

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE SAAD DAHLEB – BLIDA 1



DEPARTEMENT DE BIOLOGIE ET PHYSIOLOGIE CELLULAIRE
FACULTE DES SCIENCES ET DE LA NATURE ET DE LA VIE

*MEMOIRE DE FIN D'ETUDE EN VUE DE L'OBTENTION
DU DIPLOME DE MASTER EN SCIENCES DE LA NATURE
ET DE LA VIE*

Spécialité: Génétique et physiologie

Thème :

**DYNAMIQUE EPIGENETIQUE ET TRANSCRIPTIONNELLE
IMPLIQUEE DANS LA DIFFERENCIATION DES CELLULES T**

Présenté par: **Nesrine DAHMANE**

Devant le jury composé de:

GUEDIOURA.A	MAA	USDB	Président
GUESSAIBIA.N	MCB	USDB	Examinatrice
MOHAMED SAID.R	MAA	USDB	Promoteur
BELHOCINE.M	MAB	Inserm	Co-Promoteur

Promotion: 2011-2012

Remerciements

Aux membres du jury,

Je vous remercie vivement d'avoir accepté d'évaluer mon mémoire. Votre participation à ma soutenance a été un grand honneur pour moi.

A M.MOHAMED SAID,

Je suis reconnaissante à l'attention que vous avez bien voulu porter à ce travail et pour votre enseignement le long de notre cursus.

A M.BELHOCINE,

Je vous exprime toute ma gratitude, veuillez trouver ici l'expression de ma profonde reconnaissance pour m'avoir suggéré puis guidée sur ce passionnant thème qui est au cœur de l'actualité scientifique, ainsi que de m'avoir reçue dans le cadre de mon stage de fin d'études au sein du laboratoire Technical Advances for Genomics and Clinics (TAGC) Inserm U1090 – Aix-Marseille Université. **Au Pr S.LAURYC et Pr M. GOOSSENS,**

Merci de m'avoir donnée l'opportunité de réaliser le stage pratique au sein du **laboratoire de génétique moléculaire**, service de biochimie et génétique -**CHU Hôpital Henri Mondor à Créteil** et pour toute l'attention portée à mes démarches.

Au Dr C.PREHU,

Merci de m'avoir encadrée et guidée tout au long de mon stage, pour la confiance accordée qui m'a permis d'être autonome dans la manipulation des techniques de la génétique moléculaire et la biochimie, de m'avoir aussi appris à scruter les détails de chaque technique apprise au sein du laboratoire de génétique moléculaire.

A ma famille et ma petite sœur Sofia,

Merci, pour votre amour, votre patience et votre soutien tout le long de mes longues études.

A tonton Djamel,

Je t'exprime toute ma gratitude et tendresse pour m'avoir aidé dans mes démarches afin de pouvoir réaliser mon stage à l'hôpital Henri Mondor et d'avoir été un si gentil et génial tonton tout au long du mois passé au sein de ta famille.

A tonton Brahim et Majid,

Toute ma gratitude et ma gratitude pour votre présence, votre soutien et tout le temps consacré afin de me permettre de présenter mon projet de fin d'études. Toute mon affection pour vous.

A Fayçal,

Merci pour ton amour, ton soutien et d'avoir été présent pour tous les moments pénibles ainsi que les meilleurs.

A Nadjet,

Merci d'avoir été à mes côtés pour les meilleurs et les pires moments. Merci pour ta précieuse aide et d'être la meilleure amie qu'on puisse espérer.

A Thelushia et Mamia,

Merci les filles pour tous les bons moments passés en votre compagnie et d'avoir été d'aussi gentilles collègues durant le stage.

A mes camarades,

Merci pour toute ma promotion de génétique et physiologie, la meilleure de toutes, pour tous les moments de joie et la complicité partagée entre nous.

A mes amis,

A toutes mes amies et tous mes amis, je vous exprime mon affection et mes remerciements les plus sincères : Meriem, Manina, Manel, Mouna, Aldrina, Charlie, Yasmine, Mohamed, Romaissa, Ibrahim, Ikram, Mina, Nassim...etc.

Liste des abréviations

ADN: Acide Désoxyribo-Nucléique

ARN pol II: Acide Ribo-Nucléique Polymérase II

BED: BROWSER EXTENSIBLE DATA

CD4: Cluster de Différenciation 4

CD8: Cluster de Différenciation 8

CD44: Cluster de Différenciation 44

CD25: Cluster de Différenciation 25

ChIP-seq: Chromatine ImmunoPrécipitation-séquençage

CMH: Complexe Majeur d'Histocompatibilité

Cocas: ChIP-on-ChIP Analysis Suite

DN: Double Négative

DP: Double Positive

FT: Facteur de Transcription

GFF: General Feature Format

GREAT: Genomic Regions Enrichment of Annotations Tool

HAT: Histones Acétyl Transférases

HMT: Histones MéthylTransférases

H1: Histone 1

H3: Histone 3

H4: Histone 4

H 2A: Histone 2A

H2B: Histone 2B

H3K27ac: Histone 3 lysine 27 acétylation

H3K4me3: Histone 3 lysine 4 triméthylation

H3K4me1: Histone 3 lysine 4 monométhylation

IGB: Integrated Genome Browser

IL2: Interleukine 2

INR: Initiateur

IP: Immuno-Précipitation

KDa: *KiloDalton*

Km: Kilomètre

Nm: *Nanomètre*

N-terminale: Terminaison Amine (NH₂)

PCR: Réaction de Polymérisation en Chaîne

RAG: Recombination Activating Gene

SP: SimplePositive

RSAT: Regulatory Sequence Analysis Tools

SSRs: Séquences Signal de la Recombinaison

TCR: Récepteur des CellulesT

TCR γ : récepteur des cellules T gamma

TCR δ : récepteur des cellules T delta

SOLID: Sequencing by Oligonucleotide Ligation and Detection

TSS: Transcription Start Site (site d'initiation de la transcription)

VDJ: V pour variable, D pour diversité et J pour jonction

WIG: Wiggle

Liste des figures

Fig 1: Histoire de la chromatine.....	4
Fig 2: Schéma illustrant l'emplacement des histones au sein du nucléosome.....	5
Fig 3: Les différentes étapes d'assemblage de la chromatine.....	6
Fig 4: Exemple de modifications épigénétiques effectuées par différents groupes d'enzymes sur les résidus amine de la queue N-terminal de l'histone 3 (snapshot nature). Me (méthylation), P (phosphorylation), Ac(acétylation).....	8
Fig 5 : Eléments d'un gène eucaryote.....	9
Fig 6: Les différents éléments intervenant dans la régulation de la transcription.....	10
Fig 7: Vue d'ensemble du développement des cellules T.....	12
Fig 8: Sélection négative et positives des thymocytes.....	14
Fig 9: Etapes de la réalisation de ChIP-seq et construction d'une Bibliothèque d'ADN.....	16
Fig 10: Illustration des étapes End-repair et modification de l'extrémité 3'	17
Fig11: Etapes de ligature des adaptateurs et enrichissement des fragments par PCR	18
Fig12: Réalisation d'une emPCR pour le séquençage par SOLID.....	20
Fig 13: Illustration des étapes du séquençage par ligature des oligonucléotides et identification (SOLID).....	20
Fig 14: Interface de Cocos.....	22
Fig 15: interface d'IGB.....	22
Fig 16: interface de Galaxy.....	23
Fig 17: Interface de RSAT.....	23
Fig 18: Interface de GREAT.....	24
Fig 19: Interface de R.....	24
Fig 20: Génération des fichiers gff à l'aide de perl et le peak calling à l'aide de IGB et CoCas. me1=H3K4me1, me3=H3K4me3, ac=H3K27ac	26
Fig 21: Visualisation des fichiers gff par IGB..	28
Fig 22: Définition des paramètres d'après un gène ubiquitaire.....	28
Fig 23: Intersection des fichiers gff créés suite au peak calling.....	29

Fig 24: Paramètres d'intersect.....	30
Fig 25: Aperçu de la présentation du fichier BED.....	30
Fig 26: Représentation de la méthode d'identification par GREAT, du domaine fonctionnel d'une région cis-régulatrice à partir des deux gènes les plus proches, en amont et en aval du TSS.....	31
Fig 27: paramètres choisis dans table browser.....	31
Fig 28: Distance en Kb des régions Cis-régulatrices en du TSS pour le stade DN.....	33
Fig 29 : Distance en Kb des régions Cis-régulatrices en fonction du TSS pour le stade Commun DN/DP.....	34
Fig 30: Distance en Kb des régions Cis-régulatrices en fonction du TSS pour le stade DP.....	35
Fig 31: Radar plot présentant les voies biologiques activées durant DN, DP et commun selon leur score d'enrichissement.....	36
Fig 32: schéma démontrant la position des différents enhanceurs pour le gène Ptpn22 selon le stade concerné.....	37
Fig 33: schéma démontrant la position des différents enhanceurs pour le gène Runx2.....	37
Fig 34: schéma démontrant la position des différents enhanceurs pour le gène Fyn.....	38
Fig 35: schéma démontrant la position des différents enhanceurs pour le gène Irf1.....	38
Fig 36: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade DN.....	42
Fig 37: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade Commun.....	42
Fig 38: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade DP.....	43

Liste des tableaux

Tableau I: Les différentes classes de modifications identifiées sur les histones.....	7
Tableau II: Paramètres choisis dans Cocas après la visualisation par IGB et le nombre de piques générés par Cocas pour les trois marques épigénétiques d'intérêt pendant le stade DN et DP de la différenciation des cellules	27
Tableau III: Gène et leurs régions cis-régulatrices (enhancers) identifiés pendant le stade DN et qui sont en interaction avec les FTs détectés par GREAT	40
Tableau IV: Gène et leurs régions cis-régulatrices (enhancers) identifiés pendant le stade Commun (DN/DP) et qui sont en interaction avec les FTs détectés par GREAT	41

Résumé

La transcription des gènes est strictement contrôlée par l'interaction d'événements de régulation, des promoteurs des gènes et des éléments régulateurs géniques-distaux appelés Enhancers. Ce travail a pour but d'étudier le dynamisme épigénétique des trois marques épigénétiques H3K4me3, H3K4me1 et H3K27ac et comment elles interfèrent dans la régulation de l'expression génomique lors des différents stades de la différenciation des thymocytes T, ainsi que le dynamisme transcriptionnel régulant ce processus. Cela, grâce au séquençage à haut débit du génome murin, combiné à ChIP réalisé par l'intermédiaire des outils bioinformatiques permettant de traiter les énormes data issues de cette technologie.

Mots clés : épigénétique, cellules T, transcription, ChIP-seq, Enhancer.

Abstract

Gene transcription is strictly controlled by the interaction of gene regulatory events promoters and genic-distal regulatory elements called Enhancers. This work aims to study the epigenetic dynamics of three epigenetic marks H3K4me3, H3K4me1 and H3K27ac and how can they interfere in the regulation of genomic expression during different stages of differentiation of thymocytes, also the transcriptional dynamics regulating this process. This is possible, due to combined the high-throughput sequencing of mouse genome to ChIP and bioinformatic tools to process the huge data from this technology.

Keywords: Epigenetic, T cell, ChIP-seq, Enhancer.

تتم عملية نسخ الجينات عبر تفاعلات دقيقة متحكممة في تنظيم محفزات الجينات و في مناطق نيكلو تيديية تقع غالبا بعيدا عن الجين المراد نسخه و تسمى المحفزات. هذا العمل يهدف إلى دراسة الدينامكية اللاجينية لثلاثة علامات لاجينية H3K4me3, H3K4me1, H3K27ac و كيف تتدخل هذه العلامات في تنظيم التعبير الجيني خلال مراحل مختلفة من تمايز الخلايا اللمفاوية التائية و كذلك دراسة دينامكية نسخ الجينات التي تتدخل في تنظيم هذه العملية . هذا ممكن من جهة بفضل جمع تكنولوجيا تسلسل الحمض النووي مع الترسيب المناعي للكروماتين و من جهة أخرى بفضل الأدوات المعلوماتية الحيوية التي تسمح بالتعامل مع البيانات الضخمة الناتجة من هذه التكنولوجيا.

الكلمات المفتاحية:

رقاقة الحمض النووي , الترسيب الكروماتيني المتبع بتسلسل الحمض النووي, الخلايا التائية, نسخ الجينات .

Sommaire

Chapitre 1 : Données bibliographiques

1-1 Remodelage de la chromatine par les modifications épigénétiques des histones	4
1-1-1 La chromatine.....	4
1-1-2 Les niveaux de compaction de la chromatine.....	6
1-1-3 Rôle de la chromatine et modifications épigénétiques.....	7
1-1-3-1 Modifications d’histones.....	7
1-1-3-2 Le code Histone.....	8
1-2 Mécanisme de la transcription chez les eucaryotes	9
1-2-1 Les séquences consensus d’un gène et le mécanisme de transcription.....	9
1-2-2 Les éléments régulateurs de la transcription.....	9
1-2-3 La régulation de la machinerie transcriptionnelle.....	10
1-3 Le développement et la maturation des cellules T	
1-3-1 Les stades de différenciation des thymocytes du stade DN à DP.....	11
1-3-2 Réarrangements du TCR.....	11
1-3-3 Processus de sélections.....	13

Chapitre 2 : Matériel et méthodes

2-1 Chromatin-Immuno-Precepitation- Sequencing (ChIP-seq)

2-1-1 Les étapes de la réalisation de CHIP-seq.....	16
2-1-2 Construction d’une Bibliothèque d’ADN.....	17
2-1-3 Sequencing by Oligonucleotide Ligation and Detection (SOLID).....	18
2-1-3-1 La construction d’une librairie.....	18
2-1-3-2 La PCR par émulsion.....	19
2-1-3-3 Le séquençage.....	20

2-2 Outils bioinformatiques	22
2-2-1 Cocas.....	22
2-2-2 IGB.....	22
2-2-3 Galaxy.....	23
2-2-4 RSAT.....	23
2-2-5 GREAT.....	24
2-2-6 Le langage R.....	24

Chapitre 3 : Analyses bioinformatiques

3-1 Analyse des datas issues de ChIP-SOLID	26
3-1-1 Prétraitement des données.....	26
3-1-2 Visualisation et représentation des pics de liaison à l'aide des fichiers GFF.....	27
3-1-3 Le peak calling.....	27
3-1-3-1 Création des pics à l'aide de Cocas et IGB.....	27
3-1-3-2 Visualisation des pics et détermination des paramètres par IGB pour CoCas.....	28
3-1-4 L'intersection ou « overlapping »	29
3-1-5 Identification des séquences cis-régulatrices (enhancers) à l'aide de GREAT.....	30
3-1-6 Identification des motifs à l'aide de RSAT.....	31

Chapitre 4 : Résultats et discussion

4-1 Résultats

4-1-1 Résultats générés par GREAT.....	33
4-1-1-1 Distance à partir du TSS.....	33
4-1-1-2 Les voies biologiques.....	36
4-1-1-3 Les gènes exprimés durant DN et DP.....	37
4-1-1-4 Les facteurs de transcription.....	39
4-1-2 Résultats générés par RSAT.....	42
4-2 Discussion	44
Conclusion.....	46

Références bibliographiques.....	47
Annexes.....	51

Introduction:

Toutes les cellules d'un individu sont porteuses de la même information génétique, mais il existe une centaine de types cellulaires, aux fonctions et propriétés différentes : cela prouve que la séquence du génome ne peut pas tout expliquer. Donc, il apparaît évident qu'il existe une donnée qui vient s'ajouter à l'information fournie par les gènes: elle permet la diversification des potentialités génétiques au sein des différents types de cellules et est transmise de façon stable au cours des générations cellulaires pour constituer des organes et tissus fonctionnels pendant le développement, c'est l'épigénétique. Effectivement, l'épigénétique, c'est tout changement d'activité du génome n'impliquant pas de changement dans la séquence d'ADN et qui est héritable au cours des divisions cellulaires. **[Heard., 2012]**

Au cours de la différenciation des thymocytes du stade double négatif (DN) au stade double positif (DP), plusieurs changements ont lieu au sein de la cellule dont des variations phénotypiques impliquant par conséquent différents processus biologiques. Lors de ces changements, différents gènes sont exprimés en fonction du stade de développement de ces thymocytes et d'autres sont réprimés. L'expression de ces gènes est soumise à une régulation transcriptionnelle et épigénétique complexe. Il a été démontré par des études que les trois marques épigénétiques H3K4me1, H3K4me3 et H3K27ac seraient impliquées dans la régulation de l'expression des transcrits: La triméthylation de H3K4 serait étroitement associée à des sites d'initiation de la transcription des gènes activement transcrits **[Barski et al., 2007]** et pourrait interagir indépendamment ou en collaboration avec la TATA box afin de former le complexe de préinitiation de la transcription, afin de démarrer la transcription **[Lauberth et al., 2013]**, la H3K4me1 interagirait avec les enhancers et lorsque ces derniers portent à leur surface seulement la monométhyl H3K4, ils seraient simplement prêts à l'activation de la transcription en réponse à des stimuli externes mais pour que cette activation ait lieu, il faudrait l'intervention de la H3K427ac qui interagirait avec ces enhancers pré-actifs et activerait l'expression des gènes concernés.**[Creyghton MP et al., 2010]**.

Donc plusieurs études démontrent l'implication individuelle de ces trois marques épigénétiques dans la régulation de la transcription mais aucune étude jusqu'à maintenant n'a été portée sur le dynamisme de ces trois marques épigénétiques réunies et leur rôle dans la régulation de l'expression génomiques durant les différents stades de différenciation des cellules T. D'où le but de ce travail, qui se porte sur l'étude du dynamisme épigénétique et transcriptionnel de ces 3 marques épigénétiques lors des différents stades du développement des thymocytes, qui est de savoir est-ce-que réunies ces marques contribueraient à la régulation de la transcription des gènes lors des différents stades du développement des cellules T ? Confirme-t-on un impact sur l'expression des gènes ? Par quel processus et avec quels éléments interagiraient-elles afin de permettre cette régulation ? Quels sont les éléments et les facteurs de transcription impliqués lors de cette régulation? Et si effectivement y a un réel dynamisme épigénétique et transcriptionnel dont les auteurs sont H3K4me1, H3K4me3

et H3K27ac pendant les stades DN et DP, quels sont les éléments génomiques et transcriptionnels permettant l'expression spécifique de certains gènes durant un stade, et d'autres durant un autre stade de développement des thymocytes ?

Grâce aux données acquises de la science, on pourrait répondre à ses questions. D'une part grâce à l'utilisation de la technologie CHIP combinée au séquençage à haut débit (SOLID) réalisés par le laboratoire Technical Advances for Genomics and Clinics (TAGC) Inserm U1090 – Aix-Marseille Université et d'une autre part aux outils bioinformatiques sans lesquels il est pratiquement improbable d'analyser les myriades de données générées grâce à la technologie CHIP-seq.

1-Données bibliographiques

1-1 Remodelage de la chromatine par les modifications épigénétiques des histones:

1-1-1 La chromatine

Si l'ADN, chez un individu humain est étalé sous la forme d'une molécule linéaire, il aurait une longueur de 5×10^{10} Km, l'équivalent de la distance entre la terre et le soleil multipliée par 100. L'ADN au sein d'une cellule aurait mesuré 2 mètres, s'il n'était pas compacté. Afin d'être restreint dans le noyau eucaryotique, le matériel génomique doit adopter une certaine structure, lui permettant d'interagir stratégiquement avec son environnement, selon les besoins cellulaires du moment. [LATCHMAN, 2010]

Cette structure est la chromatine. Le terme «chromatine», est proposé pour la première fois par W. Flemming vers 1880. Suite à cela, la compréhension, de la structure de la chromatine, ses niveaux de configuration et son rôle n'ont cessé d'évoluer, ainsi que plusieurs découvertes dont la double hélice d'ADN, la relation entre les modifications d'histones et la transcription, le nucléosome, l'épigénétique et le code histone (fig1). [Olins et al., 2003]

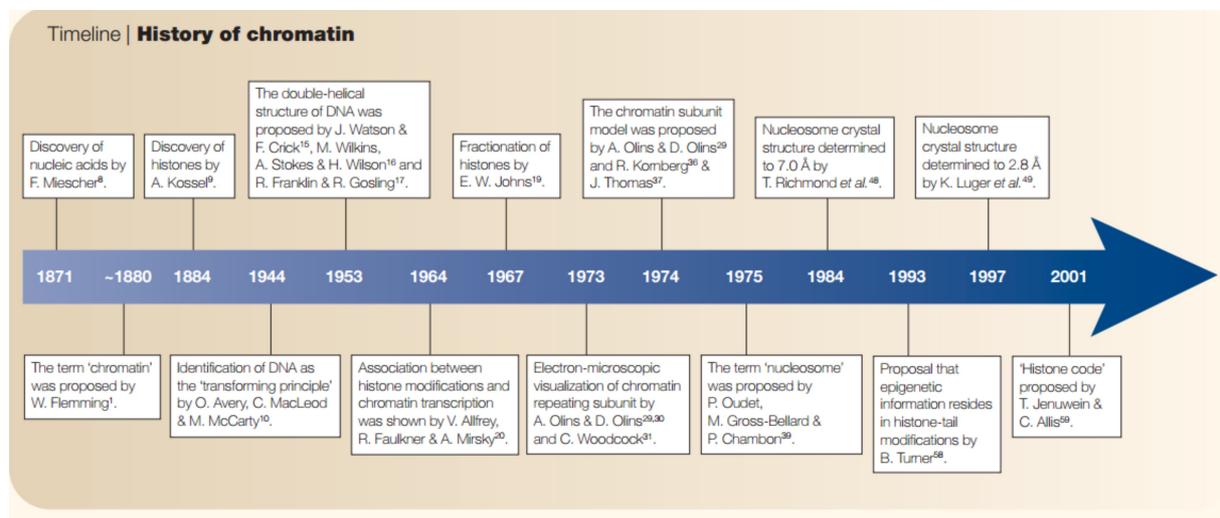


Fig1: Histoire de la chromatine. (Olins et al., 2003)

1-1-1-1 Structure et fonction de la chromatine:

Physiquement, la chromatine est un complexe ADN-protéines hautement régulé, stocké dans le noyau des cellules eucaryotes. Les protéines formant ce complexe sont de deux types : les protéines histones et les protéines non histones. [LATCHMAN, 2010]

1-1-1-1-1 Les histones et le nucléosome :

1-1-1-1-1-1 Les histones:

En 1884, A.Kossel découvre lors de ses recherches, les histones. Ce sont de petites protéines basiques, hautement conservées au cours de l'évolution, leur poids varie de 11kDa à 22kDa. Elles sont chargées positivement, ce qui leur permet de se lier à l'ADN dont la charge est négative. Cinq types d'histones ont été identifiés jusqu'à présent: H1, H2A, H2B, H3 et H4. Chacune des histones est composée de deux domaines: un domaine globulaire, médiant les interactions histone-histone ou histone-ADN et un domaine N-terminal se prolongeant au-delà de la particule nucléosomale, ciblé par des protéines régulatrices. L'association des histones contribue à la formation du nucléosome. [Olins et al., 2003], [Kouzarides,2007]

1-1-1-1-1-2 Le nucléosome:

Le nucléosome est l'unité répétitive fondamentale de la chromatine. Il est composé; d'une particule cœur constituée de 147 paires de bases d'ADN, enroulées autour d'un octamère protéique comprenant deux copies de chacune des histones H2A, H2B, H3 et H4; et d'une région internucléosomique spécifiée par la présence de l'histone H1. [Olins et al., 2003]. Dans le but de former la chromatine l'ADN se lie aux histones selon une certaine configuration.

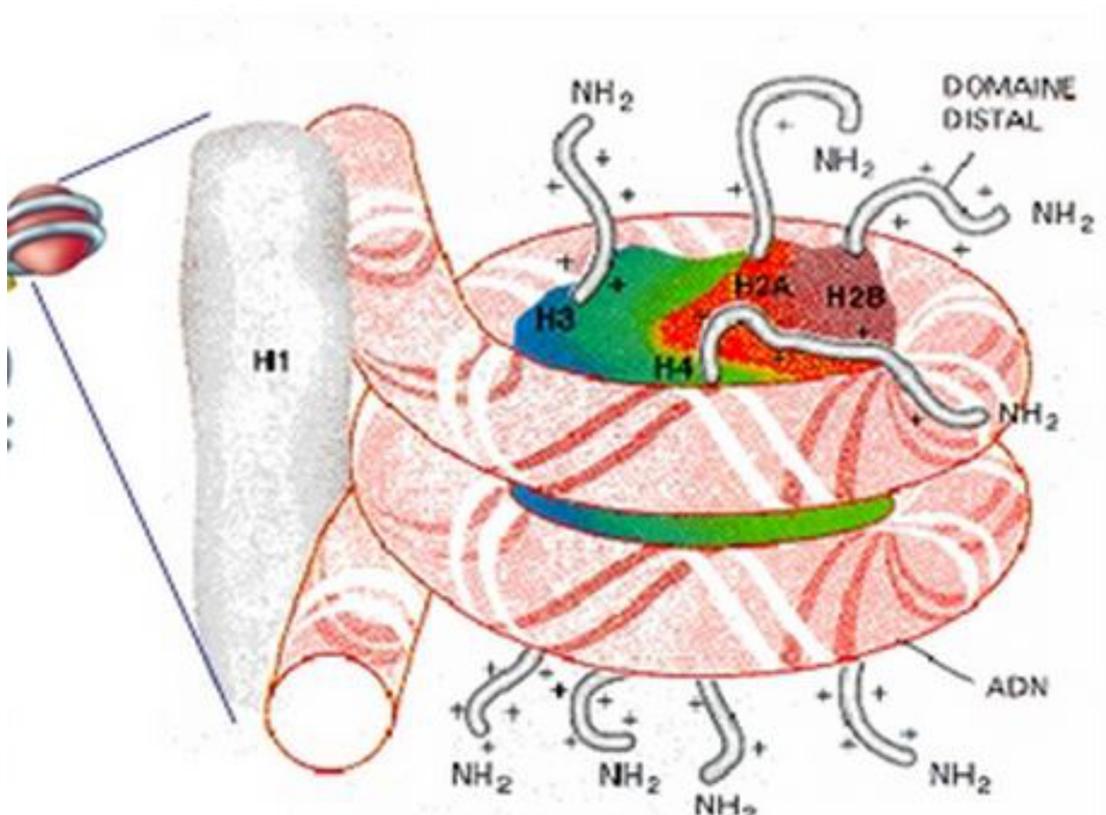
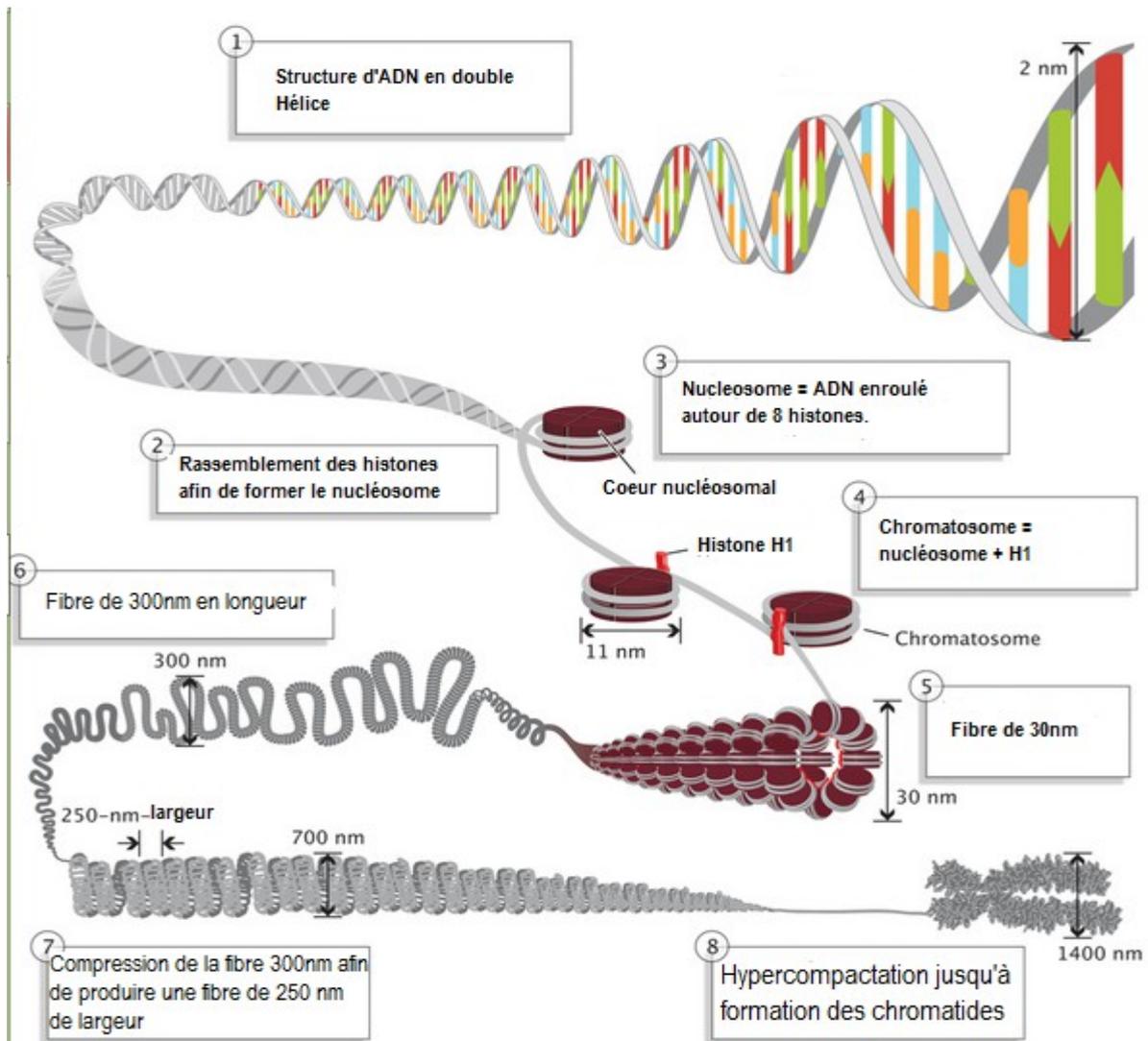


Fig2: Schéma illustrant l'emplacement des histones au sein du nucléosome. (Fortin, 2005.)

1-1-2 Les niveaux de compaction de la chromatine :

L'échafaudage de la chromatine par l'assemblage de l'ADN aux histones se déroule en plusieurs étapes et qui sont:

- 1-Sous sa forme la plus simple la chromatine se présente selon le modèle proposé par Watson et Crick c'est à dire en double hélice d'ADN.
- 2-Par la suite l'ADN s'associe aux histones (H3,H4,H2A,H2B) afin de former les nucléosomes, échafaudant ainsi un filament de 10 nm connu sous le nom de «collier de perles».
- 3- Afin de stabiliser cette structure, vient se lier à chaque nucléosome une H1 histone, formant ainsi le chromatosome et permettant l'enroulement de ce dernier en une fibre de 30 nm.
- 4-Finalement, plusieurs types de protéines sont recrutées permettant des repliements successifs de la fibre 30 nm, et ainsi l'hypercompaction de l'ADN jusqu'à la production des chromatides des chromosomes. [LATCHMAN, 2010]



F

ig 3: Les différentes étapes d'assemblage de la chromatine.
(Annunziato,A, 2008)

1-1-3 Rôle de la chromatine et modifications épigénétiques :

Par ce qui a précédé, la chromatine peut être définie comme étant l'état sous lequel l'ADN est compacté dans la cellule eucaryote. Selon l'état de configuration de la chromatine, l'expression des gènes est soit inhibée soit activée. En effet, généralement en se condensant, les nucléosomes constituent une barrière physique à l'accès de la machinerie transcriptionnelle à l'ADN et donc l'expression des gènes est réprimée. Inversement, plus les nucléosomes sont décondensés, plus le gène est activement transcrit. [LATCHMAN, 2010] Ce remodelage de la chromatine a lieu grâce à la régulation épigénétique. Cette régulation est réalisée grâce aux mécanismes épigénétiques pouvant soit favoriser la structure ouverte de la chromatine (l'euchromatine), ou bien la structure fermée de la chromatine (l'hétérochromatine). Trois mécanismes biochimiques importants sont désignés comme épigénétiques: la méthylation de l'ADN qui réprime l'expression génique, la liaison des protéines non-histones aux histones à l'ADN afin d'influencer la structure de la chromatine, et les modifications d'histones. [Bock et Lengauer, 2008]

Et ce sont ces dernières qui importent dans ce travail.

1-1-3-1 Modifications d'histones:

La queue N-terminal des histones, est sujette à plusieurs modifications covalentes, qui altèrent les charges électrostatiques des histones et changent leurs propriétés structurales et/ou modifient les interactions de leurs extrémités. Ces modifications sont effectuées par des complexes enzymatiques et ont un important impact sur l'état de compaction de la chromatine et par conséquent sur la régulation du niveau d'expression des gènes. Différentes classes de modifications d'histones sont identifiées dont les plus étudiées sont la méthylation, l'acétylation, la phosphorylation et l'ubiquitylation [LATCHMAN, 2010].

Modifications de la chromatine	Résidus modifiés	Fonction de la modification
Acétylation	K-ac	Transcription, réparation, réplication, condensation
Méthylation (Lysine)	K-me1, K-me2, Kme3	Transcription, réparation
Méthylation (Argénine)	R-me1, R-me2a, R-me2s	Transcription
Phosphorylation	S-ph, T-ph	Transcription, réparation, condensation
Ubiquitylation	K-ub	Transcription, réparation
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Déimination	R > Cit	Transcription
Proline isomérisation	P-cis > P-trans	Transcription

Tableau I: Les différentes classes de modifications identifiées sur les histones. (Kouzarides, 2007)

1-1-3-1-1 La méthylation et l'acétylation des histones:

-**La méthylation** consiste à l'ajout d'un à trois groupements méthyles (mono, di, ou triméthylation) sur l'acide aminé Lysine ou Arginine à la queue N-terminale. Les enzymes responsables de cette réaction sont les histones méthyltransférases (HMT). Selon le résidu méthylé ainsi que le nombre de groupements méthyles ajoutés par les HMT, l'activité transcriptionnelle est activée ou réprimée. [Kouzarides, 2007], [LATCHMAN, 2010]

-**L'acétylation** consiste à l'ajout de groupement acétyle à certains résidus Lysine (K). Les histones acétyl transférases (HAT) sont les enzymes responsables de cette réaction. L'acétylation favorise la structure ouverte de la chromatine et donc est associée à l'activation de la transcription. [Kouzarides, 2007], [LATCHMAN, 2010]

Les modifications des histones s'effectuent grâce à des complexes enzymatiques (protéines non histones) et sur des acides aminés spécifiques de la queue N-terminal des histones cœur, H3, H4, H2A, H2B. Ces modifications constituent des marques épigénétiques de la chromatine (Voir fig 4). A titre d'exemple la monométhylation du résidu Lysine (k) numéro 4 sur la queue N-terminal de l'Histone 3 et est notée H3K4me.



Fig4: Exemple de modifications épigénétiques effectuées par différents groupes d'enzymes sur les résidus amine de la queue N-terminal de l'histone 3 Me (méthylation), P(phosphorylation), Ac (acétylation). (Kouzarides, 2007).

1-1-3-2 Le code Histone :

L'ensemble de ces modifications et leur implication dans la modulation de la chromatine, a donné naissance à l'hypothèse du « code histone », proposée en 2001 par T.Jenuwein and all. Cette hypothèse stipule que les modifications des histones peuvent agir en synergie selon un certain programme épigénétique permettant la régulation de l'activité des gènes. [Olins et al., 2003], [LATCHMAN, 2010]

1-2 Mécanisme de la transcription chez les eucaryotes :

La transcription est le processus biologique, pendant lequel est déterminé, quels gènes seront ou non transcrits, et cela en réponse à des stimuli spécifiques comme les modifications épigénétiques de la chromatine. Ces marques épigénétiques pourraient à titre d'exemple, induire la transcription de certains gènes spécifiques du stade DN durant la différenciation des

thymocytes et réprimer leur expression avant ou après ce stade, d'où la notion de tissu spécificité génique. [RAVEN *et al.*, 2008], [LATCHMAN, 2010]

1-2-1 Les séquences consensus d'un gène et le mécanisme de transcription :

Il existe plusieurs séquences consensus chez les eucaryotes dont la boîte TATA et la boîte INR qui sont présentes au niveau d'un grand nombre de gènes. La boîte TATA (située entre -30 et -25 en amont du TSS) et la boîte INR (à partir de +1) constituent le promoteur minimum. Parmi les régions consensus y a également des séquences activatrices et amplificatrices situées jusqu'à -300 en amont du site d'initiation, parmi elles on trouve la boîte GC et la boîte CAAT. Les séquences activatrices varient selon le promoteur existant. Ces régions constituent des sites de liaison pour les facteurs de transcription, permettant ainsi la modulation de l'activité du promoteur minimum auquel se lie l'ARN pol II via les facteurs généraux de la transcription. Le mécanisme de la transcription chez les eucaryotes est régulé par les régions cis-régulatrices de type enhanceurs, silenciers et insulateurs et/ou par des promoteurs alternatifs. [LATCHMAN, 2010]

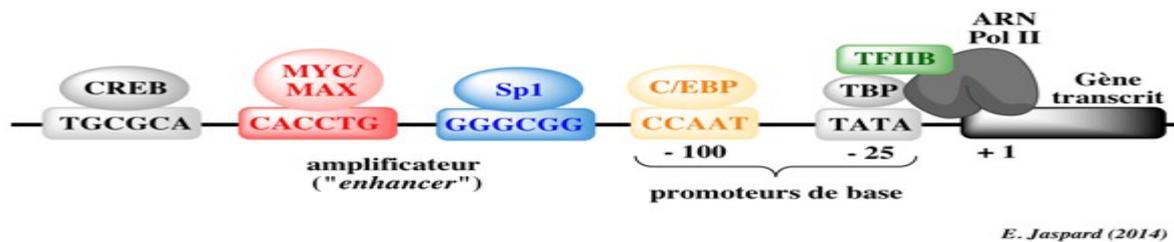


Fig 5: Eléments d'un gène eucaryote.
(E.Jaspard, 2014)

1-2-2 Les éléments régulateurs de la transcription selon LATCHMAN, Davis, 2010:

a-Les promoteur alternatifs: c'est la présence de plusieurs promoteurs pour un seul transcrit d'un gène.

b-Les éléments cis-régulateurs :

- **Silenciers ou represseurs:** C'est des séquences agissant à distance, auxquelles se lient des protéines inhibitrices de l'activité transcriptionnelle.

- **Insulators ou isolants:** C'est des séquences permettant d'isoler certaines régions génomiques.

- **Enhancers ou amplificateurs:** C'est des séquences régulatrices qui recrutent des protéines afin d'amplifier la transcription des gènes, en agissant à distance sur les promoteurs et cela dans les deux directions en fonction du Tss. Les enhanceurs peuvent se situer, en amont du site d'initiation (TSS), en aval, ou bien à l'intérieur de la région à transcrire.

1-2-3 La régulation de la machinerie transcriptionnelle:

Afin d'activer le mécanisme de la transcription, les facteurs généraux de la transcription s'associent à l'ARN pol II afin d'initier la transcription et requièrent les coactivateurs. Les activateurs s'associent aux enhancers, en réponse à certains signaux cellulaires afin d'amplifier l'activité transcriptionnelle.

Les coactivateurs s'associent, avec les activateurs et/ou les répresseurs (à une extrémité), et avec les facteurs généraux de la transcription (à l'autre extrémité), afin de transmettre le signal d'activation ou de répression de la transcription à l'ARN pol II. Les répresseurs se lient aux silencers, régulés par des signaux cellulaires afin de réduire la transcription des gènes régulés en interférant avec les activateurs. [RAVEN et al., 2008]

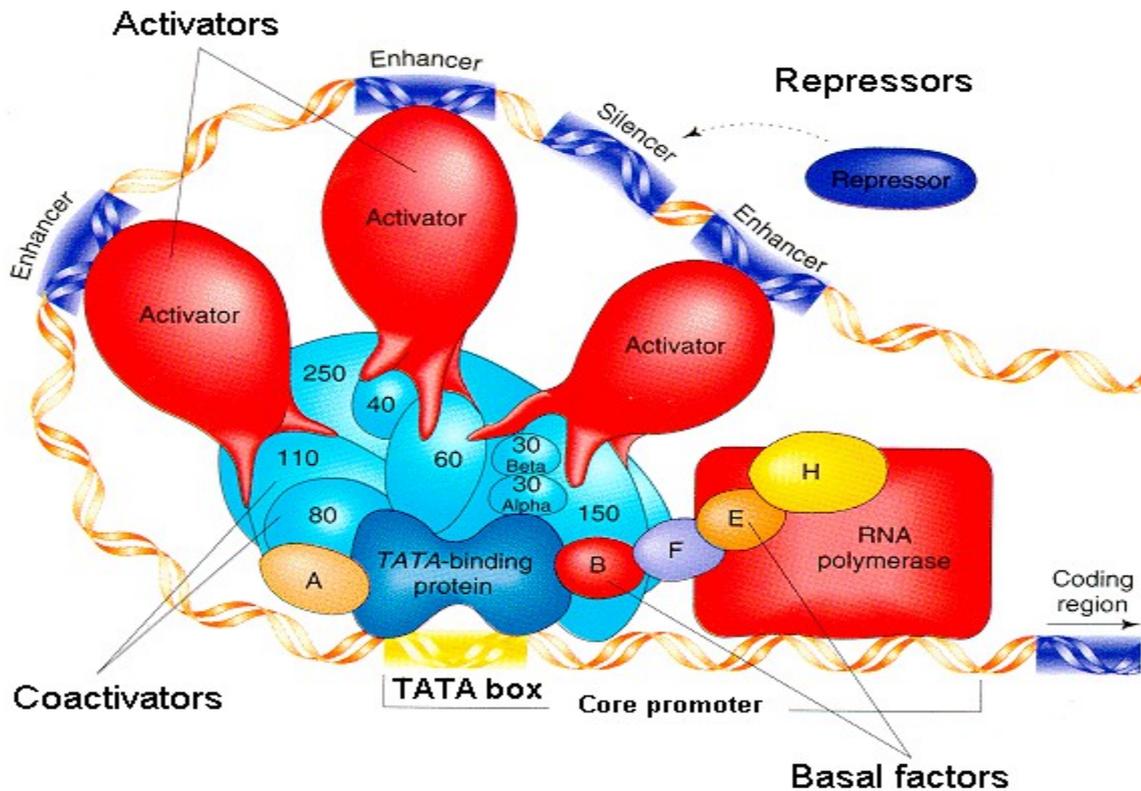


Fig6: Les différents éléments intervenant dans la régulation de la transcription. (TJIAN, 2008)

1-3 Le développement et la maturation des cellules T :

Le développement des cellules T débute par l'arrivée d'un petit nombre de cellules progénitrices T (précurseurs lymphoïdes), migrant via le système sanguin, de la moelle osseuse vers le thymus et sont appelées une fois dans ce dernier, « thymocytes ». La maturation des thymocytes implique : l'expression de divers marqueurs membranaires et d'événements intracellulaires spécifiques, caractérisant les différents stades de différenciation (DN à SP); des réarrangements des gènes du récepteur des cellules T (TCR), permettant la reconnaissance d'une grande variété des antigènes étrangers ; ainsi que des sélections thymiques, aboutissant à des cellules T restreintes au CMH du soi et auto-tolérantes. [HAYDAY *et al.*, 2007]

1-3-1 Les stades de différenciation des thymocytes du stade DN à DP :

Les précurseurs lymphoïdes, pénétrant au thymus et entamant le processus de différenciation, passent par une série de stades, où certaines molécules membranaires sont détectables. Tels que le complexe TCR-CD3, CD4 ou CD8. Les thymocytes n'exprimant pas le CD4 et/ou CD8 sont appelées cellules T double négatives (T DN). Les Thymocytes DN peuvent être cataloguées en 4 sous-populations (DN1 à DN4), chacune caractérisée par, la présence ou l'absence de molécules spécifiques à leur surface cellulaire, comme c-Kit (récepteur pour le facteur de croissance des cellules souches), CD44 (une molécule d'adhésion), CD25 (une chaîne alpha du récepteur de l'IL2). Le phénotype d'expression de surface pour chaque stade ou sous-population, se présente comme suit : DN1, c-kit⁺CD44⁺CD25⁻; DN2, c-kit⁺CD44⁺CD25⁺; DN3, c-kit⁺CD44⁻CD25⁺; et DN4, c-kit⁻CD44⁻CD25⁻. [HAYDAY *et al.*, 2007], [GERMAIN, 2002]

Les Thymocytes progressant vers le stade DN4, diminuent l'expression du CD25 et présentent à leur surface cellulaire les deux molécules CD4 et CD8. A ce stade, les thymocytes sont appelées double positives (DP), ou cellules CD4⁺8⁺. Parallèlement à l'évolution phénotypique, les thymocytes effectuent une série de réarrangements de locus génétiques codant pour les gènes du TCR. [GERMAIN, 2002]

1-3-2 Réarrangements du TCR :

Lors de leur maturation, les thymocytes expriment des protéines, tel que RAG, nécessaires aux réarrangements des gènes codant pour les chaînes du TCR. RAG est un complexe enzymatique de la recombinaison, qui reconnaît les SSRs (séquences signal de la recombinaison). Les SSRs sont situées de part et d'autres des sites subissant les réarrangements de l'ADN de type VDJ: V(variable), D(diversité), J(jonction). [KINDTET *et al.*, 2006]

Au stade DN2, débutent les réarrangements des gènes codant les chaînes TCR γ , TCR δ , TCR β (le locus TCR α est réprimé), et progressent jusqu'au stade DN3. Lors de cette transition du stade DN2 vers DN3 ; une minorité (< 5%) des thymocytes, sont destinées à devenir des cellules T $\gamma\delta$ matures sans grand changement de leur phénotype de surface ; et une majorité (~95%) à donner des cellules T $\alpha\beta$ et nécessitant pour leur maturation, de progresser au delà

du stade DN2. En acquérant le phénotype DN3, ces dernières arrêtent momentanément de proliférer et des produits issus des réarrangements du TCR β sont détectés dans le cytoplasme de ces cellules. Après que le réarrangement de la chaîne TCR β est effectué, les cellules DN3 progressent rapidement vers le stade DN4. Les chaînes β nouvellement synthétisées, s'associent à la chaîne pré-T α , une protéine de 33kDa et les molécules CD3, dans le but de former le complexe pré-TCR. La formation de ce complexe déclenche une cascade de transduction de signal, aboutissant à l'activation des réarrangements de la chaîne alpha du TCR et à la progression du développement des thymocytes vers l'état double positif CD4⁺8⁺. Les cellules double positive prolifèrent et à partir du moment de la différenciation où une cellule T a réarrangé la chaîne alpha du TCR et commence à exprimer un complexe TCR/CD3 complet, les cellules DP n'expriment plus les marqueurs précoces, tels que le CD44 et le CD25. L'expression du complexe TCR/CD3 permet aux thymocytes de subir les phases de sélection positive et de sélection négative. [Germain, 2002], [KINDTET et al., 2006]

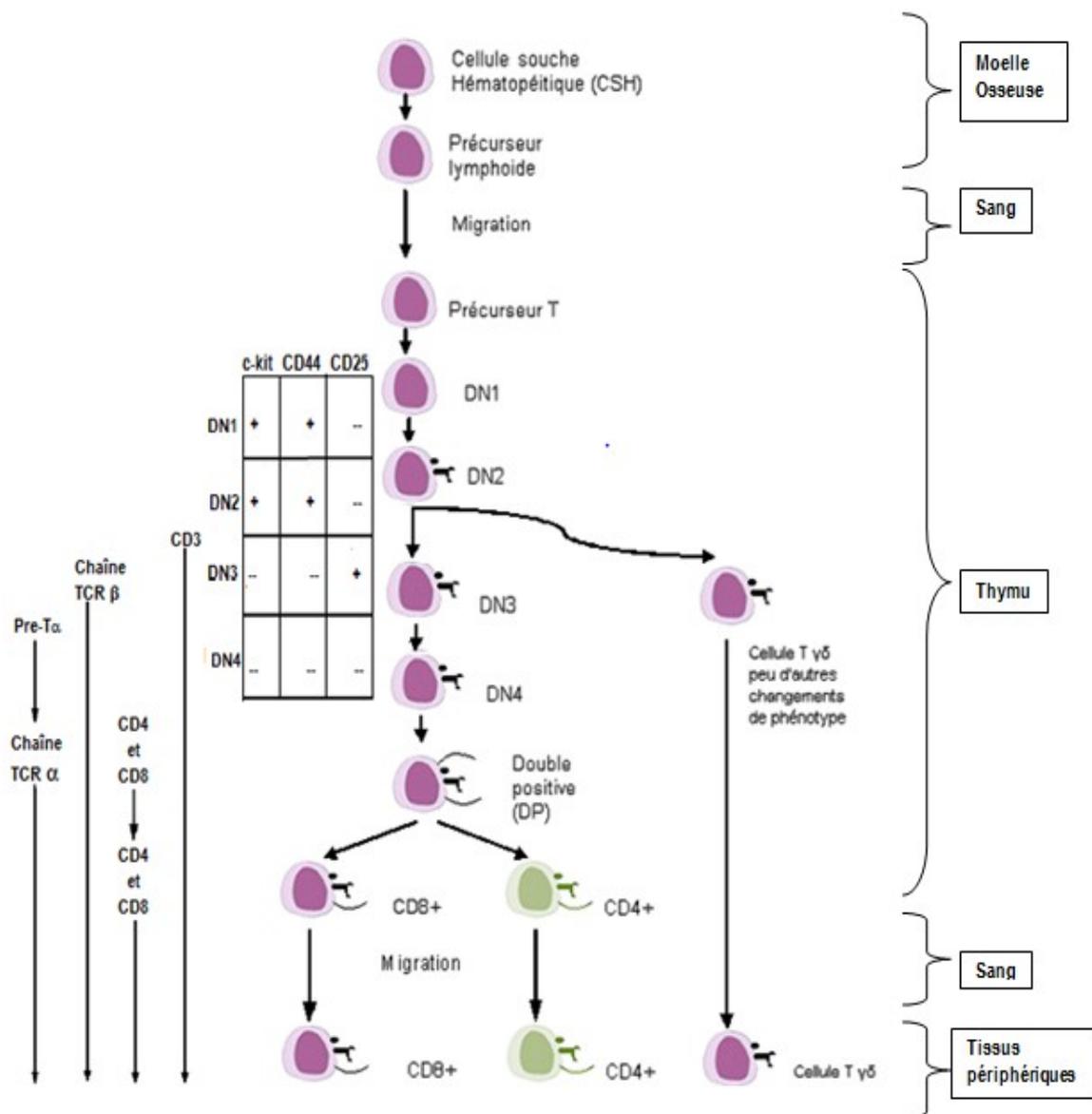


Fig7: Vue d'ensemble du développement des cellules T. (KINDTET et al., 2006)

1-3-3 Processus de sélections :

Il est estimé qu'environ 98% de l'ensemble des thymocytes n'arrivent pas à maturation, ils sont éliminés par apoptose dans le thymus, soit à cause de la formation de réarrangements non productifs des gènes du TCR, soit ils sont éliminés lors du processus de sélection positive et négative. [KINDTET et *al.*, 2006]

1-3-3-1 La sélection positive : Les thymocytes DP ($CD4^+8^+$) se trouvant dans le cortex thymique, sont mis en contact avec des antigènes présentés par des molécules du CMH du Soi. Les cellules DP, possédant des TCR incapables de se lier aux molécules du CMH du Soi, sont éliminées. Les cellules double positives survivantes, perdent l'un des deux marqueurs ($CD4, CD8$). Les $CD4^+8^+$ dont le TCR se lie avec les molécules du CMH de classe I préservent seulement le CD8 et deviennent des cellules simple positive ($SP\ CD8^+$), celle ayant un TCR interagissant avec les molécules du CMH de classe II gardent le CD4 mais perdent le CD8 et deviennent des cellules simple positives ($SP\ CD4^+$). Ainsi ce processus assure la restriction au CMH. [GERMAIN, 2002], [KINDTET et *al.*, 2006]

1-3-3-2 La sélection négative : Les cellules simple positives exprimant le $TCR\alpha\beta/CD3$ et survivant à la sélection positive, subissent une autre sélection, durant laquelle les thymocytes exprimant un récepteur ; ayant une haute affinité pour des molécules du CMH du Soi seules, ou des molécules du CMH du Soi associées à un autoantigène (peptide du soi) ; sont éliminées, assurant ainsi la tolérance aux molécules du Soi et évitant les maladies auto-immunes. Les lymphocytes survivants à la sélection négative migrent depuis le cortex vers la médulla, où elles passent dans la circulation sanguine. [GERMAIN, 2002], [KINDTET et *al.*, 2006].

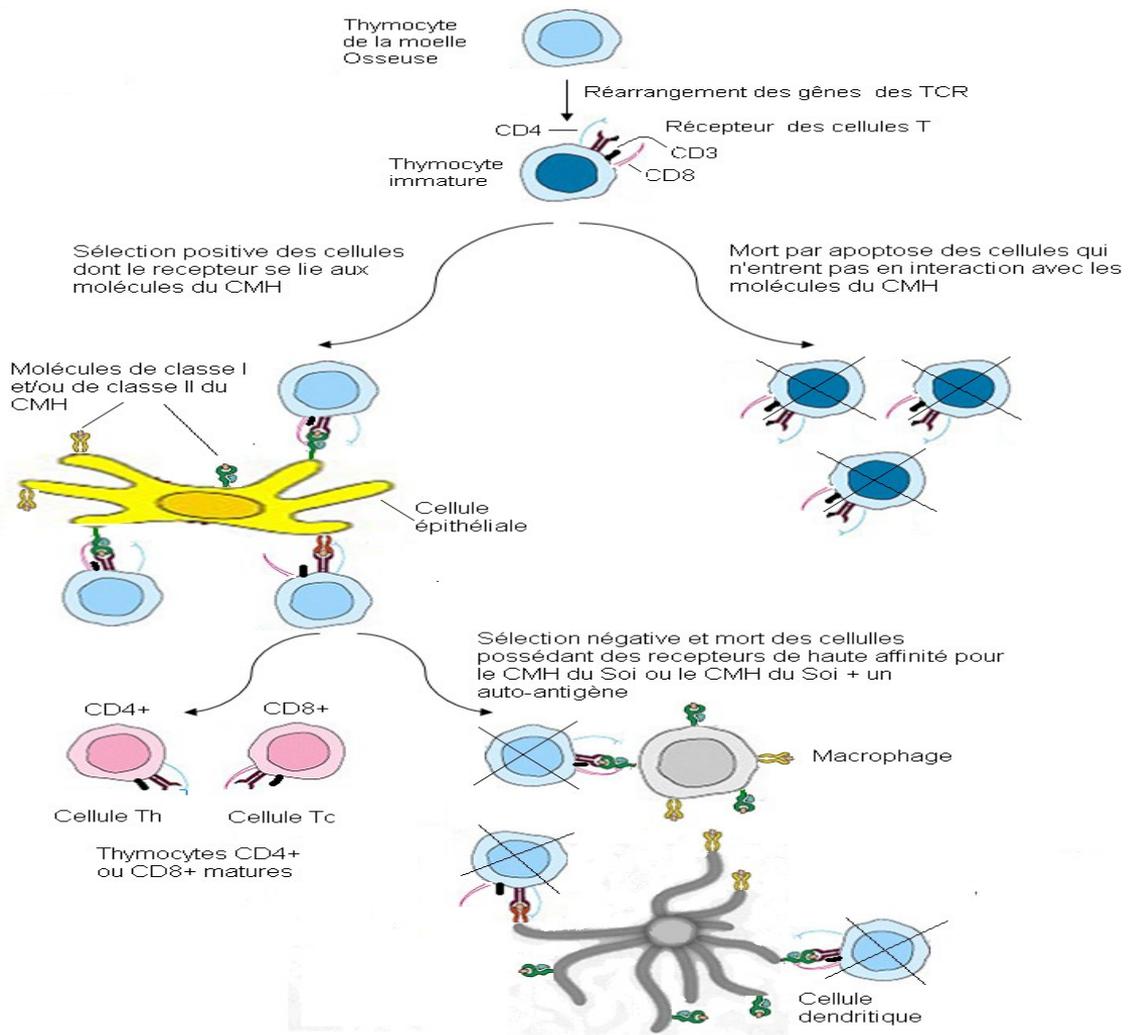


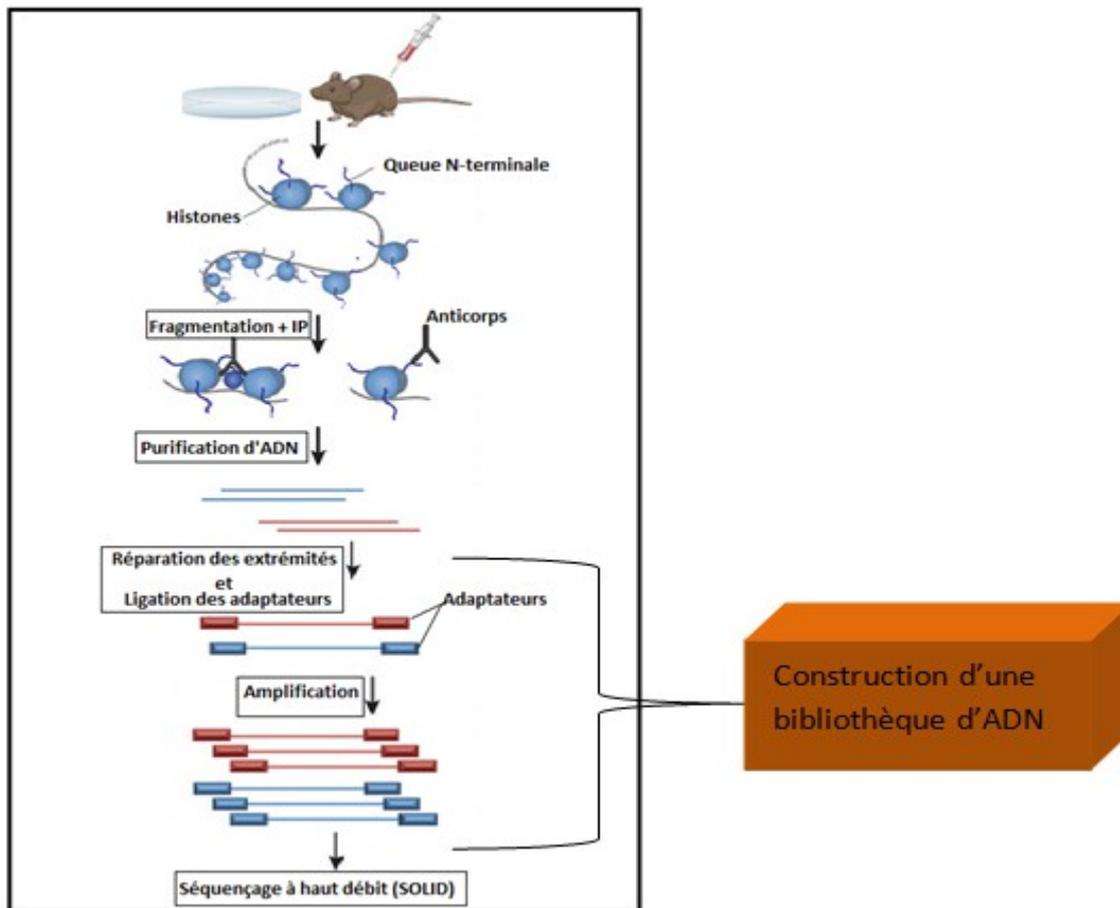
Fig8: Sélection négative et positive des thymocytes.

(KINDTET et *al.*, 2006)

2-Matériel et méthodes

2-1 Chromatin-Immuno-Precepitation- Sequencing (ChIP-seq) :

L'immunoprécipitation de la chromatine (ChIP), suivie, de séquençage à haut débit, est une puissante méthode, qui permet de cartographier la distribution des sites de liaisons, de la protéine d'intérêt sur l'ADN et les modifications d'histones sur n'importe quel génome.



Fig

9: Etapes de la réalisation de ChIP-seq et construction d'une Bibliothèque d'ADN. (KIDDER *et al.*, 2011)

2-1-1 Les étapes de la réalisation de CHIP-seq :

Après l'isolation du nombre adéquat des cellules DN et DP du thymus murin, un **crosslinking** est effectué afin de fixer les protéines sur l'ADN. Ensuite, **la chromatine est fragmentée** en petits bouts, la fragmentation est réalisée par l'enzyme MNase qui est adéquate à l'étude des modifications d'histones, cette enzyme clive les bases d'oligonucléotides au niveau 3'-phosphate. La taille des fragments générés est vérifiée par un **Bioanalyser** et leur quantité par le NanoDrope. L'étape suivante la vérification de la qualité des fragments est **l'immunoprécipitation (IP)** : lors de cette étape des anticorps attachés à des billes, sont additionnés au tampon. [kidder *et al.*, 2011]. A chaque ChIP, un type d'anticorps est utilisée pour la modification d'histone correspondante. Les anticorps dirigés contre les marques épigénétiques H3K4me1, H3K27ac et H3K4me3 sont dans l'ordre anti-H3K4me1, anti-H3K27ac et anti-H3K4me3. Suite à cela, le complexe ADN-protéine d'intérêt-anticorps précipite au fond de la solution. Ensuite, l'ADN enrichi (précipité) subit une étape de

purification. A ce stade CHIP est terminée mais il faut construire une bibliothèque d'ADN adaptée au séquençage à haut débit SOLID.

2-1-2 Construction d'une Bibliothèque d'ADN :

Suite à la purification d'ADN et afin de pouvoir séquencer, par SOLID, l'ADN enrichi par les marques épigénétiques (ADN précipité), il est nécessaire de construire une bibliothèque adaptée. La construction de cette bibliothèque, se définit par l'addition d'adaptateurs indispensables au fonctionnement du séquenceur et cela s'effectue en plusieurs étapes, pouvant être résumées comme suit selon **Mike S. Son et al., 2011**:

2-1-2-1 Réparation des extrémités des fragments d'ADN (End-repair): Les segments d'ADN générés après la fragmentation, sont clivés aléatoirement et ont donc des extrémités 5' et 3' diverses dans toutes les combinaisons possibles. Ces extrémités sont réparées en bouts francs par l'addition de nucléotides, en utilisant un mélange enzymatique contenant entre autre une ADN polymérase. Suite à cela l'extrémité 3' est modifiée.

2-1-2-2 Modification de l'extrémité 3' : Cette étape consiste à l'ajout d'une Adénine (A), aux extrémités 3' des fragments réparés précédemment. Cette étape est nécessaire pour la ligature des adaptateurs.

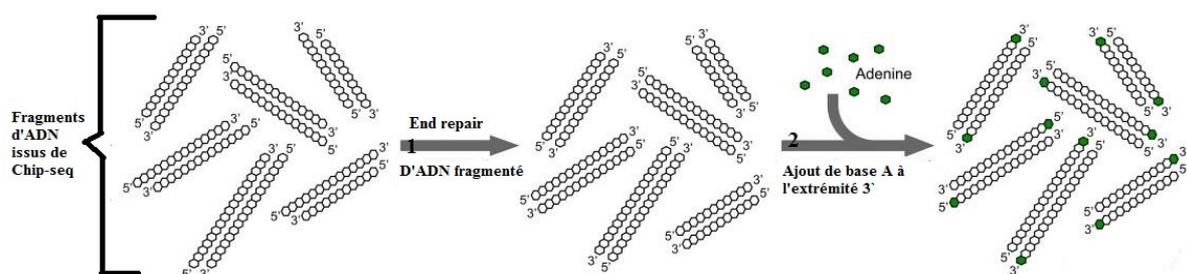


Fig10: Illustration des étapes End-repair et modification de l'extrémité 3'.
(Son et al., 2011)

2-1-2-3 La ligature des adaptateurs: La ligature des adaptateurs aux séquences d'ADN modifiées est non seulement nécessaire pour l'étape finale de la construction de la bibliothèque mais aussi indispensable pour le séquençage par SOLID. Durant cette étape chaque fragment d'ADN modifié est ligaturé à un adaptateur dans chaque extrémité. L'un des deux adaptateurs ajoutés est spécifique pour chaque librairie et est appelé code barre (barcode). A la fin les fragments résultants, sont purifiés afin d'éliminer les complexes enzymatiques et les résidus accumulés lors des précédentes étapes, puis quantifiés.

2-1-2-4 Enrichissement des fragments ADN modifiés-adaptateurs par

PCR (Amplification): L'étape finale de la construction de la bibliothèque est d'enrichir les fragments d'ADN modifiés-adaptateur, afin de permettre plus de fragments pour le séquençage. Cette étape produit une génération de clusters (plusieurs clones de fragments d'ADN) grâce à l'amplification ou l'enrichissement par PCR. Pour cela des primers sont ajoutés au mélange réactionnel dont le primer index qui est spécifique pour chaque

bibliothèque. Le produit PCR est par la suit purifié, quantifié et réservé à -20°C et prêt à être chargé dans la plateforme du séquençage SOLID.

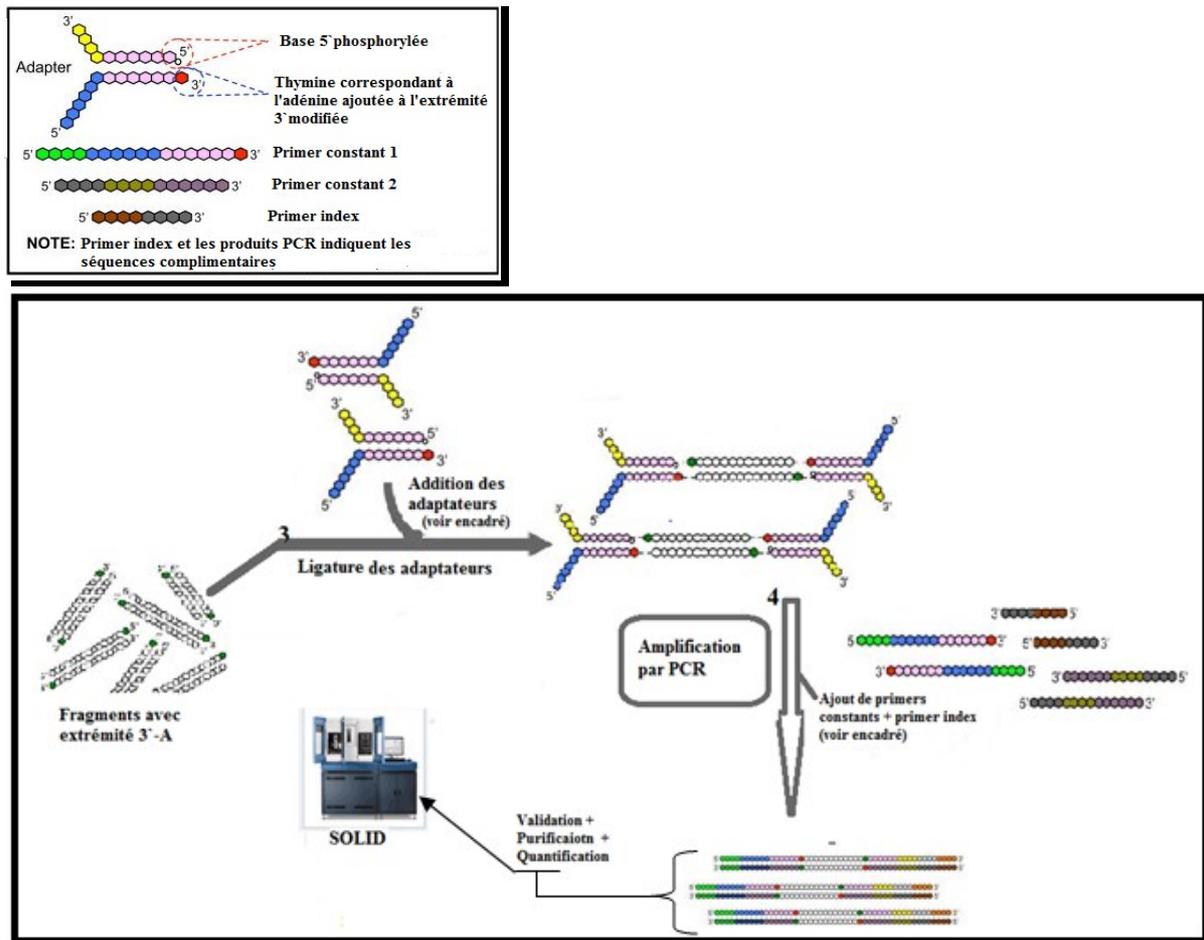


Fig11: Etapes de ligature des adaptateurs et enrichissement des fragments par PCR. (Un adapté depuis Mike S. Son et al., 2011)

2-1-3 Sequencing by Oligonucleotide Ligation and Detection (SOLID):

SOLID appartient à la deuxième génération de séquenceurs à haut débit, le principe de cette technologie est basé sur la PCR en émulsion (emPCR sur billes) et un système de ligature d'oligonucléotides. Le séquençage par SOLID s'effectue en plusieurs étapes pouvant être résumées comme suit :

2-1-3-1 La construction d'une librairie, durant laquelle des adaptateurs sont ligaturés aux fragments d'ADN issus de ChIP, expliqué en haut selon le protocole de **Mike S. Son et al., 2011**.

2-1-3-2La PCR par émulsion: Lors de cette étape chaque fragment issu de la librairie est amplifié dans une gouttelette d'eau séparée des autres gouttelettes par une solution d'huile. Chaque gouttelette est appelée microréacteur, et contient une microbille, des amorces, produits réactionnels de la PCR et un seul fragment issu de la librairie (figA,B).

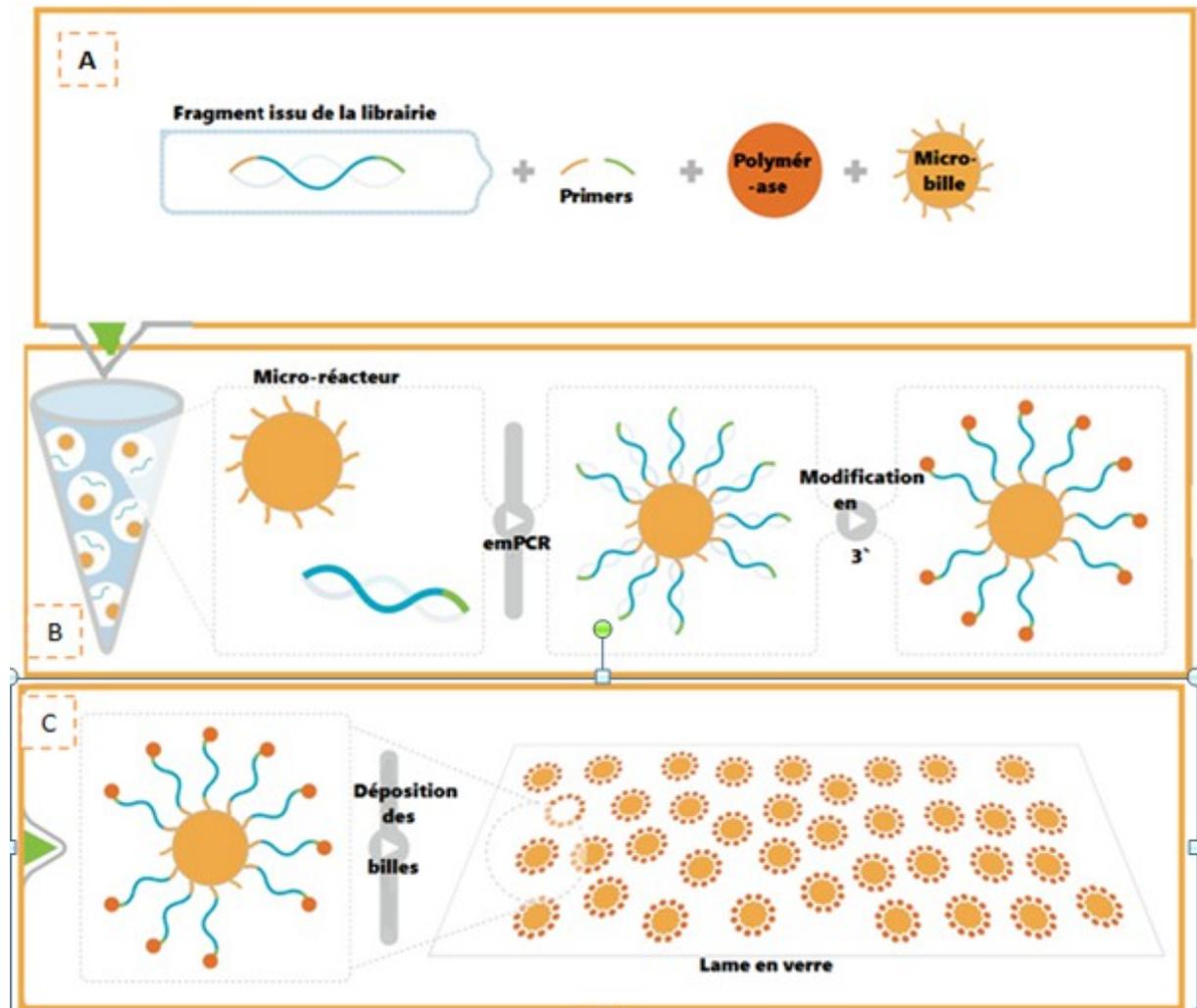


Fig12: Réalisation d'une emPCR pour le séquençage par SOLID.
(Appliedbiosystems, 2013)

-Après le lancement de la PCR, la surface de chaque microbille va être parsemée d'amorces auxquelles vont se lier les clones du fragment d'ADN ajouté via l'adaptateur. Ainsi dans chaque microréacteur, un seul fragment est amplifié. L'extrémité 3' de chaque fragment d'ADN couplé à la microbille est modifiée (Fig B). Chaque microbille est par la suite déposée via l'extrémité 3' modifiée à un puits dans une lame en verre (Fig C). [APPLIED BIOSYSTEMS, 2013]

3- Grâce à un agent chimique, un clivage est effectué au niveau du site de clivage situé entre les bases 5 et 6.

4- Des oligonucléotides sont rajoutés nouveau et les 3 premières étapes sont répétées: au 2^{ème} cycle ce sont les bases numéro 6 et 7 qui sont interrogées, puis au 3^{ème} , les bases 11 et 12 et ainsi de suite durant n cycle, jusqu'à l'obtention de la longueur de lecture voulue, ici jusqu'à 35 bases.

5- A ce stade, le brin synthétisé est dissocié ainsi que la précédente amorce (n), et un 2^{ème} tour de ligature est lancé avec une autre amorce de séquençage universelle (n-1).

6- Répétition des précédentes étapes (1 à 4), avec la nouvelle amorce n-1.

7-Puis les étapes 5 à 6 sont répétées pour chacune des ajoutées successivement, en position n-2, n-3 puis n-4.

2-2 Outils bioinformatiques :

2-2-1 Cocas: **ChIP-on-ChIP Analysis Suite** est un outil de recherche de pics correspondant aux régions génomiques enrichies, développé par l'équipe PF. L'algorithme cherche des pics à partir d'un fichier GFF. La détection des pics avec ce logiciel est variable selon les paramètres choisis. Pour chaque pic détecté, un score lui est attribué par le calcul de la surface du pic. Trois paramètres manipulables permettent de déterminer les séquences génomiques enrichies: Peak threshold (Seuil du pic), Extend threshold (seuil d'extension), Distance threshold (surface du pic). CoCas est téléchargeable depuis <http://www.ciml.univ-mrs.fr/software/ferrier.htm> .[Benoukraf et al.,2009]

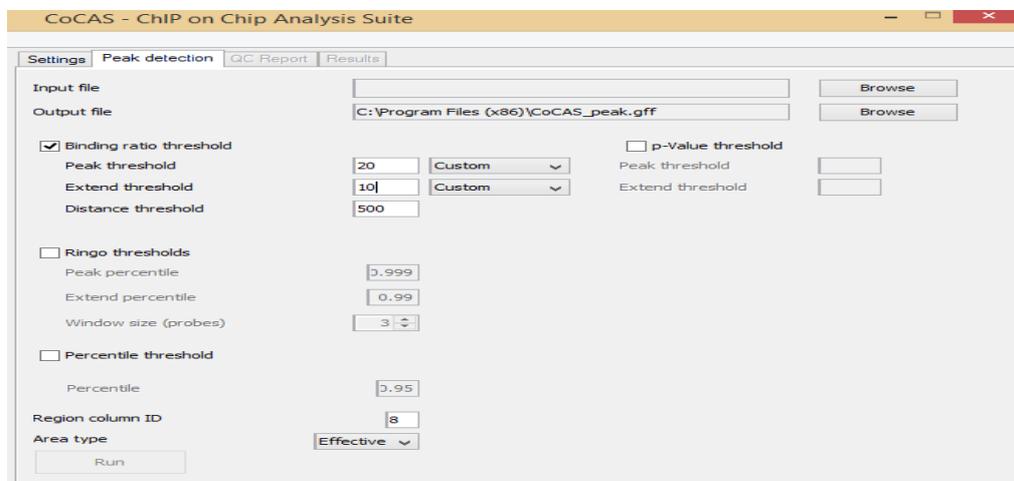
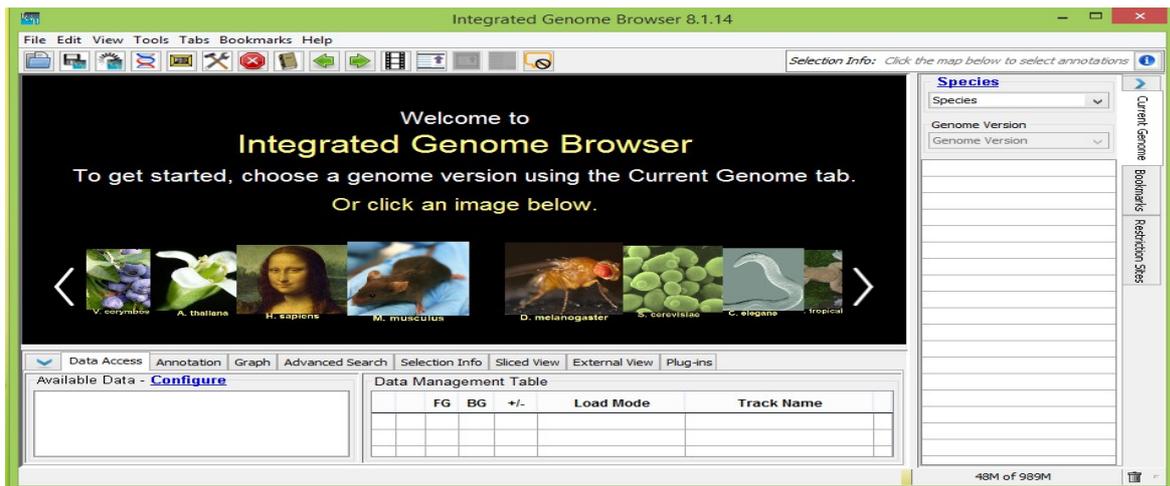


Fig 14: Interface de Cocas.

2-2-2 IGB: **Integrated Genome Browser** est un outil bioinformatique open source permettant la visualisation et l'exploration des données génomiques issues du séquençage à haut débit, notamment les données de ChIP-Seq. IGB nécessite le JAVA pour fonctionner et prend en charge l'affichage panoramique ainsi que l'option zoom à travers du génome d'intérêt. Il prend en charge des fichiers sous plusieurs formats dont BAM, BED, GFF et WIG. Téléchargeable depuis le site <http://bioviz.org/igb/> .[NICOL et al., 2009]



Fig

15: interface d'IGB.

2-2-3 Galaxy: est une plateforme open source, basée sur le web. Elle offre une interface graphique à de nombreux logiciels permettant l'analyse de datas génomiques comme l'alignement de séquences, la conversion de fichiers d'un format à un autre, l'intersection entre des datas, l'analyse ChIP-seq, etc. Les analyses et calculs effectués peuvent être conservés sur un espace de stockage sur le serveur de Galaxy et cela pour chaque utilisateur. Disponible sur <https://usegalaxy.org/>.

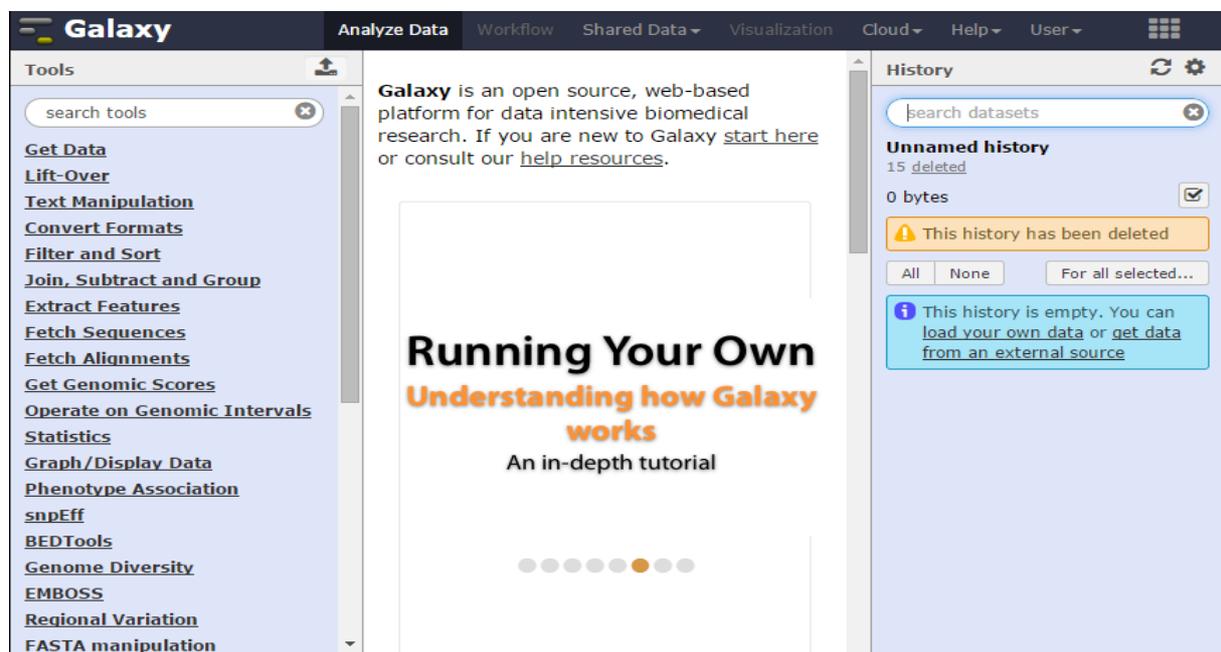


Fig16: interface de Galaxy.

2-2-4 RSAT: Regulatory Sequence Analysis Tools comprend une large collection d'outils permettant la détection des régions cis régulatrices du génome. Il permet ainsi entre autres de prédire les motifs (pic-motif) à partir des séquences génomiques issues de ChIP-seq. RSAT combine plusieurs algorithmes de découverte de motifs afin d'extraire les motifs qui sont significatifs puis ils sont comparés avec des bases de données de motifs connues. Les fichiers

chargés dans RSAT doivent être sous format FASTA. [MORGANE et al.,2011],[MORGANE et al.,2012]. Lien vers le site : <http://www.rsat.eu/>.

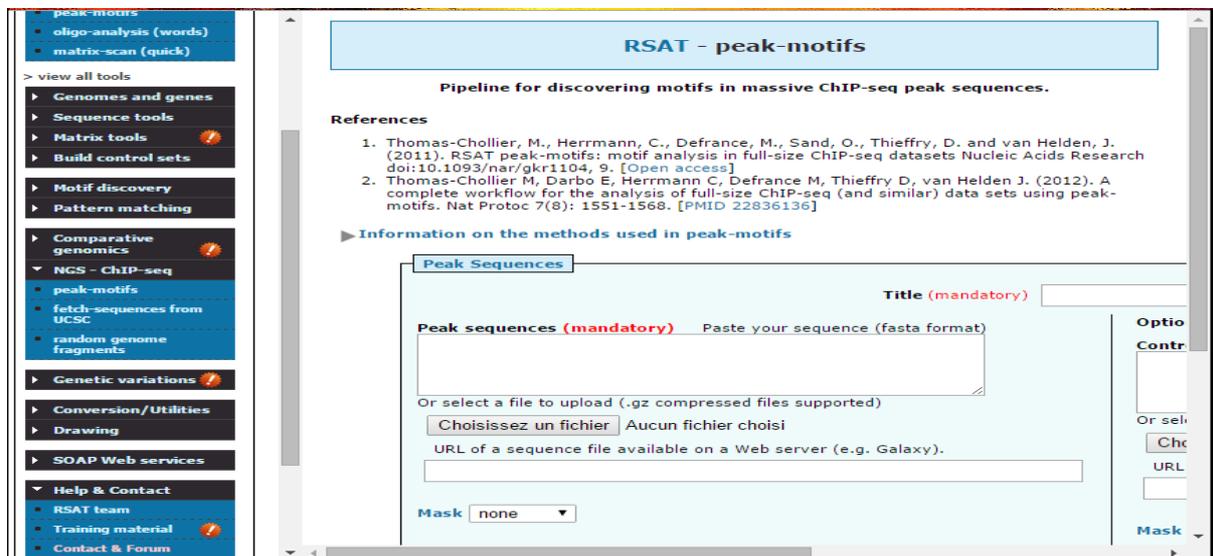


Fig17: Interface de RSAT.

2-2-5 GREAT: Genomic Regions Enrichment of Annotations Tool, est comme son nom l'indique un outil d'annotation des régions génomiques enrichies. Il permet d'analyser le rôle fonctionnel des régions cis-régulatrices, à partir de la fonction des gènes les plus proches de ces régions. La version utilisée est 2.0.2, génome:Mouse, NCBI build 37 ([UCSC mm9, Jul/2007](http://ucsc.genome.ucsf.edu/)). Great prend en charge les fichiers sous format BED [MCLEAN et al., 2010] et est disponible sur le site: <http://bejerano.stanford.edu/great/public-2.0.2/html/>

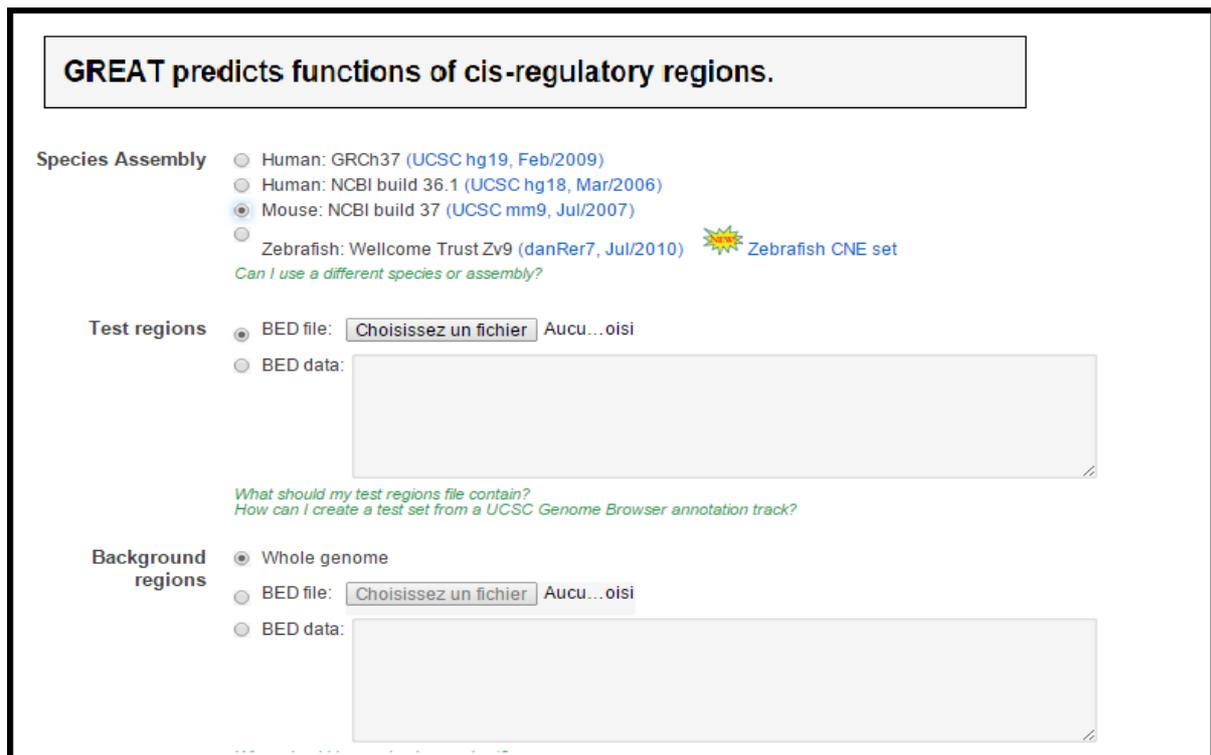


Fig18: Interface de GREAT.

2-2-6 Le langage R :

R est un langage de programmation statistique et graphique. Il est disponible librement pour Windows, MacOS et Linux. Les fonctionnalités de R sont extensibles et basées sur les contributions d'une communauté de développeurs. R a une place grandissante en biologie, de nombreux modules spécifiques sont développés pour répondre aux besoins de la biologie à grande échelle. Téléchargeable depuis le site : <http://www.r-project.org/>.

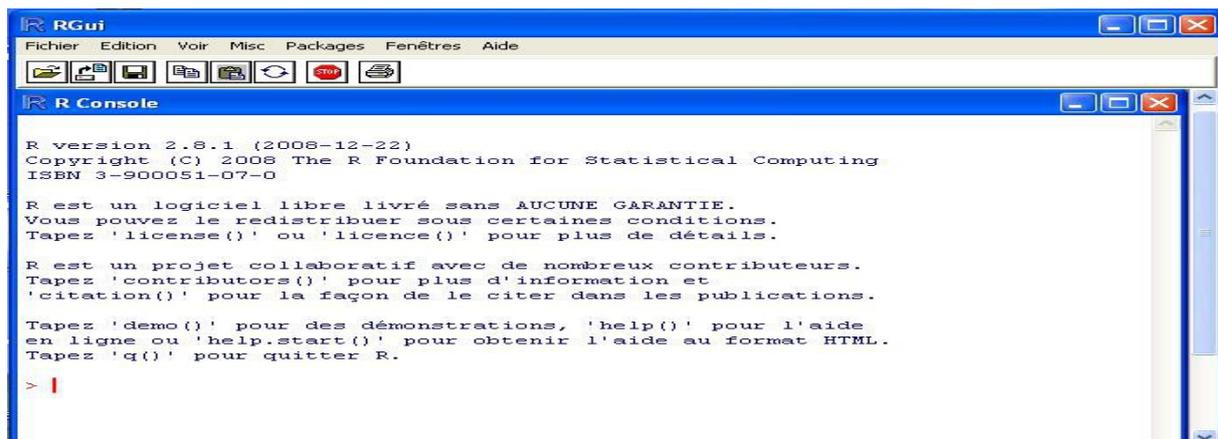


Fig 19: Interface de R.

3-Analysesbioinformatiques

3-1 Analyse des datas issues de CHIP-SOLID :

3-1-1 Prétraitement des données :

Le séquençage à haut débit SOLID, génère deux fichiers séparés, un pour les séquences et l'autre pour la qualité de celles-ci. Ces fichiers se présentent sous format Csfasta/qual (color-space), un format brut n'ayant pas d'encodage ASCII, donc les fichiers générés sont assez lourds et nécessitent un prétraitement afin de les alléger, normaliser les données et éliminer le bruit de fond qui pourrait être dû aux erreurs de séquençage. Donc les rendre manipulables par les outils bioinformatiques dont on aura besoin par la suite.

Le pipeline bioinformatique de prétraitement et de profilage des données CHIP-seq a été essentiellement développé en partie au CIML et en partie au TAGC à l'aide du langage R, avant mon stage. A la fin de la procédure du prétraitement des données, un fichier, représentant les régions génomiques où la protéine d'intérêt s'est liée, est généré sous format « wiggle » avec les scores d'enrichissement correspondants. **Wiggle** ou **WIG** est un format utilisé pour la représentation des données continues et denses comme pour les datas de transcriptome. Un fichier wig est généré pour chacune des trois marques épigénétiques : H3K4me1, H3K4me3, H3K27ac pendant le stade DN et DP (fig20).

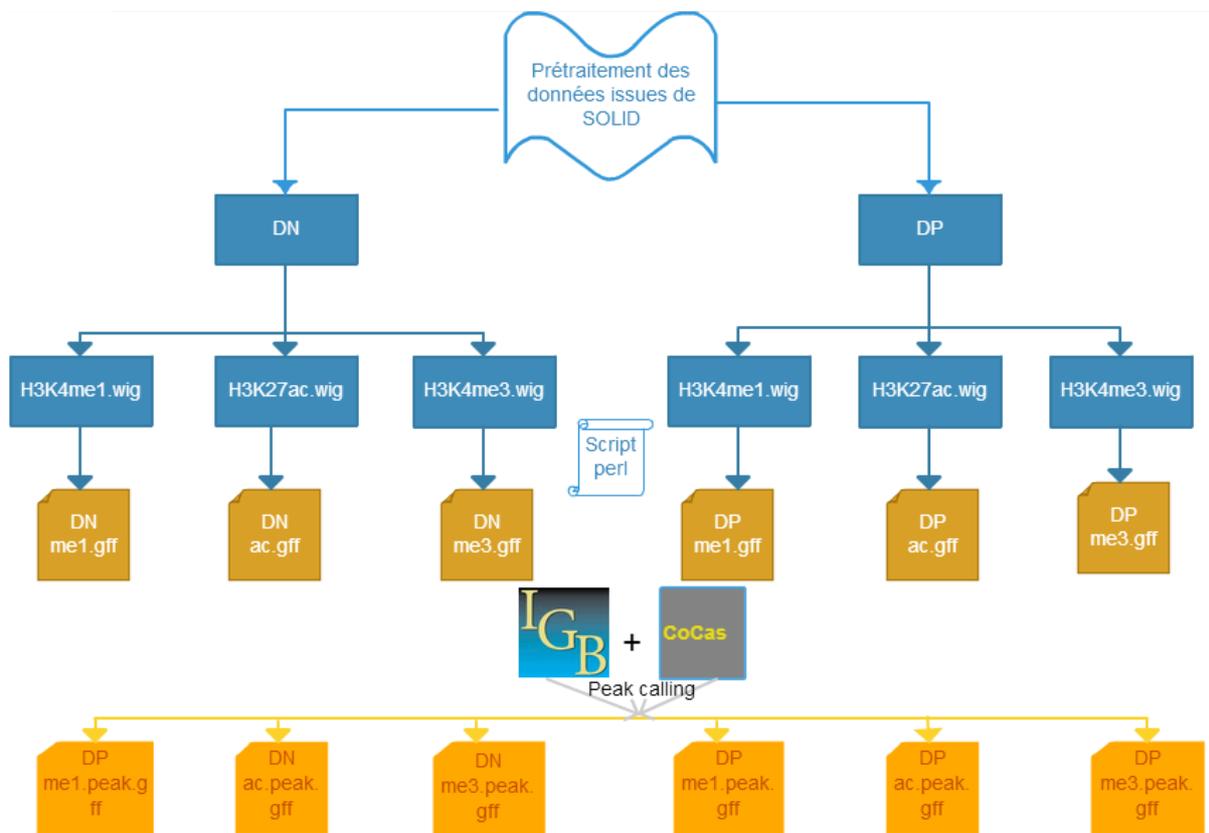


Fig20: Génération des fichiers gff à l'aide de perl et le peak calling à l'aide de IGB et CoCas. me1=H3K4me1, me3=H3K4me3, ac=H3K27ac. (Original)

3-1-2 Visualisation et représentation des pics de liaison à l'aide des fichiers GFF :

Le fichier sous format GFF (General Feature Format), est utilisé afin de localiser et de décrire des zones caractéristiques d'un génome. Il permet de créer et de visualiser les régions d'intérêt, sous forme de piques, grâce à des outils bioinformatiques comme Cocas et IGB.

Afin de rechercher les pics d'enrichissement de la protéine d'intérêt, on a tout d'abord, généré à partir d'un fichier WIG, un fichier GFF pour chaque marque épigénétique. Cela à l'aide d'un script Perl (annexe1), développé par M.M.BELHOCINE et qui permet la création des fichiers GFF en un temps minime sous LINUX. Le fichier GFF va servir de donnée d'entrée à CoCas pour le peak calling(fig20).

3-1-3 Le peak calling:

L'essentielle étape dans l'analyse des données CHIP-seq est le « peak calling » ou bien la recherche des pics. En effet, le but de cette étape est de déterminer, les régions génomiques marquées par la protéine d'intérêt (marque épigénétique). La qualité des analyses qui suivront dépendent directement du résultat trouvé lors de la recherche des pics [Epegenie, 2013].

3-1-3-1 Création des pics à l'aide de Cocas et IGB:

Afin de créer et de déterminer ces pics, j'ai utilisé l'outil CoCas. Il permet de dresser une liste des pics trouvés pour chaque marque épigénétique. Cette liste est dressée selon les paramètres établis d'après la visualisation par IGB (fig20).

Les paramètres choisis d'après la visualisation par IGB :

Marques épigénétiques	Extend threshold	Pic threshold	Distance threshold	Nombre de pics	
DN	DN-H3K4me1	3	5	400	25300
	DN-H3K4me3	400	2000	400	24875
	DN-H3K27ac	12	22	400	11375
DP	DP-H3K4me1	2	6	400	74607
	DP-H3K4me3	1	4	400	28013
	DP-H3K27ac	7	15	400	24575

Tableau II: Paramètres choisis dans Cocas après la visualisation par IGB et le nombre de pics générés par Cocas pour les trois marques épigénétiques d'intérêt pendant le stade DN et DP de la différenciation des cellules T. **(Original)**

3-1-3-2 Visualisation des pics et détermination des paramètres par IGB pour CoCas :

Afin de bien déterminer ces paramètres, j'ai utilisé un navigateur génomique IGB, qui permet la visualisation des pics. La difficulté dans le choix de ces paramètres, réside, dans la distinction des régions génomiques enrichies par la protéine d'intérêt, du bruit de fond (Fig21). Le bruit de fond, correspond aux erreurs qui pourraient être dus durant les différentes étapes du séquençage. Je me suis basée sur la visualisation d'un gène ubiquitaire pour le choix des paramètres qui est l'actine (fig22). Ainsi, après la définition des paramètres adéquats à l'aide d'IGB, la liste des pics pour chaque marque épigénétique est dressée par CoCas et sera stockée dans un fichier GFF (peak.gff). A la fin 6 listes de pics correspondant aux régions génomiques enrichies par les marques épigénétiques sont générées et stockées sous forme de fichiers peak.GFF (fig20).

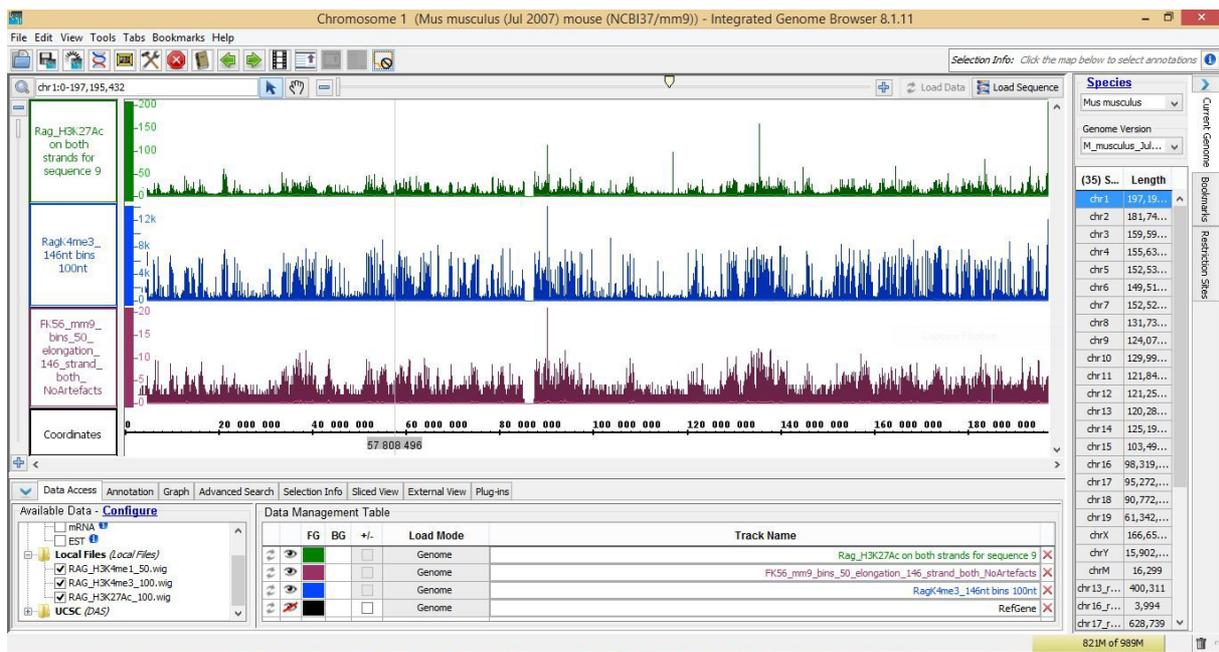


Fig21: Visualisation des fichiers gff par IGB. **(Original)**

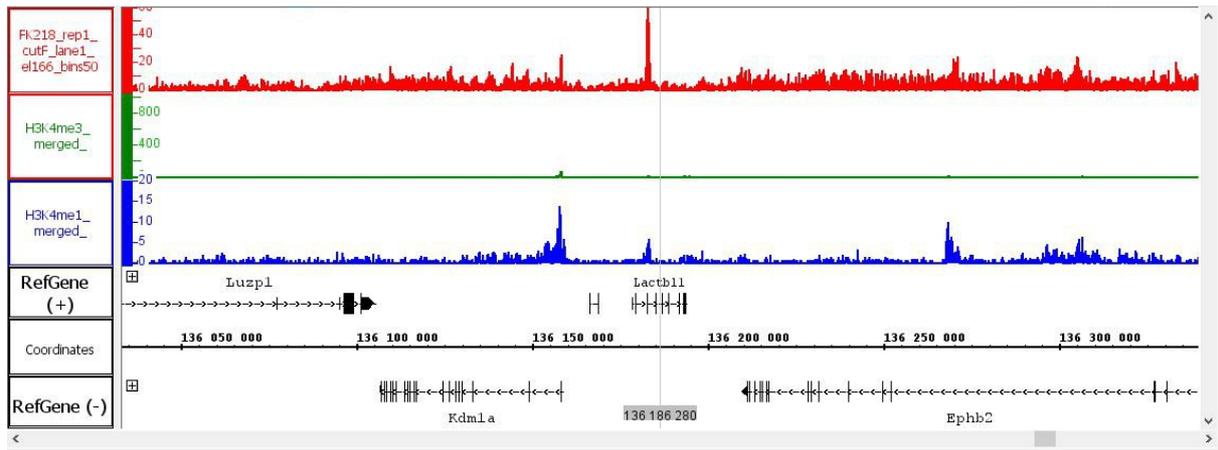


Fig22: Définition des paramètres d'après un gène ubiquitaire. **(Original)**

L'étape suivant la détermination du bruit de fond et la détection des pics pour chacune des marques épigénétiques durant les deux stades DN et DP, est l'overlapping.

3-1-4 L'intersection ou « overlapping »:

Le but de cette étape est de déterminer, les pics, représentant les régions génomiques auxquelles, les trois marques épigénétiques réunies, se lient.

Afin d'effectuer l'overlapping, j'ai utilisé l'outil Intrsect, fournit par la plate forme Galaxy. Cet algorithme permet l'identification et la visualiser des chevauchements entre deux séries de séquences génomiques (2 fichiers). Dans le cas de mon étude, je l'ai utilisé afin d'établir trois listes de pics correspondants aux régions enrichies par les 3 marques épigénétiques réunies, et cela pendant le stade DN, le stade DP ainsi que les régions génomiques enrichies communs aux deux stades (fig23)

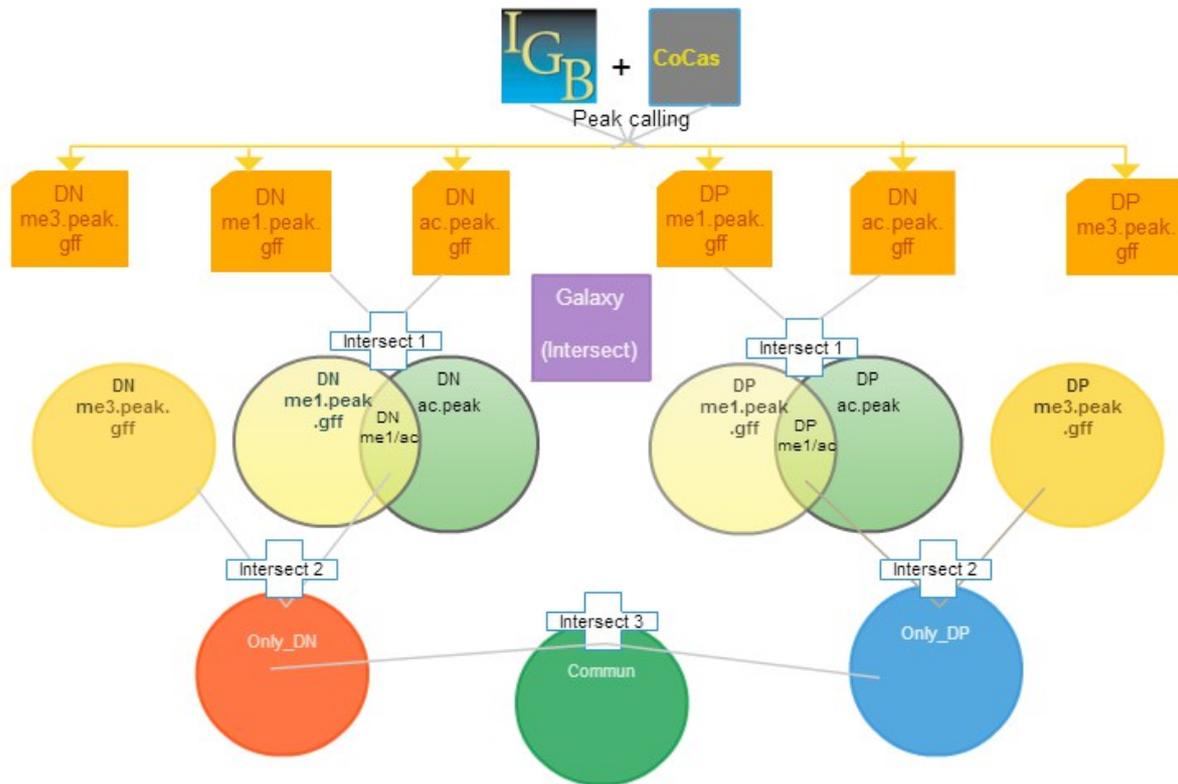


Fig23: Intersection des fichiers gff créés suite au peak calling. **(Original)**

Afin de générer ces fichiers, faut choisir les paramètres adéquats dans intersect. Deux paramètres à définir :

1-Choix entre Intersection par intervalles ou intersection par fragments d'intervalles

- a- **Intersection par intervalles:** Génère les chevauchements, entre le premier ensemble de données et le deuxième ensemble de données et cela en donnant l'intervalle génomique entier sans restriction des régions qui ne se croisent pas, ce qui ne convient pas à notre cas.
- b- **Intersection par fragments d'intervalles:** Génère les chevauchements, entre le premier ensemble de données et le deuxième ensemble, mais en affichant, seulement les régions, où l'intersection a lieu, ce qui est adapté à notre but de recherche et c'est celui là que j'ai choisis (fig24).

2-Paires de bases : définit la longueur minimale (en bases d'ADN) des chevauchements entre les éléments de données des deux ensembles. J'ai choisis 3 bases au minimum (fig24).

Grâce à Intersect 3 fichiers sont générés représentant les régions génomiques modulés par la dynamique épigénétique des protéines d'intérêt : DN (me1,ac,me3), DP(me1,ac,me3), Commun (DN-DP).

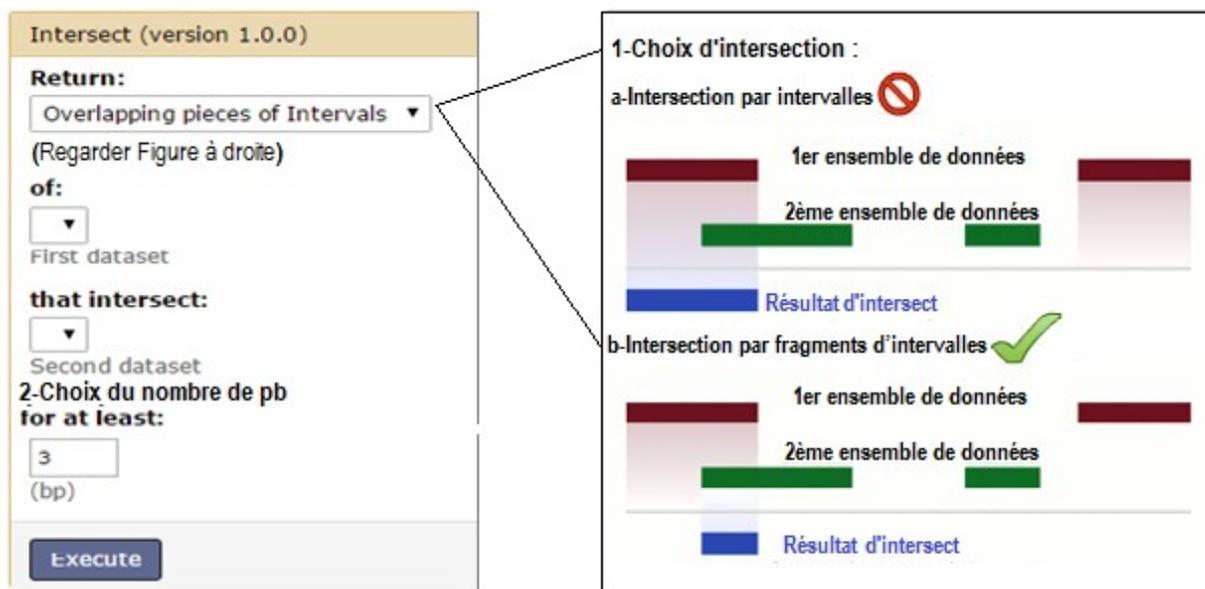


Fig24: Paramètres d'intersect.

3-1-5 Identification des séquences cis-régulatrices (enhancers) à l'aide de GREAT :

J'ai utilisé GREAT, l'outil permettant d'annoter les régions génomiques enrichies par la protéine d'intérêt, afin de définir le rôle fonctionnel des régions cis-régulatrices à partir du domaine fonctionnel de leurs gènes voisins. GREAT traite les fichiers sous format « BED ». Ce format comporte 3 champs obligatoires, contenant les données nécessaires, dont GREAT a besoin : le nom du chromosome, le début et la fin de la région génomique.

chr1	6396001	6397100	DN_only_1	0	+
chr1	6443401	6445100	DN_only_2	0	+
chr1	21819601	21820700	DN_only_3	0	+
chr1	33634301	33640100	DN_only_4	0	+

Fig25:Aperçu de la présentation du fichier BED.

Pour cela, j'ai tout d'abord généré, grâce à l'algorithme « Convert Formats » dans la plateforme GALAXY, des fichiers sous format BED à partir des 3 fichiers issus de l'overlapping [DN_only(me1,ac,me3), DP_only(me1,ac,me3), Commun (DN-DP)]. Ensuite j'ai utilisé les fichiers sous format BED (DN_only.bed, DP_only.bed, Commun.bed) comme donnée d'entrée pour GREAT. Avec les paramètres suivants :

-**Version du génome:** Mouse : NCBI build 37 ([UCSC mm9, Jul/2007](#))

- **Bruit de fond:** génome entier

- **Paramètre de définition du domaine fonctionnel des régions cis-régulatrices:** Les deux gènes les plus proches du TSS (site d'initiation de la transcription), avec une taille maximale de 1000kb, en amont et en aval de ce dernier.

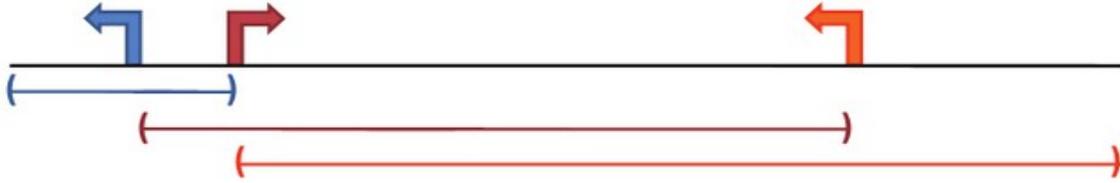


Fig26: Représentation de la méthode d'identification par GREAT, du domaine fonctionnel d'une région cis-régulatrice à partir des deux gènes les plus proches, en amont et en aval du TSS.

Pour chaque donnée d'entrée (fichier BED), GREAT dresse un ensemble de données qui seront abordés dans la partie Résultats et discussion.

3-1-6 Identification des motifs à l'aide de RSAT:

Comme décrit dans la section « outils bioinformatique », RSAT établit une liste des motifs des régions génomiques d'intérêt mais il exige pour le traitement des fichiers le format FASTA.

Convertir les listes des pics en format fasta: Fasta est le format qui affiche les séquences génomiques d'intérêt en bases d'ADN (format brut). Pour cela j'ai chargé les fichiers des pics générés avec intersect dans Table Browser dans le site UCSC GENOME BROWSER en changeant les suivants paramètres : génome (souris), version (July 2007 (NCBI37/mm9), output format (sequence) (fig 27).

clade: genome: assembly:
 output format:

Fig27: paramètres choisis dans table browser.

4- Résultats et discussion

4-1 Résultats :

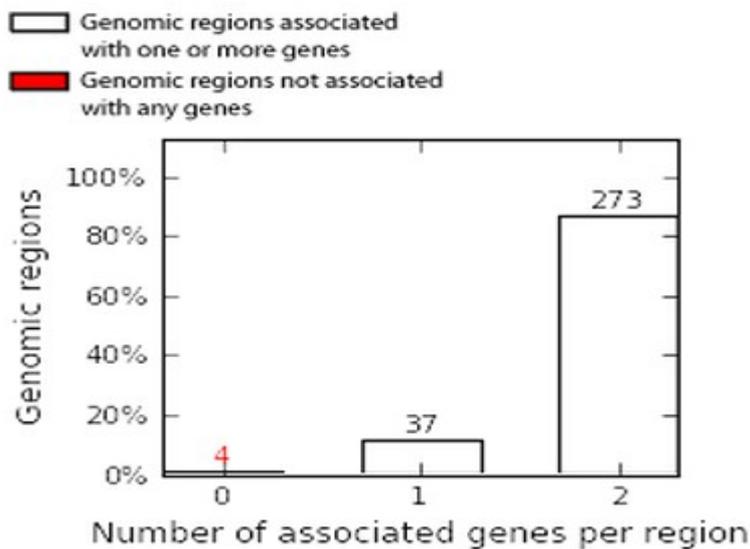
4-1-1 Résultats générés par GREAT :

4-1-1-1 Distance à partir du TSS :

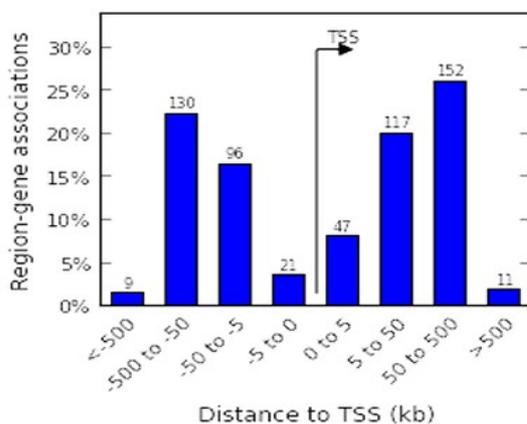
D'après les graphes dressés pour chacun des 3 fichiers bed (DN_only, DP_only, Commun) par GREAT, on observe que les régions génomiques identifiées se situent à une distance supérieure à 2 kB du TSS comme indiqué dans les graphes ci-dessous.

a-Graphes indiquant la distances entre les régions Cis et le TSS pendant le stade DN :

Number of associated genes per region



Binned by orientation and distance to TSS



Binned by absolute distance to TSS

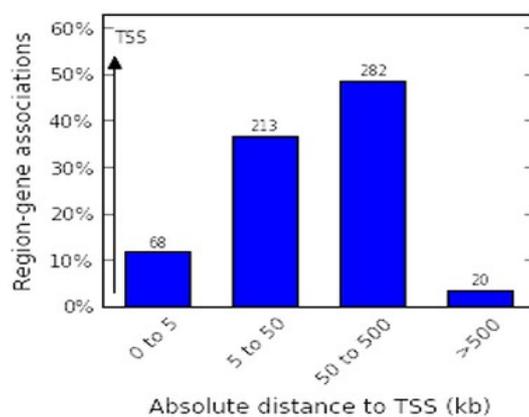


Fig 28: Distance en Kb des régions Cis-régulatrices en fonction du TSS pour le stade DN. (Original)

b-Graphes indiquant la distances entre les régions Cis et le TSS pendant le stade Commun :

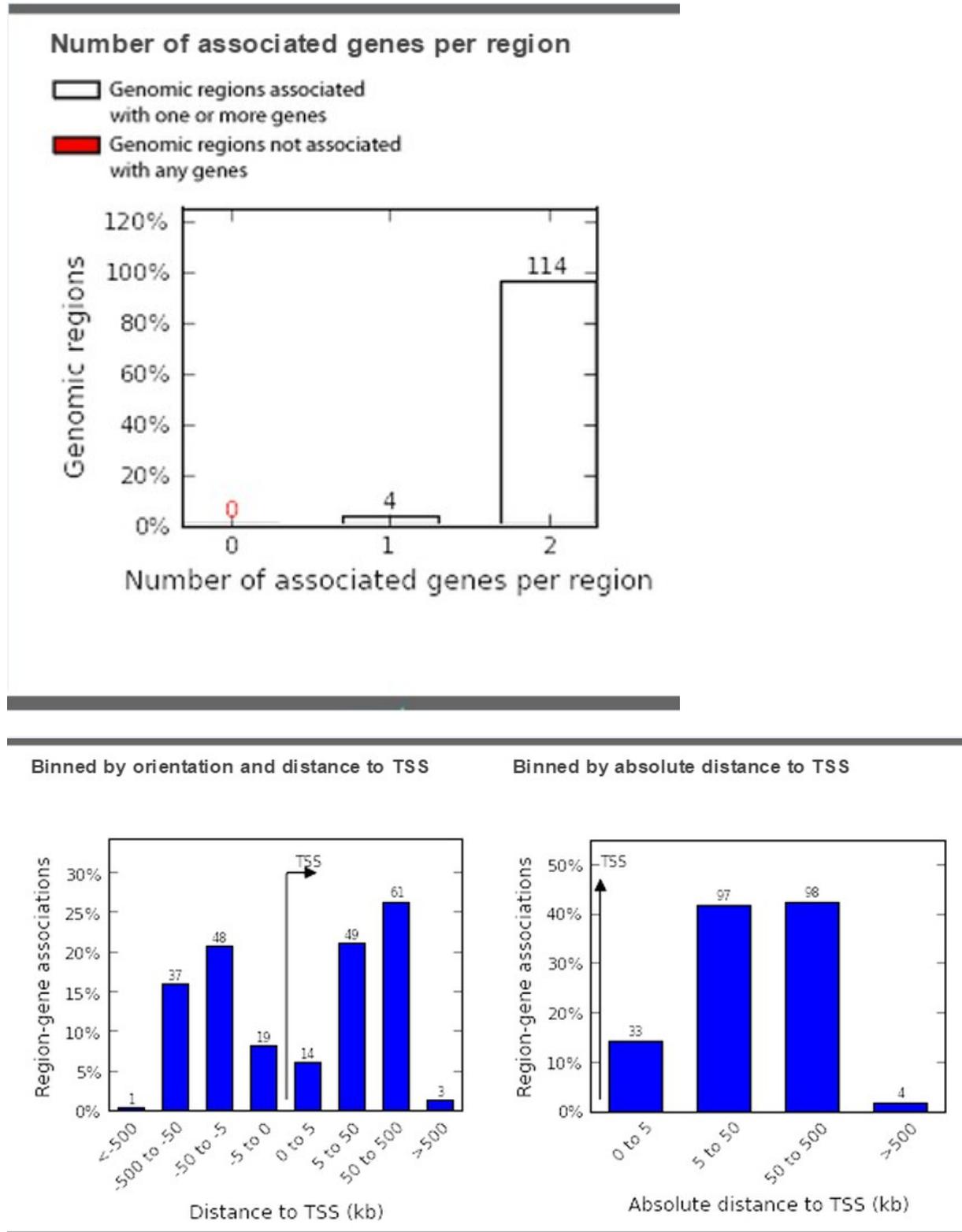


Fig29 : Distance en Kb des régions Cis-régulatrices en fonction du TSS pour le stade Commun DN/DP. (Original)

c-Graphes indiquant la distances entre les régions Cis et le TSS pendant le stade DP :

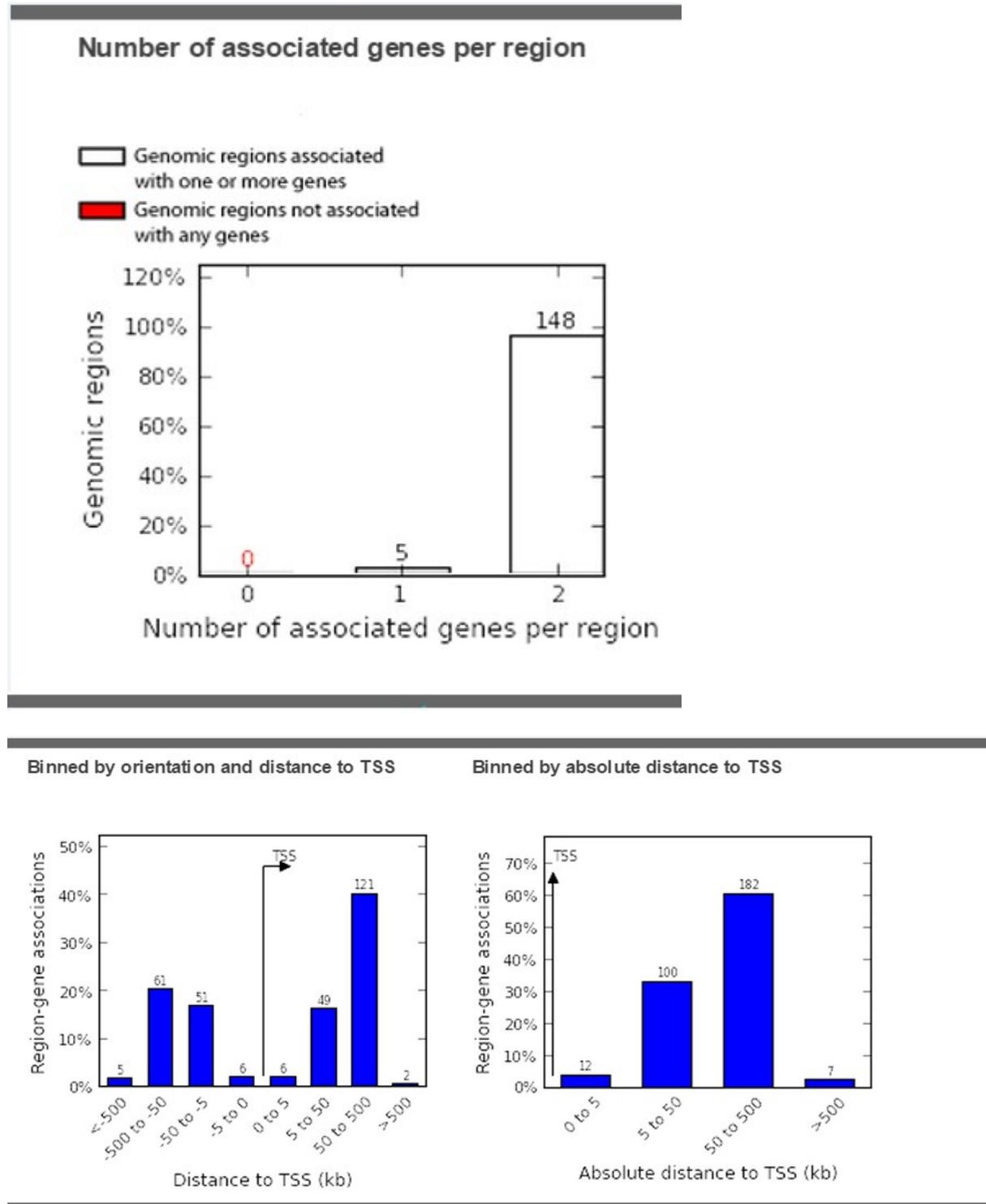


Fig 30: Distance en Kb des régions Cis-régulatrices en fonction du TSS pour le stade DP. (Original)

4-1-1-2 Les voies biologiques:

Depuis les tableaux générés de GO Biological Process (annexe 2), on remarque que les voies activées pendant le stade DN diffèrent des voies activées durant le stade DP. Les plus pertinentes d'entre elles sont représentées selon leur score d'enrichissement dans le Radar plot ci-dessous. Le radar Plot est généré en utilisant les scores d'enrichissement avec l'aide d'un script R (annexe 3).

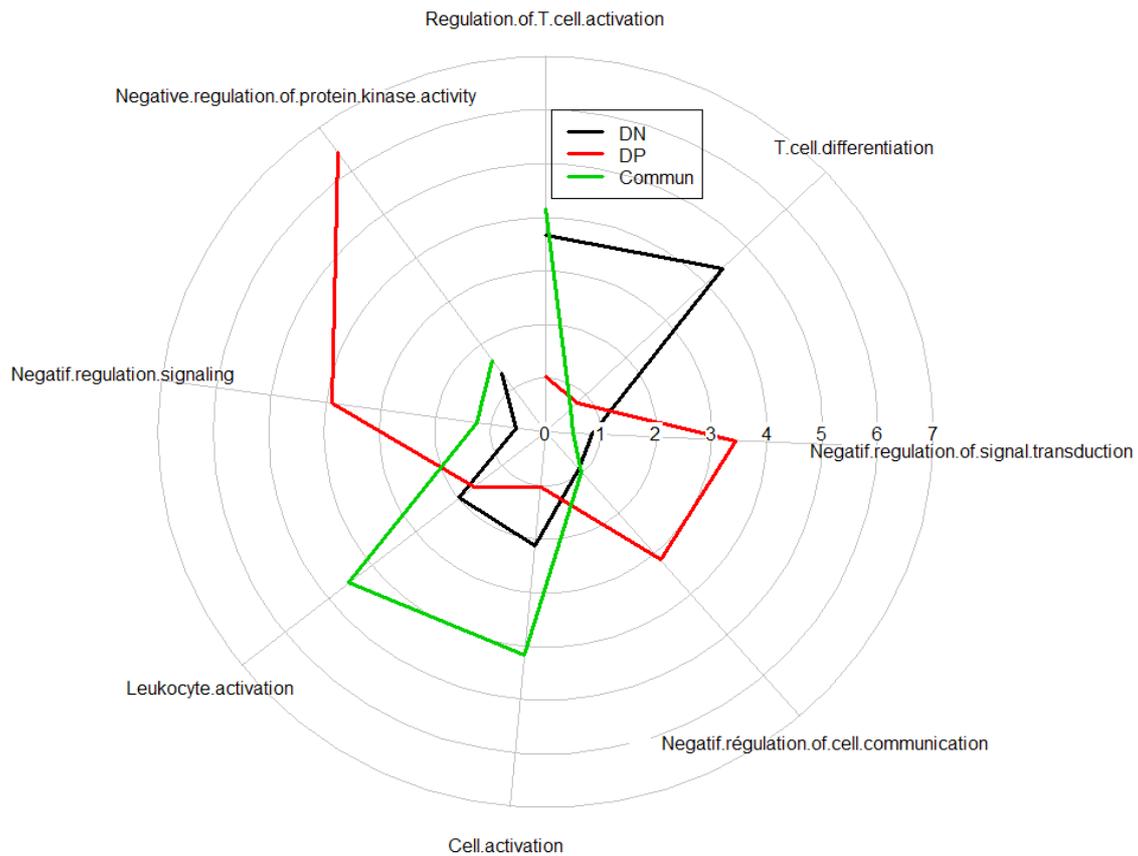


Fig31 : Radar plot présentant les voies biologiques activées durant DN, DP et commun selon leur score d'enrichissement. **(Original)**

4-1-1-3 Les gènes exprimés durant DN et DP :

Pour chacune des voies biologiques activées durant les stades DN et DP et Commun, GREAT liste, une série d'enhancers avec leurs gènes avoisinants ainsi qu'une série de gènes et les enhancers leurs correspondants, qui sont exprimés restrictivement à chaque des stades DN et DP ou communément. J'ai comparé ces listes (voir annexe 4) et j'ai observé que certains gènes sont exprimés communément dans le stade DN et dans DP mais avec des enhancers positionnés différemment :

-Ptpn 22 :

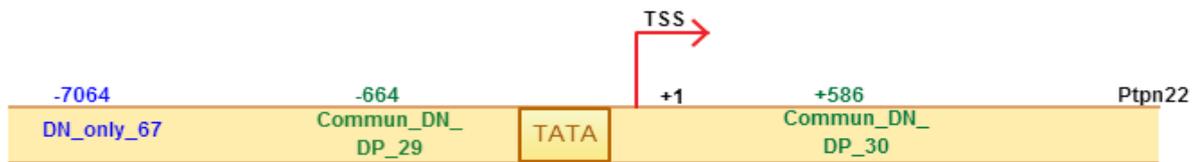


Fig32: schéma démontrant la position des différents enhancers pour le gène Ptpn22 selon le stade concerné. **(Original)**

Ptpn22 est exprimé en fonction de l'enhancer (Only_DN67) situé en amont du TSS, en position -7064 dans le stade DN ; et est lié à deux enhancers différents pour «Commun», l'enhancer (Commun_DN_DP_29) en position -664 et l'enhancer (Commun_DN_DP_30) en position +586.

-Runx2 :

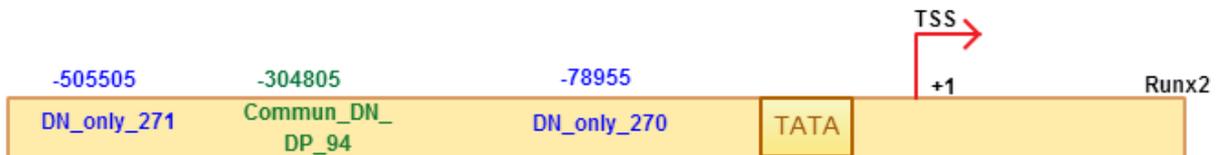


Fig33: schéma démontrant la position des différents enhancers pour le gène Runx2. **(Original)**

Runx 2 ; en DN est transcrit en fonction de l'enhancer (Only_DN_270) en position -78955 et en fonction de l'enhancer (Only_DN_271) en position -505505, en Commun ; il est transcrit en fonction de l'enhancer (Commun_DN_DP_94) en position -304804.

-Fyn :

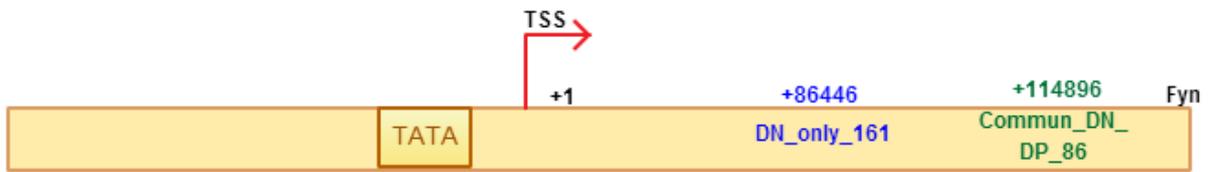


Fig34: schéma démontrant la position des différents enhanceurs pour le gène Fyn. **(Original)**

Fyn, dans DN est exprimé en fonction de l'enhancer Only-DN_161 en position +86446, dans Commun en fonction de l'enhancer Commun_DN_DP_86 en position +144,896.

-Irf1 :

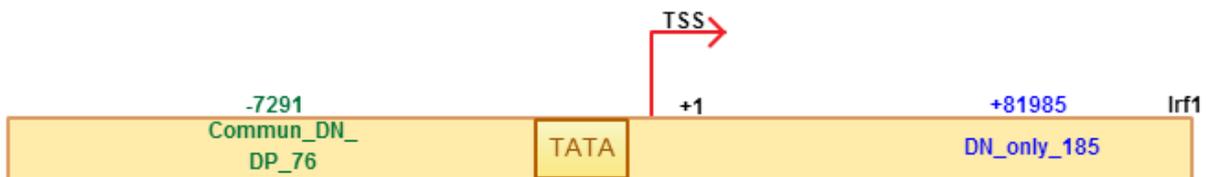


Fig35: schéma démontrant la position des différents enhanceurs pour le gène Irf1. **(Original)**

Irf1, pour DN exprimé en fonction de l'enhancer Only_DN_185 en position +81985 et en Commun en fonction de l'enhancer Commun_DN_DP_76 en position -7291.

4-1-1-4 Les facteurs de transcription:

Great a identifié des FTs impliqués dans le stade DN et impliqués notamment dans le stade Commun à DN et DP (annexe 5). Aussi, Great a listé sous forme de tableaux les gènes et leurs régions cis-régulatrices (Enhances) interagissant avec ces facteurs de transcriptions et représentés ci-dessous (Tableau 3 et 4):

Stade DN: 51 gènes ont été identifiés en relation fonctionnelle avec 62 régions cis-régulatrices (enhancers) et interagissant avec les FTs listés en annexe 5.

This term's genomic region-gene association tables (62 regions, 51 genes)	
This term's genomic region -> gene association table	
Region	Gene (distance to TSS)
DN_only_7	Mfsd6 (+96,061)
DN_only_9	Creb1 (-84,127)
DN_only_10	Cul3 (-405,846)
DN_only_11	Cab39 (-41,921)
DN_only_14	Pdcd1 (+17,229)
DN_only_15	St8sia4 (+10,170)
DN_only_25	Usp6nl (-14,781)
DN_only_26	Usp6nl (+353,669)
DN_only_27	Ith5 (+206,631)
DN_only_29	Il2ra (-6,518)
DN_only_51	Ncoa3 (-106,936)
DN_only_55	Str16 (-58,801)
DN_only_59	Tbl1xr1 (-139,923)
DN_only_67	Ptpn22 (-7,064)
DN_only_68	Cd53 (+1,116), 4933421E11Rik (+67,351)
DN_only_73	Gbp2 (-59,876)
DN_only_83	Cap1 (-47,577)
DN_only_84	Wasf2 (-747)
DN_only_85	Pnrc2 (-2,740)
DN_only_88	Tnfrsf18 (+5,600)
DN_only_92	Tmem60 (+62,830)
DN_only_95	Anbr2 (+14,030)
DN_only_97	Snmp35 (-4,313)
DN_only_99	Ing3 (+886)

This term's gene -> genomic region association table	
Gene	Region (distance to TSS)
4933421E11Rik	DN_only_68 (+67,351)
Adora2a	DN_only_162 (+3,913)
Anbr2	DN_only_95 (+14,030)
Aqp11	DN_only_124 (-10,994)
Cab39	DN_only_11 (-41,921)
Cap1	DN_only_83 (-47,577)
Cd47	DN_only_254 (-84,466), DN_only_255 (-54,616), DN_only_256 (-36,516), DN_only_257 (-7,816)
Cd53	DN_only_68 (+1,116)
Clec2d	DN_only_116 (+52,018), DN_only_117 (+55,968)
Creb1	DN_only_9 (-84,127)
Cul3	DN_only_10 (-405,846)
Elf4enif1	DN_only_179 (-6,385)
Eml4	DN_only_281 (+46,780)
Ets1	DN_only_141 (-103,726)
Evi2b	DN_only_190 (-5,787), DN_only_189 (-3,487)
Fyb	DN_only_239 (-5,020), DN_only_240 (+3,580), DN_only_241 (+12,630)
Fyn	DN_only_161 (+86,446)
Gbp2	DN_only_73 (-59,876)
Gsr	DN_only_134 (+37,141)
Hectd2	DN_only_306 (-20,028)
Il2ra	DN_only_29 (-6,518)
Ing3	DN_only_99 (+886)
Ith5	DN_only_27 (+206,631)

DN_only_116	Clec2d (+52,018)
DN_only_117	Clec2d (+55,968)
DN_only_124	Aqp11 (-10,994)
DN_only_134	Gsr (+37,141)
DN_only_141	Ets1 (-103,726)
DN_only_144	Tie3 (-259,772)
DN_only_146	Rab8b (+3,311)
DN_only_147	Mapk6 (-24,985), Leo1 (-7,180)
DN_only_158	Map3k5 (+5,119)
DN_only_161	Fyn (+86,446)
DN_only_162	Adora2a (+3,913)
DN_only_179	Eif4enif1 (-6,385)
DN_only_181	Itk (+16,466)
DN_only_189	Evi2b (-3,487)
DN_only_190	Evi2b (-5,787)
DN_only_192	Sifn2 (+13,337)
DN_only_206	Zfp361i (+281,149)
DN_only_207	Zfp361i (+270,549)
DN_only_208	Zfp361i (-116,101)
DN_only_209	Sipa11i (-443,985)
DN_only_223	Map3k1 (-175,460)
DN_only_239	Fyb (-5,020)
DN_only_240	Fyb (+3,580)
DN_only_241	Fyb (+12,630)
DN_only_254	Cd47 (-84,466)
DN_only_255	Cd47 (-54,616)
DN_only_256	Cd47 (-36,516)

DN_only_257	Cd47 (-7,816)
DN_only_264	Slc22a3 (-6,031)
DN_only_272	Kat2b (-545)
DN_only_273	Kat2b (+2,255)
DN_only_274	Kat2b (+20,355)
DN_only_281	Eml4 (+46,780)
DN_only_294	Smad7 (-104,618)
DN_only_295	Smad7 (-88,818)
DN_only_301	Lpxn (+2,352)
DN_only_302	Zfand5 (-235,517)
DN_only_306	Hectd2 (-20,028)

DN_only_181	DN_only_181 (+16,466)
Kat2b	DN_only_272 (-545), DN_only_273 (+2,255), DN_only_274 (+20,355)
Leo1	DN_only_147 (-7,180)
Lpxn	DN_only_301 (+2,352)
Map3k1	DN_only_223 (-175,460)
Map3k5	DN_only_158 (+5,119)
Mapk6	DN_only_147 (-24,985)
Mfsd6	DN_only_7 (+96,061)
Ncoa3	DN_only_51 (-106,936)
Pdcd1	DN_only_14 (+17,229)
Pnrc2	DN_only_85 (-2,740)
Ptpn22	DN_only_67 (-7,064)
Rab8b	DN_only_146 (+3,311)
Sipa11i	DN_only_209 (-443,985)
Slc22a3	DN_only_264 (-6,031)
Sifn2	DN_only_192 (+13,337)
Smad7	DN_only_294 (-104,618), DN_only_295 (-88,818)
Snmp35	DN_only_97 (-4,313)
St8sia4	DN_only_15 (+10,170)
Stx16	DN_only_55 (-58,801)
Tbl1xr1	DN_only_59 (-139,923)
Tie3	DN_only_144 (-259,772)
Tmem60	DN_only_92 (+62,830)
Tnfrsf18	DN_only_88 (+5,600)
Usp6nl	DN_only_25 (-14,781), DN_only_26 (+353,669)
Wasf2	DN_only_84 (-747)

Zfand5	DN_only_302 (-235,517)
Zfp361i	DN_only_208 (-116,101), DN_only_207 (+270,549), DN_only_206 (+281,149)

Tableau III: Gène et leurs régions cis-régulatrices (enhancers) identifiés pendant le stade DN et qui sont en interaction avec les FTs détectés par GREAT.

Commun : 21 gènes ont été identifiés en relation fonctionnelle avec 24 régions cis-régulatrices (enhancers) et interagissant avec les FTs listés en annexe 5.

☛ This term's genomic region-gene association tables (24 regions, 21 genes)

This term's genomic region -> gene association table		This term's gene -> genomic region association table	
Region	Gene (distance to TSS)	Gene	Region (distance to TSS)
Commun_DN_DP_12	Cdkn1b (+12,607)	Cdkn1b	Commun_DN_DP_12 (+12,607)
Commun_DN_DP_20	Txn1 (-788)	Cobl1	Commun_DN_DP_34 (-99,018)
Commun_DN_DP_23	Smad2 (+43,951)	Cytip	Commun_DN_DP_33 (+3,532)
Commun_DN_DP_24	Fgr (-27,959)	Fam98b	Commun_DN_DP_37 (+91,000)
Commun_DN_DP_27	Prcc (+38,133)	Fgr	Commun_DN_DP_24 (-27,959)
Commun_DN_DP_28	Prcc (+37,483)	Fyn	Commun_DN_DP_88 (+114,898)
Commun_DN_DP_29	Ptpn22 (-664)	Ggta1	Commun_DN_DP_32 (-34,356)
Commun_DN_DP_30	Ptpn22 (+586)	Hectd2	Commun_DN_DP_112 (-170,278)
Commun_DN_DP_32	Ggta1 (-34,356)	Irf6	Commun_DN_DP_51 (+47,095)
Commun_DN_DP_33	Cytip (+3,532)	Lcp2	Commun_DN_DP_75 (-500)
Commun_DN_DP_34	Cobl1 (-99,018)	Letm2	Commun_DN_DP_60 (+19,308)
Commun_DN_DP_37	Fam98b (+91,000)	Obfc2a	Commun_DN_DP_46 (-107,008), Commun_DN_DP_45 (-106,158)
Commun_DN_DP_42	Slc20a1 (-15,108)	Pou2af1	Commun_DN_DP_56 (-47,994)
Commun_DN_DP_43	Slc20a1 (-14,558)	Prcc	Commun_DN_DP_28 (+37,483), Commun_DN_DP_27 (+38,133)
Commun_DN_DP_45	Obfc2a (-106,158)	Ptpn22	Commun_DN_DP_29 (-664), Commun_DN_DP_30 (+586)
Commun_DN_DP_46	Obfc2a (-107,008)	Slc20a1	Commun_DN_DP_42 (-15,108), Commun_DN_DP_43 (-14,558)
Commun_DN_DP_51	Traf3ip3 (+1,339), Irf6 (+47,095)	Smad2	Commun_DN_DP_23 (+43,951)
Commun_DN_DP_56	Pou2af1 (-47,994)	Srp54c	Commun_DN_DP_68 (+71,586)
Commun_DN_DP_60	Letm2 (+19,308)	Traf3ip3	Commun_DN_DP_51 (+1,339)
Commun_DN_DP_68	Srp54c (+71,586)	Txn1	Commun_DN_DP_20 (-788)
Commun_DN_DP_75	Lcp2 (-500)	Zfp472	Commun_DN_DP_93 (-18,075)
Commun_DN_DP_88	Fyn (+114,898)		
Commun_DN_DP_93	Zfp472 (-18,075)		
Commun_DN_DP_112	Hectd2 (-170,278)		

Tableau IV: Gène et leurs régions cis-régulatrices (enhancers) identifiés pendant le stade Commun (DN/DP) et qui sont en interaction avec les FTs détectés par GREAT.

4-1-2 Résultats générés par RSAT :

RSAT a identifié des motifs pour le stade DN, le stade DP et les motifs communs aux deux stades ainsi que les facteurs de transcription correspondants.

DN-Only :

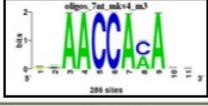
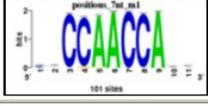
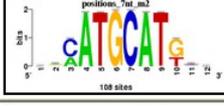
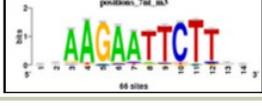
Motif	Logo	3 Top hits in databases
oligos_7nt_mk4_m1		<u>versus jaspar core vertebrates:</u> IRF1,
oligos_7nt_mk4_m2		<u>versus jaspar core vertebrates:</u> SP1, Klf4, MZF1_1-4,
oligos_7nt_mk4_m3		<u>versus jaspar core vertebrates:</u> RUNX1,
positions_7nt_m1		<u>versus jaspar core vertebrates:</u> NFIC,
positions_7nt_m2		<u>versus jaspar core vertebrates:</u> no match
positions_7nt_m3		<u>versus jaspar core vertebrates:</u> no match

Fig36: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade DN. **(Original)**

Commun :

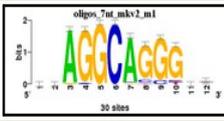
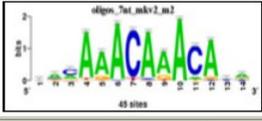
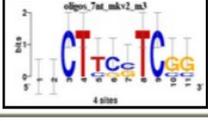
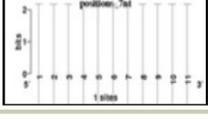
Motif	Logo	3 Top hits in databases
oligos_7nt_mk2_m1		<u>versus jaspar core vertebrates:</u> no match
oligos_7nt_mk2_m2		<u>versus jaspar core vertebrates:</u> FOXI1, Foxd3, FOXO3,
oligos_7nt_mk2_m3		<u>versus jaspar core vertebrates:</u> SPI1,
positions_7nt		<u>versus jaspar core vertebrates:</u> no match

Fig37: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade Commun. **(Original)**

DP_Only :

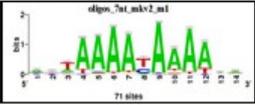
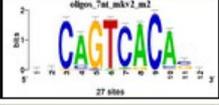
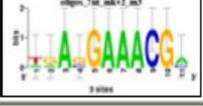
Motif	Logo	3 Top hits in databases
oligos_7nt_mkv2_m1		<u>versus jaspar_core vertebrates:</u> FOXI1, Foxd3,
oligos_7nt_mkv2_m2		<u>versus jaspar_core vertebrates:</u> no match
oligos_7nt_mkv2_m3		<u>versus jaspar_core vertebrates:</u> SPIB, ETS1,
positions_7nt_m1		<u>versus jaspar_core vertebrates:</u> HIF1A::ARNT, Arnt, Myc,
positions_7nt_m2		<u>versus jaspar_core vertebrates:</u> ARID3A,
positions_7nt_m3		<u>versus jaspar_core vertebrates:</u> no match

Fig38: Les motifs logo et les FTs correspondants identifiés par RSAT, durant le stade DP.
(Original)

4-2 Discussion:

Avec l'overlapping on a pu mettre en évidence les régions génomiques enrichies par les 3 marques épigénétiques réunies, H3K4me1, H3K4me3 et H3K27ac et cela pour le stade DN, DP puis les régions génomiques marquées communément durant DN et DP. De ce résultat on a pu démontrer qu'il y a une dynamique épigénétique lors de la différenciation des cellules T liée à nos marques épigénétiques d'intérêt.

Par la suite et d'après les résultats générés par GREAT, on a constaté que la distance entre le TSS et les régions génomiques en Cis est supérieur à 2Kb, ce qui représente un bon résultat, car il témoigne qu'effectivement les régions enrichies par H3K4me1, H3K4me3 et H3K27ac sont des régions cis-régulatrices et donc des enhancers.

Durant le stade DN, on observe d'après les résultats, que les processus biologiques activés spécifiquement à ce stade, sont ceux concernant la différenciation des cellules T et la régulation de la différenciation des Thymocytes, durant lesquels les CDN subissent des changements phénotypiques, leur permettant d'acquérir le phénotype C DP4+8+. Durant le développement des thymocytes vers le stade DP c-à-d C DP4+8+, les changements phénotypiques s'estompent afin que les CDP deviennent SP4+ ou SP8+ et donc d'autres voies biologiques de régulation entrent en jeu (Voir le chapitre 1-3).

En comparant les résultats générés concernant les gènes et leurs régions cis-régulatrices respectives (Enhancers), on a remarqué la présence de certains gènes identiques durant DN et DP mais avec des Enhancers positionnés différemment : le fait que certains gènes soient exprimés communément durant le stade DN et durant le stade DP, mais avec des enhancers positionnés différemment, comme il est le cas dans ce travail pour Runx2, Fyn, Ptpn2 et Irf, pourrait être dû au fait qu'un enhancer puisse agir, pour augmenter le niveau de transcription d'un gène donné avec un promoteur éventuellement faible, ou peut fournir des informations supplémentaires essentielles, qui ne sont pas codées dans le promoteur du gène lui-même, comme il est démontré par Yanez-Cuna JO et al., 2013 et Spitz F et al., 2012. Le répertoire des enhancers est en outre hautement spécifique de type de cellules et de l'état des gènes codant pour des protéines et cela a été abordé par l'étude de Bulger M et al., 2011 et Ong C-T et al., 2011. Par conséquent, un gène peut être régulé par différents enhancers dans différents types de cellules, à différents stades de développement, et en réponse à différents signaux ce qui correspond à la recherche faite par Levine M, 2010.

Depuis les résultats de RSAT, on peut observer que cet outil bioinformatique a détecté la présence de séquences d'ADN appelées séquences logos qui sont des séquences consensus et pourraient présenter des sites de liaison spécifiques à certains types de protéines dont les facteurs de transcription. Pour chaque stade (DN, DP et commun), des FTs sont détectés pour chaque séquence logo.

On remarque aussi que les facteurs de transcription dressés diffèrent d'un stade à un autre et donc ils pourraient agir en fonction de l'activation de certains enhancers que d'autres et

ainsi réguler le mécanisme de la transcription durant les différents stades de différenciation des thymocytes T selon un certain dynamisme transcriptionnel.

Conclusion:

En utilisant les données issues de la technologie ChIP-seq et grâce à l'exploitation d'outils bioinformatiques, on a pu démontrer que les marques épigénétiques H3K4me1, H3K4me3 et H3K27ac agissent effectivement en dynamisme afin de procéder à la régulation de la transcription.

Ces marques régulent le mécanisme transcriptionnel en interagissant avec les enhancers et les promoteurs des gènes d'intérêt et en s'association directement ou indirectement avec des facteurs de transcription, permettant ainsi un dynamisme épigénétique et transcriptionnel qui permettrait selon le stade du développement des thymocytes, l'expression de certains gènes cibles nécessaires aux besoins cellulaires durant le stade d'intérêt.

Les résultats démontrés dans ce modeste travail, s'inscrivent dans le cadre d'un ambitieux projet européen dans lequel le laboratoire d'accueil TAGC est partenaire et qui a pour ambition de caractériser l'épigénome du système hématopoïtique humain.

Les mécanismes de régulation impliquant la machinerie transcriptionnelle et le dynamisme épigénétique globale offrent encore un portail de mécanismes et processus méconnus voire inexplorés et l'avenir grâce aux progrès des technologies de séquençage à haut débit et le développement des outils bioinformatiques permettant de traiter les données générés, promet d'être fructueux en découvertes dans ce sens.

Références bibliographiques

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K.

High-resolution profiling of histone methylations in the human genome.
Cell ,129(4):823-37, Mai 2007.

Benoukraf T, Cauchy P, Fenouil R, Jeanniard A, Koch F, Jaeger S, Thieffry D, Imbert J, Andrau J C, Spicuglia S, Ferrier P.

CoCAS: a ChIP-on-ChIP analysis suite.
Bioinformatics, 25(7):954-955, 2009.

Bock C et Lengauer T.

Computational epigenetics.
Bioinformatics reviews,vol. 24 no. 1, p1–10, 2008.

Bulger M, Groudine M.

Functional and mechanistic diversity of distal transcription enhancers.
Cell 144: 327–39, 2011

[Chollier MT](#), [Darbo E](#), [Herrmann C](#), [Defrance M](#), [Thieffry D](#) et [Helden JV](#).

A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs
Nature Protocols 7, 1551–1568, 2012.

Chollier MT, Defrance m , Rivera AM , Sand O, Herrmann C, Thieffry D et Helden JV.

RSAT 2011: regulatory sequence analysis tools.
Oxford Journals,nucleic acids research,vol 39, 2011

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R.

Histone H3K27ac separates active from poised enhancers and predicts developmental state.
Proc Natl Acad Sci U S A, 107(50): 21931–21936, Dec 2010.

Cutter AR et Hayes JJ.

A brief review of nucleosome structure.
Elsevier B.V. 2015

Donald E, Olins et Olins AL.

Chromatin history: our view from the bridge.
Nature Reviews Molecular Cell Biology 4, 809-814, 2003.

Germain RN.

T-cell development and the CD4-CD8 lineage decision.
Nature immunology, vol 2, 2002.

Hayday AC et Pennington DJ.

Key factors in the organized chaos of early T cell development.
Nature immunology, vol 8, no 2, 2007

Heard E.

Epigénétique: Comment se joue la partition du génome?.
Science & nature n°11, Novembre-Décembre 2012.

Heinz Breu.

A Theoretical Understanding of 2 Base Color Codes and Its Application to Annotation, Error Detection, and Error Correction. Methods for Annotating 2 Base Color Encoded Reads in the SOLiD™ System
Applied Biosystems, 2010

Kidder BL, Hu G et Zhao K.

ChIP-Seq: technical considerations for obtaining high-quality data.
Nature immunology, vol 12, no 10, October 2011.

Kindt TJ, Osborne BA, Goldsby RA.

Immunology.
W. H. Freeman & Company, 6th edition, p 245-250, p254-256, p262, p268, p269, 2006

Kouzarides T.

Chromatin Modifications and Their Function.
Cell, Volume 128, Issue 4, p 693–705, 2007
SnapShot: Histone-Modifying Enzymes
Cell, Vol 128, Issue 4, p802.e1–802.e2, 23 February 2007.

LATCHMAN D.

Gene control (2010).
Newyork , éditions Garland Science, p29-53, p66-73, p95-115, p121-131, p155-160.

Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, Roeder RG

H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.
Cell, 152(5):1021-36, Fev 2013.

Levine M.

Transcriptional enhancers in animal development and evolution.
Curr Biol 20: R754–63. 2010.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar B T, Lowe CB, Wenger AM et Bejerano G.

GREAT improves functional interpretation of cis-regulatory regions
Nature biotechnology, vol 28, no 5, 2010.

Mike SS et Taylor RK.

Preparing DNA Libraries for Multiplexed Paired-End Deep Sequencing for Illumina GA
Sequencer.

PubMed Central, Curr Protoc Microbiol. Chapter: Unit1E.4, Fev 2011

Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE.

The Integrated Genome Browser: free software for distribution and exploration of genome-
scale datasets.

Bioinformatics, 25(20):2730-1, 2009.

Ong C-T, Corces VG.

Enhancer function: new insights into the regulation of tissue-specific gene expression.

Nat Rev Genet 12: 283–93,2011.

RAVEN, JOHNSON, MASON, LOSOS et SINGER

Biologie (2008)

France, éditions De Boeck, p369-373.

Spitz F, Furlong EEM.

Transcription factors: from enhancer binding to developmental control.

Nat Rev Genet 13: 613–26. 2012.

Tjian R.

Molecular machines that control genes

Scientific american, 2008.

Yanez-Cuna JO, Kvon EZ, Stark A.

Deciphering the transcriptional cis-regulatory code.

Trends Genet 29: 11–22, 2013.

Sites internet :

Annunziato, A.

DNA Packaging: Nucleosomes and Chromatin.

Nature Education 1(1):26,2008.

<http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310>

Consulté le 19 Septembre 2014.

Epigenie, Peak Calling for ChIP-Seq, 2013.

<http://epigenie.com/wp-content/uploads/2013/02/Peak-Calling-for-ChIP-Seq.pdf>

Consulté le 17 Octobre 2014

Fortin I, 2005.

Domaines protéiques du complexe histone acétyltransférase NuA4 impliqués dans la transcription et le maintien de l'intégrité du génome

<http://theses.ulaval.ca/archimede/fichiers/23265/ch01.html>

Consulté le 20 Septembre 2014.

Jaspard.E

<http://jaspar.genereg.net/>

Consulté le 20 Septembre 2014.

LIFE TECHNOLOGY, APPLIED BIOSYSTEMS.

Overview of SOLiD™ Sequencing Chemistry. 2013.

<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html>

Consulté le 5 Octobre 2014

<https://usegalaxy.org/> Consulté le 29 septembre 2014.

<http://genome.ucsc.edu/> Consulté le 2 Novembre 2014.

Annexes

Annex 1 : Aperçu du script perl :

```
# GREAT version 2.0.2 Species assembly: mm9 Association rule: Two
nearest genes: 1000000 bp max extension, curated regulatory domains
included
# Ontology summary statistics are provided in the footer.
#
# Ontology ID      Desc BinomRank BinomP      BinomBonfP BinomFdrQ
RegionFoldEnrichExpRegions ObsRegions GenomeFrac SetCov
HyperRank HyperP      HyperBonfP HyperFdrQ GeneFoldEnrich
ExpGenes  ObsGenes  TotalGenes GeneSetCov TermCov  Regions
Genes
GO Molecular Function GO:0070053 thrombospondin receptor activity1
5.406876e-6 1.759397e-2 1.759397e-2 3.650116e+1
1.095856e-1 4 3.489986e-4 1.273885e-2 167
7.433897e-2 1.000000e+0 1.000000e+0 1.311349e+1
7.625736e-2 1 3 1.945525e-3 3.333333e-1
DN_only_254, DN_only_255, DN_only_256, DN_only_257 Cd47
GO Molecular Function GO:0046703 natural killer cell lectin-like
receptor binding2 1.563235e-5 5.086767e-2 2.543383e-2
2.780412e+1 1.438636e-1 4 4.581642e-4
1.273885e-2 27 9.038690e-3 1.000000e+0
1.000000e+0 1.311349e+1 1.525147e-1 2 6
3.891051e-3 3.333333e-1
DN_only_114, DN_only_115, DN_only_116, DN_only_117 Clec2d, Clec2i
GO Molecular Function GO:0070737 protein-glycine ligase activity,
elongating 3 5.218025e-5 1.697945e-1 5.659818e-2
1.948018e+2 1.026684e-2 2 3.269695e-5
6.369427e-3 45 2.541912e-2 1.000000e+0
1.000000e+0 3.934047e+1 2.541912e-2 1 1
1.945525e-3 1.000000e+0 DN_only_88, DN_only_89 Ttl110
GO Molecular Function GO:0004478 methionine adenosyltransferase
activity 4 2.020739e-4 6.575485e-1 1.643871e-1
9.866287e+1 2.027105e-2 2 6.455749e-5
6.369427e-3 115 5.019333e-2 1.000000e+0
1.000000e+0 1.967023e+1 5.083824e-2 1 2.....
```

Annex 2: Voies biologiques listées par GREAT (GO Biological Process) :

Voies biologiques impliquées dans le stade DN :

GO Biological Process (13 terms) Global controls

Table controls: Shown top rows in this table: Term annotation count Min: Max: Visualize this table:

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
regulation of T cell activation	1	5.3075e-7	4.5554e-3	3.8021	20	6.37%	13	1.0588e-2	3.6653	15	161	2.92%
regulation of immune system process	3	2.6414e-6	7.5571e-3	2.2394	39	12.42%	5	6.1899e-3	2.3632	34	566	6.61%
regulation of cell-cell adhesion	5	5.3251e-6	9.1411e-3	4.7493	13	4.14%	18	1.9164e-2	6.1711	8	51	1.56%
regulation of lymphocyte activation	15	5.0754e-5	2.9041e-2	2.6912	21	6.69%	38	3.6406e-2	2.8742	16	219	3.11%
regulation of actin cytoskeleton organization	16	5.3360e-5	2.8624e-2	3.3801	15	4.78%	24	2.9875e-2	3.7767	12	125	2.33%
regulation of actin filament-based process	17	5.5175e-5	2.7857e-2	3.3699	15	4.78%	25	3.6067e-2	3.6882	12	128	2.33%
regulation of cellular component size	22	7.2320e-5	2.8215e-2	3.1358	16	5.10%	61	4.4798e-2	3.2784	12	144	2.33%
regulation of leukocyte activation	23	7.3234e-5	2.7329e-2	2.4865	23	7.32%	30	3.2991e-2	2.7447	18	258	3.50%
T cell activation	25	7.8346e-5	2.6898e-2	2.9849	17	5.41%	15	1.8470e-2	3.6475	14	151	2.72%
microtubule bundle formation	26	9.3488e-5	3.0862e-2	8.4535	6	1.91%	41	4.0096e-2	8.9410	5	22	0.97%
T cell differentiation	30	1.1724e-4	3.3544e-2	3.3092	14	4.46%	16	2.1045e-2	4.4158	11	98	2.14%
T cell receptor signaling pathway	32	1.3621e-4	3.6535e-2	6.4066	7	2.23%	29	3.3240e-2	7.6143	6	31	1.17%
regulation of cell activation	36	1.8312e-4	4.3660e-2	2.3338	23	7.32%	51	4.0958e-2	2.5844	18	274	3.50%

The test set of 314 genomic regions picked 514 (3%) of all 20,221 genes.
 GO Biological Process has 8,583 terms covering 15,210 (75%) of all 20,221 genes, and 605,197 term - gene associations.
 8,583 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Voies biologiques impliquées dans le stade DP:

GO Biological Process (9 terms) Global controls

Table controls: Shown top rows in this table: Term annotation count Min: Max: Visualize this table:

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
negative regulation of signal transduction	21	5.8705e-9	2.3993e-6	3.5343	28	18.30%	2	4.5452e-2	3.4435	17	423	7.20%
negative regulation of cell communication	26	5.0215e-8	1.6577e-5	3.1929	28	18.30%	8	3.4997e-2	3.1528	17	462	7.20%
negative regulation of signaling	27	5.2468e-8	1.6679e-5	3.1862	28	18.30%	7	3.8924e-2	3.1596	17	461	7.20%
negative regulation of protein serine/threonine kinase activity	31	1.3222e-7	3.6607e-5	9.5557	10	6.54%	4	3.7153e-2	8.5682	7	70	2.97%
negative regulation of response to stimulus	32	1.5136e-7	4.0599e-5	2.9476	29	18.95%	10	4.6577e-2	2.9100	18	530	7.63%
negative regulation of protein kinase activity	34	1.8673e-7	4.7137e-5	7.0484	12	7.84%	1	6.1658e-2	6.8242	9	113	3.81%
negative regulation of kinase activity	39	4.9871e-7	1.0975e-4	6.4178	12	7.84%	3	3.3545e-2	6.4262	9	120	3.81%
negative regulation of transferase activity	42	9.4424e-7	1.9296e-4	6.0342	12	7.84%	6	2.8208e-2	6.0245	9	128	3.81%
protein dephosphorylation	87	2.2671e-4	2.2366e-2	4.0158	10	6.54%	9	3.2277e-2	5.6288	9	137	3.81%

The test set of 153 genomic regions picked 236 (1%) of all 20,221 genes.
 GO Biological Process has 8,583 terms covering 15,210 (75%) of all 20,221 genes, and 605,197 term - gene associations.
 8,583 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Voies biologiques impliquées dans le stade Commun:

GO Biological Process (2 terms) Global controls

Table controls: Shown top rows in this table: Term annotation count Min: Max: Visualize this table:

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
cell activation	8	4.5080e-6	4.8366e-3	3.8349	16	13.56%	2	3.3268e-2	4.1673	14	333	6.86%
leukocyte activation	9	5.8060e-6	5.5370e-3	3.9870	15	12.71%	1	5.7076e-2	4.5373	13	284	6.37%

The test set of 118 genomic regions picked 204 (1%) of all 20,221 genes.
 GO Biological Process has 8,583 terms covering 15,210 (75%) of all 20,221 genes, and 605,197 term - gene associations.
 8,583 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Annexe 3 :

Le script R utilisé :

```
require(plotrix)

setwd("C:/Users/belhocine/Desktop/")

radmat<- read.delim("Great.txt", row=1, head=T)

pdf("radar.plot.pdf")

radial.plot(as.matrix(radmat),rp.type="l",lwd="3",radial.pos=seq(0,20*pi/11.1,length.out=8),

  label.pos=seq(0,20*pi/11.1,length.out=8),start=pi/2,clockwise=TRUE,

  labels=colnames(radmat),radial.lim=c(0,7))

legend(0.1,6,c("DN", "DP", "Commun"),col=1:3,lty=1,lwd="3") dev.off()
```

Le tableau des scores d'enrichissements pris depuis les tableaux Go biological process pour R :

Pathway	Regulation of T cell activation	T cell differentiation	Negatif regulation of signal transduction	Negatif régulation of cell communication	Cell activation	Leukocyte activation	Negatif regulation signaling	Negative regulation of protein kinase activity
DN	3.66	4.41	0.84	0.91	2.12	1.98	0.54	1.35
DP	1.03	0.79	3.44	3.15	1.02	1.65	3.89	6.42
Commun	4.16	0.62	0.5	0.98	4.16	4.53	1.25	1.65

Annexe 4 : Les différents gènes et enhanceurs exprimés durant les différents stades DN, Commun et DP et selon leurs voies biologiques (GO Biological process) respectives :

DN:

T cell activation:

This term's genomic region-gene association tables (20 regions, 15 genes)

This term's genomic region -> gene association table		This term's gene -> genomic region association table	
Region	Gene (distance to TSS)	Gene	Region (distance to TSS)
DN_only_6	Nck2 (+54,455)	Adora2a	DN_only_162 (+3,913)
DN_only_23	Itpkb (-16,106)	Cd47	DN_only_254 (-84,466), DN_only_255 (-54,616), DN_only_256 (-36,516), DN_only_257 (-7,816)
DN_only_28	Prkcq (+31,842)	Cd5	DN_only_298 (+40,363)
DN_only_29	Il2ra (-6,518)	Clec2i	DN_only_114 (-51,131), DN_only_115 (-48,831)
DN_only_32	Nrarp (-17,277)	Dpp4	DN_only_35 (-29,013), DN_only_34 (+16,987)
DN_only_34	Dpp4 (+16,987)	Igf2	DN_only_130 (+87,358)
DN_only_35	Dpp4 (-29,013)	Il2ra	DN_only_29 (-6,518)
DN_only_58	Pag1 (+200,278)	Itpkb	DN_only_23 (-16,106)
DN_only_67	Ptpn22 (-7,064)	Nck2	DN_only_6 (+54,455)
DN_only_114	Clec2i (-51,131)	Nrarp	DN_only_32 (-17,277)
DN_only_115	Clec2i (-48,831)	Pag1	DN_only_58 (+200,278)
DN_only_121	Tgfb1 (-3,220)	Prkcq	DN_only_28 (+31,842)
DN_only_130	Igf2 (+87,358)	Ptpn22	DN_only_67 (-7,064)
DN_only_162	Adora2a (+3,913)	Tgfb1	DN_only_121 (-3,220)
DN_only_236	Tnfsf11 (-49,151)	Tnfsf11	DN_only_236 (-49,151)
DN_only_254	Cd47 (-84,466)		
DN_only_255	Cd47 (-54,616)		
DN_only_256	Cd47 (-36,516)		
DN_only_257	Cd47 (-7,816)		
DN_only_298	Cd5 (+40,363)		

T cell activation :

This term's genomic region-gene association tables (17 regions, 14 genes)

This term's genomic region -> gene association table		This term's gene -> genomic region association table	
Region	Gene (distance to TSS)	Gene	Region (distance to TSS)
DN_only_16	Bcl2 (+113,266)	Bcl2	DN_only_16 (+113,266)
DN_only_23	Psen2 (-58,070), Itpkb (-16,106)	Dpp4	DN_only_35 (-29,013), DN_only_34 (+16,987)
DN_only_34	Dpp4 (+16,987)	Fyn	DN_only_161 (+86,446)
DN_only_35	Dpp4 (-29,013)	Hs1bp3	DN_only_199 (+31,562)
DN_only_67	Ptpn22 (-7,064)	Irf1	DN_only_185 (+81,985)
DN_only_121	Tgfb1 (-3,220)	Itk	DN_only_181 (+16,466)
DN_only_161	Fyn (+86,446)	Itpkb	DN_only_23 (-16,106)
DN_only_181	Itk (+16,466)	Ppp3cb	DN_only_224 (-71,556)
DN_only_185	Irf1 (+81,985)	Psen2	DN_only_23 (-58,070)
DN_only_199	Hs1bp3 (+31,562)	Ptpn22	DN_only_67 (-7,064)
DN_only_206	Zfp361i (+281,149)	Runx2	DN_only_271 (-505,505), DN_only_270 (-78,955)
DN_only_207	Zfp361i (+270,549)	Tgfb1	DN_only_121 (-3,220)
DN_only_208	Zfp361i (-116,101)	Zfp361i	DN_only_208 (-116,101), DN_only_207 (+270,549), DN_only_206 (+281,149)
DN_only_224	Ppp3cb (-71,556)	Zfp3612	DN_only_282 (+45,886)
DN_only_270	Runx2 (-78,955)		
DN_only_271	Runx2 (-505,505)		
DN_only_282	Zfp3612 (+45,886)		

T cell differentiation:

This term's genomic region-gene association tables (17 regions, 14 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DN_only_16	Bcl2 (+113,266)
DN_only_23	Psen2 (-58,070), Itpkb (-16,106)
DN_only_34	Dpp4 (+16,987)
DN_only_35	Dpp4 (-29,013)
DN_only_67	Ptpn22 (-7,064)
DN_only_121	Tgfb1 (-3,220)
DN_only_161	Fyn (+86,446)
DN_only_181	Itk (+16,466)
DN_only_185	Irf1 (+81,985)
DN_only_199	Hs1bp3 (+31,562)
DN_only_206	Zfp361i (+281,149)
DN_only_207	Zfp361i (+270,549)
DN_only_208	Zfp361i (-116,101)
DN_only_224	Ppp3cb (-71,556)
DN_only_270	Runx2 (-78,955)
DN_only_271	Runx2 (-505,505)
DN_only_282	Zfp361i (+45,886)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Bcl2	DN_only_16 (+113,266)
Dpp4	DN_only_35 (-29,013), DN_only_34 (+16,987)
Fyn	DN_only_161 (+86,446)
Hs1bp3	DN_only_199 (+31,562)
Irf1	DN_only_185 (+81,985)
Itk	DN_only_181 (+16,466)
Itpkb	DN_only_23 (-16,106)
Ppp3cb	DN_only_224 (-71,556)
Psen2	DN_only_23 (-58,070)
Ptpn22	DN_only_67 (-7,064)
Runx2	DN_only_271 (-505,505), DN_only_270 (-78,955)
Tgfb1	DN_only_121 (-3,220)
Zfp361i	DN_only_208 (-116,101), DN_only_207 (+270,549), DN_only_206 (+281,149)
Zfp361i	DN_only_282 (+45,886)

Regulation of cell activation :

This term's genomic region-gene association tables (23 regions, 18 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DN_only_8	Nck2 (+54,455)
DN_only_16	Bcl2 (+113,266)
DN_only_23	Itpkb (-16,106)
DN_only_28	Prikoq (+31,842)
DN_only_29	Il2ra (-6,518)
DN_only_32	Nrarp (-17,277)
DN_only_34	Dpp4 (+16,987)
DN_only_35	Dpp4 (-29,013)
DN_only_58	Pag1 (+200,278)
DN_only_67	Ptpn22 (-7,064)
DN_only_85	Cnr2 (-18,818)
DN_only_114	Clec2i (-51,131)
DN_only_115	Clec2i (-48,831)
DN_only_121	Tgfb1 (-3,220)
DN_only_129	Ptprc (+43,294)
DN_only_130	Igf2 (+87,358)
DN_only_162	Adora2a (+3,913)
DN_only_238	Tnfrsf11 (-49,151)
DN_only_254	Cd47 (-84,466)
DN_only_255	Cd47 (-54,616)
DN_only_256	Cd47 (-36,516)
DN_only_257	Cd47 (-7,816)
DN_only_298	Cd5 (+40,363)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Adora2a	DN_only_162 (+3,913)
Bcl2	DN_only_16 (+113,266)
Cd47	DN_only_254 (-84,466), DN_only_255 (-54,616), DN_only_256 (-36,516), DN_only_257 (-7,816)
Cd5	DN_only_298 (+40,363)
Clec2i	DN_only_114 (-51,131), DN_only_115 (-48,831)
Cnr2	DN_only_85 (-18,818)
Dpp4	DN_only_35 (-29,013), DN_only_34 (+16,987)
Igf2	DN_only_130 (+87,358)
Il2ra	DN_only_29 (-6,518)
Itpkb	DN_only_23 (-16,106)
Nck2	DN_only_8 (+54,455)
Nrarp	DN_only_32 (-17,277)
Pag1	DN_only_58 (+200,278)
Prikoq	DN_only_28 (+31,842)
Ptpn22	DN_only_67 (-7,064)
Ptprc	DN_only_129 (+43,294)
Tgfb1	DN_only_121 (-3,220)
Tnfrsf11	DN_only_238 (-49,151)

DP:

Negative regulation of signal transduction :

This term's genomic region-gene association tables (28 regions, 17 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DP_only_11	Ptgn6 (+5,126), Phb2 (+21,294)
DP_only_31	Dab2ip (-125,450)
DP_only_32	Dab2ip (+209,100)
DP_only_33	Dab2ip (+210,200)
DP_only_46	Cul3 (+114,979)
DP_only_52	Ptprc (+524,431)
DP_only_78	Il6st (-37,527)
DP_only_79	Il6st (-36,177)
DP_only_98	Nog (+223,872)
DP_only_99	Nog (+222,622)
DP_only_100	Socs3 (+17,379)
DP_only_101	Tnfaip3 (-49,532)
DP_only_102	Tnfaip3 (-310,232)
DP_only_108	Dusp6 (-81,664)
DP_only_115	Socs1 (+118,078)
DP_only_116	Socs1 (+116,028)
DP_only_117	Socs1 (+111,378)
DP_only_118	Socs1 (+100,878)
DP_only_119	Socs1 (+86,328)
DP_only_120	Socs1 (+85,778)
DP_only_121	Socs1 (+85,078)
DP_only_122	Socs1 (+80,428)
DP_only_124	Bcl6 (-249,553)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Bcl6	DP_only_124 (-249,553)
Cul3	DP_only_46 (+114,979)
Dab2	DP_only_125 (-872,397)
Dab2ip	DP_only_31 (-125,450), DP_only_32 (+209,100), DP_only_33 (+210,200)
Dusp5	DP_only_143 (-27,207)
Dusp6	DP_only_108 (-81,664)
Il6st	DP_only_78 (-37,527), DP_only_79 (-36,177)
Lats2	DP_only_136 (+16,059)
Nog	DP_only_99 (+222,622), DP_only_98 (+223,872)
Phb2	DP_only_11 (+21,294)
Pten	DP_only_142 (-30,466)
Ptgn6	DP_only_11 (+5,126)
Ptprc	DP_only_52 (+524,431)
Socs1	DP_only_122 (+80,428), DP_only_121 (+85,078), DP_only_120 (+85,778), DP_only_119 (+86,328), DP_only_118 (+100,878), DP_only_117 (+111,378), DP_only_116 (+116,028), DP_only_115 (+118,078)
Socs3	DP_only_100 (+17,379)
Tnfaip3	DP_only_102 (-310,232), DP_only_101 (-49,532)
Trib1	DP_only_127 (-51,908)

Negative regulation of protein kinase activity:

This term's genomic region-gene association tables (12 regions, 9 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DP_only_11	Ptgn6 (+5,126)
DP_only_31	Dab2ip (-125,450)
DP_only_32	Dab2ip (+209,100)
DP_only_33	Dab2ip (+210,200)
DP_only_52	Ptprc (+524,431)
DP_only_101	Tnfaip3 (-49,532)
DP_only_102	Tnfaip3 (-310,232)
DP_only_108	Dusp6 (-81,664)
DP_only_127	Trib1 (-51,908)
DP_only_136	Lats2 (+16,059)
DP_only_142	Pten (-30,466)
DP_only_143	Dusp5 (-27,207)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Dab2ip	DP_only_31 (-125,450), DP_only_32 (+209,100), DP_only_33 (+210,200)
Dusp5	DP_only_143 (-27,207)
Dusp6	DP_only_108 (-81,664)
Lats2	DP_only_136 (+16,059)
Pten	DP_only_142 (-30,466)
Ptgn6	DP_only_11 (+5,126)
Ptprc	DP_only_52 (+524,431)
Tnfaip3	DP_only_102 (-310,232), DP_only_101 (-49,532)
Trib1	DP_only_127 (-51,908)

Negative regulation of cell communication:

This term's genomic region-gene association tables (28 regions, 17 genes)

This term's genomic region -> gene association table		This term's gene -> genomic region association table	
Region	Gene (distance to TSS)	Gene	Region (distance to TSS)
DP_only_11	Ptpn6 (+5,126), Phb2 (+21,294)	Bcl6	DP_only_124 (-249,553)
DP_only_31	Dab2ip (-125,450)	Cul3	DP_only_46 (+114,979)
DP_only_32	Dab2ip (+209,100)	Dab2	DP_only_125 (-872,397)
DP_only_33	Dab2ip (+210,200)	Dab2ip	DP_only_31 (-125,450), DP_only_32 (+209,100), DP_only_33 (+210,200)
DP_only_46	Cul3 (+114,979)	Dusp5	DP_only_143 (-27,207)
DP_only_52	Ptprc (+524,431)	Dusp6	DP_only_108 (-81,664)
DP_only_78	Il6st (-37,527)	Il6st	DP_only_78 (-37,527), DP_only_79 (-36,177)
DP_only_79	Il6st (-36,177)	Lats2	DP_only_136 (+16,059)
DP_only_98	Nog (+223,872)	Nog	DP_only_99 (+222,622), DP_only_98 (+223,872)
DP_only_99	Nog (+222,622)	Phb2	DP_only_11 (+21,294)
DP_only_100	Socs3 (+17,379)	Pten	DP_only_142 (-30,466)
DP_only_101	Tnfaip3 (-49,532)	Ptpn6	DP_only_11 (+5,126)
DP_only_102	Tnfaip3 (-310,232)	Ptprc	DP_only_52 (+524,431)
DP_only_108	Dusp6 (-81,664)	Socs1	DP_only_122 (+80,428), DP_only_121 (+85,078), DP_only_120 (+85,778), DP_only_119 (+86,328), DP_only_118 (+100,878), DP_only_117 (+111,378), DP_only_116 (+116,028), DP_only_115 (+118,078)
DP_only_115	Socs1 (+118,078)	Socs3	DP_only_100 (+17,379)
DP_only_116	Socs1 (+116,028)	Tnfaip3	DP_only_102 (-310,232), DP_only_101 (-49,532)
DP_only_117	Socs1 (+111,378)	Trib1	DP_only_127 (-51,908)
DP_only_118	Socs1 (+100,878)		
DP_only_119	Socs1 (+86,328)		
DP_only_120	Socs1 (+85,778)		
DP_only_121	Socs1 (+85,078)		
DP_only_122	Socs1 (+80,428)		
DP_only_124	Bcl6 (-249,553)		
DP_only_125	Dab2 (-872,397)		
DP_only_127	Trib1 (-51,908)		
DP_only_136	Lats2 (+16,059)		
DP_only_142	Pten (-30,466)		
DP_only_143	Dusp5 (-27,207)		

Negative regulation of signaling :

☛ This term's genomic region-gene association tables (28 regions, 17 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DP_only_11	Ptprn8 (+5,128), Phb2 (+21,294)
DP_only_31	Dab2ip (-125,450)
DP_only_32	Dab2ip (+209,100)
DP_only_33	Dab2ip (+210,200)
DP_only_48	Cul3 (+114,979)
DP_only_52	Ptprc (+524,431)
DP_only_78	Ilf8st (-37,527)
DP_only_79	Ilf8st (-36,177)
DP_only_98	Nog (+223,872)
DP_only_99	Nog (+222,622)
DP_only_100	Socs3 (+17,379)
DP_only_101	Tnfrsf3 (-49,532)
DP_only_102	Tnfrsf3 (-310,232)
DP_only_108	Dusp6 (-81,664)
DP_only_115	Socs1 (+118,078)
DP_only_116	Socs1 (+116,028)
DP_only_117	Socs1 (+111,378)
DP_only_118	Socs1 (+100,878)
DP_only_119	Socs1 (+86,328)
DP_only_120	Socs1 (+85,778)
DP_only_121	Socs1 (+85,078)
DP_only_122	Socs1 (+80,428)
DP_only_124	Bcl8 (240,663)
DP_only_125	Dab2 (-872,397)
DP_only_127	Irf1 (-51,908)
DP_only_138	Lats2 (+16,059)
DP_only_142	Pten (-30,466)
DP_only_143	Dusp5 (-27,207)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Bcl8	DP_only_124 (-240,663)
Cul3	DP_only_48 (+114,979)
Dab2	DP_only_125 (-872,397)
Dab2ip	DP_only_31 (-125,450), DP_only_32 (+209,100), DP_only_33 (+210,200)
Dusp5	DP_only_143 (-27,207)
Dusp6	DP_only_108 (-81,664)
Ilf8st	DP_only_78 (-37,527), DP_only_79 (-36,177)
Lats2	DP_only_138 (+16,059)
Nog	DP_only_99 (+222,622), DP_only_98 (+223,872)
Phb2	DP_only_11 (+21,294)
Pten	DP_only_142 (-30,466)
Ptprn8	DP_only_11 (+5,128)
Ptprc	DP_only_52 (+524,431)
Socs1	DP_only_122 (+80,428), DP_only_121 (+85,078), DP_only_120 (+85,778), DP_only_119 (+86,328), DP_only_118 (+100,878), DP_only_117 (+111,378), DP_only_116 (+116,028), DP_only_115 (+118,078)
Socs3	DP_only_100 (+17,379)
Tnfrsf3	DP_only_102 (-310,232), DP_only_101 (-49,532)
Trib1	DP_only_127 (-51,908)

Negative regulation of protein serine/threonine kinase activity:

☛ This term's genomic region-gene association tables (10 regions, 7 genes)

This term's genomic region -> gene association table

Region	Gene (distance to TSS)
DP_only_11	Ptprn8 (+5,128)
DP_only_31	Dab2ip (-125,450)
DP_only_32	Dab2ip (+209,100)
DP_only_33	Dab2ip (+210,200)
DP_only_101	Tnfrsf3 (-49,532)
DP_only_102	Tnfrsf3 (-310,232)
DP_only_108	Dusp6 (-81,664)
DP_only_138	Lats2 (+16,059)
DP_only_142	Pten (-30,466)
DP_only_143	Dusp5 (-27,207)

This term's gene -> genomic region association table

Gene	Region (distance to TSS)
Dab2ip	DP_only_31 (-125,450), DP_only_32 (+209,100), DP_only_33 (+210,200)
Dusp5	DP_only_143 (-27,207)
Dusp6	DP_only_108 (-81,664)
Lats2	DP_only_138 (+16,059)
Pten	DP_only_142 (-30,466)
Ptprn8	DP_only_11 (+5,128)
Tnfrsf3	DP_only_102 (-310,232), DP_only_101 (-49,532)

Commun:

Cell activation:

This term's genomic region-gene association tables (16 regions, 14 genes)	
This term's genomic region -> gene association table	
Region	Gene (distance to TSS)
Commun_DN_DP_12	Apold1 (-48,993)
Commun_DN_DP_29	Ptpn22 (-664)
Commun_DN_DP_30	Ptpn22 (+586)
Commun_DN_DP_31	Lef1 (-163,638)
Commun_DN_DP_41	B2m (-323)
Commun_DN_DP_58	Tgfb2 (-190,820)
Commun_DN_DP_64	Cplx2 (-168,288)
Commun_DN_DP_65	Cplx2 (-167,512)
Commun_DN_DP_72	Bcl11b (+852,173)
Commun_DN_DP_75	Lcp2 (-500)
Commun_DN_DP_76	Irf1 (-7,291)
Commun_DN_DP_83	Tnfrsf3 (-50,532)
Commun_DN_DP_86	Fyn (+114,896)
Commun_DN_DP_94	Runx2 (-304,805)
Commun_DN_DP_105	Dhrs2 (-27,993)
Commun_DN_DP_107	Lcp1 (+41,269)
This term's gene -> genomic region association table	
Gene	Region (distance to TSS)
Apold1	Commun_DN_DP_12 (-48,993)
B2m	Commun_DN_DP_41 (-323)
Bcl11b	Commun_DN_DP_72 (+852,173)
Cplx2	Commun_DN_DP_64 (-168,288), Commun_DN_DP_65 (-167,512)
Dhrs2	Commun_DN_DP_105 (-27,993)
Fyn	Commun_DN_DP_86 (+114,896)
Irf1	Commun_DN_DP_76 (-7,291)
Lcp1	Commun_DN_DP_107 (+41,269)
Lcp2	Commun_DN_DP_75 (-500)
Lef1	Commun_DN_DP_31 (-163,638)
Ptpn22	Commun_DN_DP_29 (-664), Commun_DN_DP_30 (+586)
Runx2	Commun_DN_DP_94 (-304,805)
Tgfb2	Commun_DN_DP_58 (-190,820)
Tnfrsf3	Commun_DN_DP_83 (-50,532)

Leukocyte activation :

This term's genomic region-gene association tables (15 regions, 13 genes)	
This term's genomic region -> gene association table	
Region	Gene (distance to TSS)
Commun_DN_DP_29	Ptpn22 (-664)
Commun_DN_DP_30	Ptpn22 (+586)
Commun_DN_DP_31	Lef1 (-163,638)
Commun_DN_DP_41	B2m (-323)
Commun_DN_DP_58	Tgfb2 (-190,820)
Commun_DN_DP_64	Cplx2 (-168,288)
Commun_DN_DP_65	Cplx2 (-167,512)
Commun_DN_DP_72	Bcl11b (+852,173)
Commun_DN_DP_75	Lcp2 (-500)
Commun_DN_DP_76	Irf1 (-7,291)
Commun_DN_DP_83	Tnfrsf3 (-50,532)
Commun_DN_DP_86	Fyn (+114,896)
Commun_DN_DP_94	Runx2 (-304,805)
Commun_DN_DP_105	Dhrs2 (-27,993)
Commun_DN_DP_107	Lcp1 (+41,269)
This term's gene -> genomic region association table	
Gene	Region (distance to TSS)
B2m	Commun_DN_DP_41 (-323)
Bcl11b	Commun_DN_DP_72 (+852,173)
Cplx2	Commun_DN_DP_64 (-168,288), Commun_DN_DP_65 (-167,512)
Dhrs2	Commun_DN_DP_105 (-27,993)
Fyn	Commun_DN_DP_86 (+114,896)
Irf1	Commun_DN_DP_76 (-7,291)
Lcp1	Commun_DN_DP_107 (+41,269)
Lcp2	Commun_DN_DP_75 (-500)
Lef1	Commun_DN_DP_31 (-163,638)
Ptpn22	Commun_DN_DP_29 (-664), Commun_DN_DP_30 (+586)
Runx2	Commun_DN_DP_94 (-304,805)
Tgfb2	Commun_DN_DP_58 (-190,820)
Tnfrsf3	Commun_DN_DP_83 (-50,532)

Annex 5 : Les facteurs de transcription :

Facteurs de transcription générés par GRET pour le stade DN et Commun :

Capture écran des facteurs listés:

Back to the [Amadeus supplementary results](#) page.

Set name	Organism	TF/miRNA	Source	Target-set size	Ref.	Description
CREB_Zhang		CREB	CC	2338	[26]	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after fi
E2F4_Cam		E2F	CC	201	[5]	Genes that are bound by both E2F4 and p130 in three different growth arrest conditions, identifi
E2F_Ren		E2F	CC	79	[23]	Targets of E2F4 that are expressed during cell cycle entry, identified by ChIP-chip in quiescent
ERa_Kwon		ESR1	C-DSL	466	[13]	Targets of estrogen receptor α , identified by ChIP-DSL in MCF-7 cells. The array covered the p
ETS1_Hollenhorst		ETS1	CC	1182	[10]	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells. The microarray covered the promoter
HCC-G1S_Whitfield		E2F	Ex	286	[27]	Genes whose expression peaks periodically in the G1/S cell cycle phase.
HCC-G2M_Whitfield		NF-Y	Ex	344	[27]	Genes whose expression peaks periodically in the G2M cell cycle phase.
HNF1a_Odom		HNF1A	CC	206	[14]	Targets of HNF1a, identified by ChIP-chip in hepatocytes. The microarray covered the promote
HNF4a_Odom		HNF4A	CC	1475	[16]	Targets of HNF4a, identified by ChIP-chip in hepatocytes. The microarray covered the promote
HNF6_Odom		ONECUT	CC	212	[16]	Targets of HNF6, identified by ChIP-chip in hepatocytes. The microarray covered the promoter
HSF1_Page	Human	HSF1	CC	328	[24]	Targets of HSF1, identified by ChIP-chip in HeLa cells under heat shock. The microarray cover
ImmuneResponse_GO_Hs		IRF/NFKB	GO	588	[17]	GO Term, Accession - GO:0006955, Name - immune response, Ontology - biological process. I
Nanog_Boyer		NANOG	CC	714	[3]	Targets of Nanog, identified by ChIP-chip in embryonic stem cells. The microarray covered the j
NFKB_Schreiber		NFKB	CC	270	[26]	Genes bound by one of the five NF- κ B subunits in U937 cells before or 1 hour after lipopolysac:
Nrf1_Cam		NRF1	CC	674	[9]	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells. The microarray covered the j
Oct4_Boyer		POU5F1	CC	241	[3]	Targets of Oct4, identified by ChIP-chip in embryonic stem cells. The microarray covered the pn
p53_Kannan		TP53	Ex	38	[12]	Primary targets of p53 genes that were up-regulated under the induction of p53 and inhibition of
Sox2_Boyer		SOX2	CC	575	[3]	Targets of Sox2, identified by ChIP-chip in embryonic stem cells. The microarray covered the pr
SRF_Cooper		SRF	CC	172	[8]	Targets of SRF, identified by ChIP-chip in different cell lines: Jurkat, TIG HA-VSMC, and Be(2)
YY1_XRen		YY1	CC	713	[28]	Targets of YY1 identified by ChIP-chip. The microarray covered the promoter region from -1.3k
Foxp3_Manson		Foxp3	CC	1090	[18]	Targets of Foxp3, identified by ChIP-chip in Foxp3 ^{CD4+} T-cell hybridomas that were transduc
ImmuneResponse_GO_Mm		Irf/Nfkb	GO	329	[17]	GO Term, Accession - GO:0006955, Name - immune response, Ontology - biological process. I
MEF2_Blais		MeF2	CC	25	[2]	Targets of Myocyte Enhancer Factor 2, identified by ChIP-chip in myotubes. The microarray co
MyoD_Blais	Mouse	Myod	CC	102	[2]	Targets of MyoD, identified by ChIP-chip in growing cells or myotubes. The microarray covered
MyoD_Cao		Myod	CC	102	[6]	Targets of MyoD. The ChIP-chip assay was performed in embryonic fibroblasts derived from mi
MyoG_Cao		Myog	CC	78	[6]	Targets of MyoG. The ChIP-chip assay was performed in embryonic fibroblasts derived from mi
Myogenin_Blais		Myog	CC	106	[2]	Targets of myogenin, identified by ChIP-chip on myotubes. The microarray covered the promot
GATA_Pauli	C. elegans	GATA	Ex	1342	[22]	Genes that are expressed in the intestine.
HSF_Machin		Hsf	CC	163	[1]	Targets of heat-shock factor, identified by ChIP-chip in cells of embryos that were heat-shocke
MEF2_Sandmann	Fly	MeF2	CC+Ex	208	[25]	High confidence targets of MeF2. The confidence was measured by a score that took into consi
MSL1_Lagube		Dref	CC	116	[14]	Targets of male-specific lethal complex-1, identified by ChIP-chip on 0-14 h embryos or tightl
Myo_Oryan		Myo/Mav/Med	DamID	714	[20]	Targets of one or more of the proteins dMax, dMnt, and dMyc, identified by DamID assay on Ki
let7_Braeckner		hsa-let-7a	Ex	177	[4]	Genes that were down-regulated in L7-A549 cells following let-7a-3 transfection.
let7_Johnson		hsa-let-7b	Ex	182	[11]	Genes that show altered expression in the presence of exogenously added let-7 in both HeyG2
miR1_Lim		hsa-miR-1	Ex	65	[15]	Genes that were down-regulated in HeLa cells that were transfected with miR-1.
miR10_Linsley		hsa-miR-10	Ex	90	[16]	Genes that were down-regulated in HCT116 Dicer ^{+/+} and DLD-1 Dicer ^{+/+} cells that were transf
miR34_Chang	Human	hsa-miR-34a	Ex	89	[7]	Genes that were down-regulated in HCT116 cells that were transfected with miR-34a.
miR34_He		hsa-miR-34a	Ex	367	[9]	Genes that were down-regulated in HCT116 Dicer ^{+/+} cells that were transfected with miR-34a.
miR10b_Linsley		hsa-miR-10b	Ex	88	[16]	Genes that were down-regulated in HCT116 Dicer ^{+/+} and DLD-1 Dicer ^{+/+} cells that were transf
miR124_Lim		hsa-miR-124	Ex	116	[15]	Genes that were down-regulated in HeLa cells that were transfected with miR-124.
miR373_Lim		hsa-miR-373	Ex	43	[15]	Genes that were down-regulated in HeLa cells that were transfected with miR-373.
miR155_Rodriguez	Mouse	mmu-miR-155	Ex	95	[24]	Genes that were up-regulated in CD4+ T cells of mice deficient for bio/miR-155.

References

