

---

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 DE LA NATURE ET DE LA VIE

*MEMOIRE DE FIN D'ETUDE EN VUE DE L'OBTENTION  
DU DIPLOME DE MASTER FILIERE SCIENCES BIOLOGIQUES*

Spécialité: Génétique et physiologie

Thème :

**IDENTIFICATION DES LNCRNAS  
ASSOCIES AUX SUPERENHANCERS  
DIFFERENTIELLEMENT EXPRIMES CHEZ LES  
PATIENTS LEUCEMIQUES**

Présenté par : BEGHADID Meriem

Devant le jury composé de

<b>EDDAIKRA Atika</b>	MAA	USDB	Présidente
<b>BRADEA Maria Stella</b>	MCA	USDB	Examinatrice
<b>MOHAMED SAID Ramdane</b>	MAA	USDB	Promoteur
<b>BELHOCINE Mohamed</b>	MAB	USDB	Co-Promoteur

Promotion : 2011/2012.

---

## **Remerciements**

*Mes remerciements vont tout d'abord à Dieu le tout puissant qui a éclairé mon chemin.*

*Je voudrais aussi adresser toute ma gratitude à mes deux encadreurs, M. MOHAMED SAID Ramdane et M. BELHOCINE Mohamed pour leur précieuse aide durant le déroulement de mon projet.*

*Je remercie vivement le président du jury Melle EDDAIKRA d'avoir accepté de présider mon jury de soutenance et Mme Maria stella Bradea d'avoir accepté de faire parti du jury et d'évaluer mon mémoire de fin d'études. Votre participation à ma soutenance a été un grand honneur pour moi.*

*Mes remerciements vont aussi à tous mes professeurs, enseignants et toutes les personnes qui m'ont soutenus jusqu'au bout, et qui n'ont pas cessé de me donner des conseils très importants au cours de mon cursus.*

*Mes remerciements vont également à ma famille qui m'a aidé tout au long de mon chemin avec leur patience, encouragement et confiance.*

*Mes remerciements vont également à mes sœurs de cœurs : Yasmine, Widad, Romaiassa. Pour votre aide, écoute, présence et encouragement tout au long du chemin.*

---

## *Dédicace*

*Je dédie ce modeste travail à tous ceux qui ont été à mes cotés et m'ont encouragé dans le meilleur comme dans le pire. Surtout mes chers parents qui m'ont donné la vie et aimé inconditionnellement.*

*Papa... parti trop tôt... je te dédie ce travail en espérant que tu sois fier de ta fille qui t'aimera toujours...*

*Maman pour tous tes sacrifices, ton amour et ta patience même dans les moments les plus difficiles, tu as toujours été notre pilier.*

*Je le dédie également à mes deux petits frères : Nazim & Fateh, ma grande famille pour leurs soutient immense et présence.*

---

## Résumé

Les longs ARNs non codants ( LncRNAs) émergent comme étant des molécules clés dans la régulation de différents processus biologiques. Outre leurs implications dans la physiologie normale, l'expression aberrante des LncRNAs est souvent associée au développement du cancer et de sa progression. Leur propriété fonctionnelle dans la leucémie lymphoïde chronique B ( LLC B) reste peu connue. Afin d'identifier les LncRNAs exprimés de façon différentielle dans la LLC B comparé au cas normal, nous avons analysé leur profil d'expression dans 17 échantillons de données de puces à ADN en utilisant l'outil bioinformatique « Tigr Multi Experiment Viewer » (Tmev). Suite à cela, nous avons réalisé l'analyse fonctionnelle sur la plateforme « Database for Annotation, Visualization and Integrated Discovery » (DAVID). Nous avons trouvé 268 LncRNAs exprimés d'une façon aberrante lors des LLC B (150 surexprimés et 118 réprimés), la plupart des molécules semblent avoir un rôle au niveau des voies promoteurs ou suppresseurs de tumeurs comme la voie P53 ou l'angiogénèse tumorale. Parmi ces LncRNAs différentiellement exprimés nous avons identifié 9 enhancer RNA (eRNA) potentiels associés aux régions régulatrices des superenhancers. Ces résultats peuvent fournir des informations supplémentaires afin de mieux comprendre la maladie, mais ils pourraient aussi servir pour conférer de nouveaux biomarqueurs ou nouvelles cibles thérapeutiques.

**Mots clés :** Cancer, Long ARNs non codant, leucémie lymphoïde chronique B, superenhancer, bioinformatique.

---

## Abstract

Long non-coding RNAs emerge as being one of the key regulatory of numerous biologic processes, other than their implication in the normal physiology, their aberrant expression appear to be associated with cancer's development and progression. In spite of the functional importance, LncRNAs hold in cancer development their implication in the B-Cell chronic lymphocytic leukemia (B CLL) still poorly understood until this day. In order to identify the LncRNAs with aberrant expression in CLL B compared to normal sample, we analyzed 17 microarray datasets containing 12 CCL B patient sample and 5 check samples using the bioinformatics tool "Tigr Multiexperiment Viewer" (Tmev), next we moved to the functional analysis on "Database for Annotation, Visualization and Integrated Discovery (DAVID)". In our study 268 LncRNAs were identified to be differentially expressed in CLL B (150 UPreulated and 118 downregulated), DAVID analysis showed that these target genes were implicated in different oncogenic or tumor suppressing pathways like P53 pathway or angiogenesis. Among these LncRNAs of intrest we identified 9 potential eRNAs associated with the regulatory region of superenhancers. These findings might help the better understanding of the molecular details that occurs in CLL B as well as confers a new biomarker or new therapy target.

**keywords:** Cancer, B Cell chronic lymphocytic leukemia, bioinformatics, Long non-coding RNAs, superenhancers.

## المُلخَص

تتميز جزيئات Long non coding RNAs باعتبارها واحدة من الجزيئات التنظيمية الرئيسية للعديد من العمليات البيولوجية. بالإضافة إلى أهميتها في وظائف الأعضاء أثناء نشاطها العادي، فإن أي خلل في التعبير عن هذه الجزيئات يترافق مع تطور مرض السرطان وتقدمه. بالرغم من أهمية وظيفية هذه الجزيئات، فإن علاقتها بمرض السرطان خاصة بسرطان الدم الليمفاوي المزمن (LLC-B) وكيفية مساهمتها في تطوره لا تزال غير مفهومة تماما الى يومنا هذا.

من أجل التعرف على LncRNAs ذات ذات التعبير المتفاوت في سرطان الدم الليمفاوي المزمن ومقارنتها مع العينات الطبيعية، قمنا بتحليل بيانات 17 شريحة مجهرية تحتوي على 12 عينة لمرضى سرطان الدم الليمفاوي المزمن و 5 عينات شواهد، وذلك باستخدام أدوات البيومعلوماتية مثل Tmev، بجانب انتقالنا إلى التحليل الوظيفي على قاعدة بيانات DAVID. في دراستنا تم تحديد 268 LncRNAs أظهرت تعبيراً تفضلياً في عينات مرضى سرطان الدم الليمفاوي المزمن، كما أظهر تحليل DAVID أن هذه الجينات المستهدفة تتدخل في مختلف المسارات المسببة للورم أو الكابحة له مثل مسار P53. لقد قمنا بتعيين من بين الـ 268 LncRNAs ذات التعبير المتفاوت، 9 منها ناتجين من استنساخ منطقة الـ superenhancers .

هذه النتائج قد تساعد على فهم أفضل للتفاصيل الجزيئية التي تحدث في سرطان الدم الليمفاوي المزمن، كما يساعد على جعل هذه الجزيئات علامات بيولوجية جديدة وهدف لعلاج جديد.

الكلمات الرئيسية: السرطان, Long non coding RNAs , سرطان الدم الليمفاوي المزمن , superenhancers .

---

## **LISTE DES ABBREVIATIONS**

**ADN:** Acide **D**éoxyriboNucléique

**ARN:** Acide **R**iboNucléique

**ARNm:** Acide **R**iboNucléique **m**essenger

**ARNnc:** ARN **n**on codant

**ARNt:** ARN de **t**ransfert

**ARNr:** ARN **r**ibosomique

**ARN poly:** ARN **P**olymérase

**ARN sn:** Small **n**uclear **R**NA [ Petits ARNs Nucléaires]

**ANRIL:** Antisense non-coding **R**NA in the **I**NK4 **L**ocus

**CD:** Cluster de **D**ifférenciation

**CDK:** Cyclin**D**ependent**K**inase

**cirRNA:** circular**R**NA [ARN circulaire]

**ChIP-Seq:** **ch**romatin **I**mmunoprecipitation (ChIP) **S**equencing

**DAVID:** Database for Annotation, Visualization and Integrated Discovery

**ENCODE:** **E**ncyclopedia**O**f **D**N**A** **E**lement [Encyclopédie des Elements d'ADN ]

**eRNAs:** ARNs **e**nhancer

**H3K4me1:** monomethylated histone **H**3 lysine **4**

**HGP:** Human Genome Project [**P**rojet du **G**énome**H**umain ]

**HOTAIR:** **H**OX transcript **a**ntisense **R**NA

**HOTTIP:** **H**OXA transcript at the **d**istal **t**ip

**HOXA :**Homeobox**A**

**HOXD:** homeobox**D**

**IgH:** Immunoglobulin **H**eavy

**IL :** InterLeukine.

---

**KIR:** NK Inhibitor Receptor

**LLC:** Leucémie Lymphoïde Chronique

**LinRNA:** Long intergenic RNA [Long ARN intergénique]

**lncRNA:** Long ARN non codant

**MALAT 1:** Metastasis Associated Lung Adenocarcinoma Transcript 1 [ Transcrit1 Associé au Métastases dans les Adénocarcinomes pulmonaire]

**miRNA :** micro-ARN

**MYC:** myelocytomatosis

**Neat2:** Nuclear enriched abundant transcript 2 [Transcrit nucléaire enrichis et abondant 2 ]

**NF-κB:** Nuclear Factor-kappa B [Facteur Nucléaire kappa B]

**NRON:** Non coding Repressor Of NFAT

**ORF:** Open Reading Frame [Carte de Lecture Ouvert ]

**polyA:** Polyadénylés

**PRC2:** Polycomb Repressive Complex 2 [ Complexe Répresseur du polycomb 2]

**RNA-Seq:** RNA Sequencing

**sRNA:** Small non coding RNA

**siRNA :** Small interfering RNA [ Petit ARN interférent ]

**SAM:** Significance Analysis of Microarrays

**TNF:** Tumor Necrosis Factor

**T-MEV:** Tigr multiexperiment viewer

**Xist:** X-inactive specific transcript [Transcrit spécifique inactivateur du chromosome X]



---

## Glossaire

**cellule souche:** une cellule indifférenciée se caractérisant par la capacité à engendrer des cellules spécialisées par différenciation cellulaire et une capacité à se maintenir par prolifération dans l'organisme (auto-renouvellement) ou, indéfiniment, en culture. Les cellules souches sont présentes chez tous les êtres vivants multicellulaires. Elles jouent en effet un rôle très important dans le développement des organismes ainsi que dans le maintien de leur intégrité au cours de la vie.

**Délétion:** une mutation génétique caractérisée par la perte de matériel génétique sur un chromosome. La taille des délétions varie (d'une paire de bases à toute une région chromosomique) et les délétions peuvent survenir n'importe où sur le chromosome

**Exon:** Chez les organismes eucaryotes, les **exons** sont les fragments d'un ARN primaire qui se retrouvent dans l'ARN cytoplasmique après épissage, par opposition aux introns (fragments d'ARN primaire éliminés au cours de l'épissage).

**Enhancer ou amplificateur:** une région d'ADN (séquence régulatrice) qui peut fixer des protéines pour stimuler la transcription d'un gène. Un gène peut posséder plusieurs amplificateurs qui sont généralement situés assez loin du gène(jusqu'à 100 000 nucléotides). Le gène et l'amplificateur peuvent aussi ne pas être forcément proches l'un de l'autre et même être sur deux chromosomes différents. Cependant le repliement de l'ADN dans le noyau leur permet une proximité physique.

**Epissage:** Chez les eucaryotes, est un processus par lequel les ARN transcrits à partir de l'ADN génomique peuvent subir des étapes de coupure et ligature qui conduisent à l'élimination de certaines régions dans l'ARN final. Les segments conservés s'appellent des exons et ceux qui sont éliminés s'appellent des introns.

---

**Hématopoïèse:** processus physiologique permettant la création et le renouvellement des cellules sanguines.

**Intron:** est une portion de gène, le plus souvent non codante, qui ne se retrouve pas dans l'ARN cytoplasmique après épissage. Il s'oppose à l'exon.

**La transcriptomique:** est l'étude de l'ensemble des ARN messagers produits lors du processus de transcription d'un génome. Elle repose sur la quantification systématique de ces ARNm, ce qui permet d'avoir une indication relative du taux de transcription de différents gènes dans des conditions données. Plusieurs techniques permettent d'avoir accès à cette information, en particulier celle des puces à ADN, celle de la PCR quantitative ou encore celle du séquençage systématique d'ADN complémentaires.

**Mutation:** altération du matériel génétique (ADN ou ARN) d'une cellule ou d'un virus qui entraîne une modification durable de certains caractères du fait de la transmission héréditaire de ce matériel génétique de génération en génération.

**Oncogènes:** sont une catégorie de gènes dont l'expression favorise la survenue de cancers. Ce sont des gènes qui commandent la synthèse d'oncoprotéines (protéines stimulant la division) et déclenchent une prolifération désordonnée des cellules. Le terme vient du grec *onkos*, signifiant vrac, masse ou tumeur.

**Oncogénèse:** ensemble des facteurs et des mécanismes à l'origine des cancers ou tumeurs malignes. Les processus de cancérogenèse sont très variés mais présentent des caractères communs pour la plupart des cancers humains. Les principaux concernent la multiplicité des facteurs qui interviennent et la longue durée entre la première influence et l'apparition d'un cancer.

**Promoteur** : ou **séquence promotrice**, est une région de l'ADN située à proximité d'un gène et indispensable à la transcription de l'ADN en ARN. Le promoteur est la zone de l'ADN sur laquelle se fixe initialement l'ARN polymérase, avant de démarrer la synthèse de l'ARN. Les séquences promotrices sont en général situées en amont du site de démarrage de la transcription.

**Réarrangement de génome** : un élément de la dynamique des génomes au cours duquel un génome voit son organisation générale modifiée par le déplacement, la suppression

---

ou la duplication de parties de sa séquence. Les réarrangements de génome couvrent entre autres : au sein d'une molécule d'ADN (inversion de séquences, délétions d'une ou plusieurs séquences d'ADN, perte d'un bras chromosomique, duplications de séquences) ou entre molécules d'ADN : (duplication de chromosome, perte de chromosome, fusion de deux chromosomes par leurs extrémités, translocation entre deux chromosomes).

**RNA-Seq:** est une technique relativement récente et fait partie de ce qu'on appelle "séquençage de seconde génération" (next-generation sequencing) ou "séquençage à haut débit" (high-throughput sequencing).

**Transcriptome:** est l'ensemble des ARN (messagers, ribosomiques, de transfert et autres espèces d'ARN) issus de la transcription du génome. L'analyse transcriptomique peut caractériser le transcriptome d'un tissu particulier, d'un type cellulaire, ou comparer les transcriptomes entre différentes conditions expérimentales.

**Translocations chromosomiques:** il s'agit d'un échange de matériel entre deux chromosomes non homologues après cassure sur chacun des chromosomes impliqués. Si cet échange s'accompagne d'une perte de matériel génétique, il est déséquilibré, sinon la translocation est dite équilibrée. Dans certains cas exceptionnels, une translocation réciproque peut impliquer trois voire quatre chromosomes différents.

---

---

## Sommaire

Introduction.....	01
<b>CHAPITRE I. RAPPELS BIBLIOGRAPHIQUE .....</b>	<b>03</b>
I.1. Leucémie.....	04
I.2. Les ARNs non codants.....	11
I.3. Longs ARNs non codants. ....	14
I.4. Implication des LncRNAs dans les maladies.....	22
I.5. LncRNAs et enhancers/superenhancers.....	24
I.6. LncRNAs et cancers .....	26
<b>CHAPITRE II : MATERIELS ET METHODES .....</b>	<b>27</b>
II.1. MATERIELS.....	29
II.2. METHODES.....	34
<b>CHAPITRE III. RESULTATS ET DISCUSSION.....</b>	<b>52</b>
III. 1. RESULTATS.....	53
III. 2. DISCUSSION .....	80
<b>CONCLUSION .....</b>	<b>88</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	
<b>ANNEXE</b>	

---

---

## Liste des figures

Figure 01 : Schéma de la translocation t(14 ;18) (Larsen et al., 1994).....	08
Figure 02 : Dogme central de la biologie moléculaire ( Lodish et <i>al.</i> , 2000).....	12
Figure 03 : Chronologie de la découverte des LncRNAs en régulation (Rinn et Chang, 2012).....	15
Figure 04 : Long ARN non codant de classe intergénique ( Rinn et Chang, 2012).....	17
Figure 05 : Long ARN non codant classe Intronique ( Rinn et Chang, 2012).....	17
Figure 06 : Long ARN non codant de classe « antisense » ( Rinn et Chang, 2012).....	18
Figure 07 : Long ARN non codant de classe Divergent ( Rinn et Chang, 2012).....	18
Figure 08 : Fonctions cellulaires des LncRNAs (Gutschner et Diederichs, 2012).....	20
Figure 09 : Différentes fonction régulatrices des LncRNAs.....	21
Figure 10 : Résultat chip-sep indiquant l'enrichissement des régions superenhancers enfacteurs de transcription (Hnisz et al. 2013).....	25
Figure 11 : plateforme de travail R.....	30
Figure 12 : plateforme de travail MEV.....	31
Figure 13 : plateforme de travail DAVID.....	32
Figure 14 : plateforme de travail GALAXY.....	33
Figure 15 : Page d'accueil de la grande base de données d'expression GEO.....	34
Figure 16 : Sélection des données brutes sur GEO.....	35
Figure 17 : Téléchargement des données brutes de GEO.....	36
Figure 18 : Données sous fichier .TAR sauvegardé sur notre ordinateur.....	37
Figure 19 : Décompression du fichier initial « .tar » et placement des fichiers « .CEL » sur notre ordinateur.....	38
Figure 20 : Ouverture de la Plateforme R.....	39

---

Figure 21 : Installation de Bioconductor.....	40
Figure 22 : Réannotation des données de microarray et Refseq (Du et <i>al.</i> , 2013).....	45
Figure 23 : Reannotation de la puce type U133 plus2.....	46
Figure 24 : Les 1ères étapes de l’analyse Tmev.....	47
Figure 25 : Début de l’analyse par la soumission des listes d’intérêt.....	49
Figure 26 : Choix des différentes bases de données disponible pour notre analyse.....	50
Figure 27 : Figure représentant l’intersection entre les superenhancers et la liste des LncRNAs d’intérêt.....	51
Figure 28 : visualisation d’une puce ADN.....	53
Figure 29 : Tableau des données des puces normalisés.....	54
Figure 30 : Résultat de la réannotation des puces U133 plus2.....	55
Figure 31 : Tableau des LncRNAs totaux.....	56
Figure 32 : Comparaison de la distribution des données avant et après normalisation.....	57
Figure 33 : Niveau d’expression des genes totaux.....	58
Figure 34 : Représentation graphique de l’expression différentielle des gènes totaux.....	59
Figure 35 : Heatmap représentant l’expression différentielle des gènes.....	60
Figure 36 : Heatmap représentant l’expression des LncRNAs.....	63
Figure 37 : Représentation graphique SAM des LncRNAs différentiellement exprimés.....	64
Figure 38 : Heatmap représentant l’expression différentielle des LncRNAs.....	65
Figure 39 : Tableau résultats DAVID des voies biologique cible des LncRNAs.....	68
Figure 40 : La voie de signalisation P53.....	70
Figure 41 : La voie de signalisation calcique.....	71
Figure 42 : Résultats DAVID des voies dérégulées liées aux LncRNAs DOWN regulés.....	72
Figure 43 : Les voies de signalisation tumorales ( activation de l’angiogénèse).....	73
Figure 44 : Les voies de signalisation tumorales (blocage de la différenciation).....	74
Figure 45 : La voie de signalisation MAPK.....	75
Figure 46 : Histogramme représentant l’implication des LncRNAs réprimés dans différents processus biologiques.....	75

---

Figure 47 : Annotation fonctionnelle chart DAVID pour les DOWN régulés.....	77
Figure 48 : La voie de signalisation des chimiokines.....	77
Figure 49 : Histogramme représentant les voies renfermant la plupart des gènes réprimés d'intérêt.....	78
Figure 50 : Résultat de l'intersection des superenhancers avec les LncRNAs.....	79
Figure 51 : Implication potentiel du LncRNA « 241876 _at » dans la voie P53 lors de la LLC-B résultant de l'analyse DAVID.....	81
Figure 52 : Les gènes dérégulés lors de la LLC-B résultant de l'analyse DAVID.....	85

---

## Liste des tableaux

Tableau I : Classification de la leucémie.....	05
Tableau II : Classification anatomo-clinique de Rai (Rai <i>et al.</i> , 1975).....	09
Tableau III : Classification anatomo-clinique de Binet (Binet <i>et al.</i> , 1981).....	10
Tableau IV : Comparaison des ARNs de même taille codants et non codants.....	16
Tableau V : Liste des gènes positivement régulés chez les leucémiques.....	61
Tableau VI : Liste des gènes négativement régulés chez leucémiques.....	62
Tableau VIII : Liste des LncRNAs surexprimés dans les Leucémies chroniques B.....	66
Tableau IX : Liste des LncRNAs réprimés dans les Leucémies chroniques B.....	67
Tableau X : Tableau présentant les voies dérégulées lors de la surexpression de notre liste de gènes.....	76



---

## INTRODUCTION

L'hématopoïèse est un processus hiérarchisé impliqué dans la formation des différentes cellules sanguines, dont l'origine commune est les cellules souches hématopoïétiques qui se trouvent au niveau de la moelle osseuse (Gould and Brooker, 2000). Dans certains cas, il peut y avoir des événements mutationnels lors de ce processus qui est à l'origine de diverses pathologies malignes telles que la leucémie lymphoïde chronique de type B (Roger Lacave et *al.*, 2005).

La leucémie lymphoïde chronique de type B (LLC-B) est le type de leucémie le plus répandu chez les adultes, caractérisée par la prolifération incontrôlée des cellules B génétiquement transformées ayant un blocage d'apoptose. Les anomalies génétiques et cliniques sont bien étudiées, quant aux processus moléculaires impliqués dans cette dernière, ils sont encore peu compris (Bullrich et *al.*, 2001).

Les aberrations génomiques les plus fréquentes dans les leucémies sont généralement dues à des mutations oncogéniques. Ces oncogènes peuvent avoir plusieurs fonctions régulatrices dans le cytoplasme comme acteur d'une voie de signalisation ou dans le noyau comme facteur de transcription (Dalla et Gaidano, 2001).

Environ 70% du génome est transcrit en ARN non codants contre 2% seulement des gènes codant des protéines, laissant prédire une implication importante dans les différents événements cellulaires (Gibb et *al.*, 2011, Gutschner and Diederichs, 2012). En effet, plusieurs études ont mis en évidence l'importance du rôle que détiennent les ARN non codants dans le processus tumoral, surtout la classe des Longs ARNs non codants (LncRNAs) qui ont reçu une attention accrue (Nagano et Fraser, 2011).

Les propriétés fonctionnelles que détiennent les molécules de LncRNAs durant la pathogénèse de la LLC-B restent à ce jour peu caractérisés. Par conséquent, durant notre

---

modeste travail, nous nous sommes intéressés à la corrélation entre les LncRNAs ayant une expression aberrante lors de la LLC-B et le processus tumoral de celle-ci. Afin d'identifier les voies de signalisation dérégulées suite à la dérégulation d'expression des gènes des LncRNAs, nous utiliserons une panoplie d'outils bioinformatiques.

---

---

## **CHAPITRE I. RAPPELS BIBLIOGRAPHIQUES**

---

## I.1. Leucémies

Décrite pour la 1<sup>ère</sup> fois en 1847 par le médecin Allemand Rudolf Virchow, la leucémie est l'une des pathologies causées par une dérégulation de l'homéostasie hématopoïétique (Michaux, 2003).

Elle se détermine par la production de cellules sanguines, plus précisément les globules blancs génétiquement erronés, à cause des altérations qui se produisent durant le développement de leurs précurseurs dans la moelle osseuse au cours de l'hématopoïèse, modifiant ainsi leur identité et fonction initiale ainsi que leur caractère de différenciation et prolifération. Ces cellules, dites immatures, vont se multiplier anarchiquement et migrer dans le sang, aussi s'entasser dans la moelle osseuse causant ainsi un déséquilibre qui va conduire à la production amoindrie des autres cellules sanguines vitales à l'organisme, ayant aussi la possibilité de migrer vers d'autres organes. Il en existe quatre types (Roger Lacave et al., 2005, Marieb et al., 2014).

### I.1.1. Classification des leucémies

La leucémie regroupe différents types, en fonction du type de cellule atteinte :

**Leucémie lymphoïde** : Quand les globules blancs touchés sont les lymphocytes B ou T.

**Leucémie myéloïde** : Quand les globules blancs touchés sont les polynucléaire.

Et en fonction de la vitesse de prolifération et évolution de la maladie :

**Leucémies aiguës** : Dans ce cas, l'évolution de la maladie est rapide, les cellules altérées se multiplient au sein de la moelle osseuse et même du sang à une étape prématurée de leur différenciation (plus exactement appelée les blastes), il s'en suit une apparition rapide des symptômes.

**Leucémies chroniques** : Contrairement à la 1<sup>ère</sup>, le cas chronique est caractérisé par une évolution assez lente (voire des années), les cellules altérées qui se multiplient au sein de la

moelle osseuse parviennent presque à maturité, et il s'en suit une apparition tardive des symptômes.

La combinaison de ces deux caractéristiques nous permet de soulever quatre types de Leucémie (Peacock and Asselin, 1999, Marieb et *al.*, 2014).

Type cellulaire	Lymphoïde	Myéloïde
Vitesse d'évolution		
Rapide (aigue)	Leucémie lymphoïde aigue	Leucémie myéloïde aigue
Lente (chronique)	Leucémie lymphoïde chronique	Leucémie myéloïde chronique

**Tableau I : Classification de la leucémie**

### I.1.2 Leucémie lymphoïde chronique

La leucémie lymphoïde chronique a été différenciée vers 1880 grâce aux colorations cytologiques d'Ehrlich, qui ont permis de distinguer entre les lignées cellulaires lymphoïde et myéloïde (Triadou 2000).

La LLC est la plus répandue des leucémies chez l'adulte, absente chez l'enfant, elle n'a toujours pas de cure finale, néanmoins, elle se développe lentement chez la plupart des patients. Elle est à l'origine d'une multiplication et accumulation incontrôlée de lymphocytes monoclonaux matures dans le sang, la moelle osseuse, et les organes lymphoïdes secondaires (ganglions lymphatiques, rate) (Rozman and Montserrat, 1995, Chiorazzi et *al.*, 2005).

Dans la plupart du temps, les cellules atteintes sont du type B, nous parlerons dans ce cas de la leucémie lymphoïde chronique B (LLC-B). Sur laquelle nous allons concentrer notre étude actuelle.

Au cours de la LLC-B, les cellules B deviennent anormales, prolifèrent et elles s'échappent à la mort cellulaire programmée et ne répondent plus aux facteurs pro-apoptotiques, l'organisme continue simultanément la production de ces cellules atteintes au même titre que les lymphocytes B normaux. Suite à ce déficit apoptotique, tous ces clones du lymphocyte anormal de départ continuent de se multiplier et donc de s'accumuler dans le sang et la moelle osseuse (Harris et *al.*, 1999). Les cellules B tumorales ont un profil moléculaire résistant à l'apoptose par la surexpression des protéines anti-apoptotiques tels que : BCL-2, Bcl-xl, Mcl-1 et la sous-expression des protéines pro-apoptotiques Bax et Bcl-xs (Dighiero and Hamblin, 2008).

### **I.1.2.1 Facteurs pronostiques de la LLC**

#### **a) - Facteurs génétique**

Les facteurs génétiques semblent jouer un rôle dans la pathogénie de la maladie. Elle représente 3,5% de toutes les leucémies de l'adulte au Japon, alors qu'elle atteint jusqu'à 38% au Danemark. Ce faible taux d'incidence dans les populations orientales est maintenu dans les populations migrantes et chez leur descendance, ce qui permet d'exclure l'existence de phénomènes environnementaux ayant un effet sur les prédispositions génétiques (Weiss, 1979, Boggs et *al.*, 1987).

De plus, des études épidémiologiques démontrent l'existence, dans 5 à 10 % des cas, de prédispositions familiales (Neuland et *al.*, 1983, Linet et *al.*, 1989, Yuille et *al.*, 2000), avec au moins deux individus infectés dans une même famille. Le risque de développer la LLC quand on est parents au premier degré de patients atteints de LLC est de quatre fois plus important que la population en général, mais reste quand même faible (Capalbo et *al.*, 2000). Cette prédisposition familiale est accompagnée par ce que l'on appelle un phénomène d'anticipation (Horwitz et *al.*, 1996, Yuille et *al.*, 2000, Wiernik et *al.*, 2001), c'est-à-dire une apparition précoce de la maladie avec une évolution plus sévère pour les descendants des

patients atteints de LLC. Malgré cela, la LLC familiale est en termes clinique, moléculaire et biologique, très similaire aux cas classiques.

## b)- Facteurs cytogénétiques

En 2ème lieu, les translocations chromosomiques oncogéniques semblent être un des événements aboutissant à la formation et développement des cellules B transformés. (Dalla et Gaidano 2001).

Les Lésions cytogénétiques sont rares dans le clone leucémique au début de l'évolution de la maladie et ne sont donc pas susceptibles d'être des facteurs inducteurs de LLC. Néanmoins, certains semblent apparaître alors que la maladie progresse (Chiorazzi et al. 2005).

Quelques-unes des aberrations chromosomiques souvent observées sont citées ci-dessous :

- **Délétion 13q14** : est la plus courante aberration dans les LLC B, observée dans plus de 50% des cas (Plass et al., 2007). Cette région dénommée «minimal deletedregion » (MDR) code pour deux microARN, miR15a et miR16-1, qui sont exprimés faiblement ou complètement réprimés lors des LLC (Klein et al., 2010).
- **Délétion 11q22-q23** : contient le gène ATM (ataxia-telangiectasiamutated) (Dohner et al., 2000) impliqué dans la réparation de l'ADN endommagé, la maladie présentant cette délétion est d'évolution rapide (Zenz et al., 2010).
- **Délétion 17p13** : Cette région contient le gène de la p53, protéine répresseur des tumeurs, il en résulte donc un impact négatif sur la suppression tumorale (Zenz et al., 2010).
- **Translocation t(14 ;18)** : Cette translocation résulte en la surexpression de BCL-2 (Sen et al., 2002, Roulland et al., 2006).

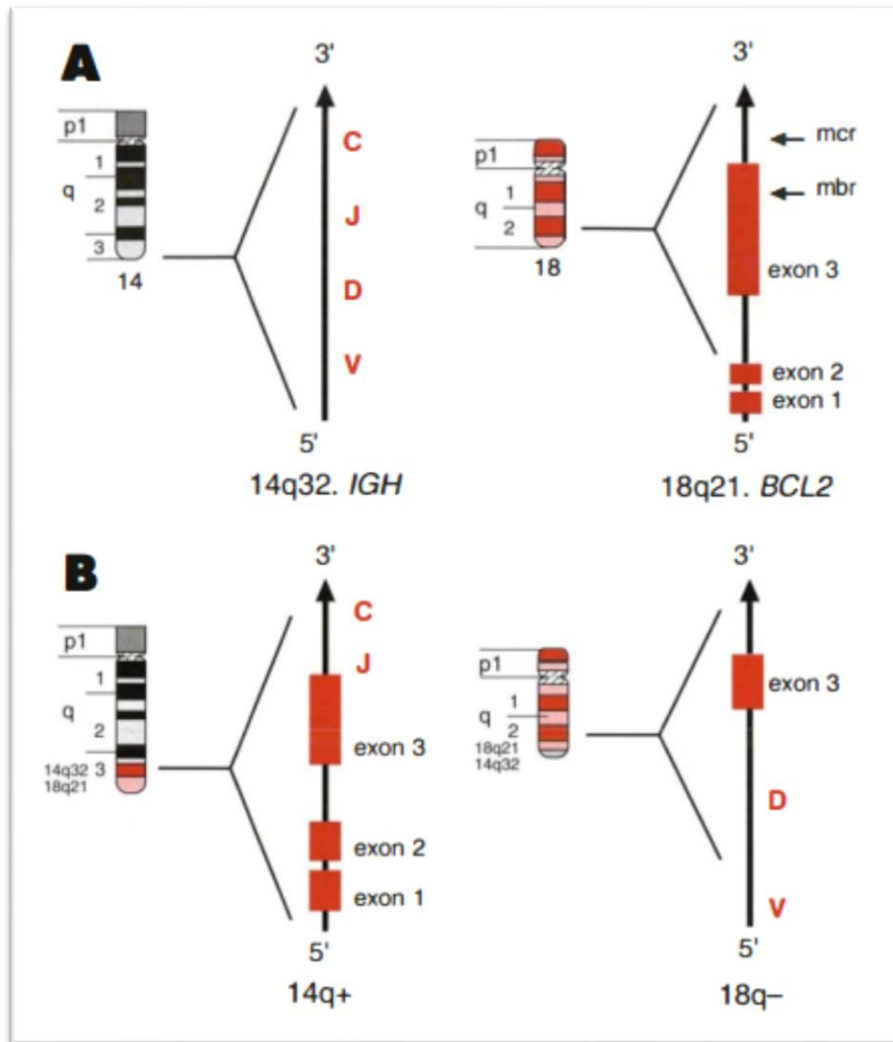


Figure 01 : Schéma de la translocation t(14 ;18) (Larsen et *al.*, 1994)

### I.1.2.2. Classification

Les premières classifications pronostiques ont été proposées par Rai (Rai et *al.*, 1975) et Binet (Binet et *al.*, 1981). Elles reposent sur les données de l'examen clinique (présence d'adénopathies, d'une hépatomégalie ou d'une splénomégalie) et de l'hémogramme (hyper lymphocytose, anémie, thrombopénie). Elles sont toujours d'actualité dans la pratique quotidienne et les essais cliniques, et sont détaillées dans les tableaux 2 et 3.



La médiane de survie des patients de pronostic intermédiaire ou défavorable décrite dans ces deux tableaux correspond à celle rapportée par Rai et Binet respectivement en 1975 et 1981. Depuis l'avènement de l'immuno-chimiothérapie, la survie de ces patients est nettement améliorée.

**Tableau II** Classification anatomo-clinique de Rai (Rai *et al.*, 1975)

		<b>Critères de définition</b>	<b>Patients (%)</b>	<b>Survie médiane (mois)</b>
Bon pronostic (0)	<b>Stade 0</b>	Lymphocytose > 5.10 <sup>9</sup> /L	31	> 150
Pronostic intermédiaire (I+II)	<b>Stade I</b>	Lymphocytose + adénopathie(s)	35	101
	<b>Stade II</b>	Lymphocytose + hépato ou splénomégalie +/- adénopathies	26	71
Mauvais pronostic (III+IV)	<b>Stade III</b>	Lymphocytose + anémie(Hb<110 g/L) +/- adénopathies et organomégalies	6	19
	<b>Stade IV</b>	Lymphocytose + thrombopénie (pq<100 G/L) +/- adénopathies et organomégalies	2	19

**Tableau III :** Classification anatomo-clinique de Binet (Binet *et al.*, 1981)

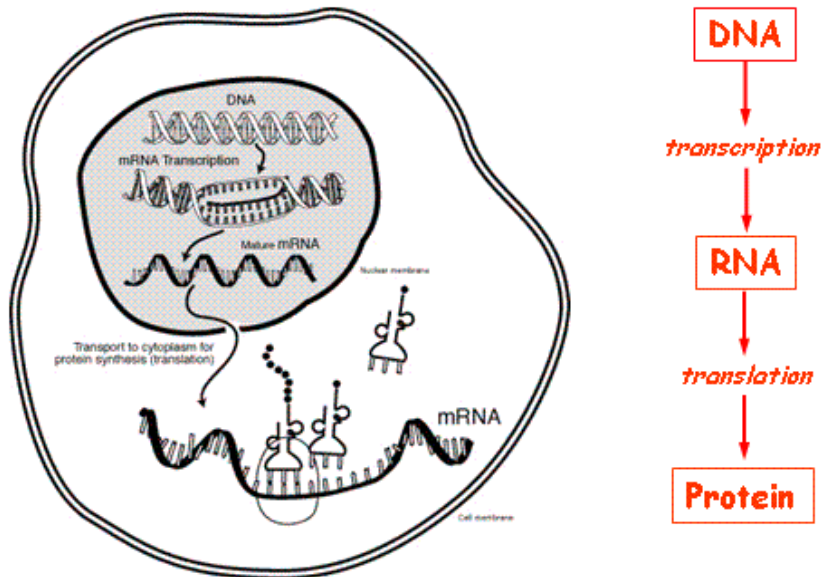
	<b>Critères de définition</b>	<b>Patients (%)</b>	<b>Survie médiane (ans)</b>
--	-------------------------------	---------------------	-----------------------------

<b>Stade A Bon pronostic</b>	Lymphocytose, Hb>100 g/L, et Pq> 100.10 <sup>9</sup> /L Moins de 3 aires lymphoïdes atteintes*	63	10 ans, comparable à la population de même âge
<b>Stade B Pronostic intermédiaire</b>	Lymphocytose, Hb> 100 g/L, et Pq> 100.10 <sup>9</sup> /L Plus de 3 aires lymphoïdes atteintes*	30	5 ans
<b>Mauvais pronostic (III+IV)</b>	Lymphocytose, Hb< 100 g/L, et/ou Pq< 100.10 <sup>9</sup> /L Quel que soit le nombre d'aires lymphoïdes atteintes	7	2 ans

## I.2. ARNs non codants

L'histoire fondamentale des êtres vivants avait été un récit des interactions entre les gènes, dans la forme d'ADN, et des protéines que les gènes codent et qui réalisent le dur travail de faire en sorte que les organismes vivants gardent l'homéostasie. En effet, le dogme central de la biologie moléculaire stipule que l'information génétique codée dans le génome est transcrite en forme d'ARN messager qui servira par la suite de matrice pour la synthèse des protéines nécessaires au bon fonctionnement cellulaire, que ce soit protéine de structure ou fonctionnelle (Raven *et al.*, 2011).

Cependant, les dernières années ont observé la montée en puissance d'une myriade d'ARN cellulaires autres que les ARNm, dits non codants, ces molécules peuvent avoir plusieurs tailles et formes d'où le nom petits ARNs non codants (sRNA: Small non coding RNA) (Costa, 2007). Aujourd'hui, le monde du Transcriptome est devenu encore plus complexe avec la découverte de l'ARN qui ressemblait à l'ARNm dans la structure, la longueur et le processus de l'épissage, mais qui ne code des protéines ; d'où le nom Long ARN non codants (lncRNAs) (Guttman and Rinn, 2012).



**Figure 02 : Dogme central de la biologie moléculaire ( Lodish et al., 2000)**

Avec l'achèvement du projet du génome humain ( HGP), il a été révélé que seulement 1.5 % du génome code pour des protéines, le 98.5 % restant est non codants considéré comme des séquences d'ADN « poubelle» accumulée au cours de l'évolution (Feng et al., 2014).

Cependant, au cours du projet ENCODE (Encyclopedia of DNA element) lancé par le NHGRI ultérieurement, il a été découvert d'une manière inattendu que 75% du génome humain est activement transcrit en ARNs (Carvunis et al., 2012). Ces données de séquençage ont donc révélé que les transcrits des longs et petits ARNs non codants dépassent le nombre des gènes codants, indiquant ainsi que le génome humain est bien plus actif qu'on ne le pensait (Fejes-Toth et al., 2009).

Les ARNs non codants sont des molécules transcrites mais non traduites en protéines (Read et al., 2008). Ils contiennent les familles d'ARNnc fonctionnels connus qui sont les ARN ribosomiques et les ARN de transfert nécessaires à la traduction (Gargaud and Despois, 2003).

Ils se divisent en deux familles selon leur taille : nous avons les petits ARNs et les longs ARNs non codants (Kapranov et al., 2002).

## **I.2.1. Types d'ARNs non codants**

### **I.2.1.1 Petits ARNs non codants**

-Les petits ARNs nucléaires ( ARN sn) : Participent au mécanisme de l'épissage des ARNm, et sont localisés au noyau.

-Les ARNs nucléaires ( ARNsno) : Localisés dans le nucléole, sont indispensables pour la maturation des ARNr .

-Les micro-ARNs (miRNA) : Ces petits ARNs effectuent une fonction de régulation de l'expression génique en se liant à un ARNm complémentaire. (Read et *al.*, 2008)

-Les petits ARN interférents (siRNA) : Ils ont un rôle dans la régulation génique à travers la dégradation de l'ARNm. (Fjose et *al.*, 2001).

### **I.2.1.2 Longs ARNs non codants**

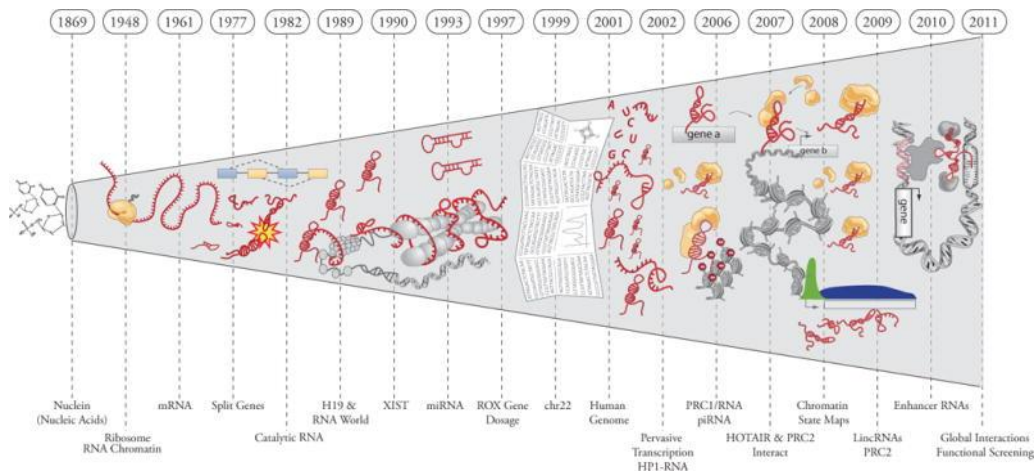
Ces molécules se trouvent impliqués dans la régulation de différentes fonctions biologiques (Gibb et *al.*, 2011).

Ce dernier type d'ARNnc fera l'objet de notre étude qui se focalisera sur son expression différentielle entre les tissus normaux et pathologiques.

### I.3. Longs ARNs non codants

Longtemps délaissés et sous-estimés pour leur incapacité à coder des protéines, ces dernières années les long ARNnc reçoivent beaucoup d'attention. Ces molécules sont, selon les récentes trouvailles, potentiellement impliquées dans beaucoup de processus cellulaires, tels que le cycle cellulaire, régulation de la transcription ...etc (Geisler and Coller, 2013).

Le 1<sup>er</sup> Long ARN non codant « *Imprinted maternally expressed transcript* » ou H19 fut le 1er à avoir été découvert en 1990, bien avant l'utilisation du terme « long ARN non codant ». Cette molécule était considérée comme différente d'un ARNm codant classique mais un ARN cytoplasmique atypique car celui-ci n'avait pas de protéine correspondante, issu d'un gène caractérisé comme inusuel (Brannan et *al.*, 1990). Même si d'autres molécules furent découvertes après l'H19 comme le XIST découvert et caractérisé au début des années 1990 (Brockdorff et *al.*, 1992, Brown et *al.*, 1992), ce n'est vraiment qu'au début des années 2000 que l'émergence de l'étude des longs transcrits s'est effectuée grâce aux avancés technologiques, et le dévoilement du séquençage complet du génome humain, ainsi que le projet ENCODE, et beaucoup d'autres études indépendantes ont permis ainsi de mieux essayer de comprendre la structure de ces molécules et leurs fonctions.



**Figure 03 : Chronologie de la découverte des LncRNAs en régulation (Rinn et Chang, 2012)**

### I.3.1. Description

La partie la plus transcrite du génome est celle des gènes codants les LncRNAs, localisée un peu partout sur les chromosomes (Fitzgerald and Caffrey, 2014). Des milliers de gènes en Long ARNnc sont transcrits chez tous les organismes eucaryotes notamment l'humain (Derrien et *al.*, 2012).

Les lncRNAs sont des transcrits de plus de 200pb de longueur (Gibb et *al.*, 2011, Ulitsky and Bartel, 2013). La plupart du temps, ils ne codent pas de protéine fonctionnelle (Guttman et *al.*, 2013). Quelques études suggèrent la possibilité que ces molécules puissent bien coder des petits peptides (Kondo et *al.*, 2010, Ingolia et *al.*, 2011). Par exemple le SRA (Lanz et *al.*, 1999).

### I.3.2. Transcription

Ces transcrits présentent des similitudes avec les ARNm, leur transcription est assurée généralement (mais pas exclusivement) par la polymérase II (Guttman et *al.*, 2009, Rinn and Chang, 2012). Subissant les modifications post transcriptionnelles telles que l'ajout d'une queue ou coiffe à leur extrémité 5' et une polyadénylation à leur extrémité 3' et un épissage (Guttman et *al.*, 2009, Djebali et *al.*, 2012). Par contre, leur niveau d'expression est moindre, comparé à ces derniers (Dong et *al.*, 2014). (Tableau III).

**Tableau III : Comparaison des ARNs de même taille codants et non codants**

mRNA	lncRNA
Expression tissu spécifique	Expression tissu spécifique
Forme une structure secondaire	Forme une structure secondaire
Subit les modifications post-transcriptionnelles la queue 5', polyadénylation, épissage	Subit les modifications post-transcriptionnelles la queue 5', polyadénylation, épissage
Rôle important dans le développement des maladies	Rôle important dans le développement des maladies
Transcrit codant	Transcrit non codant, fonction régulatrice
Bien conservé entre les espèces	Peu conservé entre les espèces
Localisé dans le noyau et le cytoplasme	Localisé principalement dans le noyau
Niveau d'expression: faible à élevé	Niveau d'expression: très faible à modéré

### I.3.3 Localisation cellulaire

L'expression des Long ARNnc est tissu spécifique (Mariner et *al.*, 2008), suggère que leur expression est soumise à un contrôle rigoureux. Contrairement aux ARNm qui se localisent au cytoplasme après leurs modifications, les LncRNAs sont localisés en permanence au noyau (Kapranov et *al.*, 2007) avec des exceptions où nous pouvons les trouver sélectivement localisés au cytoplasme (Louro et *al.*, 2009).

### I.3.4 Classification

Leurs classifications et dénomination se font respectivement par rapport à leurs gènes codant voisins et sa protéine. Jusqu'à ce jour, ce n'est pas standardisé, mais les chercheurs peaufinent leurs critères de classification avec les nouvelles études (Kung et *al.*, 2013).

Nous pouvons les diviser en cinq grandes classes : sens, anti sens, intronique, divergent et inter génique (Ponting et *al.*, 2009, Rinn and Chang, 2012).



### a)- Intergénique

Cette classe de Long ARNnc se localise entre deux régions génomiques codantes, et est nommée d'après le gène codant voisin à leur extrémité 3' (Guttman et *al.*, 2009).

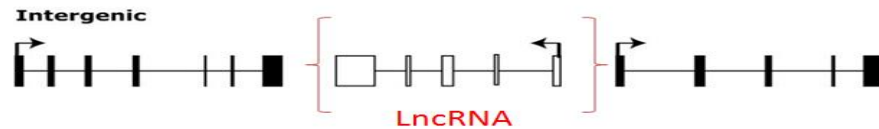


Figure 04 : Long ARN de classe intergénique (Rinn et Chang, 2012)

Ils sont généralement (5000b a 1kb) éloignés d'un gène codant évitant ainsi un chevauchement possible, et ils sont généralement bien conservés (Guttman et *al.*, 2009).

### b)- Intronique :

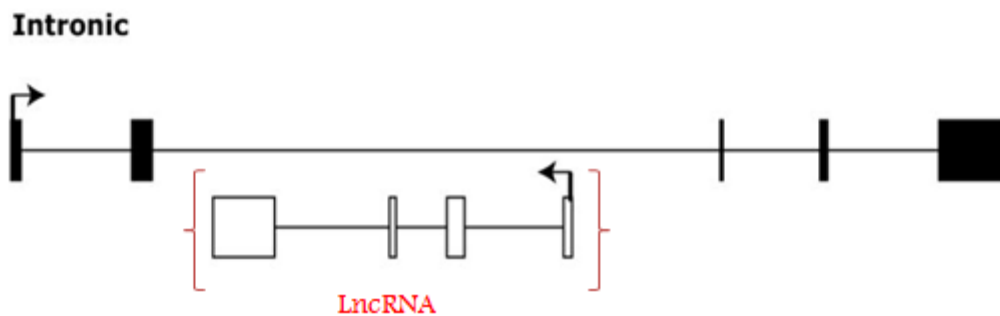


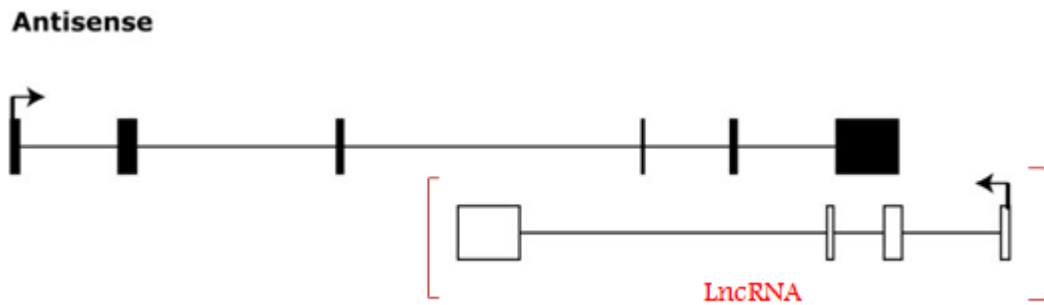
Figure 05 : Long ARN non codant classe Intronique ( Rinn et Chang, 2012)

Les LinRNAs ne chevauchent aucun exon, car ils se localisent dans une région intronique, et peuvent être initié d'une direction ou d'une autre (Guttman et *al.*, 2009).

**c)- Sens et anti sens :**

**Sens :** transcrit à partir d'un brin sens d'un gène codant une protéine ou dans une région inter génique.

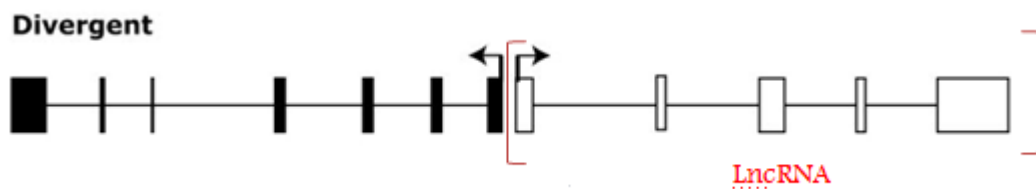
**Anti sens :** transcrit à partir d'un brin anti-sens d'un gène codant une protéine ou dans une région intergénique. (Rinn and Chang, 2012)



**Figure 06 : Long ARN non codant de classe « antisense » ( Rinn et Chang, 2012)**

**d)- Divergent :**

Transcrit à partir d'un promoteur d'un gène codant et peut aller dans les deux sens direct et indirect. (Chakalova et *al.*, 2005, Guttman et *al.*, 2009).



**Figure 07 : Long ARN non codant de classe Divergent ( Rinn et Chang, 2012)**

### I.3.5. Fonctions

Bien que des milliers de Longs ARNncs soient codés dans le génome humain, seulement quelques-uns d'entre eux ont été caractérisés en détails. Il est connu qu'ils participent à la régulation génique via différents mécanismes (Wang et *al.*, 2011b) Et ils se trouvent impliqués dans diverses voies cellulaires (Rinn and Chang, 2012) Leur faible expression ainsi que leurs tissus spécificités suggèrent un rôle régulateur plutôt qu'effecteur directe (Kaikkonen et *al.*, 2011).

Des exemples sont cités ci-dessous :

### **A. Guide:**

Les longs ARNnc peuvent servir de guide pour les complexes de modification de la chromatine vers leur gène cible comme le HOTAIR qui permet la régulation du gène HOXD via son association avec PRC2 (Gupta et *al.*, 2010, Prensner and Chinnaiyan, 2011).

Peuvent également servir au transport et localisation cellulaire des protéines, en effet l'interaction des LncRNAs aux protéines peut moduler leurs activité et localisation comme par exemple Le LncRNA NRON qui se lie au facteur de transcription cellulaire NFAT. (Gutschner et Diederichs, 2012).

### **B. Echafaudage:**

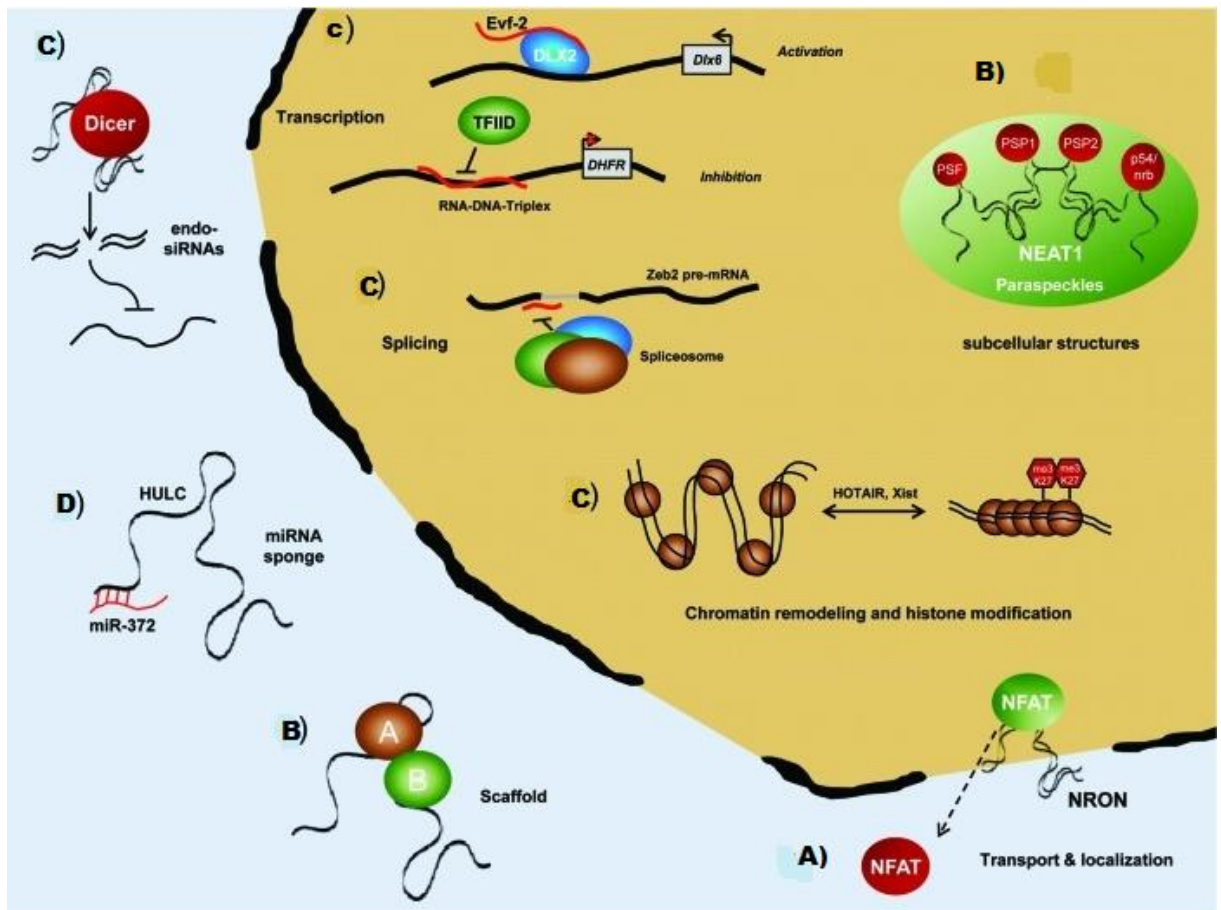
Pour les protéines n'ayant pas un domaine d'interaction commun; par ex: NEAT2 (Tripathi et *al.*, 2010, Tano and Akimitsu, 2012). (figure 08)

### **C. Régulation:**

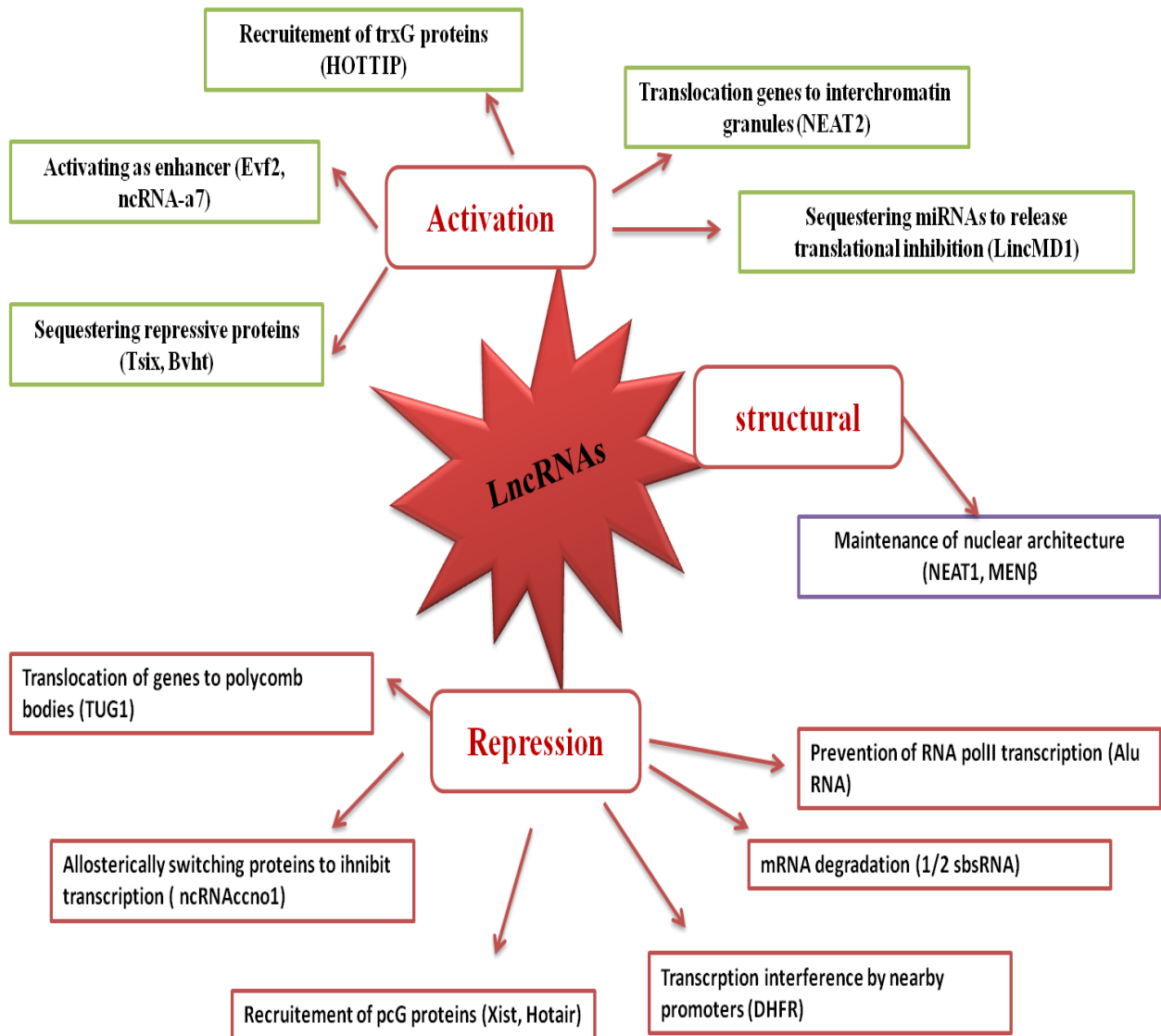
L'interaction directe avec les facteurs de transcription aboutissent à une régulation de la transcription des gènes cibles, comme par exemple : PANDA qui permet l'inhibition de l'apoptose en se liant directement au facteur de transcription Y (Hung et *al.*, 2011). Et ils peuvent agir comme répresseur comme leur implication dans le processus d'inactivation du chromosome X ( rôle des ARN non codants dans l'inactivation du X) (Gibb et *al.*, 2011). (figure 08)

**D. Eponge:**

Les LncRNAs peuvent se lier avec les microARNs (miRNAs), cette liaison empêche les miRNA d'effectuer leurs fonctions, participant ainsi à la régulation de la synthèse protéique. On appelle cette fonction « éponge à miRNA », comme par exemple : mir-333 , mir-135) (Änkö and Neugebauer, 2010). (figure 08)



**Figure 08 : Fonctions cellulaires des LncRNAs (Gutschner et Diederichs, 2012)**



**Figure 09 : Différentes fonctions régulatrices des LncRNAs**

En addition à toutes ces fonctions, les LncRNAs apparaissent comme étant d'importants régulateurs dans l'immunité. Beaucoup d'études ont démontré qu'ils sont exprimés par les cellules immunitaires aussi comme : les CD8<sup>+</sup> T cells, les CD4<sup>+</sup> TC , CD11c<sup>+</sup> ... cependant leurs fonctions exactes restent encore peu explorées (Fitzgerald and Caffrey, 2014).

## I.4. Implications des LncRNAs dans les maladies

### I.4.1 LncRNA et immunité

L'activation et la différenciation des cellules immunitaires dépendent de plusieurs mécanismes depuis le contact avec l'agent étranger, impliquant différents facteurs, parmi eux les LncRNAs. Après des études sur ces molécules impliquées dans différents processus cellulaires, il s'est avéré qu'elles pourraient avoir un rôle assez important dans l'immunité, surtout lors du changement de l'activité transcriptionnelle des cellules durant la réponse immunitaire afin de produire les molécules nécessaires à la défense, ce qui en fait des régulateurs importants dans l'immunité. Beaucoup d'études ont démontré qu'ils sont exprimés par les cellules immunitaires aussi comme: les CD8+ T cells , les CD4+ TC , CD11c+, etc. cependant leurs fonctions exactes restent encore peu explorées (Fitzgerald and Caffrey, 2014)

#### Exemples d'ARNs impliqués dans l'immunité :

##### a)- LncRNAs dans l'immunité adaptative

•**LincR-Ccr2-5'** AS pourrait participer au contrôle de la migration des Th2 aux poumons, par la régulation de la production des récepteurs (CCL2) des chimiokines (Ccr2) (Hu et *al.*, 2013).

Grâce à l'expérience de Gomez & all. qui consistait en l'insertion transgénique du gène NeST de la souche susceptible SJL/J chez les souris résistantes B10.S ( CD4+ & CD8+), NeST/ TMEVPG1 LncRNA participerait à:

► La persistance de l'infection par le virus Theiler. D'après l'expérience de Gomez & all. après l'infection par le virus, les B10.S transformé vont développer une infection persistante contrairement aux souris normales.

► La défense contre *Salmonella enterica*, les souris B10.S qui ne contiennent pas le transgène du LncRNA ne survivront pas à l'infection, contrairement aux B10.S transformés ou les SJL/J qui résistent car l'expression du NeST induit la production de l'interféron gamma, étant localisé à proximité de son gène le LncRNA joue le rôle d'un enhancer dans ce cas (Gomez et al., 2013).

### ***b)- LncRNAs dans l'immunité innée***

Différentes études ont pu démontrer à ce jour l'importance que détiennent les LncRNAs dans le mécanisme de défense innée (Carpenter et al., 2013, Rapicavoli et al., 2013) parmi eux:

► KIR LncRNA antisens du KIR, son gène chevauche l'exon 1 et 2 du gène KIR dans les cellules NK, sa transcription causerait l'extinction de l'expression du KIR par un mécanisme encore inconnu (Wright et al., 2013).

Le TNF induit l'expression de centaines d'LncRNAs, parmi eux le Lethe, transcrit sous l'action du facteur NFkB, auquel il va s'attacher (plus précisément à sa sous unité RelA) afin d'inhiber la liaison de ce dernier à l'ADN, et par cela participer à atténuer l'inflammation (Rapicavoli et al., 2013).

- HOTAIRM1 s'exprimant dans les cellules myéloïdes, a un rôle lors de la maturation granulocytaire des cellules NB4 (Zhang et al., 2009).

Ils participent aussi à la modulation des interactions des pathogènes avec leurs hôtes, par exemple ; il serait possible que la cellule hôte produise un LncRNA pour dévier un possible miRNA viral visant son ARNm (Skalsky and Cullen, 2010).

Pour conclure, Les LncRNAs ont le potentiel d'exécuter différentes fonctions au sein du système immunitaire. Leur étude reste encore un peu difficile à réaliser à cause de certains paramètres. Comprendre comment fonctionne exactement ce groupe de gènes pourrait mener à de nouvelles voies thérapeutiques dans le future (Fitzgerald and Caffrey, 2014).

## **I.5. LncRNAs et enhancers /superenhancers:**

### **a)- Enhancers**

Plusieurs études s'étaient focalisées sur les ARNs les associant à des complexes de répression de chromatine. Cependant, différentes autres études ont démontré que l'état actif de la chromatine était associé à des LncRNAs. La cartographie des modifications d'histones et des protéines de liaisons à l'enhancer sur l'échelle génomique ont fourni des informations supplémentaires pour identifier les Long ARNnc impliqués dans l'activation génique. L'analyse ChIP-seq de H3K4me1, H3K27ac et p300 (des marques associées aux enhancers) ont montré que ces régions produisent des transcrits des LncRNAs. Beaucoup de ces ARNs enhancer( eRNAs) étaient bidirectionnels, et manquaient de queue polyA et avaient un nombre restreint de copies (De Santa et *al.*, 2010, Kim et *al.*, 2010) .

Malgré le fait que beaucoup de ces transcrits aient été au départ considérés comme des dérivés de la transcription de la poly II ou de l'interaction enhancer-promoteur, plus de preuves montrent l'existence des rôles fonctionnels des LncRNAs. Les résultats d'une expérience LOF (loss of function) indiquent que 7 de 12 longs ARNncs affectent l'expression de leurs gènes voisins (Orom et *al.*, 2010).

*Exemple :*

► **HOTTIP** est un enhancer-like Long ARNnc qui interagit directement avec la protéine WDR5, une composante clé du complexe MLL/Trx qui catalyse l'activation de la marque

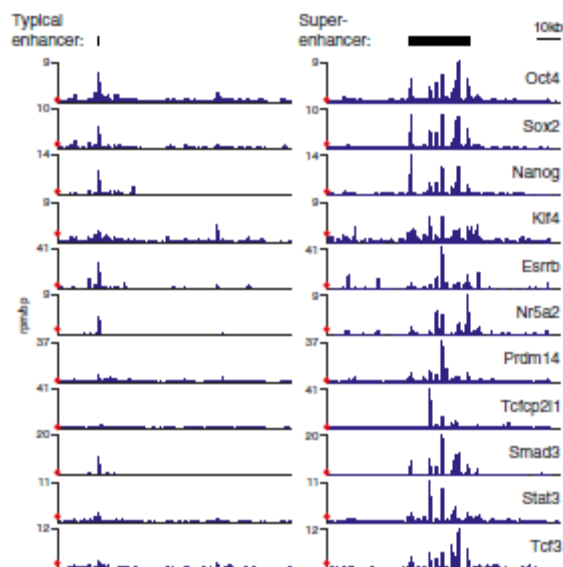


H3K4me3 (Wang et *al.*, 2011a), et est codé sur l'extrémité 5' du groupe des gènes HOXA, coordonne l'activation de plusieurs des gènes 5' Hoxa. Ce Long ARNnc est amené en étroite proximité avec les gènes chromosomiques Hoxa par le chromosome looping d'une manière similaire à un enhancer, renforçant la maintenance de H3K4me3 et l'activation du gène.

Ces études ont démontré l'importance de l'interaction des Longs ncRNA avec les machineries modifiant la chromatine et générant ainsi une activation génique « basedgene activation » et relèvent la possibilité que plusieurs autres eRNAs( enhancer RNA) pourraient agir par un mécanisme similaire. Donc, comprendre le répertoire des protéines de liaisons aux LncRANs est essentiel afin de comprendre leurs fonctions (Rinn and Chang, 2012).

### b)- Superenhancers

Les super enhancers sont un groupe d'enhancers très enrichis en facteurs, ce qui les différencie des enhancers classiques, formés par la liaison des facteurs de transcription majeurs et les co-activateurs afin de conduire à l'expression des gènes codants les régulateurs principaux de l'identité de la cellule (Hnisz et *al.*, 2013).



**Figure 10 : Résultat chip-seq indiquant l'enrichissement des régions superenhancers en facteurs de transcription (Hnisz et al. 2013)**

L'ARN polymérase II, co-facteurs et régulateurs de la chromatine qui sont enrichis dans les régions enhancers et superenhancers, peut transcrire les régions enhancers. Ces eRNA non codants résultant de cette transcription serviront comme activateurs d'enhancers (Lai et al., 2013, Li et al., 2013).

## I.6. LncRNAs et cancer

Avec cette multitude de fonctions, il n'est pas étonnant que les LncRNAs aient un rôle important dans le développement et les physiopathologies des maladies. Nous soulignons le rôle régulateur important des LncRNAs et leur émergence comme des acteurs clé dans l'étiologie de plusieurs états pathologiques (Wapinski and Chang, 2011).

Le lien le plus fort à présent est avec le cancer (Tsai et al., 2011). Il a été observé que plusieurs LncRNAs avaient une expression altérée dans les cancers humains, et sont régulés par des oncogènes et des voies suppresseurs de tumeurs spécifiques comme la P53, MYC, et NF- $\kappa$ B (Guttman et al., 2009, Hung et al., 2011).

Récemment, Hung et al ont décrit une classe des LncRNAs qui présentent une expression périodique durant le cycle cellulaire chez l'humain, et beaucoup ont une expression dérégulée dans les échantillons cancéreux humains (Hung et al., 2011).

Le HOTAIR est hautement induit dans un quart des cancers du sein, et son expression indique fortement un stade métastatique et décès (Gupta et al., 2010). Sa surexpression conduit à la métastase in vivo, en partie en interagissant avec des protéines telles *polycomb repressive complex 2* (PRC2) altérant ainsi l'identité de position des cellules cancéreuses (Gupta et al., 2010). Le niveau d'expression élevé d'HOTAIR prédit aussi un stade métastatique dans les cancers du côlon et foie, suggérant un trait oncogénique général (Kogo et al., 2011, Yang et al., 2011). Les cellules cancéreuses ont la capacité de se reprogrammer elles-mêmes, et donc la possibilité d'appartenir à un autre site anatomique (Gupta et al., 2010).

- **Concept des LncRNAs comme marqueurs des pathologies**

L'utilisation des LncRNAs comme marqueurs de pathologies est un concept qui fut renforcé par les éminentes découvertes sur les LncRNAs, peut être due à leur structure

secondaire, stable dans les fluides corporels et permet un diagnostic non invasif (Prensner and Chinnaiyan, 2011).

Chinnaiyan et ses collègues ont découvert une large gamme de LncRNAs présentes dans les cancers de la prostate humaine en utilisant la RNA-Seq. Ils purent identifier par la suite le PCAT1 qui est un LncRNA impliqué dans la répression génique, grâce auquel il est possible d'identifier les patients au mauvais pronostic en se basant sur le niveau d'expression dans les urines (Prensner and Chinnaiyan, 2011).

► **ANRIL** est un LncRNAs localisé en amont du locus CDKN2A qui est un répresseur tumoral, codant la protéine P16 inhibitrice de CDK, sa mutation est associée au cancer et les maladies cardiovasculaires, et conduit à l'expression des transcrits d'ANRIL aberrants et de la perte de la répression P16 (Burd et *al.*, 2010).

Tous ces exemples illustrent divers mécanismes pathogènes : à commencer par l'altération de l'état chromatinien ( HOTAIR , ANRIL ) , la modulation de la voie p53 ( linc-p21 et PANDA). Ainsi que l'épissage alternative qui augmente une production de protéines oncogénique ( Zeb2 ARN antisens) (Beltran et *al.*, 2008).

Le cancer fut le plus étudié mais il est probable que les LncRNAs soient impliqués dans plusieurs autres maladies. Conformément à cette notion, les centaines de régions génomiques non codantes sont largement associées à un large spectre de maladies humaines. Les études futures devront identifier et/ou préciser les transcrits LncARNs dans ces régions et comprendre si et comment le génome non codant contribue aux maladies humaines (Rinn and Chang, 2012).

---

## **CHAPITRE II. MATERIEL ET METHODES**

---

## **Le modèle d'étude :**

Le monde des ARNs non codant ne cesse de nous surprendre jour après jour, et surtout après la découverte de la nouvelle classe des Longs ARNs non codant, qui crée en elle-même un autre monde très vaste, devenu ces dernières années le sujet de nombreuses recherches les révélant être acteurs dans différents processus physiologiques et cellulaires ainsi que dans les pathologies.

Divers outils bioinformatiques s'offrent à nous pour ce travail afin d'identifier les LncRNAs impliqués dans la leucémie chronique Lymphoïde de type B. En analysant les données d'expression des Microarray (annexe I) des patients leucémiques et contrôles sains.

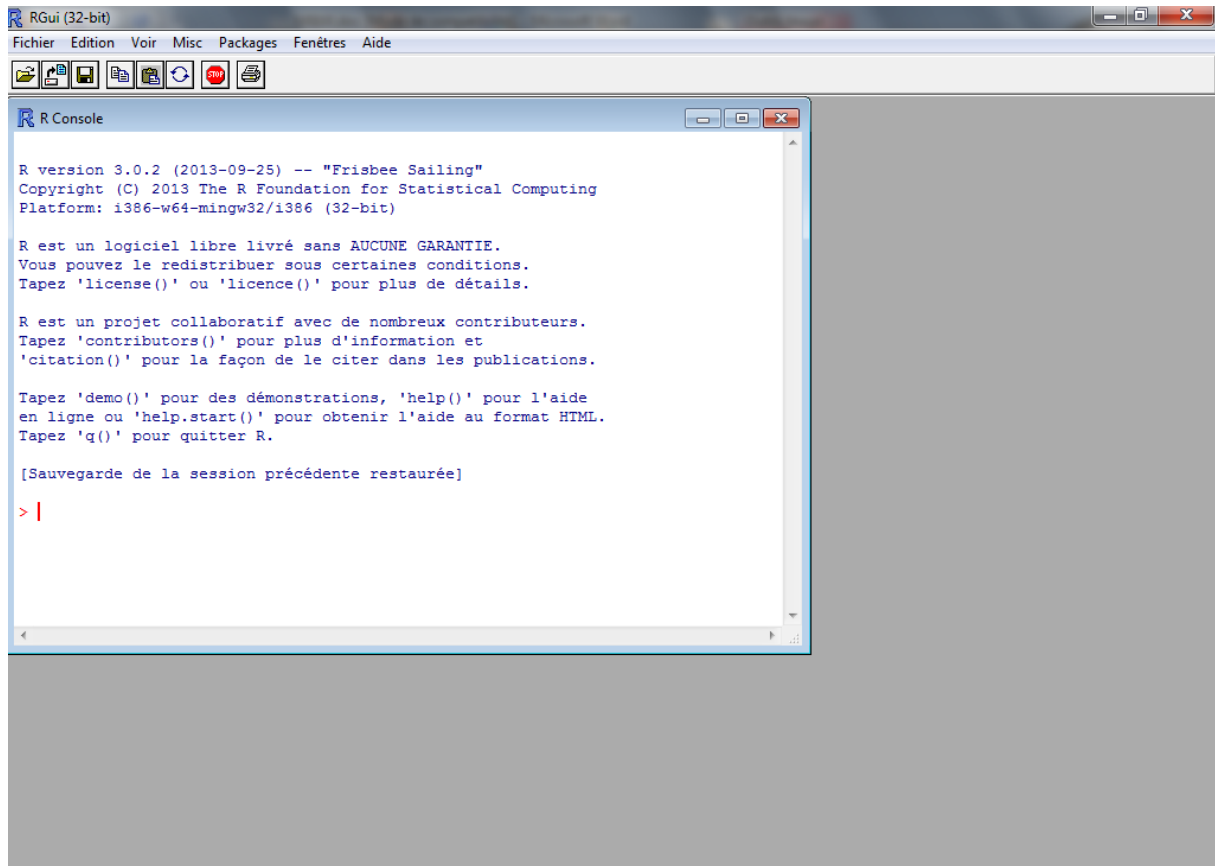
### **II.1. MATERIEL**

Différents outils bioinformatiques étaient à notre porté afin d'atteindre notre but, après avoir téléchargé les données brutes sur la base de données publique « GEO », nous les avons utilisées pour le traitement de ces dernières. Nous allons présenter ces outils ci-dessous.

#### **1)- Logiciel R**

Pour notre 1ère étape de travail qui est le prétraitement des données brutes nous avons utilisé le logiciel R.

**R** est un logiciel Open Source (libre) de traitement statistique de données et un langage de programmation en même temps, disposant d'un large spectre de fonctions disponibles au sein de ses bibliothèques.

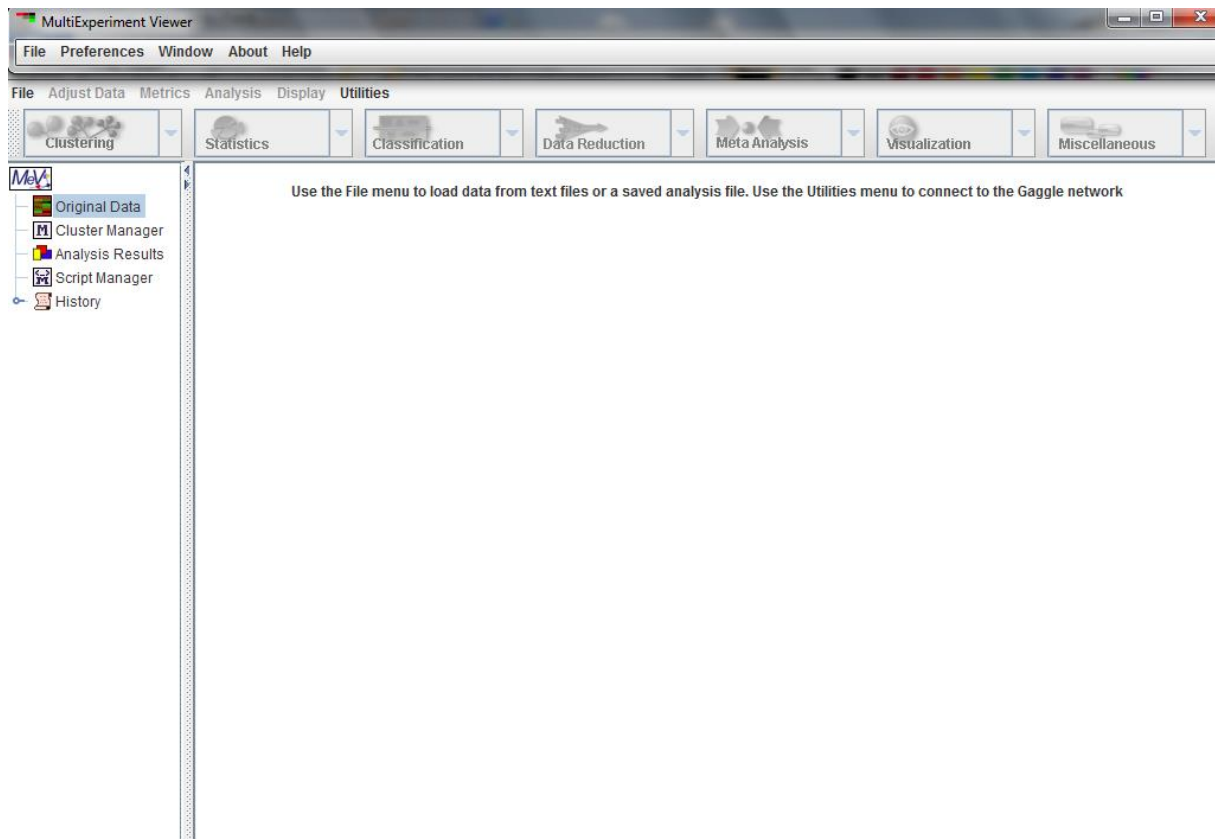


**Figure 11 : plateforme de travail R**

## 2)- Tmev :

Après avoir récupéré notre fichier normalisé sous format ".txt", nous avons pu l'ouvrir avec l'outil de visualisation graphique Tmev (Multi Experiment Viewer) afin de voir les gènes différentiellement exprimés dans les cellules testées.

MEV est un outil bioinformatique Open Source d'analyse des données microarray nécessitant JAVA. Sa plateforme est utilisable en mode bureau.



**Figure 12 : plateforme de travail MEV**

### 3)- DAVID :

L'annotation fonctionnelle se fait sur la plateforme DAVID (the Database for Annotation, Visualization and Integrated Discovery).

DAVID est une ressource bioinformatique qui consiste en un ensemble de connaissances biologiques intégrées et outils analytiques, visant à donner systématiquement une signification biologique à une liste de gènes ou de protéines (Huang et *al.*, 2008).

DAVID Functional Annotat x david.abcc.ncifcrf.gov

**DAVID Bioinformatics Resources 6.7**  
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home Start Analysis Shortcuts to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

**Shortcut to DAVID Tools**

**Functional Annotation**  
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)

**Gene Functional Classification**  
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)

**Gene ID Conversion**  
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)

**Gene Name Batch Viewer**  
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

**Welcome to DAVID 6.7**

2003 - 2015

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an [update to the sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch

**What's Important in DAVID?**

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supporte](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

**Statistics of DAVID**

DAVID Bioinformatic Resources Citations

3797

**Figure 13 : plateforme de travail DAVID**

#### 4)- GALAXY TOOLS :

Le « Galaxy tools » est une plateforme de travail en ligne (<https://usegalaxy.org/>), contenant divers outils pour effectuer un grand nombre d'opérations sur des données génomiques telle que : merge, intersection, BLAST...etc. En effet on peut réaliser, reproduire ou partager des analyses complètes ou alors visualiser nos données. Que ce soit sur le serveur publique ou notre propre instance.

Il existe différentes approches pour commencer à manipuler avec GALAXY, son interface est assez simple ce qui facilite son utilisation, mais il existe aussi des pages spéciales contenant des tutoriels pour chaque outils (Wikigalaxy : <https://wiki.galaxyproject.org/>).



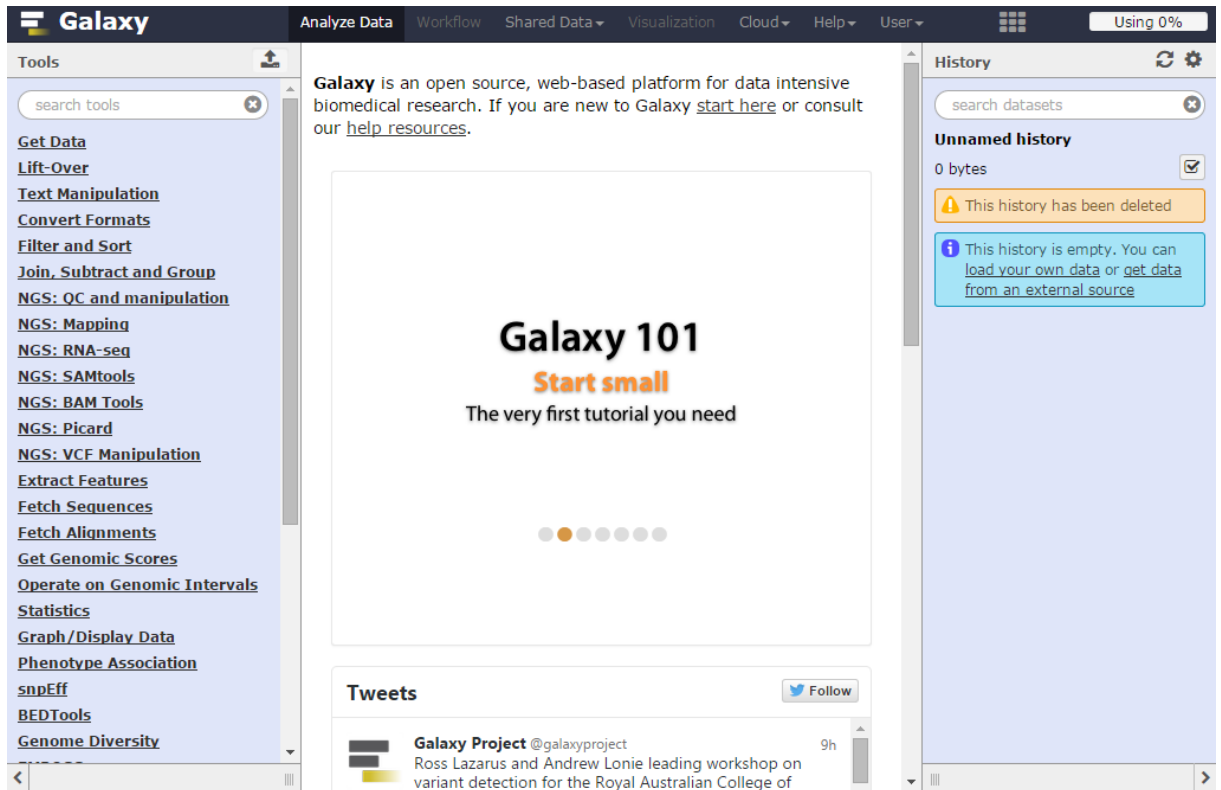


Figure 14 : plateforme de travail GALAXY

L'équipe GALAXY fait partie du centre de génomique comparative et bioinformatique de l'université de Pennsylvanie.

## II.2. METHODES

### Analyse du transcriptome

#### II. 2. 1. Récupération des données brutes sur la base NCBI GEO

En premier lieu, il nous fallait trouver des données adéquates disponibles sur la LLC-B, pour cela, nous avons effectué une recherche sur NCBI qui abrite plusieurs bases de données, parmi elles : GEO (Gene Expression Omnibus) contenant d'importantes données génomiques brutes générées par plusieurs expériences de haut débit comme des puces ou des séquençages.

NCBI Resources How To Sign in to NCBI

GEO Home Documentation Query & Browse Email GEO

### Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 3848
About GEO DataSets	Search GEO Documentation	Series:  57345
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 14292
About GEO2R Analysis	GEO BLAST	Samples: 1398420
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	

Information for Submitters		
Login to Submit	Submission Guidelines	MIAME Standards
	Update Guidelines	Citing and Linking to GEO
		Guidelines for Reviewers

Figure 15 : Page d'accueil de la grande base de données d'expression GEO

Notre recherche a abouti à la sélection des données brutes suffisantes pour notre travail, issues de la technique puce à ADN à partir de la base d'expression GEO en utilisant la plateforme [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array.

NCBI > GEO > [Accession Display](#) Not logged in | [Login](#)

Scope:  Format:  Amount:  GEO accession:

**Series GSE26725** [Query DataSets for GSE26725](#)

Status	Public on Mar 01, 2011
Title	Gene expression analysis of 12 B-cell Chronic Lymphocytic Leukemia samples and 5 CD19+ control samples
Organism	<a href="#">Homo sapiens</a>
Experiment type	Expression profiling by array
Summary	Elevated levels of microRNA miR-155 represent a candidate pathogenic factor in chronic B-lymphocytic leukemia (B-CLL). In this study, we present evidence that MYB (v-myb myeloblastosis viral oncogene homolog) is overexpressed in a subset of B-CLL patients. MYB physically associates with the promoter of MIR155 host gene (MIR155HG, also known as BIC, B-cell integration cluster) and stimulates its transcription. This coincides with the hypermethylated histone H3K4 residue and spread hyperacetylation of H3K9 at MIR155HG promoter. Our data provide evidence of oncogenic activities of MYB in B-CLL that include its stimulatory role in MIR155HG transcription. We have studied differentially expressed genes in the whole genome and also in the preselected groups of MYB target genes and miR-155 microRNA predicted targets.
Overall design	This study was intended to investigate differentially expressed genes mRNA in 12 B-CLL patient peripheral blood samples in comparison with 5 CD19+ healthy donor peripheral blood samples.
Contributor(s)	<a href="#">Stopka T</a> , <a href="#">Trněný M</a> , <a href="#">Vargová K</a> , <a href="#">Zavadil J</a> , <a href="#">Kulvait V</a>
Citation(s)	Vargova K, Curik N, Burda P, Basova P et al. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. <i>Blood</i> 2011 Apr 7;117(14):3816-25. PMID: 21296997

**Figure 16 : Sélection des données brutes sur GEO**

En effet, 17 puces ont été sélectionnées et sont décrites comme suit:

- **12 échantillons de cellules B leucémique :**

PB C104 B-CLL / PB P155 B-CLL / PB P138 B-CLL / B-CLL PB P137 / B-CLL  
PB P143 / B-CLL PB P271 / B-CLL PB P136 / B-CLL PB P249 / B-CLL PB  
P254 / B-CLL PB P33L / B-CLL PB P250 / B-CLL PB P255.

- **5 échantillons de cellules B normales de contrôle :**

Control PB C106 Control / PB C103 Control / PB C107 Control / PB C105 Control.

Après la sélection, nous avons commencé le téléchargement de nos données brutes (non traitées), qui ont été générées sous format de fichier «CEL ». Ces fichiers sont enregistrés sous le numéro d'accession « GSE26725 ».


Platforms (1) [GPL570 \[HG-U133\\_Plus\\_2\]](#) Affymetrix Human Genome U133 Plus 2.0 Array

Samples (17) [GSM658010](#) Control PB C106  
[More...](#) [GSM658011](#) Control PB C103  
[GSM658012](#) Control PB C107

**Relations**  
 BioProject [PRJNA136137](#)

**Analyze with GEO2R**

Download family	Format
<a href="#">SOFT formatted family file(s)</a>	SOFT <a href="#">?</a>
<a href="#">MINiML formatted family file(s)</a>	MINiML <a href="#">?</a>
<a href="#">Series Matrix File(s)</a>	TXT <a href="#">?</a>



Supplementary file	Size	Download	File type/resource
<a href="#">GSE26725_RAW.tar</a>	77.6 Mb	<a href="#">(http)(custom)</a>	TAR (of CEL)

*Raw data provided as supplementary file*  
*Processed data included within Sample table*

**Figure 17 : Téléchargement des données brutes de GEO**

Les données brutes de chaque puce ont été générées sous un même dossier compressé avec l'extension ".tar", qui contient les 17 fichiers format « .CEL ».

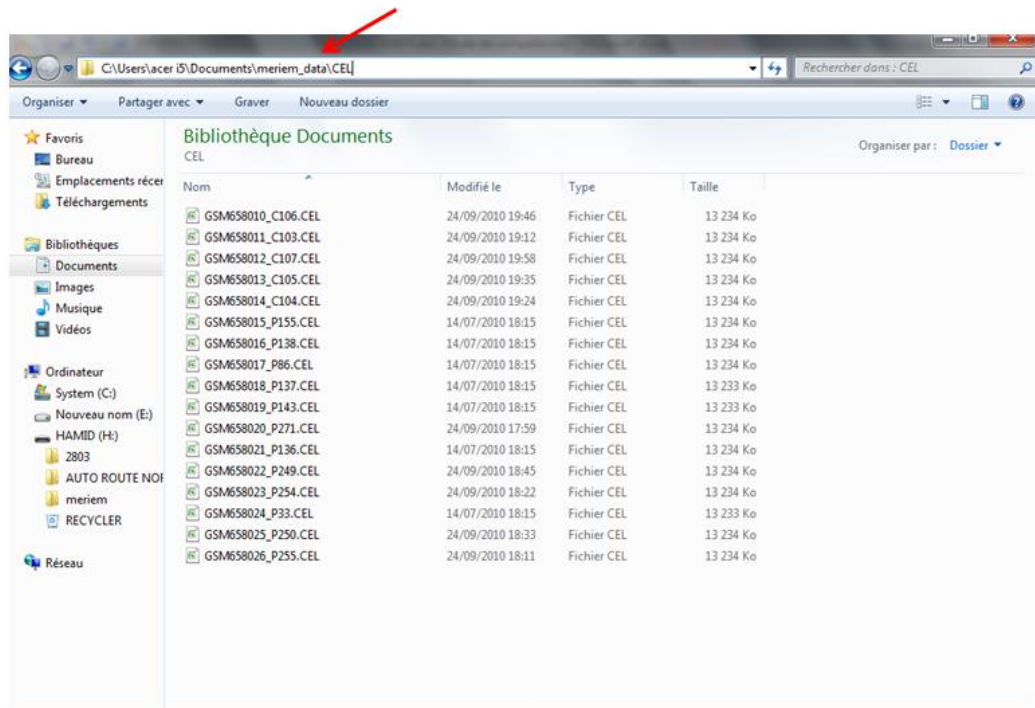
Nom	Taille	Compressé	Type	Modifié	CRC32
Dossier de fichiers					
GSM658010_C106.CEL.gz	4 520 587	4 520 587	Archive WinRAR	19/01/2011 21:37	
GSM658011_C103.CEL.gz	4 661 599	4 661 599	Archive WinRAR	19/01/2011 21:37	
GSM658012_C107.CEL.gz	4 687 226	4 687 226	Archive WinRAR	19/01/2011 21:37	
GSM658013_C105.CEL.gz	4 613 562	4 613 562	Archive WinRAR	19/01/2011 21:37	
GSM658014_C104.CEL.gz	4 675 844	4 675 844	Archive WinRAR	19/01/2011 21:37	
GSM658015_P155.CEL.gz	4 754 437	4 754 437	Archive WinRAR	19/01/2011 21:37	
GSM658016_P138.CEL.gz	4 696 755	4 696 755	Archive WinRAR	19/01/2011 21:37	
GSM658017_P86.CEL.gz	4 991 862	4 991 862	Archive WinRAR	19/01/2011 21:37	
GSM658018_P137.CEL.gz	4 905 705	4 905 705	Archive WinRAR	19/01/2011 21:37	
GSM658019_P143.CEL.gz	5 075 170	5 075 170	Archive WinRAR	19/01/2011 21:38	
GSM658020_P271.CEL.gz	4 726 582	4 726 582	Archive WinRAR	19/01/2011 21:38	
GSM658021_P136.CEL.gz	4 774 579	4 774 579	Archive WinRAR	19/01/2011 21:38	
GSM658022_P249.CEL.gz	4 622 670	4 622 670	Archive WinRAR	19/01/2011 21:38	
GSM658023_P254.CEL.gz	4 850 051	4 850 051	Archive WinRAR	19/01/2011 21:38	
GSM658024_P33.CEL.gz	4 815 404	4 815 404	Archive WinRAR	19/01/2011 21:38	
GSM658025_P250.CEL.gz	5 013 647	5 013 647	Archive WinRAR	19/01/2011 21:38	
GSM658026_P255.CEL.gz	4 964 481	4 964 481	Archive WinRAR	19/01/2011 21:38	

Figure 18 : Données sous fichier .TAR sauvegardé sur notre ordinateur

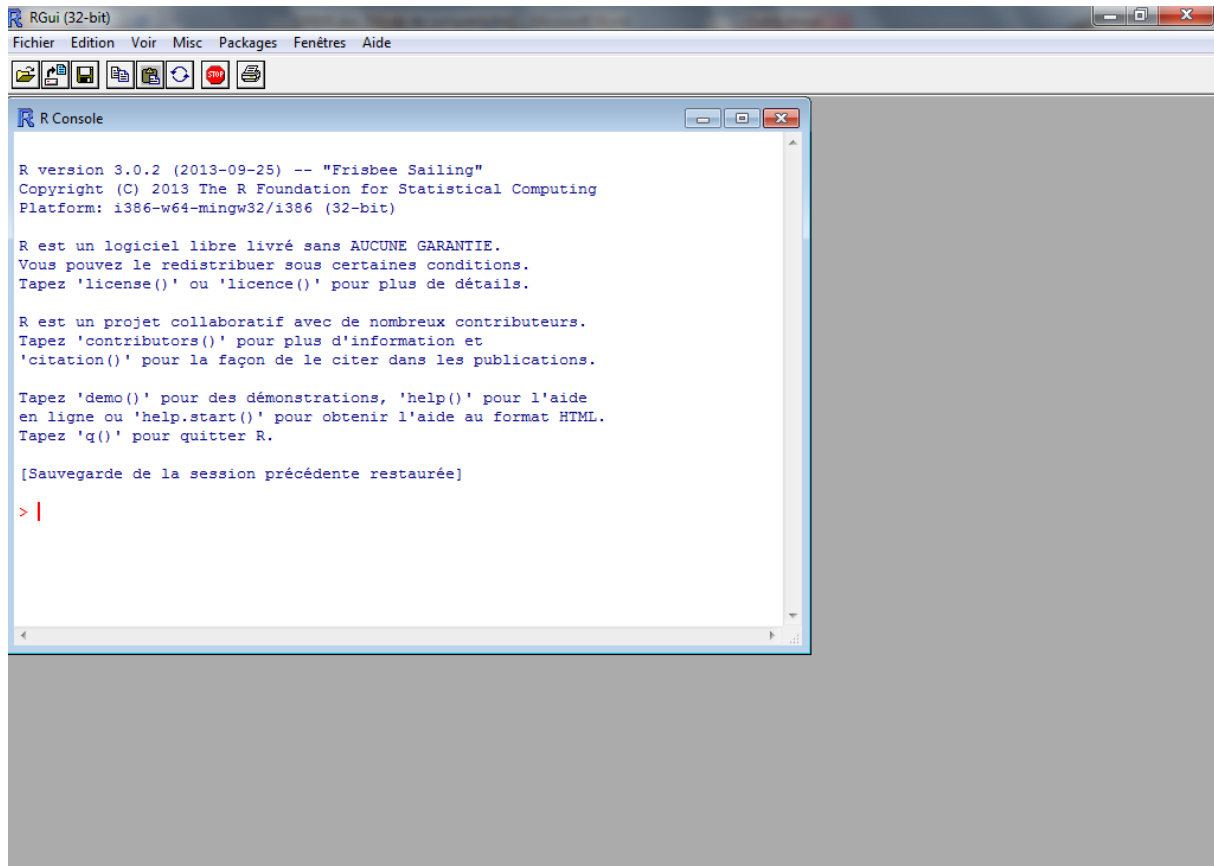
## II. 2. 2. Normalisation des données brutes ( pré-analyse)

Le but de cette étape est d'ajuster ou de corriger les données brutes afin d'enlever les effets systématiques et corriger les différences entre les mesures qui ne représentent pas de véritables variations biologiques. Cela permet d'examiner les réplicas contre le même échantillon où de vraies différences d'expression ne doivent pas apparaître.

Après la décompression, nous avons eu les fichiers « .CEL » séparés chacun à part, nous les avons placés sur le même répertoire de travail "mes documents" dans un dossier appelé « Meriem\_data » afin d'en faciliter l'accès et l'exploitation en utilisant le même chemin d'accès lors de l'analyse sous R.



**Figure 19 : Décompression du fichier initial « .tar » et placement des fichiers « .CEL » sur notre ordinateur**



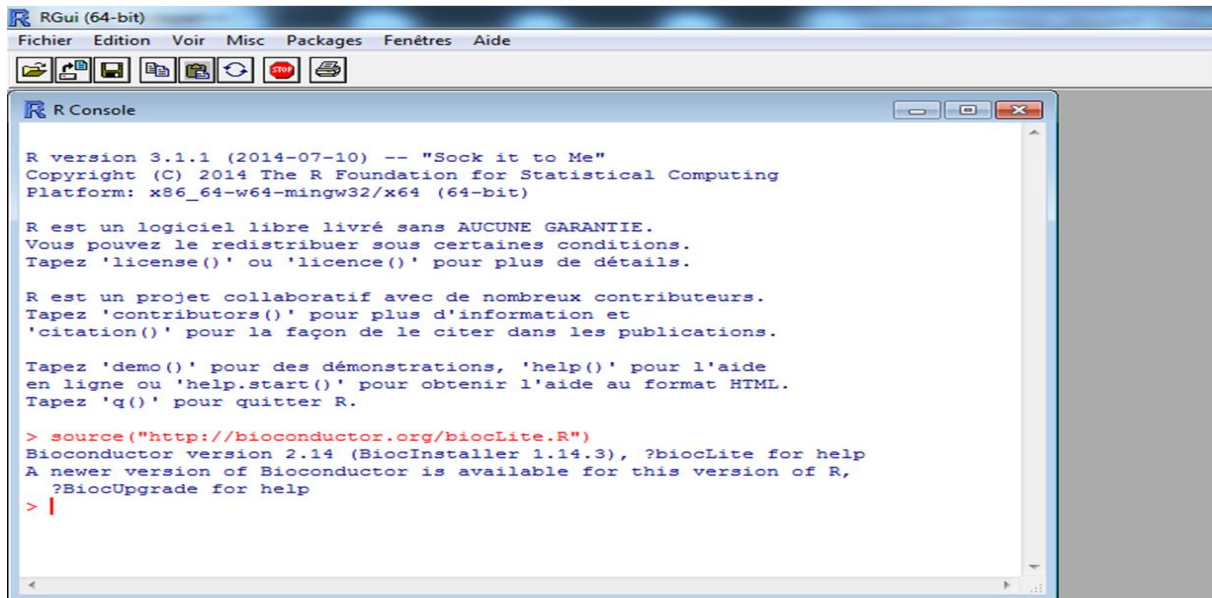
**Figure 20 : Ouverture de la Plateforme R**

Tout d'abord, nous avons commencé par l'installation du package *Bioconductor*. En utilisant la commande suivante :

```
# package installation: do it just once
source("http://bioconductor.org/biocLite.R")
biocLite()
```

*Bioconductor*, utilise le langage de programmation statistique R, il est Open Source et à développement ouvert. Il a été lancé en 2001, ayant pour but d'offrir à la communauté scientifique une panoplie d'outils pour l'analyse et la compréhension des données génomiques (biologiques) à haut débit. Il en existe deux versions chaque année, 934 logiciels, et une communauté active d'utilisateurs (Huber et *al.*, 2015).

Nous avons ouvert l'environnement R, puis lancé le script de téléchargement de "Bioconductor" suivant :



**Figure 21 : Installation de Bioconductor**

Ensuite, nous avons lancé d'autres lignes de commandes pour télécharger des packages nécessaires pour le traitement des données des microarray via *"bioconductor"*.

```
biocLite("affy")
biocLite("arrayQualityMetrics")
biocLite("tkWidgets")
```

Le package "affy" contient des fonctions pour l'analyse exploratoire de réseau d'oligonucléotides.

La 1ère fonction téléchargée « Library(Affy) » nécessaire à la lecture des puces.

```
library(affy) ### charger l'outil de lecture des puces
```

La lecture des fichiers ".CEL" en lançant cette commande, R créera un objet « AFB » contenant nos données « .CEL »



```
setwd("C:/Users/acer i5/Document/meriem_data/CEL") #définir le repertoire de travail  
afb <- ReadAffy()
```

Sur le logiciel R, la visualisation des puces avant prétraitement ainsi que l'enregistrement de l'image est possible à l'aide de la commande suivante :

```
image (afb[,1])  
  
jpeg("puce1.jpeg")  
image (afb[,1])  
dev.off ()
```

Avant de commencer le pré-traitement, nous avons vérifié la qualité des fichiers « .CEL » avant normalisation, pour cela, nous lançons une commande de contrôle de qualité dans le logiciel R :

```
dir.create("ControlNormalisation")  
arrayQualityMetrics(expressionset=afb , outdir="Q_without_normalisation")
```

Nous avons testé quatre types de normalisation avec des paramètres différents à chaque fois : set1, set2, set3 et set4.

```

### Realisation de 3 normalisations :

setwd("C:/Users/acer i5/Documents/meriem_data/ControlNormalisation/")

#set1 <- exprsso(afb, widget=TRUE) #####choisir les parametre (voir le tableau word)
#arrayQualityMetrics(set1 , do.logtransform=TRUE, outdir="QM1") #####controlez et visualiser

#set2 <- exprsso(afb, widget=TRUE)
#arrayQualityMetrics(set2 , do.logtransform=TRUE, outdir="QM2")

#set3 <- exprsso(afb, widget=TRUE)
#arrayQualityMetrics(set3 , do.logtransform=TRUE, outdir="QM3") ## on choisira la normalisation la plus performante

set4 <- rma(afb)
arrayQualityMetrics(set4 , do.logtransform=TRUE, outdir="QM4")

```

Chaque type de normalisation va générer un dossier appelé QM « qualitymetrics » contenant divers figures ainsi qu'une page HTML .

Le choix de la meilleure normalisation se fera en comparant leurs QM, et le set avec le meilleur QM sera choisi. Pour notre cas, le set4 a été sélectionné : robust multi array average (RMA)

```

set4 <- rma(afb)
arrayQualityMetrics(set4 , do.logtransform=TRUE, outdir="QM4")

```

Élimination des mauvaises probes, à faible affinité avec la commande ci-dessous. En effet, la fonction « grep » nous détecte les mauvaises probes x,s ( ayant faible affinité) sous l'objet « indexes to remove » qui sont par la suite supprimés de nos données finales.

```

set <- data.frame(set4)

set <- t(set)

set <- set[-nrow(set),]

rownames(set) <- rownames(set4)

indexesToRemove <- grep("[x|s]_at", rownames(set) ,perl=T) ### i

Data <- set[-indexesToRemove,] ### supprimer les probes x et s

```

Sauvegarder le fichier final normalisé sur notre ordinateur.

```
write.table(Data, file = "Table_Transcriptome.txt", sep = "\t", row.names = TRUE, col.names = TRUE)
```

Le fichier généré après normalisation « table\_transcriptome.txt » ne contient pas les noms des gènes (non annoté). Afin d'y ajouter l'annotation, nous utiliserons la fonction «merge» sous R, pour joindre notre fichier avec « HG-U133\_Plus\_2\_annotation.txt » qui a été ultérieurement récupéré du site Affymetrix.

```
#####  
# ajout de l'annotation des gènes  
setwd("C:/Users/acer i5/Documents/meriem_data")  
a = read.delim("HG-U133_Plus_2_annotation.txt", sep="\t", header=TRUE)  
  
a = a[,c(1,15,14)]  
setwd("C:/Users/acer i5/Documents/meriem_data/Resultats")  
  
tab=read.delim("Table_Transcriptome.txt", sep="\t", header=TRUE)  
  
m= merge(a , tab, by.x="Probe.Set.ID" , by.y="Probe_ID")
```

## Contrôle de qualité après normalisation

### Réalisation des boxplot

Une série de représentation graphique est disponible sous le logiciel R qui permet d'observer la distribution de nos données, parmi elles, nous avons les « boxplot ». Pour notre travail, nous allons réaliser deux boxplot différents, avant la normalisation ainsi qu'après la normalisation afin d'observer et comparer la différence de distribution de nos données des microarray avant et après la normalisation et ainsi confirmer le bon déroulement de notre pré-traitement.

Ceci se fait en lançant une commande sur R, comme cet exemple :

```
« boxplot ( log(pm) , col=c(rep("blue",5),rep("red",12)) , main="distribution des données  
avant la normalisation"  
,names=c("C1","C2","C3","C4","C5","P1","P2","P3","P4","P5","P6","P7","P8","P9","P10","  
P11","P12")) »
```

Chaque commande est expliquée ci-dessous :

PM = notre tableau de données à utiliser pour cette représentation graphique.

Col = ajouter les couleurs souhaitées qui sont pour notre boxplot bleu et rouge, en distinguant les échantillons sains en bleu et les échantillons malades en rouge.

Main = titre de la représentation graphique.

Names = pour nommer chaque boxplot.

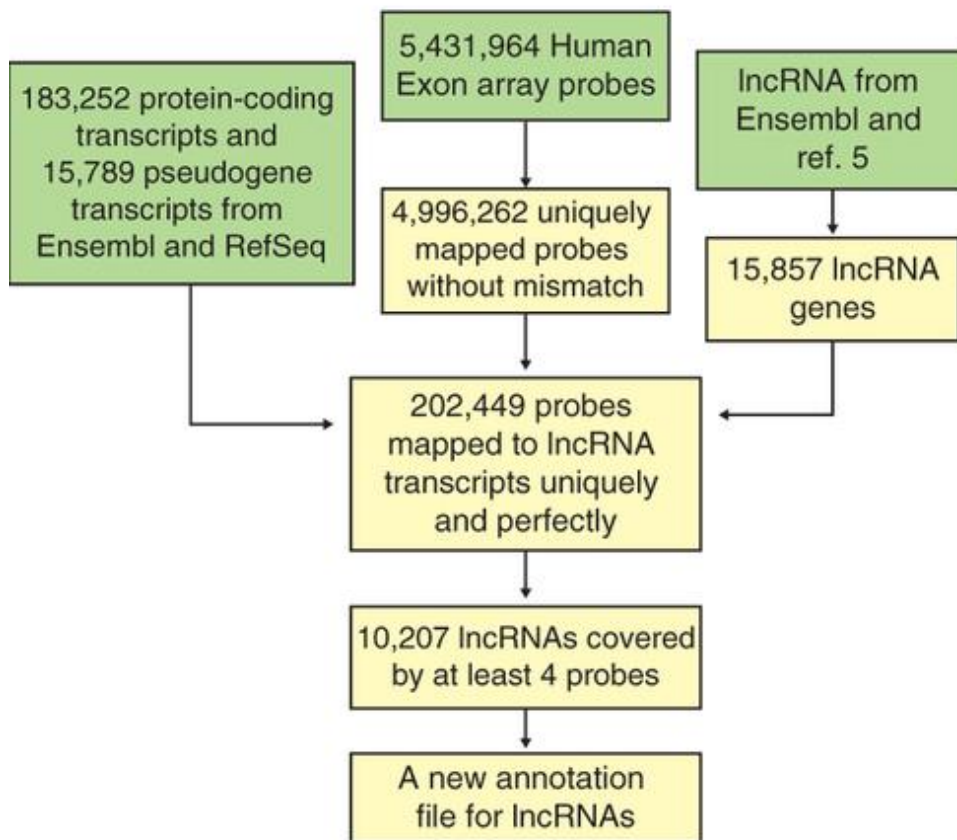
Pour enregistrer le boxplot résultant autant qu'image (.jpeg), nous lançons une autre commande :

```
« jpeg("box3.jpeg")
```

```
boxplot ( log(pm) , col=c(rep("blue",5),rep("red",12)) , main="distribution des données avant  
la normalisation" ,  
names=c("C1","C2","C3","C4","C5","P1","P2","P3","P4","P5","P6","P7","P8","P9","P10","P  
11","P12"))  
dev.off() ».
```

### II.2.3. Récupération de sondes associées au LncRNAs

Dans la publication "Integrative genomic analyses reveal clinically relevant long non codingRNAs in human cancer", Zhou et ses collaborateurs ont ré-annoté des probes set de puce à ADN Affymetrix afin de quantifier l'expression des LncRNAs impliqués pendant les étapes de cancérisation.



**Figure 22 : Réannotation des données de microarray et Refseq (Du et al., 2013)**

Dans le but de procéder à l'analyse des lncRNAs putatifs présents dans les leucémies chroniques B, nous suivons la méthode de Zhou Du et ses collaborateurs (Du et al., 2013) En effet, nous avons les puces du type U133 2 qui ont été ré-annotées par M. BELHOCINE afin d'en sortir les probes codant uniquement les lncRNAs, et ce, comme suit:

- Un fichier CDF ( Chip Description File) a été récupéré du site Affymetrix, qui contient les informations de chaque probe des puces (les fragments) contenant 100 . 000 probes, dont 80.000 représentent des gènes codants.
- Un BLAST sera effectué entre le fichier contenant tous les fragments du CDF avec les fragments de bases de données GENCODE V19 qui contient un catalogue de lncRNAs.

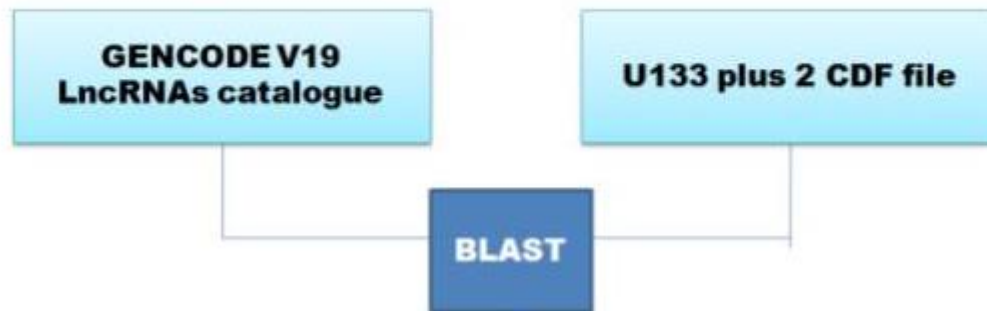


Figure 23 : Reannotation de la puce type U133 plus2

Sur R nous allons lancer une commande appelée « merge » entre le tableau contenant notre transcriptome totale et le tableau contenant les LncRNAs présents dans les puces U133 plus2, le but étant d'en sortir les lignes qui codent pour les LncRNAs seulement présentes dans nos échantillons de travail.

```

## recupere les lignes seulement des non codant
setwd("C:/Users/acer i5/Documents/meriem_data/")

b = read.delim("All_LincRNA(3293)_in_Affy_U133plus2_with_annotated.txt" ,sep="\t", header=TRUE)
b=b[,c(1,5,8)]

lnc= merge(b , tab, by.x="probe" , by.y="Probe_ID")

boxplot(lnc[,4:ncol(lnc)])

setwd("C:/Users/acer i5/Documents/meriem_data/Resultats/")
write.table(lnc, file = "LncRNA_Transcriptome.txt", sep = "\t", row.names = FALSE, col.names = TRUE)
  
```

## II.2. 4. Identification des gènes différentiellement exprimés entre contrôle et LLC-B

Notre analyse Tmev se fera en deux fois :

- Analyse de l'ensemble du transcrit. (codant + non codant)
- Analyse des LncRNAs seulement.

La 1ère étape est de charger notre dataset dans Tmev « selection « file == load file » du menu de recherche, après cela une nouvelle fenêtre s’ouvre afin qu’on puisse sélectionner nos fichiers d’intérêt et lancer l’analyse.

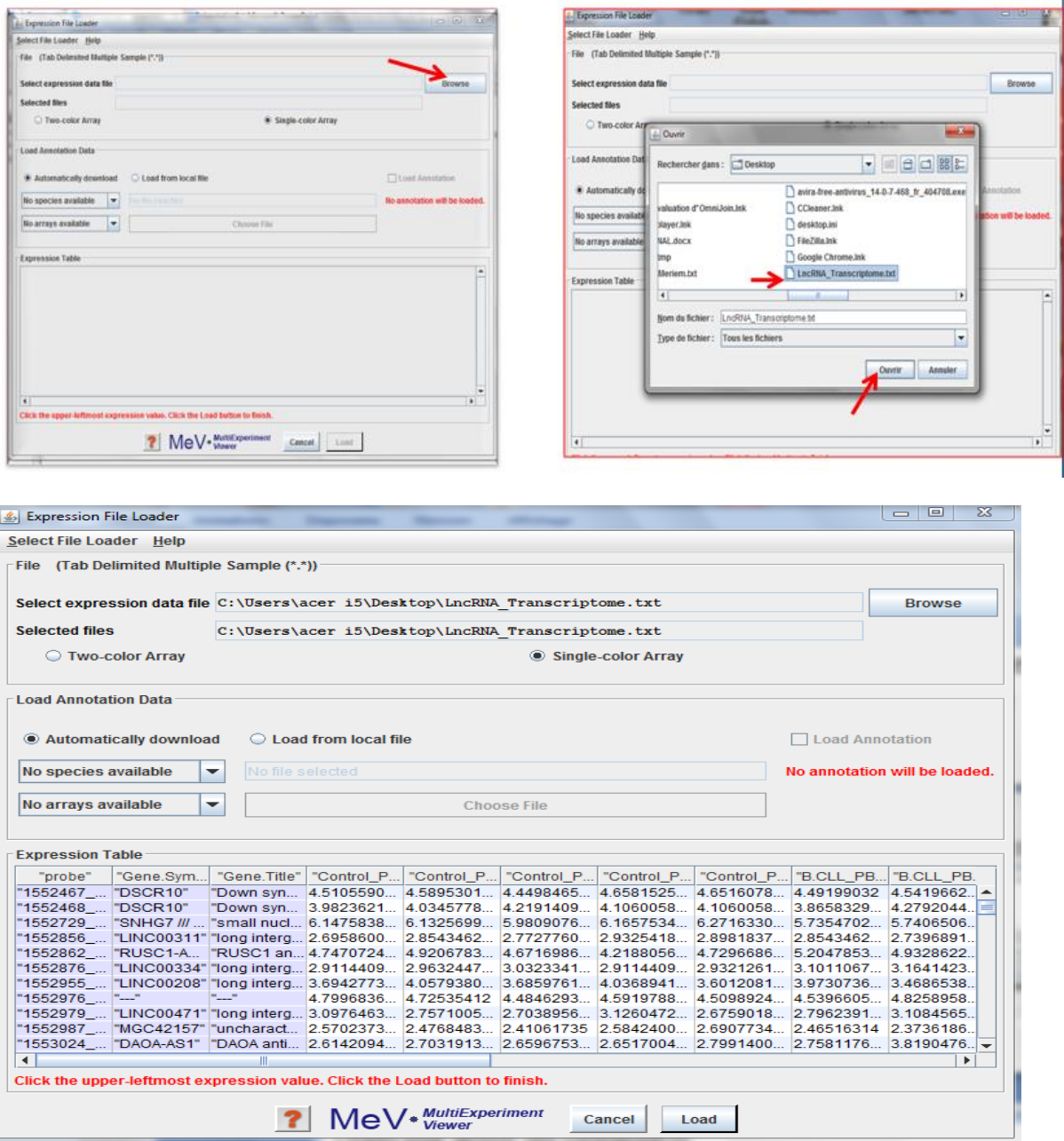


Figure 24 : Les 1ères étapes de l’analyse Tmev

Suite à cela, nous sommes passés à l’analyse Statistique SAM ‘Significant Analysis for Microarrays’, qui servira à identifier les différences d’expressions des gènes avec les listes des gènes surexprimés (Up\_regulated) et réprimés (Down\_regulated).

Si le nombre de données est très volumineux nous devons filtrer nos données « filter data ».

## **II. 2. 5. Analyse fonctionnelle des gènes différentiellement exprimés**

L'annotation fonctionnelle se fait sur la plateforme DAVID (the Database for Annotation, Visualization and Integrated Discovery).

Le début d'analyse est le chargement des listes à annoter. Nous avons fait le travail pour quatre listes de gènes obtenues à partir de l'analyse Statistique SAM (Significance Analysis of Microarrays) dans tmev.



The screenshot shows the 'Analysis Wizard' interface. On the left is a blue sidebar titled 'Upload Gene List' with tabs for 'Upload', 'List', and 'Background'. The sidebar contains the following steps:

- Step 1: Enter Gene List**
  - A: Paste a list: A text area containing '227765\_at', '227431\_at', and '236555\_at'. A blue arrow points to this area.
  - Or B: Choose From a File: A file selection button with the text 'Choisissez un fichier' and 'Aucun fichier choisi', and a 'Multi-List File' checkbox.
- Step 2: Select Identifier**: A dropdown menu showing 'AFFYMETRIX\_3PRIME\_IVT\_ID'.
- Step 3: List Type**: Radio buttons for 'Gene List' (selected) and 'Background'.
- Step 4: Submit List**: A 'Submit List' button with a blue arrow pointing to it.

On the right, the main area is titled 'Analysis Wizard' and contains the following text:

- Links: [Tell us how you like the tool](#) and [Contact us for questions](#).
- Instruction: 'Step 1. Submit your gene list through left panel.' with a blue arrow pointing left.
- Example: 'An example:' followed by a list of gene IDs: '1007\_s\_at', '1053\_at', '117\_at', '121\_at', '1255\_g\_at', '1294\_at', '1316\_at', '1320\_at', '1405\_i\_at', '1431\_at', '1438\_at', '1487\_at', '1494\_f\_at', '1598\_g\_at'.
- Instructions: 'Copy/paste IDs to "box A" -> Select Identifier as "Affy\_ID" -> List Type as "Gene List" -> Click: "Submit" button'.

**Figure 25 : Début de l'analyse par la soumission des listes d'intérêt**

Une fois notre liste à analyser est rentrée dans le site et enregistrée, nous allons avoir diverses options pour l'analyse fonctionnelle, et le choix entre différentes bases de données ainsi que voies de signalisation ( pathways) pour voir la voie de signalisation dérégulée suite à une dérégulation dans cette liste de gènes (up ou down).

**Gene List Manager**

Select to limit annotations by one or more species [Help](#)

- Use All Species -  
 Homo sapiens(106)  
 Unknown(44)

Select Species

**List Manager** [Help](#)

List\_1

Select List to:

Use Rename  
 Remove Combine  
 Show Gene List

[View Unmapped Ids](#)

**Annotation Summary Results** [Help and Tool Manual](#)

Current Gene List: List\_1 100 DAVID IDs  
 Current Background: Homo sapiens Check Defaults  Clear All

Disease (0 selected)  
 Functional\_Categories (0 selected)  
 Gene\_Ontology (0 selected)  
 General\_Annotations (0 selected)  
 Literature (0 selected)  
 Main\_Accessions (0 selected)  
 Pathways (1 selected)

<input type="checkbox"/> EC_NUMBER	7.0%	7	Chart
<input checked="" type="checkbox"/> KEGG_PATHWAY	5.0%	5	Chart
<input type="checkbox"/> PANTHER_PATHWAY	6.0%	6	Chart
<input type="checkbox"/> REACTOME_PATHWAY	5.0%	5	Chart

Protein\_Domains (0 selected)  
 Protein\_Interactions (0 selected)  
 Tissue\_Expression (0 selected)

\*\*\*Red annotation categories denote DAVID defined defaults\*\*\*

**Combined View for Selected Annotation**

Functional Annotation Clustering  
 Functional Annotation Chart  
 Functional Annotation Table

Figure 26 : Choix des différentes bases de données disponible pour notre analyse

## II. 2. 6. Identifications des LncRNAs associés aux superenhancers :

Notre objectif en dernier lieu est de déterminer la présence parmi nos LncRNAs d'intérêt des « eRNA » qui sont le produit de la transcription des régions régulatrices superenhancers et qui participent par la suite à la fonction d'activation d'enhancer (Li et al., 2014).

Afin d'atteindre notre objectif nous allons utiliser une fonction disponible sur le site GALAXY TOOLS nommée « Intersect ».

Cet outil réalise l'intersection entre deux séquences génomiques. Pour notre travail nous avons croisé les LncRNAs différemment exprimés lors de la LLC B ( UP & DOWN) avec les superenhancers présent dans les cellules sanguines et le thymus identifiés et caractérisés auparavant par Hnise et al.



**Figure 27 : Figure représentant l'intersection entre les superenhancers et la liste des LncRNAs d'intérêt.**

---

## **CHAPITRE III. RESULTATS ET DISCUSSION**

---

### III.1. RESULTATS

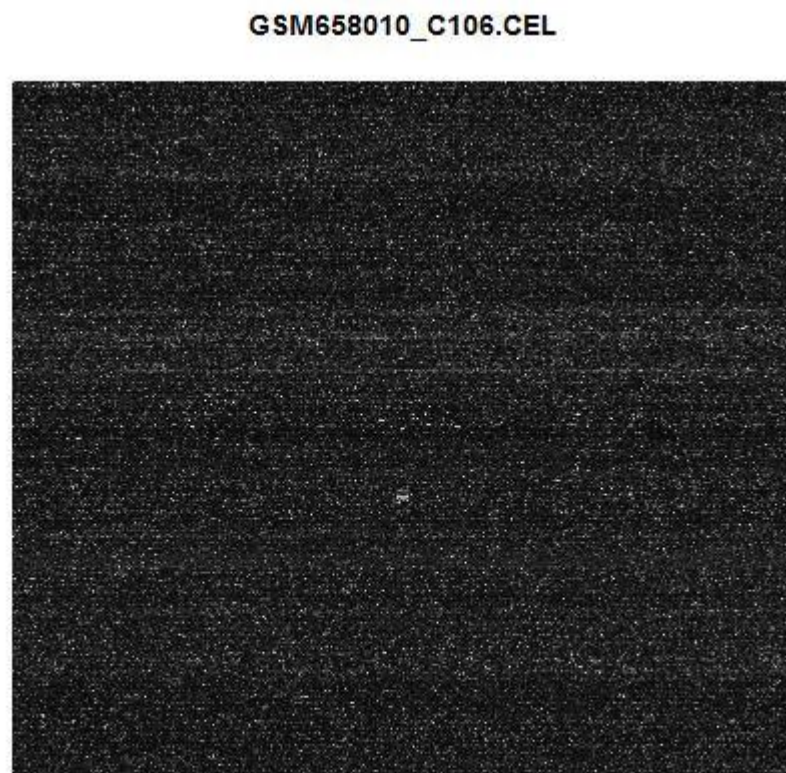
Dans cette partie, nous allons présenter les résultats de nos analyses bioinformatiques ultérieures.

#### III.1.1. Récupération des données brutes

Nous avons récupéré les 17 puces et avons pu visualiser sous R.

#### Exemple d'une Image de puce à ADN :

Cette figure représente un affichage graphique d'un de nos fichiers « .CEL » à leur état natif avec les spots et chaque spot correspond à un gène en blanc et le background en noir.



**Figure 28 : visualisation d'une puce ADN**

### III.1.2. Pré-analyse des données brutes

Après la normalisation et l'ajout de l'annotation des gènes, nous avons obtenu un tableau contenant l'ensemble du transcriptome annoté avec les gènes sur les lignes et les conditions en colonnes normalisé en format « .txt », prêt à être utilisé pour le reste des analyses.

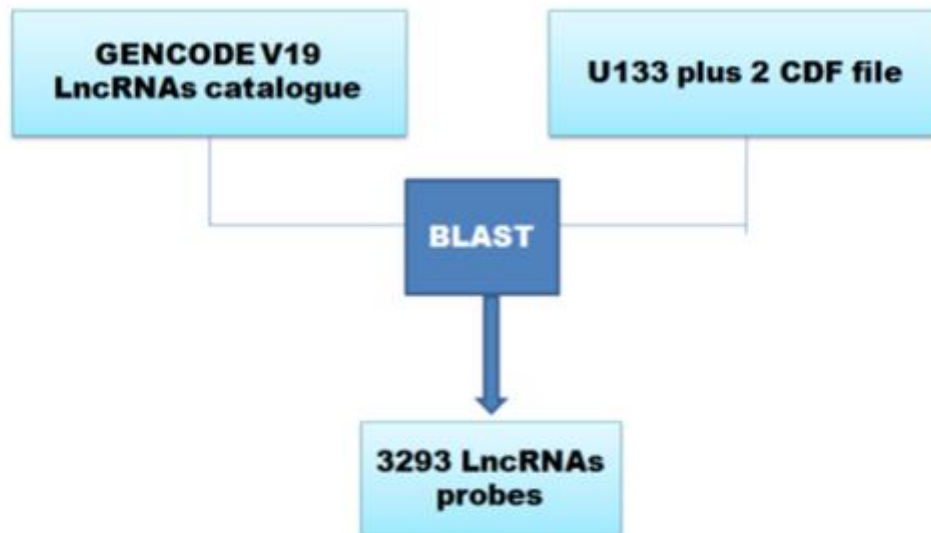
**Les transcrits totaux représentant l'ensemble du transcriptome (gènes codant et non codant) « Table\_with\_names\_ALLtranscritome ».**

1	Probe.Set.ID	Gene.Symbc	Gene.Title	Control_PB_1	Control_PB_2	Control_PB_3	Control_PB_4	Control_PB_5	B.CLL_PB_P1_1	B.CLL_PB_P1_2	B.CLL_PB_P1_3	B.CLL_PB_P1_4	B.CLL_PB_P1_5
2	1053_at	RFC2	replication f	7.223721858	6.942547009	7.039851033	7.532958637	6.985192949	6.257770566	6.268871435	7.585231214	6.758153816	6.901111111
3	117_at	HSPA6	heat shock	7.6.318957122	6.835975382	7.953726629	6.603966068	6.277539037	7.825031768	7.301566502	7.148294989	4.692757388	4.800000000
4	121_at	PAX8	paired box 8	7.760756502	7.518482019	7.084905884	7.393243023	7.232118914	7.333679144	7.270398679	7.197832953	7.191004093	7.200000000
5	1255_g_at	GUCA1A	guanylate cy	2.863608022	2.90525337	2.628568771	2.804188626	2.806307753	2.777693537	2.916812311	2.760383537	2.631295092	2.800000000
6	1294_at	UBA7	ubiquitin-lik	8.514357812	8.682510874	8.705587822	8.754975767	8.549077128	8.111555944	7.924720724	7.930574142	7.40371518	7.600000000
7	1316_at	THRA	thyroid horn	5.462603902	5.273602434	5.354390888	5.324265475	5.32890367	5.720600354	5.962973127	5.599701538	5.753905719	5.200000000
8	1320_at	PTPN21	protein tyros	3.695418649	3.605899411	3.720085309	3.694967566	3.794109428	3.544268095	3.556816825	3.435769427	3.378519467	3.500000000
9	1405_i_at	CCL5	chemokine (	10.07305384	10.00936637	10.14789534	9.864607312	8.577655504	12.62223145	11.3221705	10.36377591	9.177051947	7.500000000
10	1431_at	CYP2E1	cytochrome	3.708155965	3.885936626	3.748289155	4.149883535	3.803254687	3.611047149	3.875561103	3.536721821	4.021977266	3.600000000
11	1438_at	EPHB3	EPH receptor	4.665414874	4.318634264	4.186457756	4.162477613	4.279365974	4.39014061	4.05030399	4.164535728	4.179885186	4.100000000
12	1487_at	ESRRA	estrogen-rel	6.678820341	6.620881023	6.708228592	6.364808459	6.733944931	6.74690353	6.536485177	6.764438363	6.942885123	6.500000000
13	1494_f_at	CYP2A6	cytochrome	4.872716793	4.743100274	4.893591136	5.011989766	5.042167389	4.947802434	4.838606314	4.844607304	4.656801468	4.400000000
14	1552256_a_a	SCARB1	scavenger re	6.869049324	6.609693384	7.214465275	7.068401599	7.377097245	5.728146224	5.223609478	6.217387197	4.953424302	5.000000000
15	1552257_a_a	TTL12	tubulin tyros	7.031253893	7.242270492	7.091078305	7.247080822	7.110803363	6.953700704	6.087317978	7.113499731	6.760381155	6.700000000
16	1552258_at	LINC00152	long interge	4.790042398	4.49126984	4.911782852	5.069891224	4.358108899	5.328208708	4.842527079	5.228052789	4.238347994	4.400000000
17	1552261_at	WFDC2	WAP four-di	3.849844249	3.669914685	3.486528773	3.740447607	3.565804096	3.520228995	3.788441123	3.675059012	3.603945449	3.500000000
18	1552263_at	MAPK1	mitogen-act	7.868801442	7.857488623	7.79657569	7.548121756	7.38575454	8.294757318	7.955307453	8.469200341	7.88512549	7.700000000
19	1552264_a_a	MAPK1	mitogen-act	8.055051265	8.053523385	8.079376268	7.842668146	7.530279561	8.744001104	8.388436276	9.306542639	8.567724986	8.000000000
20	1552266_at	ADAM32	ADAM metal	2.61782305	2.563973491	2.560911347	2.546430874	2.799140007	2.745399126	2.715451695	2.605795888	2.715451695	2.400000000
21	1552269_at	SPATA17	spermatoger	2.750807393	2.654224678	2.761480229	2.997879016	2.997879016	3.037024655	2.90857659	2.524635431	2.512677495	2.400000000
22	1552271_at	PRR2	proline rich	4.509633184	4.310220646	4.531212724	4.165446369	4.742662303	4.320288203	4.520835996	4.469497189	3.862574659	4.400000000
23	1552272_a_a	PRR2	proline rich	3.442337839	3.54776003	3.429584424	3.633181987	3.816270134	3.697621618	3.659979376	3.652695509	3.618742126	3.600000000
24	1552274_at	PXK	PX domain c	7.670584687	7.719128723	7.817841858	7.504127854	8.087777611	7.926088117	7.961389209	6.928375528	8.218598419	8.600000000
25	1552276_a_a	VPS18	vacuolar pro	5.411063236	5.066050515	4.868534185	4.870314381	5.0490606	4.893058259	4.369026719	4.686980949	4.893818748	4.800000000

Figure 29 : Tableau des données des puces normalisés

### III.1.3 Récupération de sondes associées aux LncRNAs

La réannotation des puces U133 plus a abouti à la récupération de 3293 probes de LncRNAs parmi les 100.000 probes totaux de cette même puce.



**Figure 30 : Résultat de la réannotation des puces U133 plus2**

Après avoir obtenu notre tableau du transcriptome total des puces U133 plus 2 normalisés ainsi que les probes des LncRNAs après réannotation, nous avons exécuté une commande « merge » sous R afin de sortir de notre tableau initial les lignes des gènes codant les LncRNAs non codants seulement.

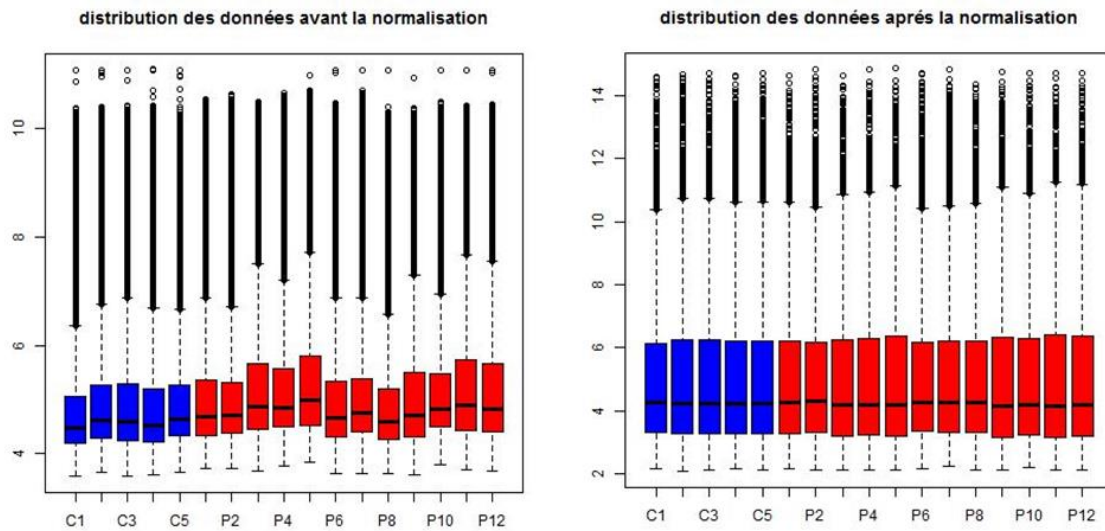
**Les transcrits des LncRNAs « LncRNA\_transcriptome ».**

	A	B	C	D	E	F	G	H	I	J	K	L	M
	probe	Gene.Symbol	Gene.Title	Control_PB	Control_PB	Control_PB	Control_PB	Control_PB	B.CL	B.CL	B.CL	B.CL	B.C
1	1552467_at	DSCR10	Down syndr	4.510559008	4.589530163	4.449846517	4.658152554	4.651607869	4.49199032	4.541966254	4.510559008	4.510559008	4.2
2	1552468_a_a	DSCR10	Down syndr	3.982362119	4.034577823	4.219140936	4.106005828	4.106005828	3.865832982	4.279204447	3.864432355	4.045460446	4.2
3	1552729_at	SNHG7 /// S	small nucleo	6.147583898	6.132569985	5.980907623	6.165753409	6.271633093	5.735470226	5.740650607	5.77958743	5.284290452	5.7
4	1552856_at	LINC00311	long intergei	2.695860039	2.854346231	2.772776059	2.932541821	2.898183751	2.854346231	2.739689148	3.180981251	2.998310252	2.7
5	1552862_at	RUSC1-AS1	RUSC1 antisense	4.747072448	4.920678366	4.671698863	4.218805688	4.729668653	5.204785329	4.932862295	5.401679717	5.183810777	5.3
6	1552876_at	LINC00334	long intergei	2.911440971	2.963244793	3.032334197	2.911440971	2.932126108	3.101106709	3.164142351	2.639485625	2.712541265	2.7
7	1552955_at	LINC00208	long intergei	3.694277349	4.057938074	3.685976191	4.036894151	3.601208169	3.973073619	3.468653889	3.780143948	3.685976191	3.3
8	1552976_at	---	---	4.799683688	4.72535412	4.484629302	4.591978855	4.509892425	4.539660594	4.825895838	4.574315416	4.574315416	4.3
9	1552979_at	LINC00471	long intergei	3.097646384	2.757100553	2.703895629	3.126047217	2.675901891	2.796239139	3.108456586	2.823702207	2.843807459	2.7
10	1552987_a_a	MGC42157	uncharacteri	2.570237348	2.476848364	2.41061735	2.584240054	2.690773481	2.46516314	2.373618618	2.454300138	2.493601917	2.5
11	1553024_at	DAOA-AS1	DAOA antisense	2.614209467	2.703191343	2.659675362	2.651700409	2.799140007	2.758117663	3.819047632	2.528038922	2.530714348	2.5
12	1553053_at	LINC00521	long intergei	3.173435148	3.187251933	3.179613601	3.157155353	3.01247298	3.40336754	3.173435148	3.156281682	3.173435148	3.5
13	1553226_at	LINC00052	long intergei	3.213454862	3.49728492	3.437325229	3.200476195	3.394438354	3.237869494	3.460066073	3.047674641	3.026673456	3.0
14	1553269_at	ZNF718	Zinc finger p	5.319827028	4.554652163	4.096062618	4.600982924	4.852760627	4.672293528	5.790601834	4.35374632	4.543131784	7.9
15	1553390_at	ATG9B	autophagy re	3.652678927	3.20778746	3.097535934	3.105273281	3.233238317	3.350636428	2.972466947	3.048097268	3.146835709	3.0
16	1553420_at	SATB2-AS1	SATB2 antisense	3.326130476	2.854346231	2.813421211	2.799140007	2.799140007	2.71475666	2.930347721	2.593645154	2.684634038	2.6
17	1553428_at	LY86-AS1	LY86 antisense	2.961514833	2.961514833	3.030538663	2.970066548	2.915746602	3.00427705	3.424303245	2.74515341	2.885015948	3.9
18	1553440_at	CHST9-AS1	CHST9 antisense	2.739689148	2.654224678	2.634837852	2.869727314	2.784641048	2.779635443	2.691316129	2.669460566	2.935139319	2.7
19	1553443_at	FER1L6-AS1	FER1L6 antisense	3.255991376	3.213761083	3.110917004	2.905955107	3.292066023	2.947079086	3.175068515	2.869767744	3.272759381	2.8
20	1553448_at	FLJ34503	uncharacteri	3.267982359	3.123707069	3.123707069	3.147672164	3.253962174	3.123707069	3.118791179	3.02593342	3.070816025	3.1
21	1553449_at	LINC00304	long intergei	3.738045914	3.362349045	3.774060797	3.499883462	3.629171814	3.774060797	3.648227329	3.774060797	3.917891683	3.9
22	1553457_at	LINC00269	long intergei	5.025929809	4.970321607	4.786766576	4.833028434	5.167316035	4.845031168	4.690878851	4.573151133	4.769608723	4.5
23	1553472_at	FLJ32955	uncharacteri	3.036884907	2.989219328	3.155750034	2.852510888	2.962408556	2.816158962	3.010858472	2.762380332	2.693399128	2.8
24	1553478_at	KIRREL3-AS3	KIRREL3 antisense	3.581889462	3.653413164	3.619935142	3.522483071	3.986255723	3.380203075	3.719968882	3.229569202	3.42372602	3.3

Figure 31 : Tableau des LncRNAs totaux

Par la suite, nous avons obtenu deux boxplot représentant la distribution des données avant et après normalisation. Ce qui confirme l'élimination de variation ainsi que l'homogénéité des données qui sont, à présent, prêtes à être utilisées pour la suite.



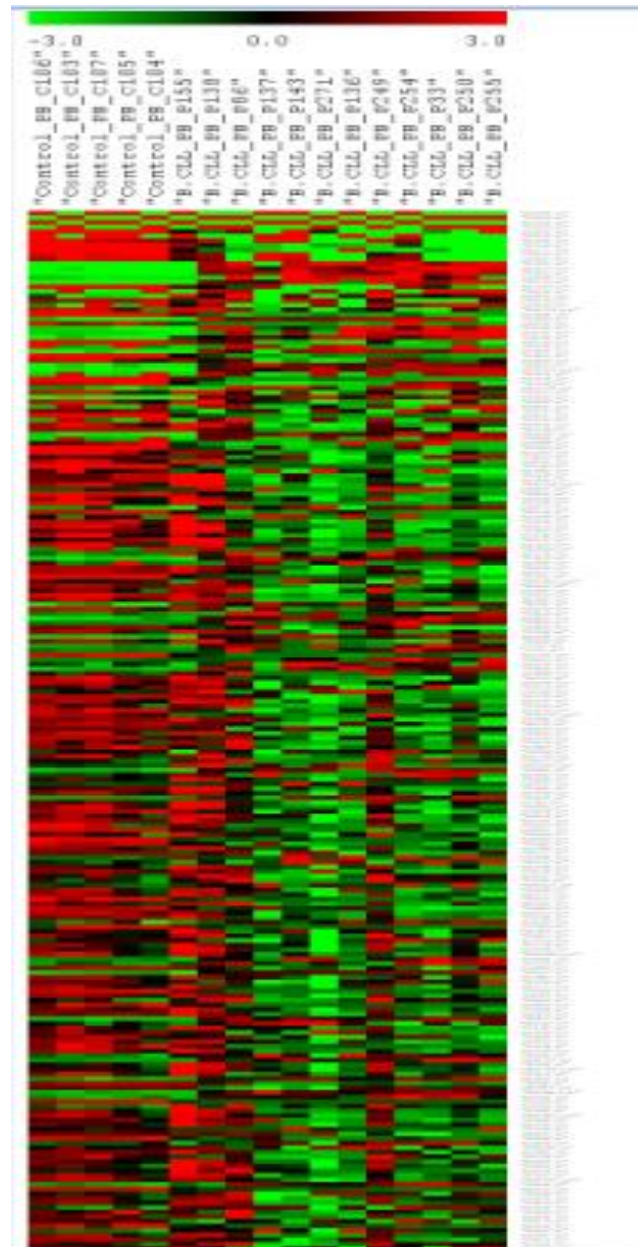


**Figure 32 : Comparaison de la distribution des données avant et après normalisation**

#### II.1.4. Identification des gènes différentiellement exprimés entre contrôle et LLC-B

L'analyse Tmev vise à identifier les gènes différentiellement exprimés entre la leucémie chronique B et les contrôles sains.

##### 1)- Analyse du transcriptome total (codant + non codant)

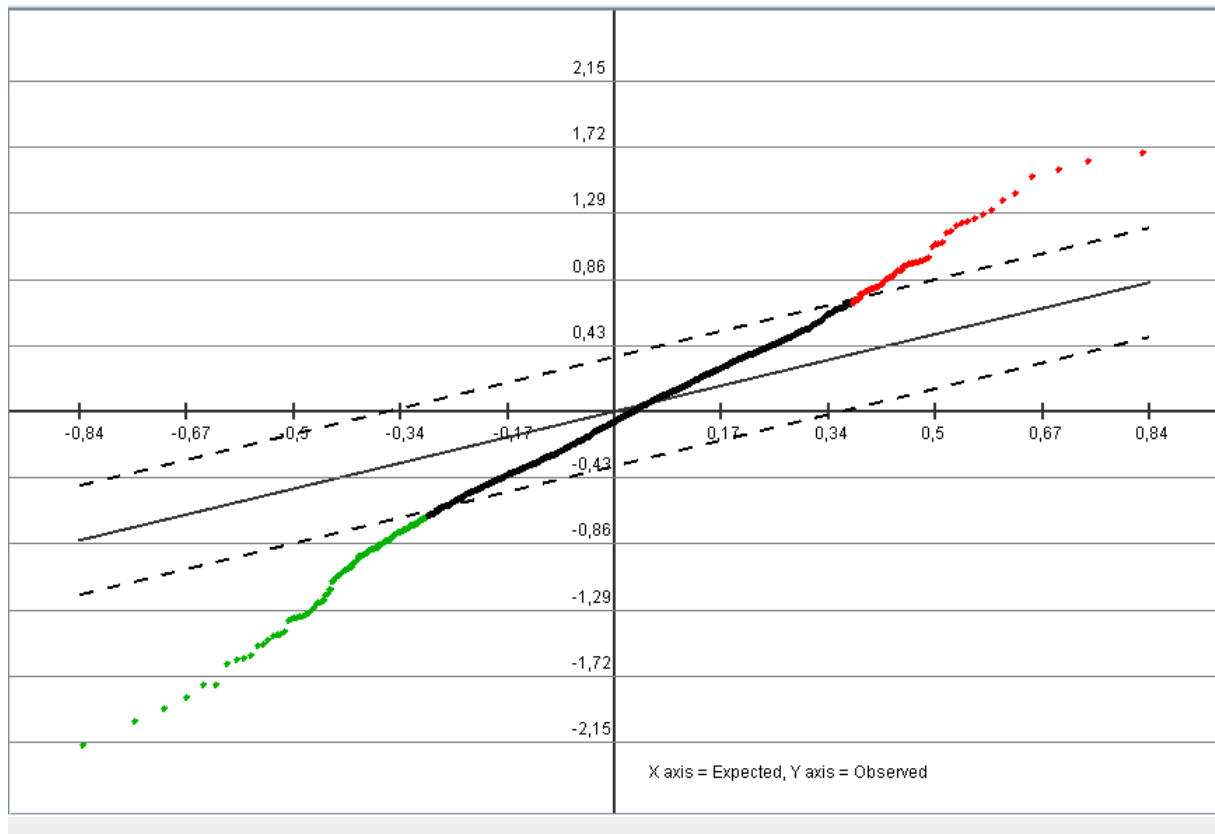


**Figure 33 : Niveau d'expression des genes totaux.**

Ce heatmap représente l'expression totale des gènes entre surexprimés, réprimés et normaux.

### **GRAPHE SAM ( Significant Analysis of Microarrays)**

Afin de cibler les gènes significatifs, nous avons lancé « SAM », dont a resulté ceci :

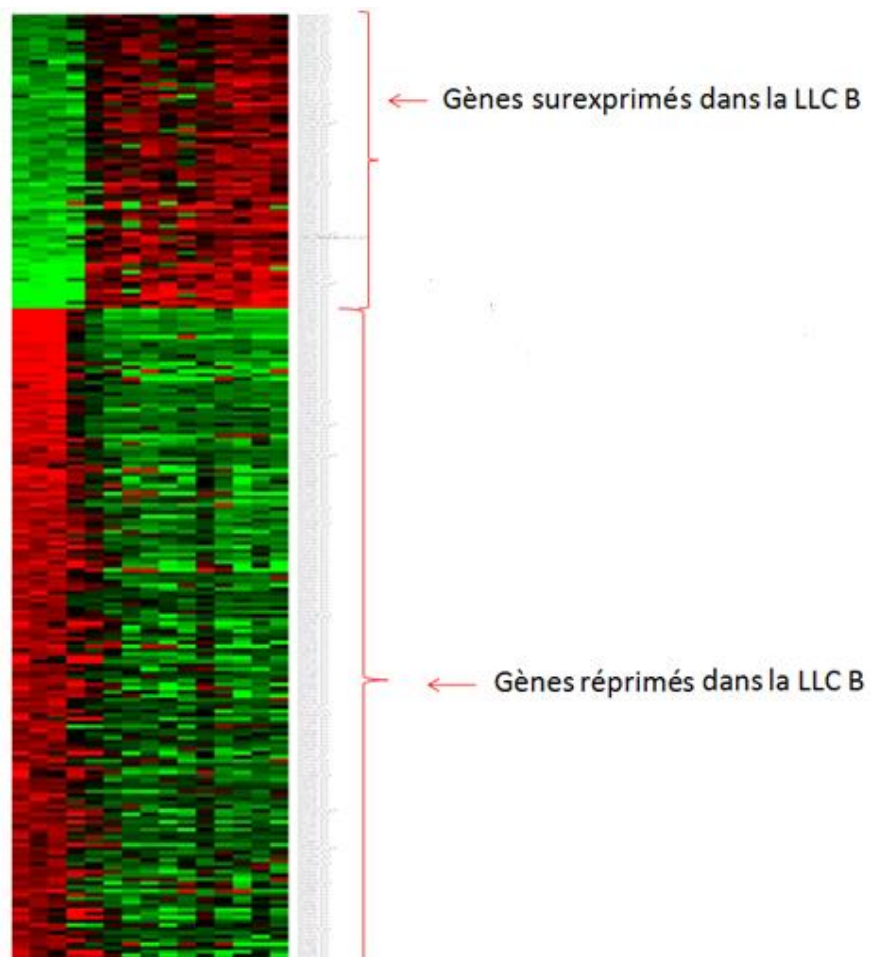


**Figure 34 : Représentation graphique de l'expression différentielle des gènes totaux**

Le seuil de signification est de 274 gènes.

Les rouges représentent les gènes surexprimés ou « upregulated » qui sont en nombre de 77 gènes.

Les verts représentent les gènes réprimés ou « down-regulated » qui sont au nombre de 197 gènes.



**Figure 35 : Heatmap représentant l'expression différentielle des gènes**

Les rouges représentent les gènes surexprimés ou « upregulated », lors de la LLC-B comparé à leur expression dans les cas normaux.

Les verts représentent les gènes réprimés ou « down-regulated » lors de la LLC\_B comparé à leur expression normale.

En noir, ce sont les gènes qui s'expriment de la même manière pour les patients et les contrôles sains.

En plus des heatmaps, l'analyse SAM confère des tableaux pour chaque cas :

- Liste des gènes surexprimés.
- Liste des gènes réprimés.

**Tableau V : Liste des gènes positivement régulés chez les leucémique : (surexpression)**

"Probe.Set.ID"	"Gene.Symbol"	"Gene.Title"	Expected sco...	Observed sc...	Numerator(r)	Denominator...	Fold change(...
"225238_at"	"MSI2"	"musashi ho...	-0.17443664	0.73003566	2.1362934	2.9262865	5.9135756
"1570021_at"	"---"	"---"	-0.025012588	0.731416	1.9502299	2.6663759	4.1228175
"209920_at"	"BMPR2"	"bone morph...	-0.033482328	0.7342905	1.9632695	2.673696	4.177408
"227260_at"	"---"	"---"	-0.060353447	0.74487305	2.0136347	2.703326	4.349855
"1555904_at"	"---"	"---"	-0.14396577	0.74665797	2.127006	2.848702	4.911614
"238604_at"	"---"	"---"	-0.13130704	0.7468029	2.1118178	2.8278115	4.947204
"230793_at"	"LRRC16A"	"leucine rich ...	-0.13285682	0.7506982	2.1229115	2.827916	4.6163774
"231152_at"	"INO80D"	"INO80 comp...	-0.066213384	0.7514901	2.033393	2.7058146	4.327611
"225062_at"	"LOC389831"	"uncharacteri...	-0.1481985	0.7526942	2.1468801	2.8522608	5.252759
"221765_at"	"UGCG"	"UDP-glucos...	-0.12916614	0.7577031	2.1313908	2.812963	5.1005726
"212970_at"	"--"	"--"	-0.18027052	0.78393483	2.2732356	2.8997762	5.999991
"1564996_at"	"---"	"---"	-0.19986916	0.7867593	2.3154821	2.9430628	6.0617585
"209469_at"	"GPM6A"	"glycoprotein ...	-0.320942	0.78702354	2.5256515	3.2091181	6.3247423
"1565804_at"	"---"	"---"	-0.06088107	0.7870325	2.0962448	2.6634793	4.4421396
"1557910_at"	"HSP90AB1"	"heat shock ...	-0.13364902	0.7910549	2.212995	2.797524	5.0598803
"230939_at"	"---"	"---"	-0.17333701	0.79195684	2.2824469	2.8820343	5.864257
"226408_at"	"TEAD2"	"TEA domain...	-0.25051066	0.80127746	2.4422364	3.0479286	6.615901
"238323_at"	"TEAD2"	"TEA domain...	-0.28354642	0.8066579	2.5130334	3.1153646	8.6081295
"47553_at"	"DFNB31"	"deafness, a...	-0.36617297	0.80684364	2.669717	3.3088408	13.099091
"242774_at"	"SYNE2"	"spectrin rep...	-0.18603486	0.81026393	2.3444877	2.8934865	6.095533
"238759_at"	"CCDC88A"	"coiled-coil d...	-0.1918839	0.8118726	2.3578947	2.9042668	6.0703835
"228737_at"	"TOX2"	"TOX high m...	-0.40015894	0.8146804	2.7584715	3.3859553	11.019518
"239317_at"	"CEACAM21"	"carcinoembr...	-0.2190889	0.8192647	2.4301784	2.9662921	6.9964795
"1562754_at"	"LOC339260"	"uncharacteri...	-0.30404267	0.82017976	2.5871837	3.1544106	10.169219
"236458_at"	"---"	"---"	-0.20460802	0.8257914	2.4153986	2.92495	6.2694297
"227666_at"	"DCLK2"	"doublecortin...	-0.2518576	0.8259047	2.50737	3.035907	7.070748
"226560_at"	"---"	"---"	-0.2084433	0.8278292	2.4294512	2.9347253	6.4801517
"230689_at"	"---"	"---"	-0.13679875	0.8288285	2.2976563	2.7721734	5.3574576
"1555731_a...	"AP1S3"	"adaptor-rela...	-0.21636158	0.83016235	2.4513047	2.9528015	6.689707
"227921_at"	"---"	"---"	-0.2539	0.83022976	2.5223975	3.038192	7.980313
"229569_at"	"---"	"---"	-0.26881737	0.83779204	2.5652456	3.061912	8.942162
"230551_at"	"KSR2"	"kinase supp...	-0.29992923	0.84600234	2.648231	3.1302881	9.932384
"230110_at"	"MCOLN2"	"mucolipin 2"	-0.105704725	0.85004497	2.2804313	2.6827185	5.149171
"230312_at"	"---"	"---"	-0.23134051	0.8507807	2.531323	2.975294	7.2888713
"242541_at"	"ABCA9"	"ATP-binding...	-0.33884868	0.86853325	2.7892938	3.2114992	11.513324
"40148_at"	"APBB2"	"amyloid bet...	-0.23433018	0.87405795	2.5959148	2.9699574	7.6733336
"213419_at"	"APBB2"	"amyloid bet...	-0.278029	0.8892914	2.7137682	3.0516074	8.88756

Tableau VI : Liste des gènes négativement régulés ( downregulated ou réprimés)

"Probe.Set.ID"	"Gene.Symbol"	"Gene.Title"	Expected sco...	Observed sc...	Numerator(r)	Denominator...	Fold change(...
"227646_at"	"EBF1"	"early B-cell f...	-0.5986825	-2.1542242	-6.70924	3.1144576	0.020548474
"232204_at"	"EBF1"	"early B-cell f...	-0.4641133	-1.995194	-5.4208407	2.7169492	0.025970588
"233261_at"	"EBF1"	"Early B-cell f...	-0.494709	-1.9160179	-5.531197	2.8868191	0.031150522
"204963_at"	"	"	-0.41347235	-1.8398776	-4.771692	2.5934832	0.037172608
"244876_at"	"--"	"--"	-0.4508895	-1.7656525	-5.0428944	2.856108	0.040312942
"229487_at"	"EBF1"	"early B-cell f...	-0.43318847	-1.7635492	-4.8788404	2.7664895	0.03992612
"244387_at"	"--"	"--"	-0.3815515	-1.623692	-4.3634553	2.6873662	0.051208798
"214777_at"	"--"	"--"	-0.5862298	-1.5991063	-5.6773305	3.5503147	0.26492804
"227006_at"	"PPP1R14A"	"protein phos...	-0.36428234	-1.5903606	-4.2658834	2.682337	0.05553365
"226932_at"	"SSPN"	"sarcospan"	-0.33246663	-1.5670849	-4.0613112	2.5916345	0.06039148
"209789_at"	"CORO2B"	"coronin, acti...	-0.2994226	-1.507205	-3.8254833	2.5381308	0.06867309
"216517_at"	"IGKC"	"immunoglo...	-0.32403925	-1.4960316	-3.9579897	2.645659	0.06342345
"215768_at"	"	"	-0.3987948	-1.4740323	-4.278082	2.9022987	0.059871286
"222450_at"	"PMEPA1"	"prostate tran...	-0.29030782	-1.4419711	-3.7166183	2.577457	0.07764878
"219049_at"	"CSGALNAC...	"chondroitin ...	-0.4732013	-1.429071	-4.639698	3.2466533	0.090586886
"203939_at"	"NT5E"	"5'-nucleotid...	-0.42017448	-1.4289738	-4.285181	2.9987822	0.07956412
"212592_at"	"IGJ"	"immunoglo...	-0.6513165	-1.4103572	-5.4380646	3.8558066	0.13253202
"228854_at"	"--"	"--"	-0.40442187	-1.3452222	-4.0819707	3.034421	0.09849191
"217979_at"	"TSPAN13"	"tetraspanin ...	-0.5232925	-1.3281746	-4.6439342	3.4964788	0.122620344
"229568_at"	"MOB3B"	"MOB kinase ...	-0.2954875	-1.3224835	-3.6169815	2.7349918	0.09016525
"213122_at"	"TSPYL5"	"TSPY-like 5"	-0.28665853	-1.320901	-3.5717773	2.7040462	0.08051702
"225809_at"	"PARM1"	"prostate and...	-0.2743424	-1.3176275	-3.5214543	2.672572	0.09360249
"222449_at"	"PMEPA1"	"prostate tran...	-0.24460734	-1.3165877	-3.3913555	2.5758677	0.095625214
"235652_at"	"--"	"--"	-0.31212726	-1.3044747	-3.6464767	2.7953603	0.10079222
"1560520_at"	"LOC401312"	"uncharacteri...	-0.2340244	-1.2917972	-3.3289232	2.5769706	0.09963782
"226844_at"	"MOB3B"	"MOB kinase ...	-0.33385998	-1.2855952	-3.703366	2.8806627	0.09938953
"235278_at"	"MACROD2"	"MACRO do...	-0.3536494	-1.2647538	-3.740323	2.9573526	0.091237925
"239806_at"	"--"	"--"	-0.31047726	-1.2586336	-3.563947	2.8316	0.07530209
"202796_at"	"SYNPO"	"synaptopodin	-0.22074795	-1.2346598	-3.1932323	2.5863256	0.10684575
"233969_at"	"--"	"--"	-0.21799782	-1.2256951	-3.175549	2.5908148	0.10734229
"1563209_a...	"MACROD2"	"MACRO do...	-0.2616793	-1.223501	-3.33905	2.7290947	0.09942527
"228434_at"	"BTNL9"	"butyrophilin-...	-0.2392321	-1.2137862	-3.2527244	2.6798165	0.11169398
"235802_at"	"PLD4"	"phospholipa...	-0.21745332	-1.2115656	-3.157745	2.6063342	0.11141865
"203038_at"	"PTPRK"	"protein tyros...	-0.44845992	-1.178654	-3.9265556	3.3313897	0.16942464
"201417_at"	"SOX4"	"SRY (sex de...	-0.37720525	-1.1703857	-3.6110497	3.08535	0.15602827
"205987_at"	"CD1C"	"CD1c molec...	-0.33177736	-1.144623	-3.4277487	2.994653	0.1414665
"243231_at"	"SLC38A11"	"solute carri...	-0.30561188	-1.1401519	-3.3408656	2.9301934	0.11131989

## 2)-analyse graphique des LncRNAs via l'outil de visualisation TMEv

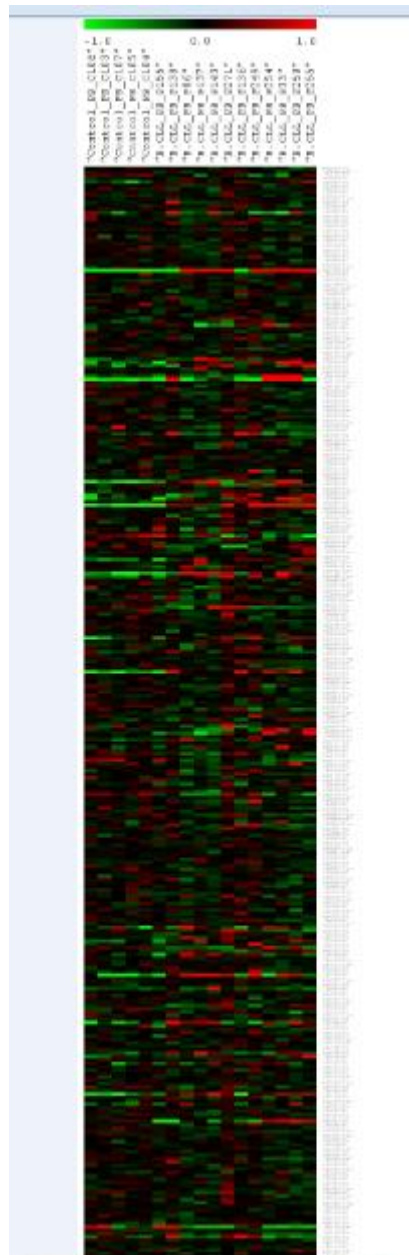
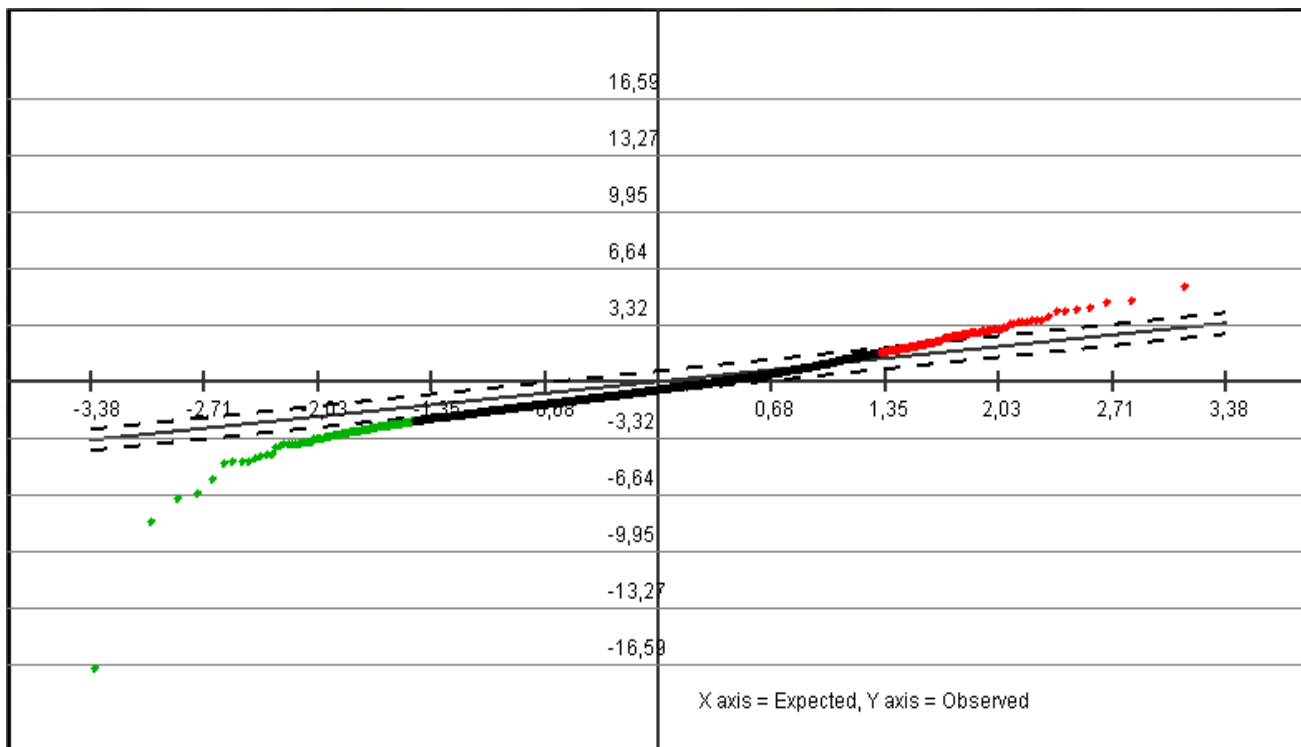


Figure 36 : Heatmap représentant l'expression des LncRNAs

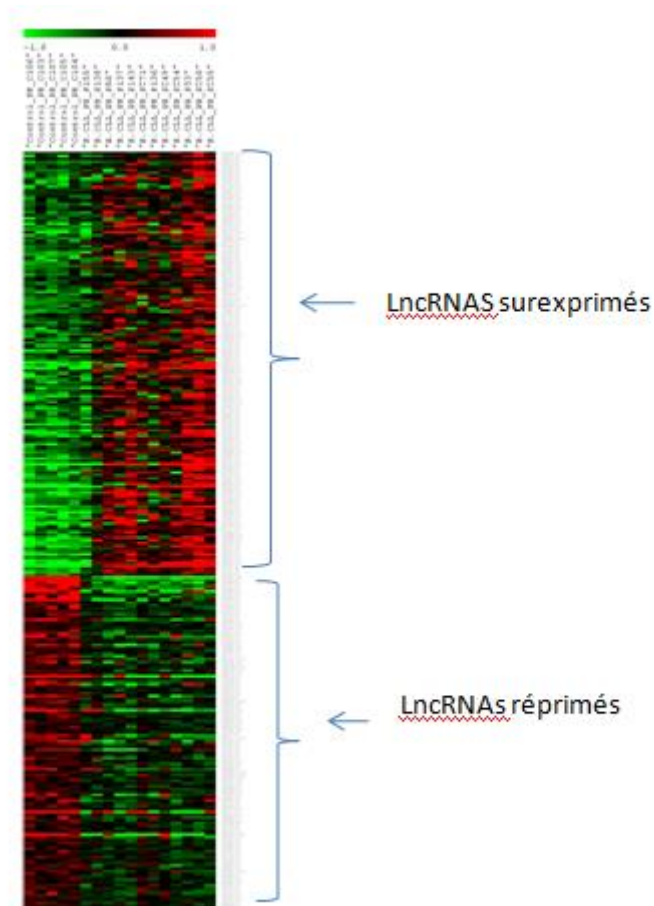
- ANALYSE SAM :



**Figure 37 : Représentation graphique SAM des LncRNAs différentiellement exprimés**

Avec l'outil statistique SAM, nous avons obtenu approximativement 268 gènes significatifs.





**Figure 38 : Heatmap représentant l'expression différentielle des LncRNAs**

Les rouges représentent les gènes surexprimés ou « upregulated », lors de la LLC-B comparé à leur expression dans les cas normaux. ( au nombre de 150)

Les verts représentent les gènes réprimés ou « down-regulated » lors de la LLC\_B comparé à leur expression normale. ( au nombre de 118)

En noir, ce sont les gènes qui s'expriment de la même manière pour les patients et les contrôles sains.

En plus des heatmaps, l'analyse SAM confère des tableaux pour chaque cas :

- Liste des gènes surexprimés. ( down-regulated)
- Liste des gènes réprimés.( up-regulated)

- Liste des gènes up-regulated lors de la leucémie :

**Tableau VIII : Liste des LncRNAs surexprimés dans les Leucémies chroniques B**

"probe"	"Gene.Symbol"	"Gene.Title"	Expected sco...	Observed sc...	Numerator(r)	Denominator...	Fold change(...
"1562997_a...	"--"	"--"	-0.27905846	1.9428866	0.33730918	0.17361239	1.2918729
"230678_at"	"RGS5"	"regulator of ...	0.28759405	1.9473834	0.6491022	0.3333202	1.5748059
"228440_at"	"--"	"--"	0.19652344	1.9529122	0.14912663	0.07636116	1.1109747
"238139_at"	"--"	"--"	0.7208883	1.9529504	0.76935905	0.39394706	1.9897527
"215344_at"	"--"	"--"	0.06191247	1.9656799	0.4778297	0.24308622	1.4595965
"236042_at"	"LOC100130...	"uncharacteri...	0.535467	1.9670527	0.24594346	0.12503146	1.1846918
"240561_at"	"LOC100652...	"uncharacteri...	0.9464604	1.9675683	0.55662334	0.2828991	1.5535221
"227893_at"	"LINC00476"	"long interge...	0.18352751	1.9784112	0.61832964	0.31253847	1.6829726
"239846_at"	"LOC100506...	"uncharacteri...	0.85794204	1.9796364	0.70117533	0.354194	1.7777635
"243048_at"	"CECR7"	"cat eye synd...	1.2531838	1.9884851	0.7851258	0.39483616	1.9212946
"241834_at"	"IPW   LOC...	"imprinted in ...	1.1246864	1.9906106	0.64985704	0.32646117	1.629654
"227745_at"	"LOC100507...	"uncharacteri...	0.181467	1.9915361	0.7235071	0.36329097	1.8396329
"1557293_at"	"--"	"--"	-1.1328019	1.994622	0.73883945	0.37041578	1.7999356
"240194_at"	"LOC441204"	"uncharacteri...	0.8994565	1.9967258	0.29811484	0.14930184	1.2446191
"228567_at"	"MIR4720"	"microRNA 4...	0.20066056	1.9979051	0.23798737	0.11911845	1.1853622
"244322_at"	"LOC646329"	"uncharacteri...	1.6560699	2.0038853	0.30405962	0.15173505	1.2357626
"1568813_at"	"LOC100506...	"uncharacteri...	-0.09925245	2.0068545	0.38781708	0.19324623	1.3222595
"237249_at"	"KCNQ10T1"	"KCNQ1 opp...	0.6234916	2.0144536	0.2955635	0.14672141	1.220475
"225225_at"	"OIP5-AS1"	"OIP5 antise...	0.14330181	2.0252252	0.5359508	0.26463762	1.5316435
"1562718_at"	"--"	"--"	-0.33004925	2.0312276	0.3994407	0.19664991	1.351805
"240589_at"	"LINC00313"	"long interge...	0.95332146	2.0328164	0.22036852	0.10840552	1.1641703
"235134_at"	"--"	"--"	0.49670076	2.038877	0.61789584	0.30305696	1.6409885
"223628_at"	"--"	"--"	0.12094762	2.043959	0.43662432	0.21361698	1.3938856
"222128_at"	"NSUN6"	"NOP2/Sun d...	0.11551808	2.0525901	0.599624	0.2921304	1.5750403
"1558794_at"	"LOC728190"	"uncharacteri...	-0.8872618	2.0530264	0.43844232	0.21355902	1.3996603
"228387_at"	"--"	"--"	0.1938238	2.0552728	1.1444386	0.5568305	2.5696013
"232772_at"	"LOC221272"	"uncharacteri...	0.3975168	2.0657012	0.6271819	0.30361694	1.6770372
"244429_at"	"--"	"--"	1.7010934	2.0677717	0.87748396	0.42436212	2.1590776
"1557167_at"	"HCG11"	"HLA comple...	-1.1603637	2.0801427	1.1367486	0.54647624	2.5570376
"229227_at"	"DICER1-AS1"	"DICER1 anti...	0.22956067	2.0818067	0.60574216	0.29096946	1.5844328
"1569100_a...	"--"	"--"	-0.07796811	2.084131	0.6452617	0.3096071	1.7240942
"228909_at"	"--"	"--"	0.2150727	2.1002698	0.5438347	0.25893563	1.4972211
"220703_at"	"IDI2-AS1"	"IDI2 antisen...	0.10326286	2.100806	0.32535133	0.15486977	1.2722338
"229388_at"	"LOC100507...	"uncharacteri...	0.23438577	2.1045856	0.49718028	0.23623666	1.4709327
"244749_at"	"--"	"--"	2.0601153	2.1060483	0.31272814	0.14849049	1.2443752
"243699_at"	"LOC100507...	"uncharacteri...	1.3915929	2.1098003	0.8774215	0.41587892	2.1188145
"232504_at"	"LOC285628...	"uncharacteri...	0.38713223	2.1121955	1.0047694	0.47569907	2.8274913

- Liste des gènes up-regulated lors de la leucémie :

**Tableau XI : Liste des LncRNAs down-regulated lors des leucémies chronique B.**

"probe"	"Gene.Symbol"	"Gene.Title"	Expected sco...	Observed sc...	Numerator(r)	Denominator...	Fold change(...
"1560520_at"	"LOC401312"	"uncharacteri...	-0.68732125	-16.590004	-3.3289232	0.20065838	0.09963782
"244280_at"	"LOC100507..."	"uncharacteri...	1.6501707	-7.9353557	-1.9522972	0.24602517	0.2323768
"227985_at"	"LOC100506..."	"uncharacteri...	0.1883109	-6.6831136	-1.6573035	0.24798374	0.32511306
"243977_at"	"LOC541472"	"uncharacteri...	1.4988853	-6.307063	-1.3187172	0.20908579	0.38450614
"243185_at"	"LOC100507..."	"uncharacteri...	1.2699608	-5.476652	-1.6410879	0.29965165	0.2982471
"1568838_at"	"WASIR2"	"WASH and I...	-0.09720293	-4.6130576	-2.0564766	0.4457947	0.2716255
"237115_at"	"--"	"--"	0.6102267	-4.4484544	-0.7768774	0.17463985	0.57744247
"241535_at"	"LOC100292..."	"uncharacteri...	1.1048125	-4.442656	-1.0608819	0.2387945	0.43517813
"239186_at"	"MGC39372"	"serpin pepti...	0.7902054	-4.4217978	-1.9053756	0.4309052	0.27513346
"242747_at"	"--"	"--"	1.2037392	-4.2400784	-0.48679197	0.1148073	0.7142626
"1559221_at"	"LOC100507..."	"uncharacteri...	-0.833082	-4.1673794	-0.6567473	0.15759239	0.61639076
"230451_at"	"LOC100506..."	"uncharacteri...	0.2713845	-4.076695	-0.64553165	0.15834682	0.620994
"236973_at"	"LOC100131..."	"uncharacteri...	0.5944887	-4.010891	-0.33204257	0.08278524	0.7943977
"232111_at"	"LINC00617"	"long interge...	0.36949685	-3.5955646	-0.5844474	0.16254677	0.6634487
"243742_at"	"LOC100505..."	"uncharacteri...	1.4176304	-3.5100262	-0.38134226	0.10864371	0.76750946
"235696_at"	"--"	"--"	0.521612	-3.472374	-1.0641875	0.30647263	0.4948924
"1556536_at"	"LOC729224"	"uncharacteri...	-1.3476303	-3.4594853	-0.30719027	0.088796526	0.8096041
"1557778_at"	"--"	"--"	-1.0195585	-3.4015403	-0.26404974	0.07762652	0.8341889
"1562655_at"	"--"	"--"	-0.33996865	-3.3988547	-0.26202065	0.07709087	0.8317192
"236977_at"	"LOC646588"	"uncharacteri...	0.5953678	-3.3836508	-0.5156729	0.15240134	0.69994515
"1560978_at"	"--"	"--"	-0.6120324	-3.3815043	-0.74854445	0.22136433	0.5789202
"219871_at"	"FLJ13197"	"uncharacteri...	0.092441246	-3.3134	-0.9584	0.28924972	0.52957463
"233398_at"	"--"	"--"	0.42385352	-3.3064606	-0.46785522	0.14149728	0.7285281
"1561333_at"	"--"	"--"	-0.53165585	-3.3052952	-0.28326705	0.08570098	0.818896
"232589_at"	"LOC100506..."	"uncharacteri...	0.39009023	-3.2689602	-1.0612162	0.3246342	0.5089844
"1558123_at"	"FLJ35390   ..."	"uncharacteri...	-0.96960807	-3.1212432	-0.37715957	0.120836325	0.77474946
"242860_at"	"--"	"--"	1.2256397	-3.1211948	-0.27628246	0.08851817	0.828155
"1557483_at"	"LOC284788"	"uncharacteri...	-1.097004	-3.0793545	-0.42611602	0.13837835	0.74838793
"237772_at"	"--"	"--"	0.6936114	-3.0610085	-0.6994917	0.22851674	0.63304913
"1558147_a..."	"FLJ90757"	"uncharacteri...	-0.9665891	-3.060013	-0.5044703	0.16485886	0.7151944
"1559787_at"	"--"	"--"	-0.7672497	-3.0275354	-0.8620827	0.28474736	0.54829717
"1563685_at"	"--"	"--"	-0.20280287	-3.0021553	-0.32893884	0.10956756	0.79302907
"1560582_a..."	"--"	"--"	-0.6737919	-2.9812696	-0.26637924	0.08935094	0.83405274
"1552729_at"	"SNHG7   S..."	"small nucle..."	-2.8835104	-2.9581652	-0.48599213	0.16428837	0.7258707
"1557564_at"	"LOC100507..."	"uncharacteri...	-1.0754192	-2.9506812	-0.25503477	0.08643251	0.83703315
"229040_at"	"LOC100505..."	"uncharacteri...	0.22267641	-2.9219675	-2.1085548	0.7216216	0.4022373
"243098_at"	"--"	"--"	1.2603133	-2.9207737	-0.2603567	0.08913963	0.8362563

**III.1.5 Analyse fonctionnelle des gènes différentiellement exprimés**

Après l'obtention des listes de gènes dérégulés lors de la leucémie chronique B, nous sommes passés à l'annotation fonctionnelle des gènes différentiellement exprimés sur DAVID, afin de fournir une interprétation fonctionnelle sur ces gènes dérégulés dans les diverses voies biologiques, et leur implication dans ces dernières.

- **Annotation fonctionnelle des LncRNAs différentiellement exprimés dans les LLC-B :**

- **Upregulated LncRNAs (surexprimés)**

Tout d'abord, nous avons commencé par chercher les voies biologiques ou les LncRNAs régulés positivement ou surexprimés qui pourraient être impliqués.

Les résultats après l'analyse effectuée sur la base de données « KEGG PATHWAY » (**Kyoto Encyclopedia of Genes and Genomes**) présentent différentes voies potentielles, mais pour notre travail, deux voies liées au processus cancéreux sont particulièrement intéressantes qui sont :

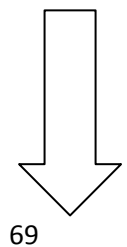
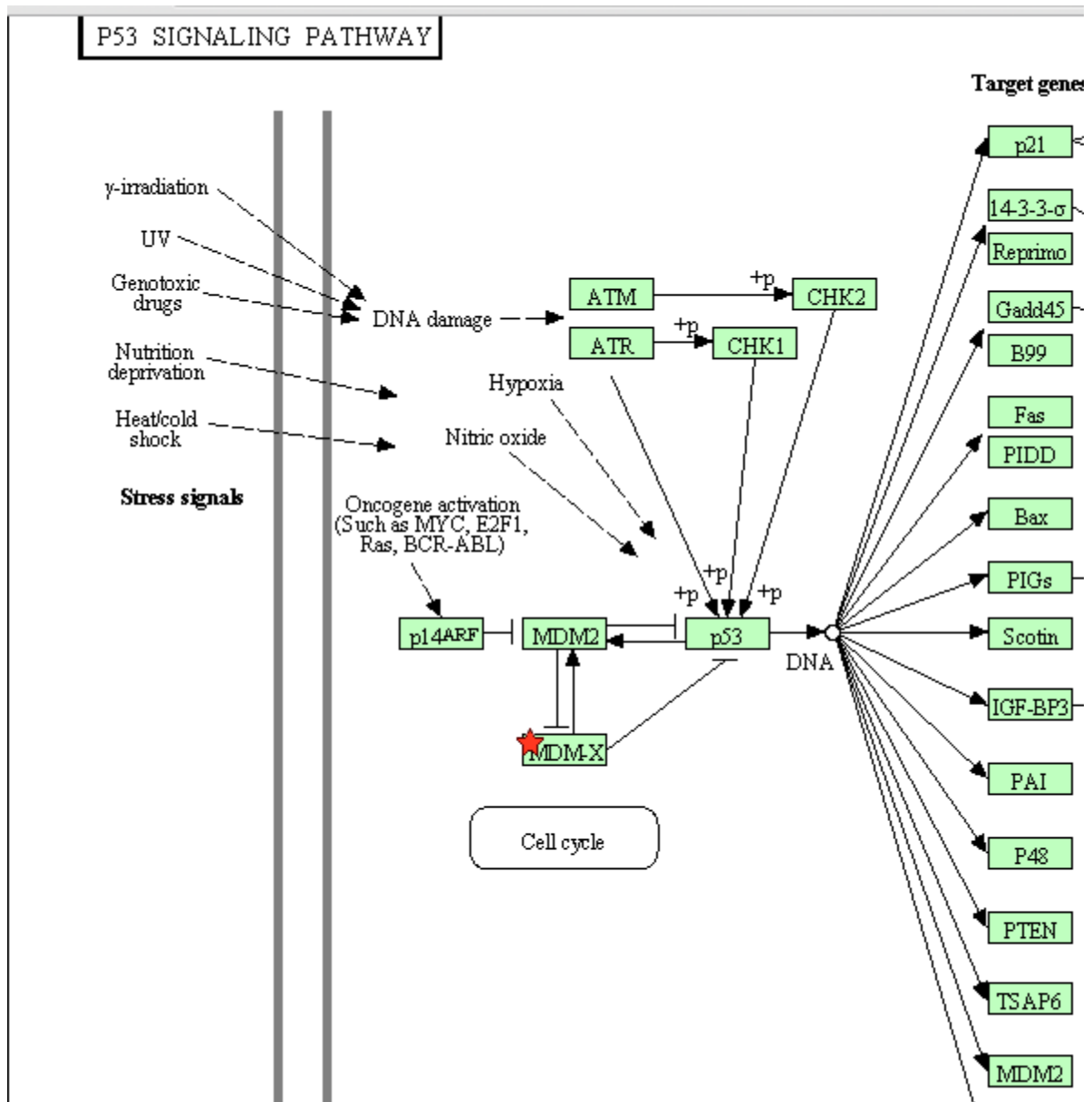
- « P53 signaling pathway »
- « Calcium signaling pathway »

241876_at	<a href="#">Mdm4 p53 binding protein homolog (mouse)</a>
KEGG_PATHWAY	<a href="#">p53 signaling pathway,</a>
238089_at	<a href="#">mannosidase, alpha, class 2C, member 1</a>
KEGG_PATHWAY	<a href="#">Other glycan degradation,</a>
239846_at	<a href="#">methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase</a>
KEGG_PATHWAY	<a href="#">Glyoxylate and dicarboxylate metabolism, One carbon pool by folate,</a>
243996_at	<a href="#">natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)</a>
KEGG_PATHWAY	<a href="#">Purine metabolism, Vascular smooth muscle contraction,</a>
239242_at	<a href="#">solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5; solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5 pseudogene 8</a>
KEGG_PATHWAY	<a href="#">Calcium signaling pathway, Parkinson's disease, Huntington's disease,</a>

**Figure 39 : Tableau résultats DAVID des voies biologique cible des LncRNAs**

- **La voie de signalisation P53 :**

Cette voie est l'un des mécanismes de suppression tumorale (Lu, 2005), nous observons que notre molécule sous le probe ID « 241876 » régule l'expression du facteur MDM-X qui est dans le cas normal un inhibiteur de la P53 (Shadfian *et al.*, 2012).



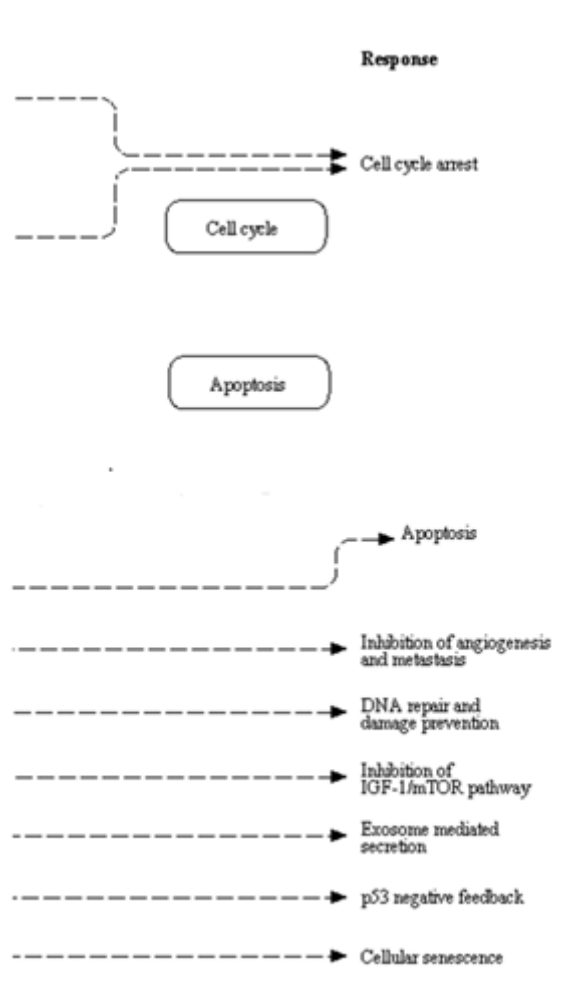


Figure 40 : La voie de signalisation P53

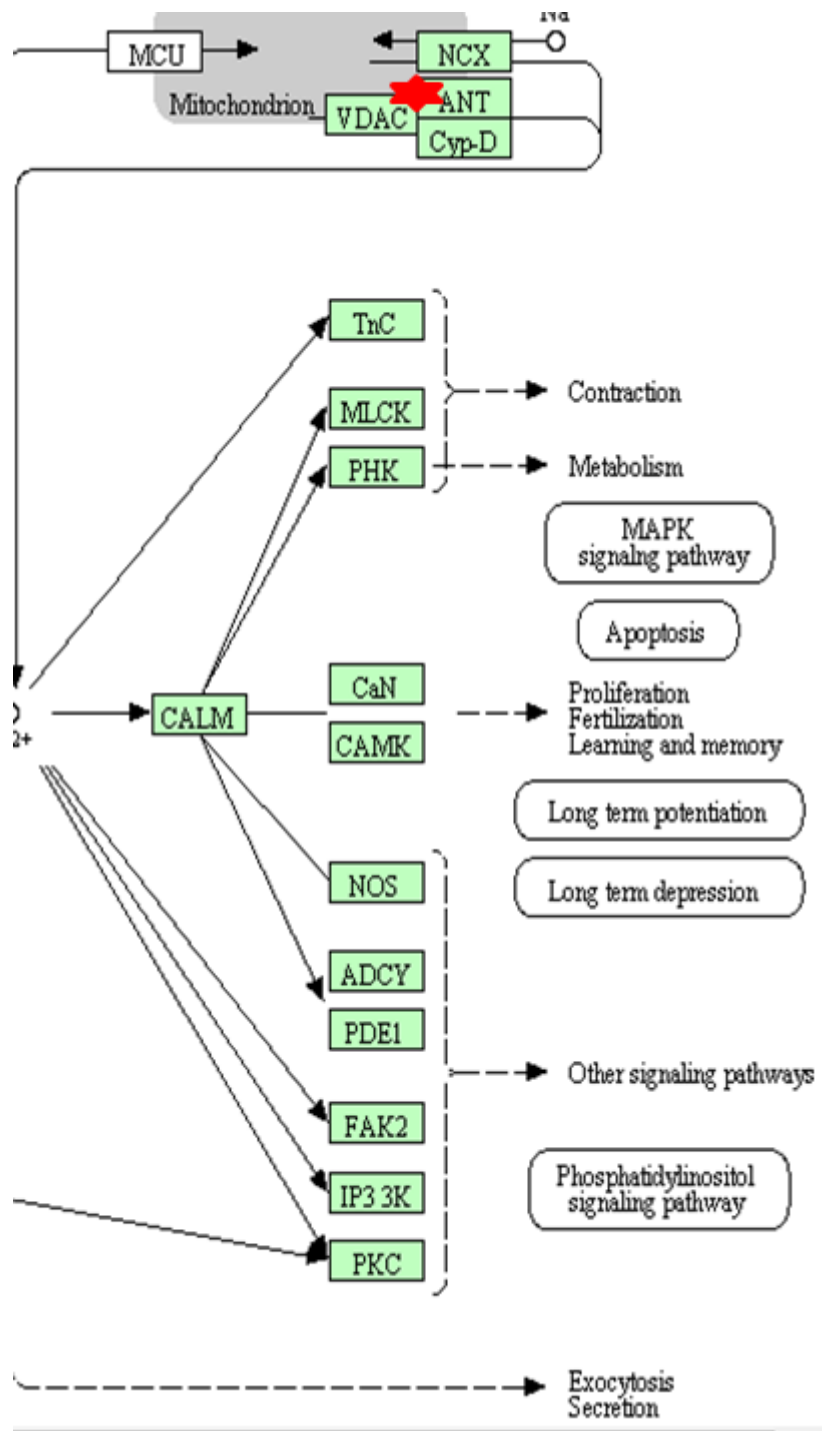


Figure 41 : La voie de signalisation calcique

- Calcium signaling pathway :** La voie de signalisation calcique régule plusieurs processus biologiques par le biais des protéines  $Ca^{2+}$  dépendantes, parmi elles, la régulation de l'équilibre des cellules B (Véronique, 2000). Selon notre résultat, il semblerait que la molécule « SLC25A5-AS1 » régule l'« Adénine nucléotide translocator ou ANP » responsable de la translocation de l'ADP/ATP.

### ▪ Down regulated LncRNAs dans la LLC-B :

Suite à l'annotation des LncRNAs surexprimés, nous sommes passés à l'annotation des LncRNAs négativement régulés ou réprimés.

Comme résultats, nous avons aussi diverses voies potentielles, où nos molécules de LncRNAs semblent être impliquées.

La recherche effectuée sur la base de données « KEGG PATHWAY » nous a donné les résultats compris dans ce tableau ci-dessous :

1565731_at	aldehyde dehydrogenase 3 family, member B1	Related Genes	Homo sapiens
KEGG_PATHWAY	Glycolysis / Gluconeogenesis, Histidine metabolism, Tyrosine metabolism, Phenylalanine metabolism, Metabolism of xenobiotics by cytochrome P450, Drug metabolism,		
241055_at	endothelial PAS domain protein 1	Related Genes	Homo sapiens
KEGG_PATHWAY	Pathways in cancer, Renal cell carcinoma,		
236990_at	eukaryotic translation initiation factor 2-alpha kinase 3	Related Genes	Homo sapiens
KEGG_PATHWAY	Alzheimer's disease,		
243977_at	interleukin 6 (interferon, beta 2)	Related Genes	Homo sapiens
KEGG_PATHWAY	Cytokine-cytokine receptor interaction, Toll-like receptor signaling pathway, NOD-like receptor signaling pathway, Cytosolic DNA-sensing pathway, Jak-STAT signaling pathway, Hematopoietic cell lineage, Intestinal immune network for IgA production, Prion diseases, Pathways in cancer, Graft-versus-host disease, Hypertrophic cardiomyopathy (HCM),		
236785_at	mitogen-activated protein kinase kinase kinase 5	Related Genes	Homo sapiens
KEGG_PATHWAY	MAPK signaling pathway, Neurotrophin signaling pathway, Amyotrophic lateral sclerosis (ALS),		
238085_at	phosphodiesterase 8B, cGMP-specific, rod, beta	Related Genes	Homo sapiens
KEGG_PATHWAY	Purine metabolism,		

**Figure 42 : Résultats DAVID des voies dérégulées liées aux LncRNAs DOWN régulés**

Deux voies ont particulièrement retenu notre attention, qui sont : les voies de signalisation liées aux cancers ainsi que la voie des MAPK.

- **Voies liées aux cancers :**

Dans ce résultat, on trouve différentes voies de développement tumoral :

**L'angiogenèse** ou la formation de nouveaux vaisseaux sanguins lors de la prolifération tumorale (Roger Lacave et *al.*, 2005). Notre LncRNA identifié sous le probe ID « 241055\_at » régule au niveau de l'Apoptotic Signal Regulated Kinase « ASK1 ».



- au niveau du facteur « *HIF-1α* ». responsable de la stimulation de la production des globules rouges
- **Le blocage de la différenciation** ou la molécule dérégulée suite à la dérégulation du LncRNA sous le probe ID« 241055\_at » est l'interleukine 6 « IL-6 » qui participe à la différenciation des cellules B également.

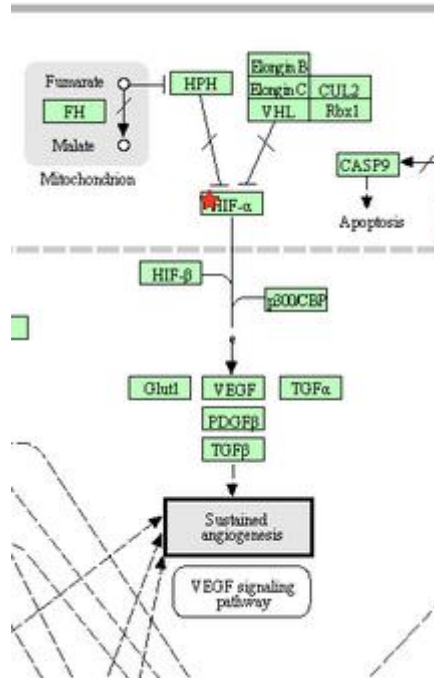


Figure 43 : Les voies de signalisation tumorales ( activation de l'angiogénèse)

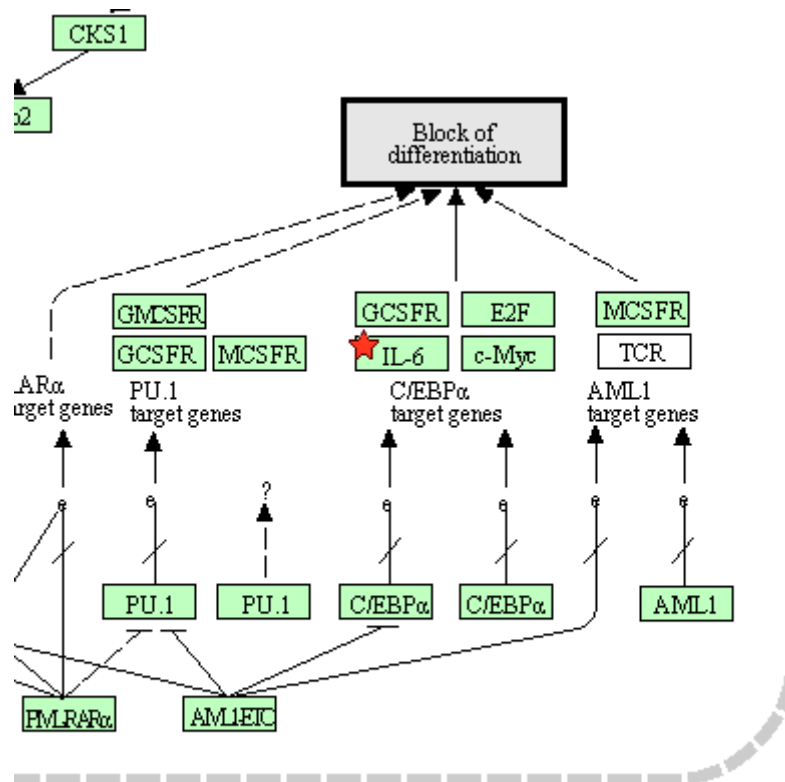


Figure 44 : Les voies de signalisation tumorales (blocage de la différenciation)

- **Voie de signalisation MAPK :**

La voie de signalisation MAPK détient un rôle important dans le contrôle de la prolifération cellulaire (Meyers, 2007). Le lncRNA réprimé est le « LOC100507390 » régulant au niveau de l'Apoptotic Signal Regulated Kinase « ASK1 ».

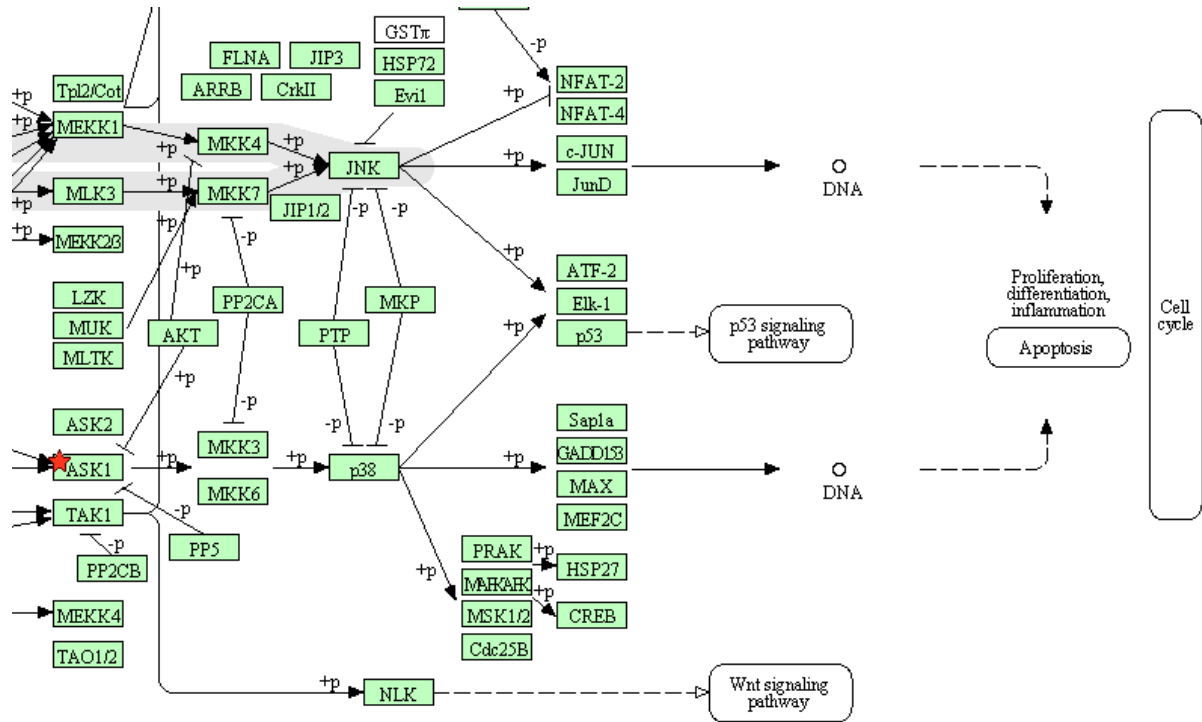


Figure 45 : La voie de signalisation MAPK

Le résultat de recherche effectué sur l'ontologie « GO TERM » regroupe différentes fonctions biologiques que nous avons représentées dans les histogrammes suivant :

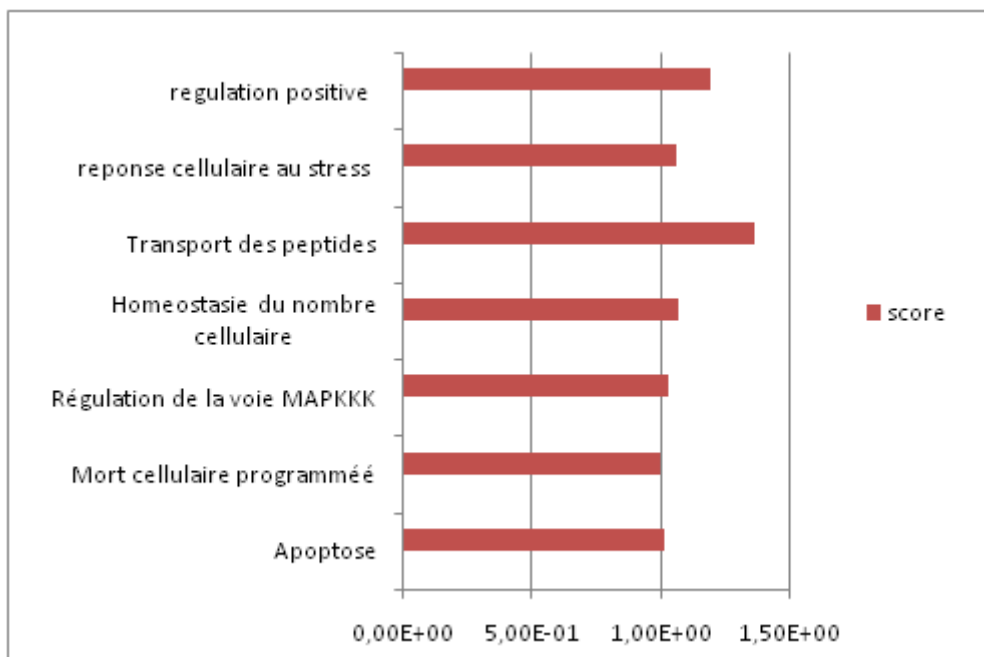


Figure 46 : Histogramme représentant l'implication des LncRNAs réprimés dans différents processus biologiques

D'après cette représentation, nous remarquons l'implication des LncRNAs réprimés dans différents processus tels que l'apoptose, régulations des voies de signalisation...etc, et ils semblent avoir le plus grand score pour le transport des peptides.

- **Annotation fonctionnelle des gènes différentiellement exprimés dans les LLC-B :**

Après l'annotation des LncRNAs seulement, nous utiliserons notre 2ème liste générée à partir des analyses SAM afin de situer fonctionnellement nos gènes différentiellement exprimés lors des leucémies chroniques B.

Nos variantes recherches sur les bases de données qu'abritent DAVID nous exposent les différentes voies et processus biologiques dans lesquels ces gènes exprimés différentiellement ont un rôle.

- **Upregulated genes (surexprimés) :**

Il en résulte de nos recherches dans « KEGG PATHWAY » ainsi que « GO TERM » le tableau suivant, qui regroupe les voies et rôles biologiques les plus importants des gènes surexprimés lors de la leucémie chronique B.

**Tableau X : Tableau présentant les voies dérégulées lors de la surexpression de notre liste de gens.**

Type de régulation	Processus	Les gènes surexprimés.
surexprimé	Voies liées aux cancers	-WNT3 (wingless-type MMTV integration site family, member 3) -HSP90AB1(heat shock protein 90kDa alpha (cytosolic), class B member 1).
surexprimé	WNT signaling pathway	WNT3 (wingless-type MMTV integration site family, member 3)
surexprimé	Régulation de l'apoptose	APBB2 (amyloid beta (A4) precursor protein-binding, family B, member 2)

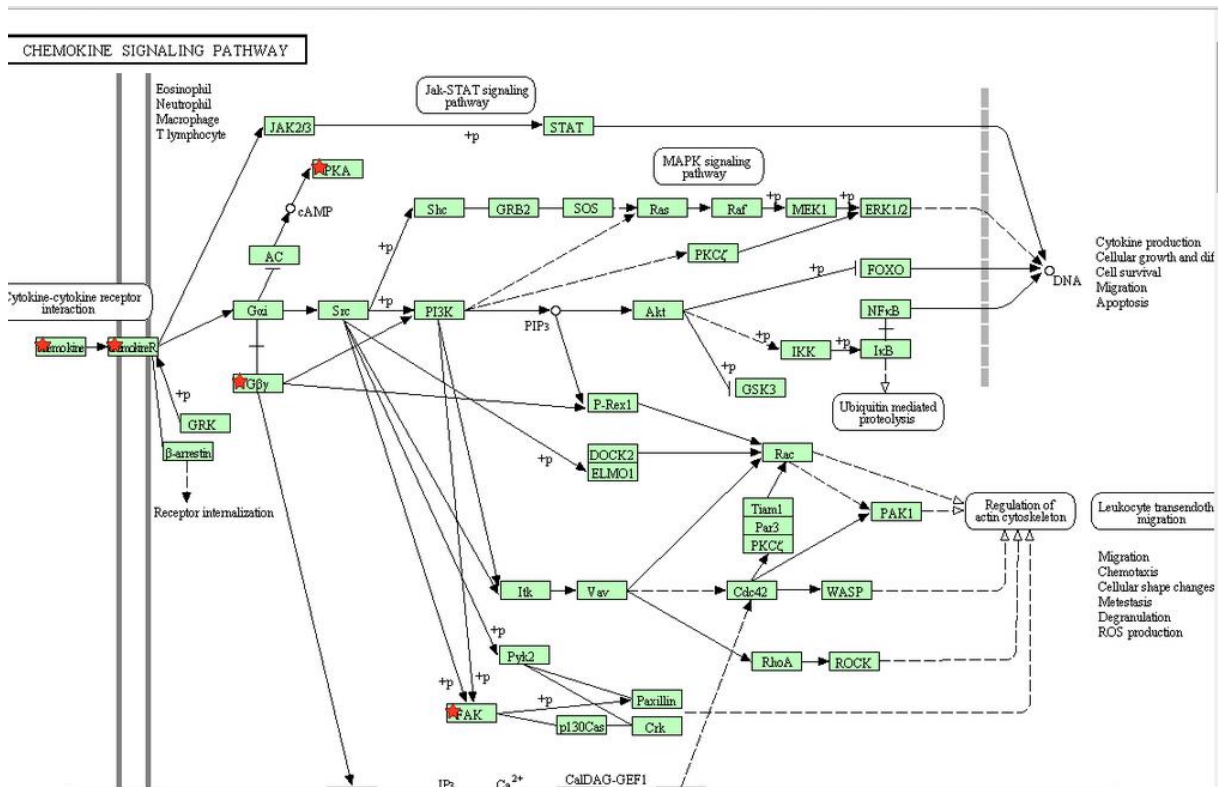
➤ **Downregulated genes( reprimés)**

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Hematopoietic cell lineage</a>	RT		10	6,9	8,8E-8	7,6E-6
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Intestinal immune network for IgA production</a>	RT		6	4,1	9,4E-5	4,1E-3
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Cell adhesion molecules (CAMs)</a>	RT		7	4,8	1,5E-3	4,3E-2
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Cytokine-cytokine receptor interaction</a>	RT		8	5,5	1,2E-2	2,3E-1
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Chemokine signaling pathway</a>	RT		6	4,1	3,3E-2	4,4E-1
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Antigen processing and presentation</a>	RT		4	2,8	4,5E-2	4,9E-1
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Graft-versus-host disease</a>	RT		3	2,1	5,4E-2	5,0E-1
<input type="checkbox"/>	KEGG_PATHWAY	<a href="#">Pathways in cancer</a>	RT		7	4,8	9,3E-2	6,5E-1

**Figure 47 : Annotation fonctionnelle chart DAVID pour les DOWN régulés**

Nos gènes d'intérêt se trouvent impliqués dans des voies telles que :

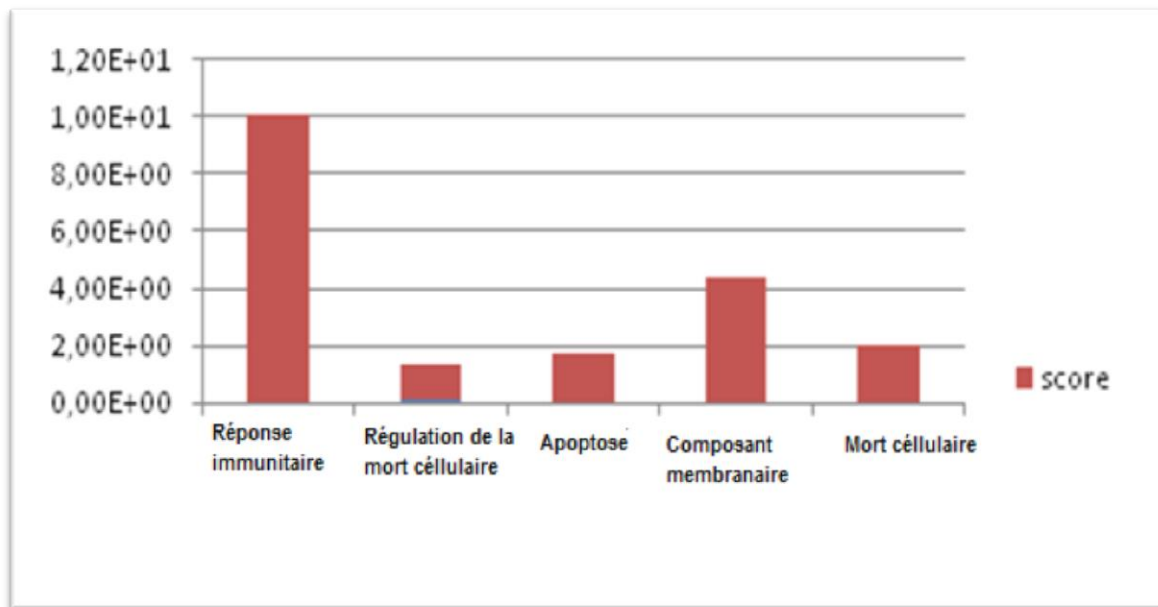
- Voies liées aux cancers : SMAD3 , PTK2 ...etc
- Voies de signalisation des chimokines : Cycline L2 , PTK2 ... etc



**Figure 48 : La voie de signalisation des chimiokines**

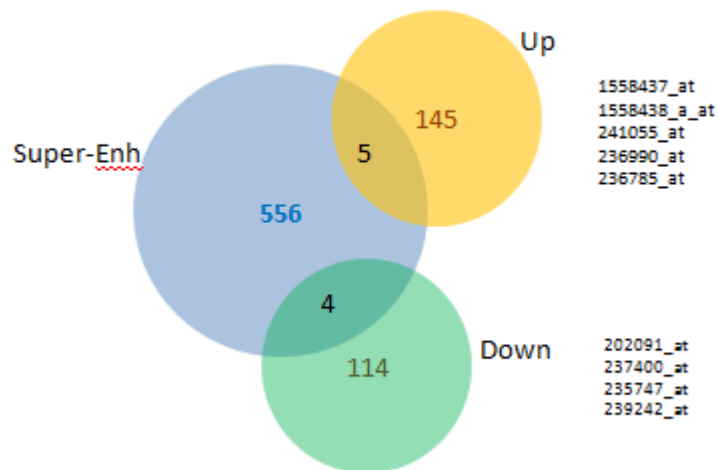
Pour la seconde recherche sur l'ontologie « GO TERM » nous présentons les résultats comme suit :

Les voies qui renferment la majorité de nos molécules dérégulées sont représentées ci-dessous, parmi elles : des acteurs de l'apoptose, composants de la membrane plasmique, acteurs de la réponse immunitaire.



**Figure 49 : Histogramme représentant les voies renfermant la plupart des gènes réprimés d'intérêt**

### **II.1. 6 Identification des LncRNAs associés aux superenhancers :**



**Figure 50 : Résultat de l'intersection des superenhancers avec les LncRNAs**

Parmi nos LncRNAs d'intérêt on retrouve 9 LncRNA qui semblent avoir des similitudes avec les régions régulatrices des superenhancers et sont ainsi des « eRNA » potentiels :

- 4 down regulated.
- 5 UP regulated.

### III.2. DISCUSSION

La leucémie lymphoïde chronique de type B est la forme de leucémie humaine la plus fréquente. Bien que les caractéristiques cliniques et les anomalies génétiques de la LLC-B soient bien connues, les détails moléculaires sous-jacents la maladie sont encore en étude (Goldman and Masson, 2013).

Même si multiples études ont démontré l'importance des lncRNAs dans la pathogénèse du cancer et leur abondance pendant cette dernière, ce qui laisse prédire un rôle potentiel dans son développement. (Nagano and Fraser, 2011), les propriétés fonctionnelles de la majorité des lncRNAs impliqués dans la leucémie lymphoïde chronique B restent pour la plupart inconnues. Par conséquent, nous avons analysé le profil d'expression des ARNm & lncRNAs dans nos échantillons leucémiques et sains, afin de révéler le rôle potentiel des lncRNAs durant la pathogénèse tumorale.

L'analyse des données des puces à ADN a révélé la présence de 267 lncRNAs significativement différentiellement exprimés dont 150 UP regulated ou surexprimés et 117 DOWN regulated ou réprimés.

L'annotation « KEGG » de ces deux listes de gènes a démontré la présence de nos molécules cibles dans 6 voies pour chaque type de régulations des lncRNAs (down & up).

Pour les lncRNAs surexprimés, nous les trouvons impliqués dans la voie de P53 ainsi que la voie calcique.

- L'inactivation de la voie de **P53** est pratiquement présente dans tous les types de cancers. Notre lncRNA sous l'identifiant « 241876\_at », qui est non caractérisé à ce jour, semble être impliqué dans cette voie, plus exactement en régulant l'expression du facteur MDM2 qui en temps normal est un inhibiteur de l'expression de la P53. Nous suggérons que la surexpression de cet lncRNA dans la LLC -B pourrait avoir un rôle d'activateur d'expression du MDM2 ainsi inhiber constamment la P53. Les travaux d'Ali ZHANG et al.





différenciation cellulaire. Notre molécule identifiée sous « 241055\_at » est non caractérisée à ce jour.

Pour ces deux voies citées plus haut, ce LncRNA semble agir au niveau de « HIF- $\alpha$  » et « IL-6 » respectivement.

Ce LncRNA réprimé dans la LLC-B nous laisse penser qu'il puisse avoir un rôle de régulation négative sur le « HIF- $\alpha$  » car ce dernier active l'angiogenèse et la néovascularisation en continu qui favorise la prolifération tumorale. Sülтан et Diedrichs dans leurs publication «LNC's to cancer » (Sülтан et Diedrichs, 2014) ont cité l'implication des LncRNAs dans les divers processus promoteur des tumeurs comme notamment l'angiogenèse, ce qui confirme leur implication dans cette dernière.

D'un autre côté, l'IL-6 inhibe la différenciation des cellules B en activant la voie de blocage de différenciation, ce qui nous laisse prédire que le LncRNA réprimé dans ce cas pourrait être un activateur de la cytokine IL-6.

A notre connaissance, aucune caractérisation de cet LncRNA « 241055\_at » n'est disponible à ce jour, nous ne pouvons que suggérer son rôle en observant son gène cible dans la voie étudiée.

- **Voie MAPK :** cette voie de signalisation détient un rôle important dans le contrôle de la prolifération cellulaire (Meyers, 2007). Notre LncRNA d'intérêt impliqué dans celle-ci est le « LOC100507390 » qui est une molécule non encore caractérisée, régule au niveau du « ASK1 » ou l'Apoptotic Signal Regulated Kinase, fait partie de la famille des MAP kinase kinasekinase, c'est un important médiateur de signal notamment le signal apoptotique.

Nous suggérons alors que notre LncRNA « LOC100507390 » réprimé durant la LLC –B pourrait en temps normal effectuer une régulation positive sur l'ASK1 qui est un transducteur du signal pour l'apoptose, contrairement au cas tumoral où il y a blocage de l'apoptose.

Il n'y a pas de travail à ce jour qui caractérise cette molécule ou son mode d'action, ce qui n'est pas étonnant dans le monde des nouvelles molécules de LncRNAs.

En second lieu, les résultats « GO TERM » associent les LncRNAs réprimés dans la LLC-B à différents rôles comme : transport peptidique, régulation positive, Apoptose, régulation de la voie MAPK ou encore maintien de l'équilibre cellulaire. Ce qui les montre comme étant d'importants acteurs dans les voies répresseurs de tumeurs, et sont donc réprimés durant la cancérogénèse.

Nos résultats concordent avec plusieurs travaux effectués auparavant, montrant les rôles que détiennent les molécules des LncRNAs dans les voies comme par exemple :

- **Transport des peptides** : Pour les protéines n'ayant pas un domaine d'interaction commun; par exemple: NEAT2 (Tripathi et *al.*, 2010, Tano and Akimitsu, 2012).
- **Apoptose** : le mécanisme de la mort programmée sert à éliminer les cellules endommagée et arrêter leur prolifération, Le LncRNA réprimé « LOC100507360 », non encore caractérisé, semble avoir un rôle lors de l'apoptose. Tenant compte de sa répression lors du processus, nous suggérons qu'il pourrait jouer un rôle activateur dans la mort cellulaire programmée. Nous pouvons trouver dans la revue « New players of apoptosis control » de Marianna Nicoletta Rossi and Fabrizio Antonangeli, les différentes molécules de LncRNAs surexprimés ou réprimés impliqués dans l'apoptose tumorale, confirmant ainsi l'implication de ces molécules dans ce processus.

La 2ème étape de l'analyse fut pour le transcriptome complet, incluant les gènes codant. L'analyse de l'expression différentielle des gènes a révélé la présence de 273 significativement différentiellement exprimés dans les leucémies chronique B comparé à leurs expressions dans les échantillons de contrôles sains.

Les gènes différentiellement exprimés sont de :

- 77 Gènes surexprimés (UPregulated)
- 196 Gènes réprimés (DOWN regulated)

La suite d'analyse fonctionnelle sur DAVID a abouti à divers résultats nous permettant de fournir des explications quant aux rôles potentiels des gènes surexprimés et réprimés.

- **Gènes surexprimés :**

Parmi les gènes surexprimés lors de la LLC-B, nous trouvons le « WNT3 » , « ABC transporters » ou bien « HSP90AB1 ».

- La famille des « WNT » sont des protéines impliquées dans la prolifération et la différenciation cellulaire dans le cas normal, mais aussi dans le cas tumorale lors de leur dérégulation. L'activation aberrante de ces dernières contribue à la tumorigénèse (Barker and Clevers, 2006).

Les travaux d'Ali Memarian et al. sur l'expression des « wnt » dans la leucémie lymphoïde chronique ont démontré une surexpression significative des WNT3 lors des LLC comparé au cas normal, ce qui confirme notre résultat indiquant que la surexpression des « wnt3 » dans la voie de signalisation WNT pourrait avoir un rôle dans la cancérisation des cellules B.

« **HSP90AB1** » gène codant la protéine de la famille heatshockprotein 90 « *Heatshockprotein HSP 90-beta* » l'une des 5 isoformes de la HSP90, impliquée dans différents processus cellulaires tels que la prolifération, la différenciation et l'apoptose (Garrido et al., 2006). La surexpression de la HSP90-beta est également présente dans d'autres types de cancers tels que le cancer du sein (Yano et al., 1996), leucémie myéloïde (Chant et al., 1995). La surexpression de l'HSP90 semble être un événement commun lors des cancers (Ciocca and Calderwood, 2005).

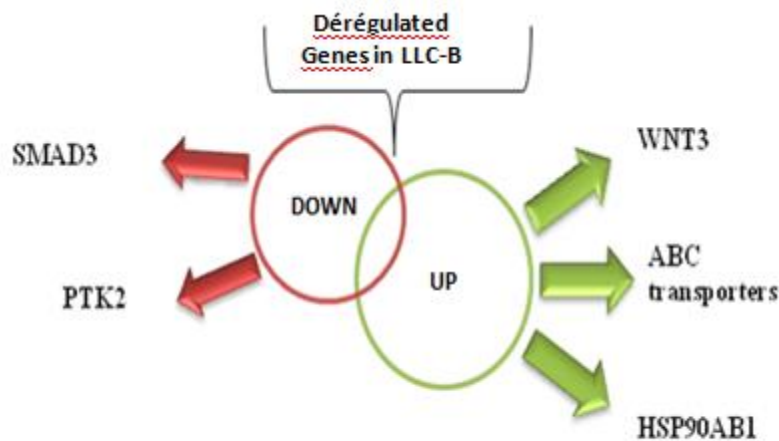
D'après notre résultat, nous suggérons que la surexpression protéine HSP90AB1 lors des LLC-B contribue à l'évasion des cellules tumorales de l'apoptose, ce qui concorde avec d'autres travaux faits auparavant ayant prouvé la corrélation entre la surexpression de la HSP90-beta avec la tumorigénèse comme ce fut prouvé dans le travail de Biaoxue et al. (Biaoxue et al., 2012).

- **Gènes réprimés :**

Impliqués dans des voies telles que : signalisation des chemokines ou voies liées aux cancers, ils semblent être des acteurs de l'apoptose, composants de la membrane plasmique, acteurs de la réponse immunitaire.

**SMAD3** ou bien **PTK2** sont tous deux des gènes réprimés lors de la leucémie, et sont deux acteurs dans les voies liées aux cancers, induisant dans le cas normal l'apoptose et la réponse aux signaux d'anti prolifération. Dans le cas cancéreux, ces gènes sont réprimés et la cellule ne répond plus aux signaux d'arrêt de division et se divise d'une manière incontrôlée.

Nous observons comme dans la plupart des cancers la surexpression dans la leucémie des gènes codants des protéines promoteurs de tumeurs comme « wnt , hsp90-beta » et la répression des gènes anti tumoraux comme « SMAD3 , PTK2 » , ce qui confère à la cellule B le pouvoir de la division illimitée et l'évasion de l'apoptose.



**Figure 52 : Les gènes dérégulés lors de la LLC-B résultant de l'analyse DAVID.**

La dernière étape de notre travail se présente en la recherche et l'identification dans nos LncRNAs d'intérêt des potentiels « eRNA ». Après l'analyse sur GALAXY, nous en sommes sortis avec 9 LncRANs associés aux superenhancers.

Les superenhancers sont des régions contenant un très grand complexe de transcription afin de réguler la transcription d'une façon fine et très contrôlée (Hnisz et al., 2013). Ces régions régulatrices jouent un rôle très important dans la détermination de l'identité cellulaire ainsi que l'expression génique (Bulger and Groudine, 2011).

Nos résultats concordent avec les travaux de Lai et al en 2013 et Orom et al. en 2012 (Orom et al., 2010), qui ont identifié les LncRNAs ayant une activité et propriétés propres à celles d'un enhancer « enhancer-like ».

Ces 9 LncRNAs différenciellement exprimés sont des candidats très importants dans le développement de cette maladie, car ce sont des super-enhancers et peuvent créer une dérégulation assez importante pour induire la leucémie.

---

---

## **CONCLUSION**

---

---

La leucémie lymphoïde chronique B constitue l'une des formes de leucémies les plus fréquentes chez l'adulte. Elle se caractérise par la prolifération illimitée des cellules B endommagées (Goldman and Masson, 2013)

Durant les dix dernières années, la découverte de la classe des LncRNAs ; l'une des types ARNs non codants, a révolutionné le monde des ARNs, même si la compréhension biologique de ces molécules avance lentement. Nous sommes à présent sûrs de leur importance dans de nombreux processus cellulaires, parmi eux le processus tumoral (Rinn and Chang, 2012).

En effet, de nombreuses études ont prouvé la corrélation entre l'expression des LncRNAs et les cancers. Dans notre travail, nous avons pu identifier des LncRNAs qui sont exprimés différemment et sont associés aux superenhancers lors des leucémies chroniques B comparé aux cas normaux, surexprimés pour certains et réprimés pour d'autres, impliqués dans les voies promotrices ou bien suppresseurs de tumeurs. Ces molécules LncRNAs restent encore non caractérisées jusqu'à ce jour.

L'importance de l'identification et l'analyse fonctionnelle de ces LncRNAs est de contribuer à une meilleure définition de leurs rôles durant le développement tumorale des cellules B. De plus, ces molécules pourraient avoir des applications ; comme être utilisé comme biomarqueurs pour le diagnostic et/ou le pronostic des cancers avec le changement de leurs expressions associées aux différents types de leucémies, ainsi que le développement de nouvelles stratégies thérapeutiques.



---

## REFERENCES BIBLIOGRAPHIQUES

---

Änkö M-L, Neugebauer KM (2010) Long Noncoding RNAs Add Another Layer to Pre-mRNA Splicing Regulation. *Molecular Cell* 39:833-834.

Barker N, Clevers H (2006) Mining the Wnt pathway for cancer therapeutics. *Nature reviews Drug discovery* 5:997-1014.

Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, Bonilla F, de Herreros AG (2008) A natural antisense transcript regulates *Zeb2/Sip1* gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes & development* 22:756-769.

Biaoxue R, Xiling J, Shuanying Y, Wei Z, Xiguang C, Jinsui W, Min Z (2012) Upregulation of Hsp90-beta and annexin A1 correlates with poor survival and lymphatic metastasis in lung cancer patients. *Journal of Experimental & Clinical Cancer Research* : CR 31:70.

Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Jacquillat C, Boivin P, Lesty C, Duault MT, Monconduit M, Belabbes S, Gremy F (1981) A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* 48:198-206.

Boggs DR, Chen SC, Zhang ZN, Zhang A (1987) Chronic lymphocytic leukemia in China. *American journal of hematology* 25:349-354.

Brannan CI, Dees EC, Ingram RS, Tilghman SM (1990) The product of the H19 gene may function as an RNA. *Molecular and Cellular Biology* 10:28-36.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992) The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71:515-526.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527-542.

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144:327-339.

Bullrich F, Fujii H, Calin G, Mabuchi H, Negrini M, Pekarsky Y, Rassenti L, Alder H, Reed JC, Keating MJ, Kipps TJ, Croce CM (2001) Characterization of the 13q14 tumor suppressor locus in CLL: identification of ALT1, an alternative splice variant of the *LEU2* gene. *Cancer research* 61:6640-6648.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS genetics* 6:e1001233.

Capalbo S, Trerotoli P, Ciancio A, Battista C, Serio G, Liso V (2000) Increased risk of lymphoproliferative disorders in relatives of patients with B-cell chronic lymphocytic leukemia: relevance of the degree of familial linkage. *European journal of haematology* 65:114-117.

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-Bezy M, Lawrence JB, O'Neill LA, Moore MJ, Caffrey DR, Fitzgerald KA (2013) A long noncoding RNA mediates both activation and repression of immune response genes. *Science (New York, NY)* 341:789-792.

Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloreaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A, Thierry-Mieg N, Cusick ME, Vidal M (2012) Proto-genes and de novo gene birth. *Nature* 487:370-374.

Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P (2005) Replication and transcription: Shaping the landscape of the genome. *Nat Rev Genet* 6:669-677.

Chant ID, Rose PE, Morris AG (1995) Analysis of heat-shock protein expression in myeloid leukaemia cells by flow cytometry. *British journal of haematology* 90:163-168.

Chiorazzi N, Rai KR, Ferrarini M (2005) Chronic lymphocytic leukemia. *The New England journal of medicine* 352:804-815.

Ciocca DR, Calderwood SK (2005) Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress & Chaperones* 10:86-103.

Costa FF (2007) Non-coding RNAs: Lost in translation? *Gene* 386:1-10.

Dalla-Favera R, Gaidano G. (2001). *Cancer* DeVita VT, Hellman S and Rosenberg SA. (eds). Lippincott Williams and Wilkins: Philadelphia, USA, pp. 2215-2235.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS biology* 8:e1000384.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R (2012) The GENCODE v7 catalog of human

long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research* 22:1775-1789.

Dighiero G, Hamblin TJ (2008) Chronic lymphocytic leukaemia. *Lancet* (London, England) 371:1017-1029.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR (2012) Landscape of transcription in human cells. *Nature* 489:101-108.

Dong, R., Du, J., Wang, L., Wang, J., Ding, G., Wang, S., et al. (2014) Comparison of Long Noncoding RNA and mRNA Expression Profiles in Mesenchymal Stem Cells Derived from Human Periodontal Ligament and Bone Marrow. *BioMed Research International*, 2014, 12.

Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, Dohner K, Bentz M, Lichter P (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *The New England journal of medicine* 343:1910-1916.

Du Z, Fei T, Verhaak RGW, Su Z, Zhang Y, Brown M, Chen Y, Liu XS (2013) Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* 20:908-913.

Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, Foissac S, Willingham AT, Duttagupta R, Dumais E, Gingeras TR (2009) Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs: Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project\*. *Nature* 457:1028-1032.

Feng Y, Fan Y, Huiqing C, Zicai L, Quan D (2014) [The emerging landscape of long non-coding RNAs]. *Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji* 36:456-468.

Fitzgerald KA, Caffrey DR (2014) Long noncoding RNAs in innate and adaptive immunity. *Current opinion in immunology* 26:140-146.

Fjose A, Ellingsen S, Wargelius A, Seo HC (2001) RNA interference: mechanisms and applications. *Biotechnology annual review* 7:31-57.

Gargaud M, Despois D (2003) *Les traces du vivant*: Presses universitaires de Bordeaux.

Garrido C, Brunet M, Didelot C, Zermati Y, Schmitt E, Kroemer G (2006) Heat shock proteins 27 and 70: anti-apoptotic proteins with tumorigenic properties. *Cell cycle (Georgetown, Tex)* 5:2592-2601.

Geisler S, Collier J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews Molecular cell biology* 14:699-712.

Gérard R, 2012 : *Application de techniques de fouille de données en Bio-informatique*, thèse de doctorat, Université de Nantes, France, 135P.

Gibb EA, Brown CJ, Lam WL (2011) The functional role of long non-coding RNA in human carcinomas. *Molecular cancer* 10:38.

Goldman L, Masson PL (2013) *Goldman's Cecil Medicine Cancérologie*: Elsevier Health Sciences France.

Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, Chang HY, Brahic M, Kirkegaard K (2013) The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* 152:743-754.

Gould D, Brooker C (2000) *Applied Microbiology for Nurses*: Macmillan.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071-1076.

Gutschner T, Diederichs S (2012) The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biology* 9:703-719.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223-227.

Guttman M, Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. *Nature* 482:339-346.

Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154:240-251.

Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA, Bloomfield CD (1999) World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 17:3835-3849.

Hnisz D, Abraham Brian J, Lee Tong I, Lau A, Saint-André V, Sigova Alla A, Hoke Heather A, Young Richard A (2013) Super-Enhancers in the Control of Cell Identity and Disease. *Cell* 155:934-947.

Horwitz M, Goode EL, Jarvik GP (1996) Anticipation in familial leukemia. *American Journal of Human Genetics* 59:990-998.

Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, Zhu J, Zhao K (2013) Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 14:1190-1198.

Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* 4:44-57.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nature methods* 12:115-121.

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerod A, Borresen-Dale AL, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature genetics* 43:621-629.

Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147:789-802.

Justine K. P, Peter J V. (2005) Growing Applications and Advancements in Microarray Technology and Analysis Tools, department de Bioinformatique, Rotterdam, Cell Biochemistry and Biophysics , 43:149–166.

Kaikkonen MU, Lam MT, Glass CK (2011) Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research* 90:430-440.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science (New York, NY)* 296:916-919.

Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, NY)* 316:1484-1488.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465:182-187.

Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, Ambesi-Impiombato A, Califano A, Migliazza A, Bhagat G, Dalla-Favera R (2010) The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer cell* 17:28-40.

Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, Miyano S, Mori M (2011) Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer research* 71:6320-6326.

Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y (2010) Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science (New York, NY)* 329:336-339.

Kumaravel., S, Sathich., K, Narundra., W. 2002: Dna microarray technology and its applications in cancer biology, Bangalore (India), 4 : 1-9.

Kung JTY, Colognori D, Lee JT (2013) Long Noncoding RNAs: Past, Present, and Future. *Genetics* 193:651-669.

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494:497-501.

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97:17-27.

Larsen C, Seité P, Berger R (1994) Données récentes sur les mécanismes moléculaires de la translocation chromosomique t (14; 18)(q21; q32) activant le gène BCL2.

Li W, Lam MTY, Notani D (2014) Enhancer RNAs. *Cell cycle (Georgetown, Tex)* 13:3151-3152.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498:516-520.

Linnet MS, Van Natta ML, Brookmeyer R, Khoury MJ, McCaffrey LD, Humphrey RL, Szklo M (1989) Familial cancer history and chronic lymphocytic leukemia. A case-control study. *American journal of epidemiology* 130:655-664.

Lodish H, Berk A, Zipursky SL, et al. (2000) *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman.

Louro R, Smirnova AS, Verjovski-Almeida S (2009) Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93:291-298.

Lu X (2005) p53: a heavily dictated dictator of life and death. *Current opinion in genetics & development* 15:27-33.

Marieb E, Hoehn K, Moussakova L, Lachaine R (2014) *Anatomie et physiologie humaines: Livre + eText + plateforme numérique MonLab - Licence étudiant 60 mois: ERPI*.

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 29:499-509.

Meyers JN (2007) *Trends in Signal Transduction Research*: Nova Science Publishers.

Michaux JL (2003) *La solitude Bartok: une leucémie cachée: Age d'homme*.

Nagano T, Fraser P (2011) No-Nonsense Functions for Long Noncoding RNAs. *Cell* 145:178-181.

Neuland CY, Blattner WA, Mann DL, Fraser MC, Tsai S, Strong DM (1983) Familial chronic lymphocytic leukemia. *Journal of the National Cancer Institute* 71:1143-1150.



Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143:46-58.

P. Triadou. Quelques éléments d'histoire de l'hématologie biologique. *Annales de Biologie Clinique*. 2000;58(1):19-28.

Peacock J, Asselin B (1999) *Leukemia: LifeMatters*.

Plass C, Byrd JC, Raval A, Tanner SM, de la Chapelle A (2007) Molecular profiling of chronic lymphocytic leukaemia: genetics meets epigenetics to identify predisposing genes. *British journal of haematology* 139:744-752.

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629-641.

Prensner JR, Chinnaiyan AM (2011) The emergence of lncRNAs in cancer biology. *Cancer discovery* 1:391-407.

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS (1975) Clinical staging of chronic lymphocytic leukemia. *Blood* 46:219-234.

Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY (2013) A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* 2.

Raven PH, Johnson GB, Mason KA, Losos JB, Bouharmont J (2011) *Biologie: De Boeck Supérieur*.

Read A, Donnai D, Sznajer Y, Verloes A (2008) *Génétique médicale: De la biologie à la pratique clinique: De Boeck Supérieur*.

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annual review of biochemistry* 81:145-166.

Roger Lacave, Larsen CJ, Robert J (2005) *Cancérologie fondamentale*.

Rotival., M. 2011 : *Approches intégrées du génome et du transcriptome dans les maladies complexes humaines*, thèse de doctorat, Université Paris 1, 256P.

Roulland S, Navarro JM, Grenot P, Milili M, Agopian J, Montpellier B, Gauduchon P, Lebailly P, Schiff C, Nadel B (2006) Follicular lymphoma-like B cells in healthy individuals: a novel intermediate step in early lymphomagenesis. *The Journal of experimental medicine* 203:2425-2431.

Rozman C, Montserrat E (1995) Chronic lymphocytic leukemia. *The New England journal of medicine* 333:1052-1057.

Sen F, Lai R, Albitar M (2002) Chronic lymphocytic leukemia with t(14;18) and trisomy 12. *Archives of pathology & laboratory medicine* 126:1543-1546.

Shadfan M, Lopez-Pajares V, Yuan Z-M (2012) MDM2 and MDMX: alone and together in regulation of p53. *Translational Cancer Research* 1:88-99.

Sultmann H, Diederichs S (2014) Long noncoding RNA: "LNCs" to cancer. *European urology* 65:1152-1153.

Skalsky RL, Cullen BR (2010) Viruses, microRNAs, and host interactions. *Annual review of microbiology* 64:123-141.

Tano K, Akimitsu N (2012) Long non-coding RNAs in cancer progression. *Frontiers in Genetics* 3.

Thibault, C. 2007 : Impression de Biomolécules par Lithographie douce, applications pour les Biopuces, de l'échelle Micrométrique à Nanométrique, thèse de doctorat, université de toulouse, France, 205P.

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol Cell* 39:925-938.

Tsai M-C, Spitale RC, Chang HY (2011) Long Intergenic Noncoding RNAs: New Links in Cancer Progression. *Cancer research* 71:3-7.

Ulitsky I, Bartel DP (2013) lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* 154:26-46.

Véronique L (2009) La Leucémie lymphoïde chronique et la maladie de Waldenström: John Libbey Eurotext.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY (2011a) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472:120-124.

Wang X, Song X, Glass CK, Rosenfeld MG (2011b) The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. *Cold Spring Harbor perspectives in biology* 3:a003756.

Wapinski O, Chang HY (2011) Long noncoding RNAs and human disease. *Trends in cell biology* 21:354-361.

Weiss NS (1979) Geographical variation in the incidence of the leukemias and lymphomas. *National Cancer Institute monograph* 139-142.

Wiernik PH, Ashwin M, Hu XP, Paietta E, Brown K (2001) Anticipation in familial chronic lymphocytic leukaemia. *Br J Haematol* 113:407-414.

Wright PW, Huehn A, Cichocki F, Li H, Sharma N, Dang H, Lenvik TR, Woll P, Kaufman D, Miller JS, Anderson SK (2013) Identification of a KIR antisense lncRNA expressed by progenitor cells. *Genes Immun* 14:427-433.

Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F, Zheng SS (2011) Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Annals of surgical oncology* 18:1243-1250.

Yano M, Naito Z, Tanaka S, Asano G (1996) Expression and Roles of Heat Shock Proteins in Human Breast Cancer. *Japanese Journal of Cancer Research* 87:908-915.

Yuille MR, Matutes E, Marossy A, Hilditch B, Catovsky D, Houlston RS (2000) Familial chronic lymphocytic leukaemia: a survey and review of published studies. *Br J Haematol* 109:794-799.

Zenz T, Mertens D, Kuppers R, Dohner H, Stilgenbauer S (2010) From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature reviews Cancer* 10:37-50.

Zhang, A., Xu, M., & Mo, Y.-Y. (2014). Role of the lncRNA-p53 regulatory network in cancer. *Journal of Molecular Cell Biology*, 6(3), 181-191. <http://doi.org/10.1093/jmcb/mju013>.

Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE (2009) A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113:2526-2534.

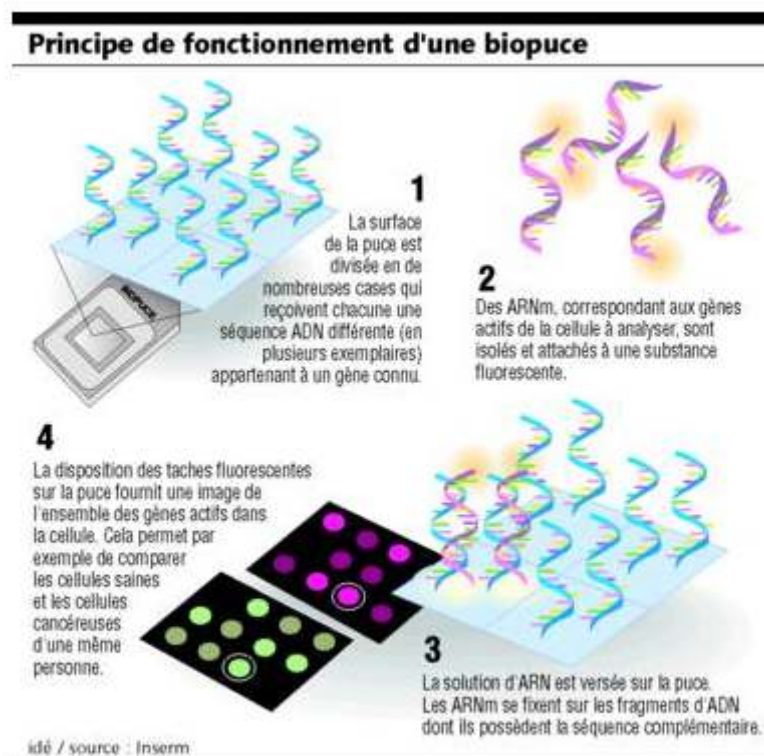
Wikigalaxy : <https://wiki.galaxyproject.org/>.

# ANNEXE I

## tDNA Microarray

Cette technologie des biopuces a été connue sous le nom 'macroarrays' elle est constituée de fragments d'ADN immobilisés sur un support solide selon une disposition ordonnée (Thibault, 2007).

Des expériences effectuées sur une grande variété de membranes recouvertes avec l'ADNc pour l'hybridation comparative des espèces d'ARN ont permis l'étude d'expression du génome en entier ayant besoin d'une seule expérience, avec un signal quantifiable étant généré lors de l'hybridation, ce dernier est directement proportionnel au niveau d'expression génique dans les cellules ou les tissus d'intérêt. (Justine et Peter, 2005).



Principe d'une biopuce.

## **Types de puces a ADN :**

Une puce comporte quelques centaines à plusieurs dizaines de milliers d'unités d'hybridation appelées « spots », chacune contenant un dépôt de fragments d'ADN ou d'oligonucléotides correspondant à des sondes de séquences données.

L'hybridation de la puce avec un échantillon biologique, marqué par un radioélément ou par une molécule fluorescente est conçue sur des membranes poreuses de nylon (appelées parfois « macroarrays » par opposition aux « microarrays »), les puces à ADN ont été progressivement mises au point sur lames de verre à la fin des années 90. La miniaturisation, rendue possible par l'utilisation d'un support solide, de marqueurs fluorescents et par les progrès de la robotique, permet aujourd'hui de fabriquer des puces comportant une très haute densité de spots, susceptibles de recouvrir l'intégralité du génome d'un organisme sur une simple lame de microscope.

Nous observons différents types de puces suivant des paramètres comme : la densité des spots, le mode de fabrication, nature des fragments fixés à la surface ou encore les méthodes d'hybridation utilisées (Rotival, 2011).

## **Applications**

La technique des puces à ADN est devenue l'une des techniques les plus répandues ces dernières années ciblant différents domaines d'études, ciblant plusieurs buts de recherche ou diagnostic comme par exemple :

- L'étude des pathologies humaines.
- Cancérologie : comme les lymphomes, les leucémies, le cancer du sein, de la prostate et le mélanome (Gérard, 2012).
- L'identification des bio-marqueurs tumoraux, les gènes responsables de la chimiorésistance aux drogues et la classification de tumeurs ainsi que la détection des mutations (Kumaravel et al. 2002). Elle a aussi affiné la définition de celles-ci, en établissant des sous-types tumoraux

qui étaient difficilement analysables par les voies traditionnelles utilisées jusqu'alors par les cliniciens (Gérard, 2012).

- L'étude de profil de l'expression en ciblant un large spectre de gènes.