT	cell cycle	RT		51	7.0	1.2E-7	7.9E-5
<input type="checkbox"/>	GOTERM_BP_FAT	chromosome organization	RT		39	5.3	1.4E-7	6.7E-5
<input type="checkbox"/>	GOTERM_CC_FAT	chromosomal part	RT		30	4.1	5.7E-7	2.8E-5
<input type="checkbox"/>	GOTERM_BP_FAT	chromatin organization	RT		31	4.2	2.3E-6	8.9E-4
<input type="checkbox"/>	GOTERM_MF_FAT	zinc ion binding	RT		119	16.3	3.8E-6	1.1E-3
<input type="checkbox"/>	GOTERM_MF_FAT	nucleotide binding	RT		122	16.7	4.7E-6	8.9E-4
<input type="checkbox"/>	GOTERM_CC_FAT	nucleoplasm part	RT		38	5.2	5.0E-6	2.2E-4
<input type="checkbox"/>	GOTERM_MF_FAT	transcription regulator activity	RT		77	10.5	5.7E-6	8.0E-4
<input type="checkbox"/>	GOTERM_CC_FAT	chromatin	RT		19	2.6	7.3E-6	2.8E-4
<input type="checkbox"/>	GOTERM_MF_FAT	structure-specific DNA binding	RT		13	1.8	2.6E-5	2.9E-3
<input type="checkbox"/>	GOTERM_MF_FAT	ribonucleotide binding	RT		100	13.7	5.6E-5	5.3E-3
<input type="checkbox"/>	GOTERM_MF_FAT	purine ribonucleotide binding	RT		100	13.7	5.6E-5	5.3E-3
<input type="checkbox"/>	GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT		93	12.7	6.1E-5	2.1E-3
<input type="checkbox"/>	GOTERM_CC_FAT	non-membrane-bounded organelle	RT		93	12.7	6.1E-5	2.1E-3
<input type="checkbox"/>	GOTERM_BP_FAT	chromatin modification	RT		23	3.2	7.1E-5	2.3E-2

IV.2. Voies d'enrichissement des gènes codant_codant en PolII

331 chart records [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_CC_FAT	mitochondrion	RT	158	14.7	3.4E-25	1.4E-22	
<input type="checkbox"/>	GOTERM_BP_FAT	RNA processing	RT	73	6.8	3.8E-20	6.9E-17	
<input type="checkbox"/>	GOTERM_CC_FAT	membrane-enclosed lumen	RT	127	11.8	3.1E-16	6.9E-14	
<input type="checkbox"/>	GOTERM_BP_FAT	ncRNA metabolic process	RT	42	3.9	5.2E-15	4.7E-12	
<input type="checkbox"/>	GOTERM_CC_FAT	intracellular organelle lumen	RT	120	11.2	1.3E-14	1.8E-12	
<input type="checkbox"/>	GOTERM_CC_FAT	organelle lumen	RT	120	11.2	1.6E-14	1.7E-12	
<input type="checkbox"/>	GOTERM_CC_FAT	ribonucleoprotein complex	RT	66	6.1	8.4E-14	7.0E-12	
<input type="checkbox"/>	GOTERM_BP_FAT	ncRNA processing	RT	35	3.3	2.0E-13	1.2E-10	
<input type="checkbox"/>	GOTERM_CC_FAT	mitochondrial part	RT	67	6.2	9.4E-12	6.6E-10	
<input type="checkbox"/>	GOTERM_BP_FAT	ribonucleoprotein complex biogenesis	RT	29	2.7	9.8E-11	4.5E-8	
<input type="checkbox"/>	GOTERM_BP_FAT	ribosome biogenesis	RT	26	2.4	1.4E-10	5.0E-8	
<input type="checkbox"/>	GOTERM_BP_FAT	translation	RT	46	4.3	1.5E-10	4.6E-8	
<input type="checkbox"/>	GOTERM_BP_FAT	RNA splicing	RT	35	3.3	2.2E-10	5.8E-8	
<input type="checkbox"/>	GOTERM_BP_FAT	response to DNA damage stimulus	RT	42	3.9	6.9E-10	1.6E-7	
<input type="checkbox"/>	GOTERM_BP_FAT	DNA metabolic process	RT	53	4.9	7.6E-10	1.5E-7	
<input type="checkbox"/>	GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT	156	14.5	1.9E-9	1.1E-7	
<input type="checkbox"/>	GOTERM_CC_FAT	non-membrane-bounded organelle	RT	156	14.5	1.9E-9	1.1E-7	
<input type="checkbox"/>	GOTERM_CC_FAT	nuclear lumen	RT	87	8.1	4.7E-9	2.5E-7	
<input type="checkbox"/>	GOTERM_CC_FAT	mitochondrial inner membrane	RT	42	3.9	6.3E-9	2.9E-7	
<input type="checkbox"/>	GOTERM_MF_FAT	RNA binding	RT	66	6.1	7.4E-9	4.9E-6	
<input type="checkbox"/>	GOTERM_BP_FAT	rRNA processing	RT	19	1.8	1.3E-8	2.3E-6	
<input type="checkbox"/>	GOTERM_BP_FAT	rRNA metabolic process	RT	19	1.8	1.6E-8	2.6E-6	
<input type="checkbox"/>	KEGG_PATHWAY	Spliceosome	RT	23	2.1	2.1E-8	3.1E-6	
<input type="checkbox"/>	GOTERM_CC_FAT	nucleolus	RT	42	3.9	2.4E-8	1.0E-6	

IV.3. Voies d'enrichissement des gènes unidirectionnelles

191 chart records [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_FAT	carbohydrate binding	RT	36	3.4	3.5E-5	2.7E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	polysaccharide binding	RT	20	1.9	4.6E-5	1.8E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	pattern binding	RT	20	1.9	4.6E-5	1.8E-2	
<input type="checkbox"/>	GOTERM_BP_FAT	response to wounding	RT	38	3.6	5.1E-5	1.2E-1	
<input type="checkbox"/>	GOTERM_MF_FAT	phosphoric diester hydrolase activity	RT	14	1.3	9.6E-5	2.5E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	glycosaminoglycan binding	RT	18	1.7	1.1E-4	2.1E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	heparin binding	RT	15	1.4	1.1E-4	1.8E-2	
<input type="checkbox"/>	GOTERM_BP_FAT	inflammatory response	RT	27	2.5	1.8E-4	2.0E-1	
<input type="checkbox"/>	GOTERM_MF_FAT	3',5'-cyclic-nucleotide phosphodiesterase activity	RT	8	0.8	1.8E-4	2.4E-2	
<input type="checkbox"/>	GOTERM_CC_FAT	extracellular region part	RT	66	6.2	2.4E-4	8.4E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	cyclic-nucleotide phosphodiesterase activity	RT	8	0.8	2.4E-4	2.7E-2	
<input type="checkbox"/>	GOTERM_BP_FAT	homeostasis of number of cells within a tissue	RT	7	0.7	6.3E-4	4.0E-1	
<input type="checkbox"/>	GOTERM_CC_FAT	extracellular space	RT	45	4.2	1.5E-3	2.4E-1	
<input type="checkbox"/>	GOTERM_MF_FAT	lipase activity	RT	13	1.2	1.6E-3	1.4E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	defense response	RT	40	3.8	1.9E-3	7.0E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of coagulation	RT	6	0.6	3.4E-3	8.1E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	cell-cell signaling	RT	28	2.6	3.7E-3	7.8E-1	
<input type="checkbox"/>	KEGG_PATHWAY	Purine metabolism	RT	18	1.7	4.1E-3	5.0E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	multicellular organismal homeostasis	RT	11	1.0	4.4E-3	7.8E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	morphogenesis of an epithelial sheet	RT	4	0.4	4.6E-3	7.5E-1	
<input type="checkbox"/>	GOTERM_CC_FAT	extracellular matrix	RT	29	2.7	5.3E-3	4.7E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	regulation of blood coagulation	RT	5	0.5	5.3E-3	7.6E-1	
<input type="checkbox"/>	GOTERM_CC_FAT	extrinsic to membrane	RT	40	3.8	5.6E-3	4.0E-1	
<input type="checkbox"/>	GOTERM_CC_FAT	proteinaceous extracellular matrix	RT	28	2.6	5.8E-3	3.4E-1	
<input type="checkbox"/>	GOTERM_BP_FAT	positive regulation of multicellular organismal process	RT	18	1.7	6.5E-3	8.0E-1	