

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



FACULTE DES SCIENCES DE LA NATURE ET DE LA VIE

Département de Biologie et Physiologie Cellulaire

Mémoire de fin d'études

En vue de l'obtention du Diplôme de

MASTER EN SNV

Filière Biologie

OPTION : Bio- informatique

Thème

Analyse bioinformatique du rôle des LncRNAs (long non codingRNAs) et leur contribution dans le diabète type 1

Présenté par :

M^{elle} DIB Asmaa

M^{elle} NEHAR Hadjer

Devant le jury composé de :

Mme GUESSAIBYA N	M.C.B	UB1	Président
Mr MOHAMED SAID R	M.C.B	UB1	Examineur
Mme EDDAIKRA A	M.A.A	UB1	Promotrice

Promotion 2016 -2017

Remerciement

On remercie Dieu, le tout puissant de nous avoir donné la santé, la volonté, le courage d'achever ce mémoire de fin d'étude.

Ces quelques lignes ne suffiront pas à exprimer toute la gratitude et tous les moments que nous avons partagés ensemble, mais parfois un simple merci vaut bien plus que tous les discours du monde ; Ce mémoire nous a permis de rencontrer également des personnes merveilleuses et d'une rare générosité

Nous exprimons notre profond remerciement et gratitude à notre promotrice Mme Eddakira A ; Maître de conférences au département de $\mathcal{B}\mathcal{R}\mathcal{G}$ de l'université de Blida 1, pour bien vouloir de diriger ce travail, et pour ses conseils et ses remarques constructive. nous la remercions particulièrement pour sa disponibilité, ses encouragements. Que dieu la protège.

Nos remerciements les plus sincères s'adressent à Mme Guaisaibya N, d'avoir accepté de présider notre jury de soutenance.

Nous tenons à remercier Mr. Mohamed said R pour l'honneur qu'elle nous a fait en acceptant d'examiner notre travail. Nous tenons à le remercier d'avoir été à notre écoute durant toute notre formation en qualité de chef d'option de la licence génétique puis du Master de Bioinformatique.

Nous exprimons notre profonde gratitude à Mr. Beninet M en qualité de co-promoteur et qui nous a accompagnés dans le départ de ce travail.

Nous exprimons notre profonde gratitude à Mme le chef de département d'avoir été à notre écoute pendant toute la durée de notre formation.

Je tiens à remercier tous mes enseignants actuels et passés du département des sciences biologique de l'université Saad Dahlab de Blida.

Dédicace



Je dédie ce modeste travail

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude ; je t'aime maman.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Quisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Une spéciale dédicace pour mes chères sœurs:

Karima et son mari Ibrahim

Fatima, Djamilia et Malika.

A mes chers frères : Noureddine, Kamal, Khalad, Hakim et Hamza.

Ames chères nièces et neveux : Fassine, Maroua, Ikram, Houssam, Rahil, Salah et Amir. Que dieu les protège.

Mon adorable amie, ma sœur Amina et sa famille, je vous souhaite tout le bien, le bonheur, et beaucoup de succès dans ta vie personnelle et professionnelle.

Mes sincères copines : Selma, Abir, Boutaina et Phaima.

Et à toute ma famille et à tous ceux que j'aime.

Hadjer Nihar



Dédicaces

Je m'incline devant Dieu tout puissant qui m'a ouvert la porte du savoir et m'a aidé à la franchir.

Je dédie ce modeste travail :

A mon père source de respect, en témoignage de ma profonde reconnaissance pour tout l'effort et le soutien incessant qui m'a toujours apporté.

A ma chère et tendre mère, source d'affection de courage et d'inspiration qui a autant sacrifié pour me voir atteindre ce jour.

A mes beauxsœurs : Hadjer et Sarah .

A mon frère : Oussama.

A mes chères amies : Romaiassa; Hadjer ; Sarah et Salima

Aussi, je dédie ce travail à toute ma famille DIB.

A tous les membres de ma promotion.

A tous mes enseignants depuis mes premières années d'études.

A tous ceux qui m'ont aidé et soutenus pendant tout mon cursus universitaire. Que dieu les protège et leur réalise leurs rêves et souhaits.

Asma

ملخص

أظهرت التحليلات الجينية على نطاق واسع وجود عدد كبير من ARN طويل غير دال (Inc RNA) على الرغم من أن وظيفة غالبية هذه Inc RNA لا تزال غير مستكشفة في الوقت الحاضر، مؤخرا تحديد هوية ARN طويل غير الدال وبداية وصف وظيفتها في مجموعة متنوعة من الأنسجة تشكل تقدما علميا كبيرا يبدو واضحا أن العديد من هذه القطع المستنسخة تلعب دورا هاما في تنظيم التعبير الجيني وتشارك في مختلف الأمراض، بما في ذلك مرض السكري.

ويمثل مرض السكري مجموعة من الأمراض الأيضية المعقدة التي تؤدي إلى تدمير انتقائي ومتقدم لخلايا البنكرياس β المنتجة للأنسولين

على الرغم من الاهتمام القوي في وصف الترانسكريبتوم لمختلف أنواع الخلايا البشرية لفهم الأساس الجزيئي لمرض السكري،

قد تم إيلاء القليل جدا من الاهتمام لدور Inc RNA والمساهمة في مرض السكري من النوع 1

في هذه الدراسة، تم إجراء تحليلات المعلوماتية الحيوية التي تسمح بالتنبؤ بالأدلة المتزايدة للدور المحتمل لهذه Inc RNA في وظيفة الخلايا β والتفاوت في داء السكري من النوع 1، مع التركيز على الجينومية.

تتوفر العديد من الأدوات المعلوماتية الحيوية لنا لهذا العمل مثل: Galaxy, David, GeneMANIA, String, Blast, Uniprot

الكلمات المفتاحية: Inc RNA ، السكري من النوع 1 ، الخلايا β ، الأدوات المعلوماتية الحيوية

Liste des outils

- NCBI : <https://www.ncbi.nlm.nih.gov>
- BLAST : <https://www.ncbi.nlm.nih.gov/BLAST/>
- *UNIPROT* : www.uniprot.org/
- Galaxy : <https://usegalaxy.org/>
- GeneMANIA : <https://genemania.org/>
- String : (<http://string-db.org>)
- Phyre2 : www.sbg.bio.ic.ac.uk/~phyre2/

LISTE DES ABREVIATIONS

ADN : Acide **D**ésoxyribo**N**ucléique

ARN : Acide **R**ibo**N**ucléique

ARNm : Acide **R**ibo**N**ucléique **m**essenger

ARNnc : **ARN** non codant

ARN poly : **ARN** **P**olymérase

ARNsn : Small **n**uclear **R**NA [petits **ARNs** Nucléaires]

ARNt : **ARN** de transfert

ATP : Adénosine **T**ri**P**hosphate

BLAST : **B**asic **L**ocal **A**lignement **S**earch **T**ool

CD : Cluster de **D**ifférenciation

CTD : Carboxy **T**erminal **D**omain

DID : Diabète **I**nsulino **D**épendant

DT1 : Diabète **T**ype 1

GEWAS : **G**Enome-**W**ide **A**ssociation **S**tudies

H3K4me3 : trimethylated histone **H3** lysine **4**

H3K36me3 : trionomethylated histone **H3** lysine **36**

HLA : **H**uman **L**eukocyte **A**ntigene

INS : **I**nsuline

LinRNA : Long **i**ntergenic **R**NA [Long **ARN** ntergénique]

LINE : Long **I**nterspersed **N**uclear **E**lements

LncRNA : Long **ARN** non codant

mi ARN : **micro-ARN**

NCBI : National Center for **Biotechnology Information**

PDX1 : **Pancreas/Duodenum homeobox protein 1**

Pi ARN : **Piwi- interacting RNA**

Phyre 2 : **Protein homology/ analogy recognition engine V2.0**

RNA-seq : **RNA Sequencing**

sRNA : **Small non coding RNA**

SINE : **Short Interspersed Nuclear Elements**

Uniprot : **Universal Protein Resource**

ZnT8 : **Zinc Transporter 8 protein**

Liste des figures

Figure	Titre	Page
1	Apparition du diabète type 1- Intervention du système immunitaire	4
2	Schéma de la molécule d'insuline	5
3	le dogme central de la biologie moléculaire	6
4	Représentation schématique de l'initiation de la transcription	8
5	la phase d'élongation de la transcription	8
6	la terminaison de la transcription	9
7	Processus de maturation des transcrits primaires	10
8	La traduction chez les eucaryotes	11
9	Acétylation / Méthylation des histones	12
10	la régulation transcriptionnel chez les eucaryotes	13
11	représentation schématique des principales catégories de LncRNA	15
12	Approches expérimentales menant à la caractérisation fonctionnelle des lncRNA	16
13	Exemples de mécanismes d'action des LncRNA	17
14	Rôles potentiels des lncRNAs dans le développement d'ilots pancréatique et la fonction de cellules endocrines	20
15	L'organigramme de notre travail sur Galaxy	28
16	Les pré- résultats d'alignement multiple	32
17	les résultats d'alignement multiple	33
18	Schéma représentatif des interactions géniques	34
19	la structure 3D de ZnT8 visualisé par RasMol	36
20	le nombre de résidus de la protéine ZnT8	37
21	Résultats d'interaction protéique par String	38
22	les gènes de coexpression d'interaction	39
23	comparaison de la qualité des séquences	41
24	Comparaison de la distribution de score de qualité par séquences	42

25	Représentation graphique le contenu de séquence par base	43
26	Le nombre de GC contenu par séquence	43
27	Le pourcentage de reads.	44
28	La distribution de la taille des séquences	
29	Le degré de duplication des RNA-seq	45
30	Les séquences contaminées de reads	45
31	Le nombre des adaptateurs dans les séquences	46
32	Résultats d'alignement.	46
33	visualisation des transcrits analysés.	56

Liste des tableaux

tableau	Titre	page
I	Description des outils bioinformatiques	22
II	Description des molécules biologiques	23
III	Résultats du blast	26
IV	Les données statistiques de RNA-seq	39
V	Résultats des insertions des séquences	46

Sommaire

INTRODUCTION	01
CHAPITRE I : GENERALITES	
I. Généralités sur le diabète	02
I.1. Le diabète insulino-dépendant (DID).....	02
I.2. Hémostase de la glycémie et diabète.....	03
II. Le dogme central de la biologie moléculaire.....	04
II.1. La transcription chez les eucaryotes	05
II.1.1. L'initiation de la transcription.....	06
II.1.2. Elongation de la transcription.....	07
II.1.3. La terminaison.....	08
II.1.4. La maturation de l'ARN	08
II.2. La traduction chez les eucaryotes	09
II.3. La régulation chez les eucaryotes.....	10
II.3.1. Au Niveau chromatinien.....	10
II.3.2. Au niveau transcriptionnel.....	11
II.3.3. Au niveau post-transcriptionnel	12
II.3.4. Au niveau traductionnel.....	12
II.3.5. La régulation post-traductionnelle.....	12
III. ARNs non codants	13
IV. Une panoplie d'ARN longs non codants.....	13
V. Fonctions des lncRNA dans les processus physiologiques.....	16
VI. lncRNA - cellules β.....	17
VII. Expression des lncRNAs dans les cellules pancréatiques humaines	18
VIII. Rôles potentiels des lncRNAs dans le développement, la fonction et maintenance des cellules β pancréatiques.....	18
CHAPITRE II : Matériel et méthodes	

II.1. Matériel	21
II.1.1. Matériel non biologique	21
II.1.1.1. Outils bio-informatiques	21
II.1.2. Matériel biologique	23
II.2. Etude Moleculaire	24
II.2.1. Traitement des données transcriptomiques	24
II.2.1.1. Outil Galaxy	24
II.2.2. Etude d'interactions des données moléculaires	28
II.2.2.1. Outil Blast	28
II.2.2.2. OutilPhyre2	29
II.2.2.3. Outil UNIPROT	30
II.2.2.4. Outil GeneMANIA	30
II.2.2.5. Etude des interactions protéiques par String	30

Chapitre III : Résultats et discussion

III.1. Résultats	32
III.1. 1. Alignement multiple par la base de données UniProt	32
III.1.2. Etude d'interaction entre les gènes	34
III.1. 3. Annotation génomique par Blast	35
III.1. 4. La modélisation 3D de ZnT8	36
III.1. 5. Etude d'interaction protéique (String)	37
III.1.6. Outil Glaxy	40

III.1.2. discussion

Conclusion

Références bibliographiques

Annexe

Résumé

Des analyses génomiques à grande échelle ont mis en évidence l'existence d'une multitude de longs ARN non codants (lncRNA). Bien que la fonction de la majorité de ces lncRNA reste à l'heure actuelle inexplorée, la récente identification des longs ARN non codants et le début de leur caractérisation fonctionnelle dans une variété de tissus constituent des avancées scientifiques majeures.

Il semble clair que plusieurs de ces transcrits jouent un rôle important dans la régulation de l'expression génique et sont impliqués dans diverses pathologies, notamment le diabète.

Malgré un fort intérêt à caractériser le transcriptome des différents types de cellules humains pour comprendre la base moléculaire du diabète, on a accordé très peu d'attention au rôle des lncRNAs et leur contribution dans le diabète type 1.

Pour cela nous nous sommes intéressés dans cette étude *in silico* à travers l'utilisation des données de transcriptome d'une lignée cellulaire de cellules beta pancréatiques à comprendre les bases moléculaire du diabète et le rôle potentiel des lncRNAs dans la dysrégulation des fonctions des cellules β dans diabète type 1, en mettant l'accent sur les loci génomiques

Divers outils bioinformatique s'offrent à nous pour ce travail telques : Galaxy, GeneMANIA, String, Blast et Uniprot.

Nos résultats nous offrent la possibilité de confirmer le rôle des lnc RNA dans l'émergence du diabète type 1

Mots clés: lncRNAs – Diabète type 1 – Cellules β – outils bioinformatiques.

Summary

Large-scale genomic analyzes have demonstrated the existence of a multitude of long non-coding RNAs (lncRNAs). Although the function of the majority of these lncRNA remains at the present time unexplored, the recent identification of long non-coding RNAs and the beginning of their functional characterization in a variety of tissues constitute major scientific advances.

It seems clear that many of these transcripts play an important role in the regulation of gene expression and are involved in various pathologies, including diabetes.

Diabetes mellitus represents a group of complex metabolic diseases that result in selective and progressive destruction of insulin-producing pancreatic β cells.

In spite of a strong interest in characterizing the transcriptome of the different types of human cells to understand the molecular basis of diabetes, very little attention has been paid to the role of lncRNAs and their contribution to type 1 diabetes.

In this study have done bioinformatic analyzes to predict growing evidence of the potential role of these lncRNAs in the function of β cells and dysregulation in type 1 diabetes, with an emphasis on genomic loci.

Various computer tools are available to us for this work: Galaxy, David, GeneMANIA, String, Blast et Uniprot.

Key words: lncRNAs - Type 1 diabetes - β Cells - bioinformatics tools.

Introduction

Les progrès technologiques récents dans le domaine du séquençage du génome ont ouvert la voie à une nouvelle appréciation des ARN non codants dans la régulation des gènes. L'analyse de transcriptome à très haut débit a révélé que la grande majorité du génome est transcrite, avec les deux tiers du génome humain. Ce dernier couvre par des transcrits traités, dont une petite fraction (< 2%) est traduite en protéines (**Djebali et al., 2012**). L'identification de plusieurs caractéristiques génomiques et fonctionnelles communes de ces ARN non traduits a conduit à leur catégorisation en diverses catégories d'ARN non codants. Une telle classe qui a fait l'objet d'une vaste recherche est celle des longs ARN non codants (lncRNAs).

Les lncRNAs sont des transcrits de plus de 200 nucléotides qui ne codent pas pour des protéines (**Fatica et Bozzoni, 2014**). Comme pour les gènes codant pour les protéines, les lncRNAs sont marqués par des signatures de chromatine typiques de la transcription active dans les types de cellules où ils sont exprimés. Ils sont composés de **H3K4me3** (Lysine triméthylée 4 dans l'histone H3) au promoteur, et suivis par **H3K36me3** le long des régions transcrites (appelées "K4-K36 domaines") (**Guttman et Rinn, 2012**).

Grâce à de nombreuses études, plusieurs principes généraux de la fonction des lncRNAs ont été apparus. Les lncRNAs ont été montrés à fonctionner en cis, c'est-à-dire localement proches du site de leur production et en trans, c'est-à-dire sur des sites sur d'autres chromosomes. En raison de leurs diverses fonctions, une mauvaise régulation des lncRNAs peut entraîner une défaillance de développement normal et par conséquent à la maladie.

Ces résultats soulèvent la possibilité que les lncRNAs et autres ARN non codants puissent être des candidats moléculaires passionnants à prendre en compte le risque génétique non résolu en maladies complexes telles que le diabète (**Medici et al., 1999; Hyttinen et al., 2003**). Dans le cas du diabète de type 1 (DT1), l'altération métabolique est le résultat d'une destruction auto-immune des cellules β pancréatiques productrices de l'insuline.

Des efforts de grande envergure identifier les variantes d'ADN associées à des formes plus courantes de diabète grâce à des études d'association à l'échelle du génome (GWAS) ont principalement identifié des variantes candidates qui se situent dans des régions non codantes et avec des fonctions encore inconnues (**McCarthy, 2010**). Ainsi, pour améliorer notre compréhension actuelle de la base moléculaire du diabète sucré et pour développer un meilleur traitement, nous devons caractériser soigneusement le transcriptome de Cellules β pancréatiques en mettant l'accent sur l'élucidation des fonctions de Transcripts non codantes. Dans cette étude, nous présentons des preuves récentes du rôle des LncRNAs (long non coding RNAs) dans la régulation de la cellule β et leur contribution dans le diabète type 1.

I. Généralités sur le diabète

Le diabète est une maladie fréquente, connue depuis fort longtemps, très répandue. C'est une pathologie chronique, caractérisée par une hyperglycémie. En effet, lorsque la glycémie dans le sang, mesurée à jeun, devient supérieure à 1,26 g par litre, la personne est considérée comme diabétique. Cette maladie est incurable, mais peut néanmoins être traitée efficacement (**Buyschacrt et al., 1998 ; Raccah, 2004**). Il est aussi un désordre du métabolisme lipidique, glucidique et protéique attribué à la production diminuée de l'insuline ou à une résistance anormale à cette hormone qui entraîne une hausse du taux de glucose.

L'hyperglycémie chronique du diabète est associée à long terme à des complications, à des dysfonctions et à des insuffisances de différents organes notamment les yeux, les reins, les nerfs, le cœur et les vaisseaux sanguins.

On peut distinguer plusieurs types de diabète, à savoir :

Le diabète insulindépendant (DT1) ou diabète insulino-prive ou diabète maigre est actuellement appelé diabète type 1. C'est un diabète qui apparaît souvent de façon brutale, il touche préférentiellement les jeunes (diabète juvénile).

Le pancréas étant incapable de produire de l'insuline, la survie de ces malades dépend entièrement d'injection quotidienne de cette hormone, d'où son nom de diabète insulindépendant. Ce traitement est associé à un régime alimentaire en glucose finement contrôlé (**Grimaldi et al., 2001**).

C'est une maladie auto-immune conduisant à une destruction sélective et progressive des cellules β pancréatique, productrices de l'insuline. Le processus auto-immun des cellules β débute plusieurs années (5 à 10 ans voir plus) avant le début de diabète (**figure 1**). L'évaluation de la glycémie suppose une destruction de 80 à 90 des cellules β (**Grimaldi et al., 2001**).

Cette destruction résulte de la production d'auto-anticorps dirigés contre les antigènes des cellules β . Elle semble apparaître chez des sujets génétiquement prédisposés, C'est-à-dire possédant des gènes de susceptibilité liées au système HLA. Cependant, le processus auto-immun qui serait déclenché par un facteur environnemental encore mal connu (**Boitard et al., 2002**).

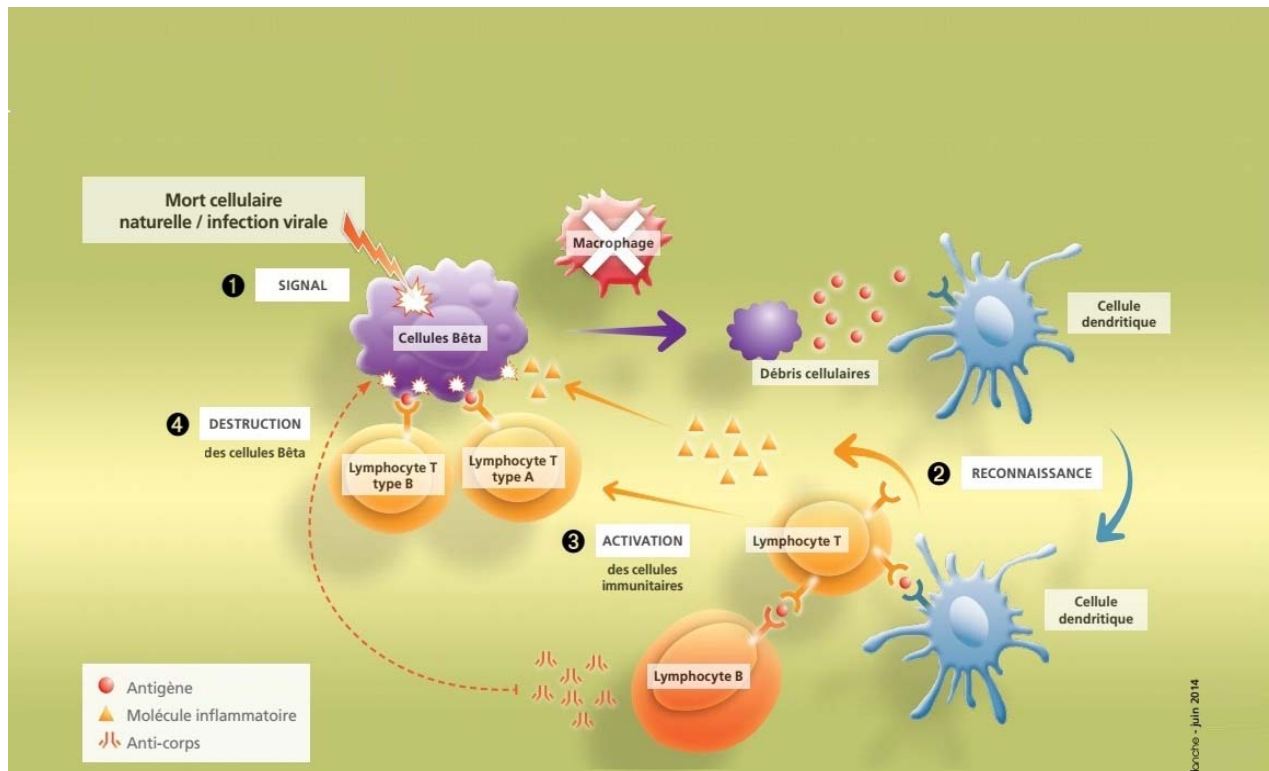


Figure 1 : Apparition du diabète type 1- Intervention du système immunitaire (Boitard et al., 2002).

I.2. Hémostase de la glycémie et diabète

Un contrôle très étroit de l'homéostasie glucidique est assuré par les hormones, au 1^{er} rang desquelles figure l'insuline. L'insuline est la seule hormone hypoglycémiante de l'organisme, elle est sécrétée par les cellules β du pancréas endocrine et participe au contrôle du métabolisme énergétique, en particulier, du métabolisme du glucose. La molécule de l'insuline est un polypeptide d'un poids moléculaire d'environ 6kDa. C'est un hétéro-dimère constitué de deux chaînes polypeptidiques, reliées entre elles par deux ponts disulfures (Figure 2). Le gène de l'insuline humaine est situé sur le bras court du chromosome 11 (Meslier et al., 2013).

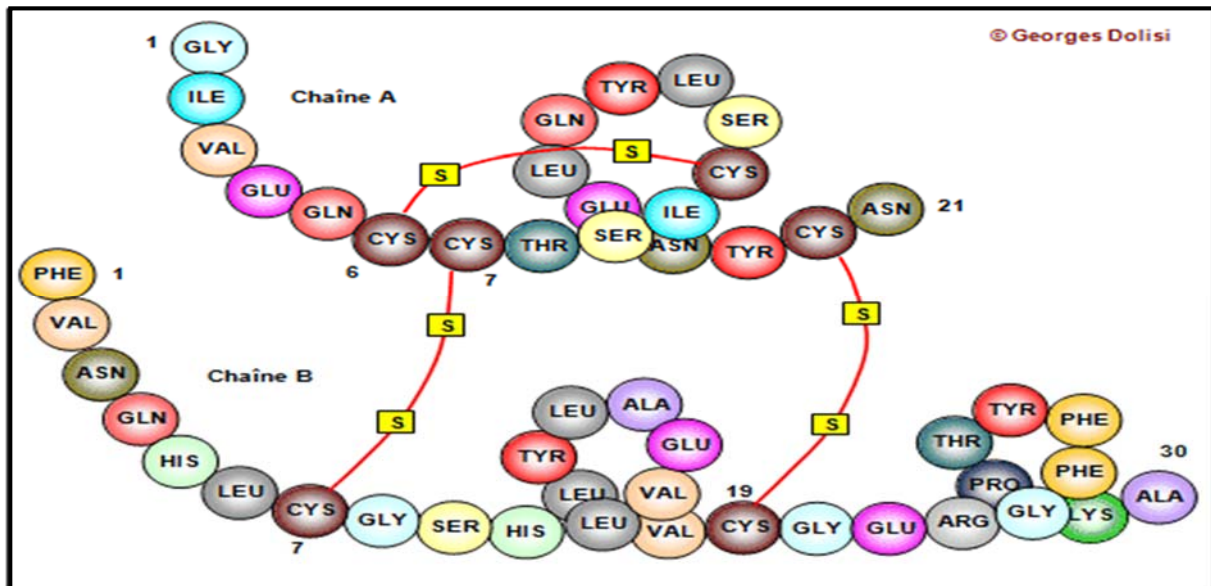


Figure2 : Schéma de la molécule d'insuline.

II. Le dogme central de la biologie moléculaire

Le dogme central de la biologie moléculaire stipule que l'information génétique codée dans le génome est transcrite sous 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)(figure 3).

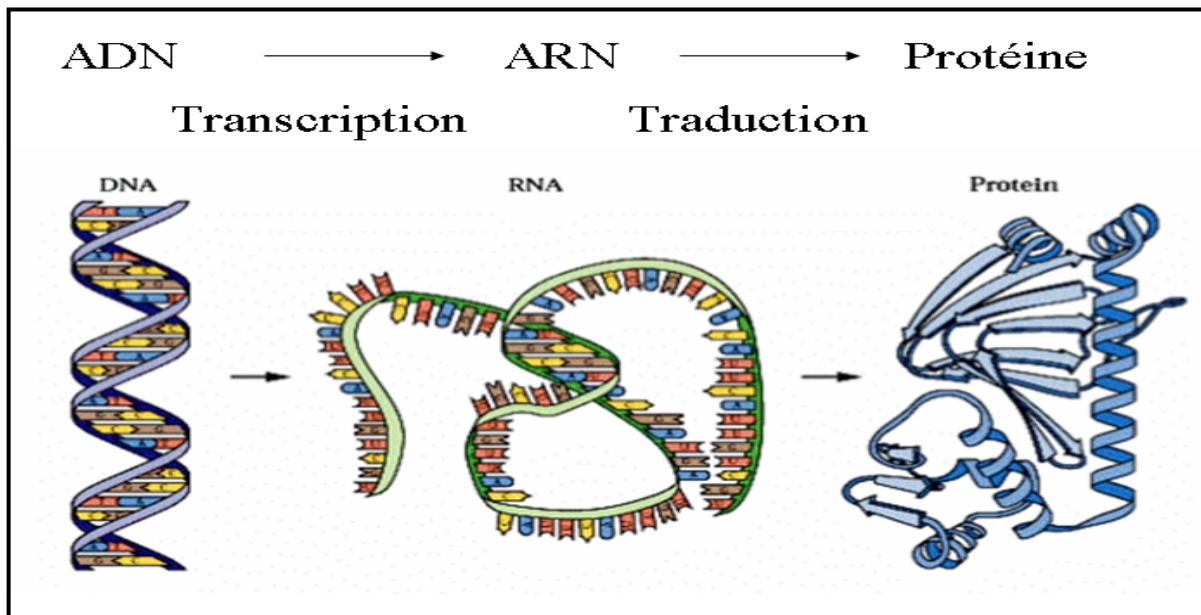


Figure3 : le dogme central de la biologie moléculaire (Raven et al., 2011).

II.1. La transcription chez les eucaryotes

L'expression du génome aboutit à la synthèse dans les cellules de macromolécules acides nucléiques et protéines, dont la structure primaire est déterminée par celle de l'ADN.

Cette expression se fait par deux mécanismes principaux :

- La structure primaire de l'ADN s'exprime d'abord par la synthèse d'acides ribonucléiques dont la structure primaire est parallèle à celle de l'ADN. C'est la transcription.
- La structure transcrite sur certains ARN, dits « messenger », s'exprime enfin par la synthèse de protéines dont la structure primaire traduit en acides aminés l'information portée par la structure primaire d'ADN. C'est la traduction.

Transcription qualifie le processus par lequel l'information contenue dans l'ADN est convertie en équivalent ARN. Une partie de ces ARN, les ARN messager, ARNm, servira à la synthèse des protéines. La régulation de la transcription est plus complexe chez les eucaryotes que chez les procaryotes, ceci est dû en particulier à la compartimentalisation qui existe chez les eucaryotes, l'élément régulateur majeur est constitué par des protéines, les facteurs de transcription lesquels sont synthétisés dans le cytoplasme, ils doivent pénétrer dans le noyau pour agir. Par ailleurs l'ADN des eucaryotes est beaucoup plus condensé en hétérochromatine que ne l'est celui des procaryotes. Cette condensation gêne l'accès des facteurs de transcription et de l'ARN polymérase et empêche la transcription (**Swynghedauw et Silvestre, 2008**).

La transcription est conduite par plusieurs enzymes :

- ARN polymérase I qui synthétise les ARN cytoplasmiques : ARN ribosomiques (18 S – 5.8 S – 28 S).
- ARN polymérase II qui synthétise les ARN messagers qui contiennent l'information destinée à la traduction et certains des ARN sn.
- ARN polymérase III qui synthétise les petits ARN (tRNA, rRNA 5 S, snRNA, 7SL-RNA) (**Housset et Raisonnier, 2009**).

II.1.1. L'initiation de la transcription

Tout l'ADN n'est pas transcrit, seules les régions correspondant à des gènes le sont. Dès lors, un acteur doit intervenir pour déterminer à quel endroit une région d'ADN doit être transcrite : c'est le rôle du promoteur.

❖ Promoteur

Le promoteur, est situé en amont de la séquence codante du brin matrice . Le point de départ de la transcription est le nucléotide à partir duquel l'ARN polymérase commencera à transcrire le brin matrice. L'initiation s'effectue donc au niveau du promoteur, auquel l'ARN polymérase se fixe.

❖ Le complexe d'initiation

L'ARN polymérase II des eucaryotes ne reconnaît pas seul le promoteur proximal. Elle effectue ce travail en compagnie de nombreux cofacteurs protéiques qui se recrutent les uns les autres et qui forment avec elle un complexe d'initiation (**Hochheimer, Tjan, 2003**).

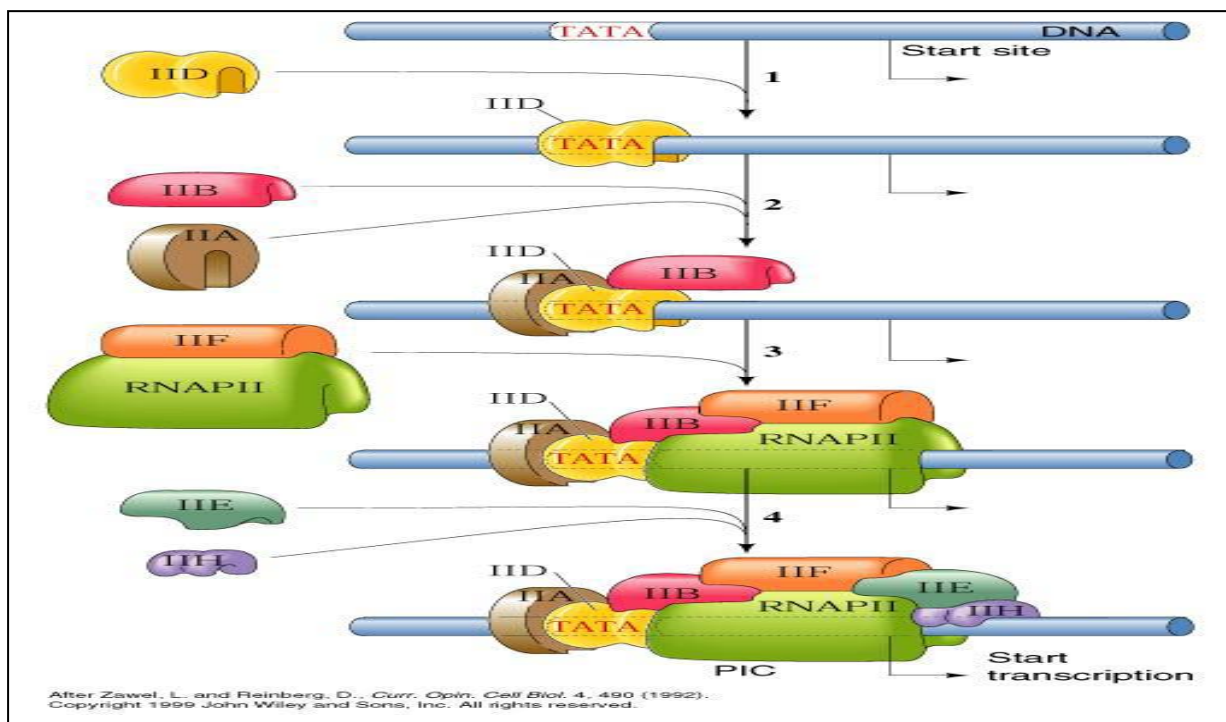


Figure 4 : Représentation schématique de l'initiation de la transcription.

II.1.2. Elongation de la transcription

L'élément central de l'élongation est la phosphorylation du domaine CTD, domaine spécifique l'ARN polymérase II. La phosphorylation par TF_{IIH} (du domaine CTD) qui est une protéine Kinase en présence d'ATP va déplacer l'ARN polymérase jusqu'au lieu d'origine de la transcription (**Saunders et al., 2006**)(figure 5).

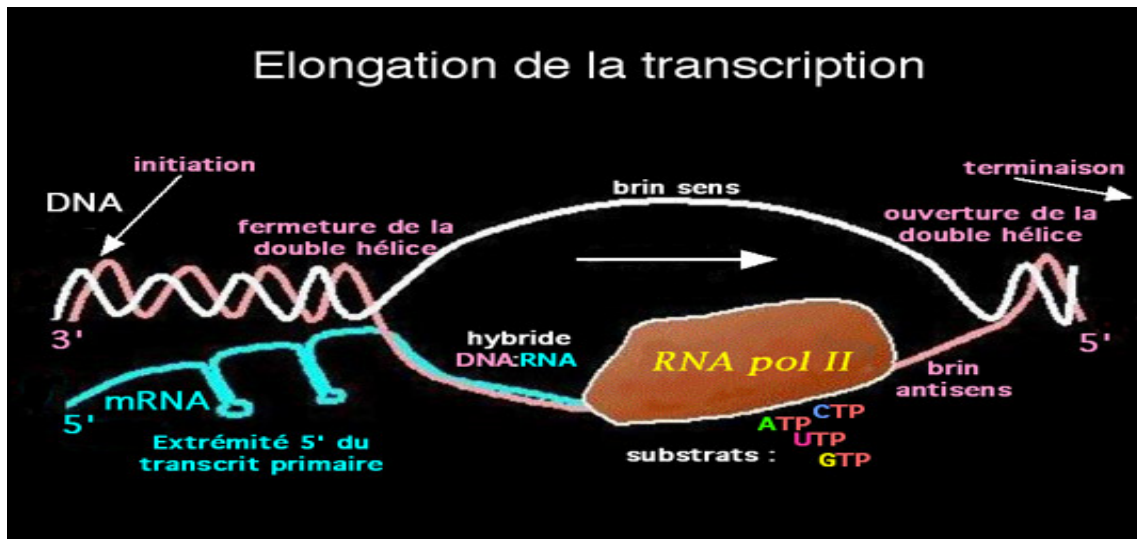


Figure 5 : la phase d'élongation de la transcription (Saunders et al., 2006).

II.1.3. La terminaison

L'ARN polymérase II reconnaît en aval des signaux pour arrêter la transcription. Ex : TTATTT, parfois aussi plus en aval ATACAA la déphosphorylation de l'ARN polymérase II entraîne la libération de l'ARN pré messagé et l'ADN (figure 6). La transcription proprement dite est terminée mais l'ARN obtenu n'est pas fonctionnel pour autant et doit subir 3 étapes de maturation.

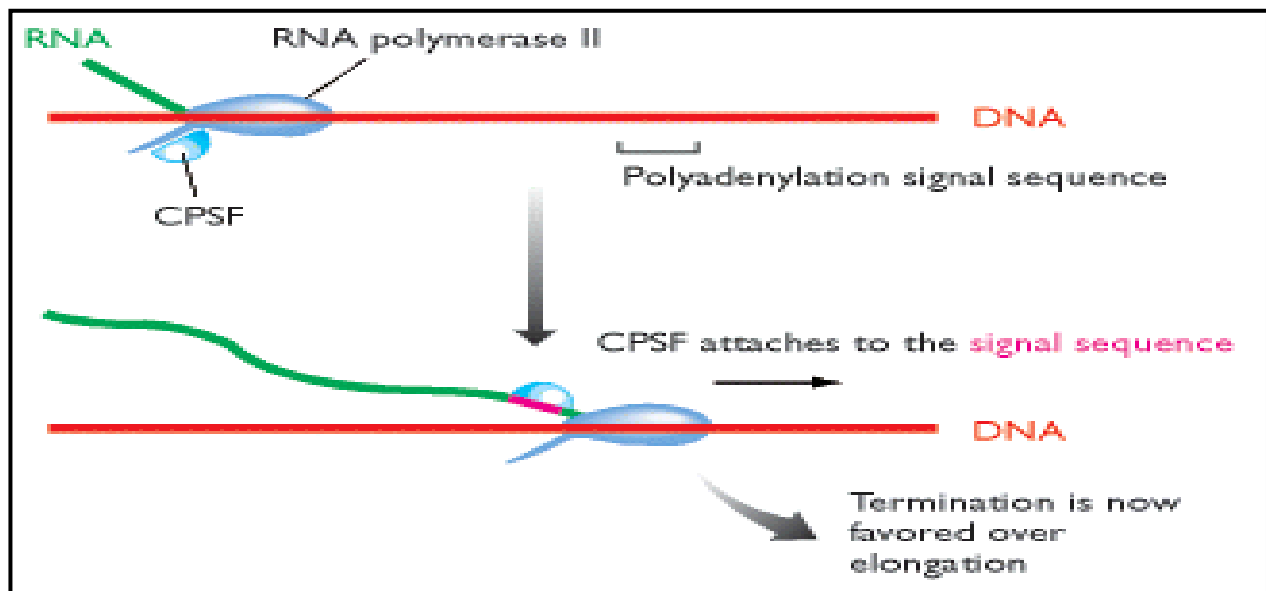


Figure 6: la terminaison de la transcription.

II.1.4. La maturation de l'ARN

Un ARN nouvellement transcrit est appelé transcrit primaire. Chez les eucaryotes et les transcrits primaires d'ARN subissent quelques transformations post-transcriptionnelles. Cette phase dite de maturation des transcrits comporte trois étapes: l'addition d'une coiffe en 5' du transcrit, l'addition d'une queue poly A en 3' et enfin l'épissage du transcrit primaire (**figure7**). L'épissage ; l'ADN chez les eucaryotes possède des séquences codantes (Exons) et des séquences non codantes (Introns). Seuls les exons participent donc à la synthèse des protéines. L'ARN des eucaryotes est d'abord produit sous forme de pré-ARNm (introns + exons). Il subit ensuite une opération d'épissage : un complexe nucléoprotéique (le splicéosome) reconnaît les introns et les élimine5 (**Couée et al.,2010**).

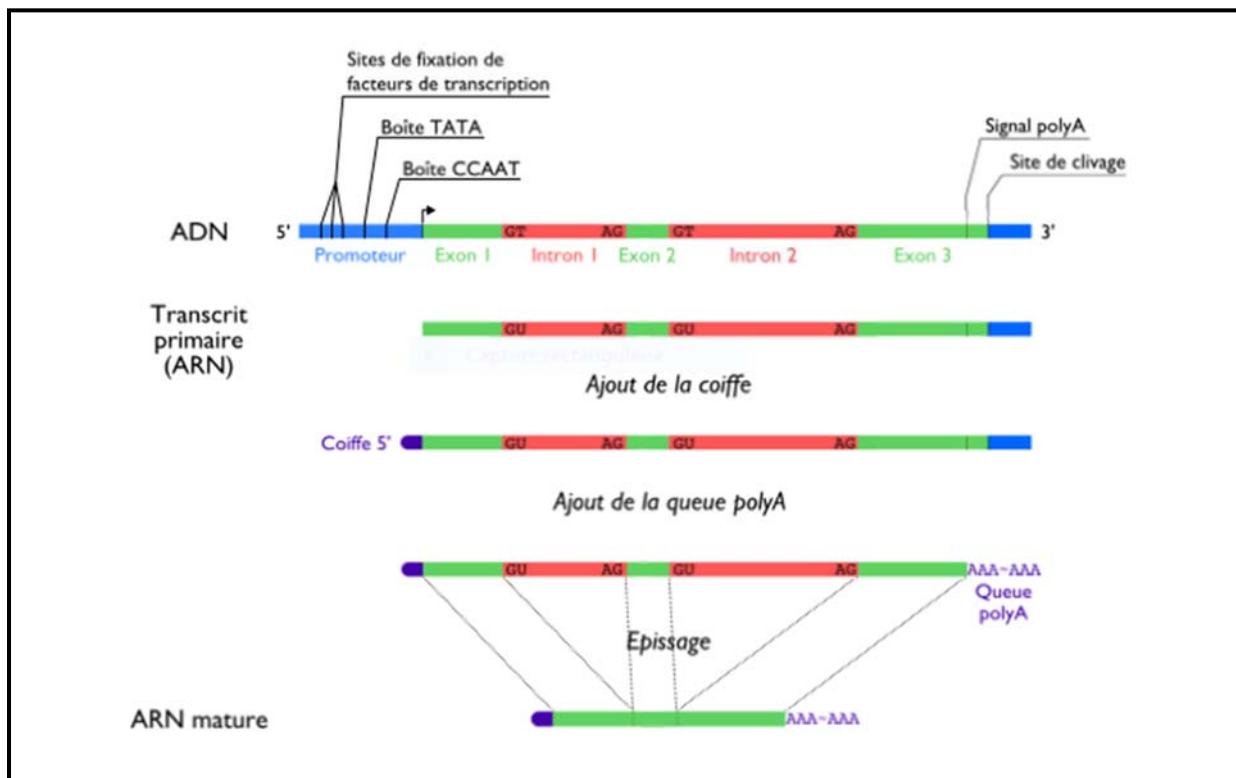


Figure 7 : Processus de maturation des transcrits primaires (**Couée et al.,2010**).

II.2. La traduction chez les eucaryotes

La traduction est le passage de l'ARNm en une séquence d'acides aminés représentant une protéine. Autrement dit, il faut donner une signification à une suite de bases ; c'est ce que fait le code génétique en associant à un triplet de bases un acide aminé ; il devient donc

possible de traduire une molécule d'ARNm en une séquence d'acides aminés c'est-à-dire une protéine.

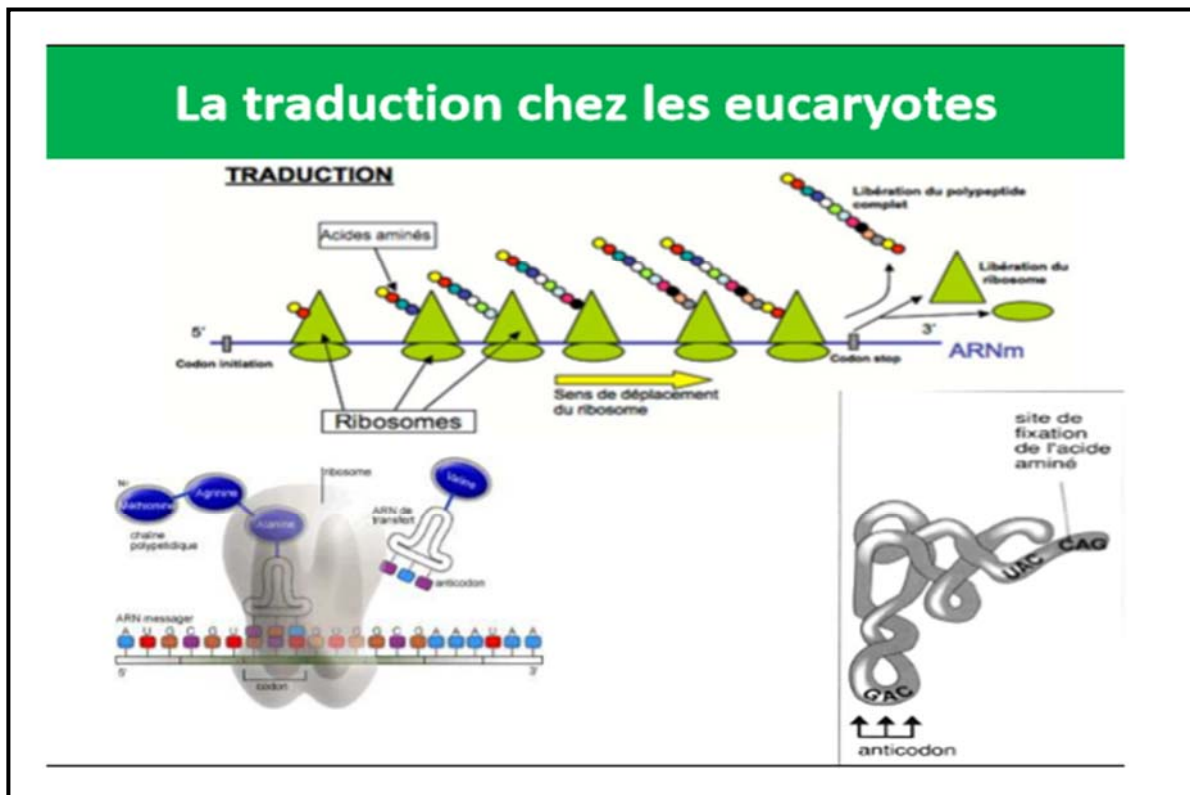


Figure 8 : La traduction chez les eucaryotes (Swynghedauw et Silvestre, 2008).

La machine de synthèse est formée d'acides aminés, de l'ARNm, de ribosomes (**figure8**) et de l'ARNt qui est formé d'une chaîne nucléotidique associée à des protéines et qui présente la forme d'une feuille de trèfle avec quatre pôles dont un qui forme un site de fixation de l'acide aminé et un autre qui est formé d'un triplet complémentaire au codon correspondant à l'acide aminé en question et qu'on appelle anticodon (Swynghedauw et Silvestre, 2008).

II.3. La régulation chez les eucaryotes

II.3.1. Au Niveau chromatinien

La régulation au niveau chromatinien se fait soit par :

- ❖ Acétylation des histones
- ❖ Méthylation des histones

L'acétylation des histones est l'addition d'un groupement $\text{CH}_3\text{C}=\text{O}$ sur des résidus lysine. Elle stimule la transcription en diminuant la compacité de la chromatine et la force d'interaction entre les histones et l'ADN et en créant des sites de fixation pour les complexes

d'activation de la transcription (**Krebs et al.,1999**). La méthylation C'est l'ajout d'un groupement méthyle à des bases d'ADN. Une forte méthylation inhibe la transcription (**figure 9**). Chez les eucaryotes, lorsque le promoteur d'un gène est méthylé, le gène en aval est en général réprimé et n'est donc plus transcrit en ARNm.

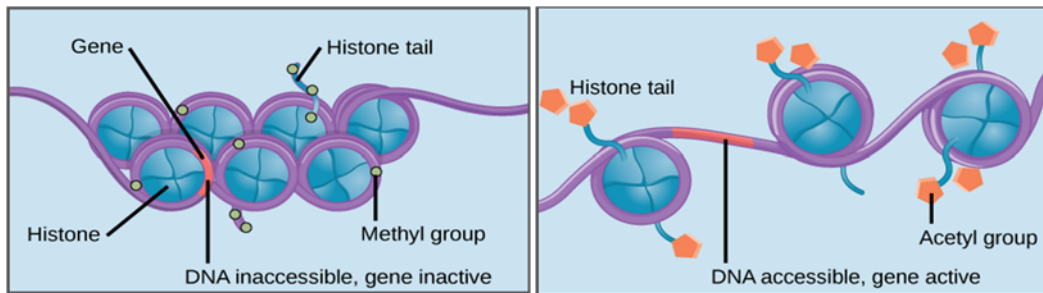


Figure9 : Acétylation / Méthylation des histones (**Krebs et al., 1999**).

II.3.2. Au niveau transcriptionnel

- Fixation des activateurs au Enhancers (zone amplificatrice).
- Une protéine s'installe sur l'ADN (DNA bending).
- Sa déformation force l'ADN à se courber.
- Rapprochement des activateurs du promoteur.
- Liaison d'une protéine médiatrice et des facteurs de transcription à l'ADN.
- Fixation de l'ARN polymérase au promoteur ce qui va permettre la transcription de l'ARN prés messenger (**figure10**).

Des répresseurs (silencers) peuvent se fixés sur la zone amplificatrice empêchant ainsi la fixation des activateur inhibant le processus de transcription (**Kevin, 1999**).

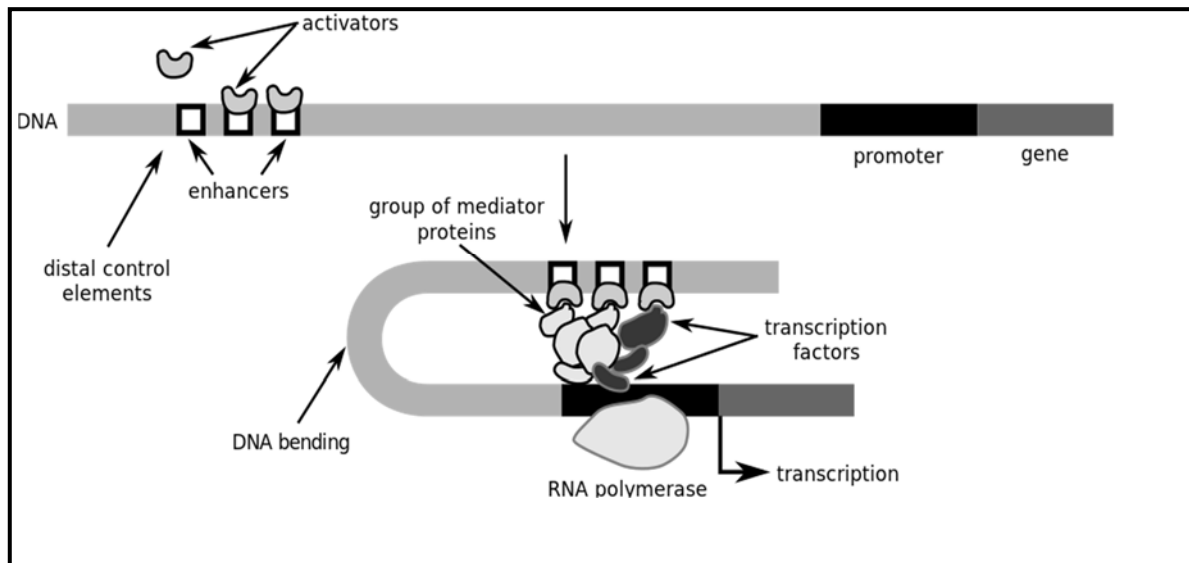


Figure10 : la régulation transcriptionnel chez les eucaryotes (Kevin, 1999).

II.3.3. Au niveau post-transcriptionnel

La régulation à ce niveau se fait par :

La polyadénylation (AAAAAA extrémité 3'), l'épissage, l'édition, ARN interférent et la dégradation de ARNm non sens.

II.3.4. Au niveau traductionnel

La Coiffe en 5' qui résulte de l'addition du 7-méthylguanosine sur la 1ere nucléotide de l'ARN, elle protège contre la dégradation de l'ARN. Cette dernière sera ensuite reconnu par des facteurs d'initiations (eIF4E) qui permet le recrutement du ribosome et la traduction en protéine.

II.3.5. La régulation post-traductionnelle

Modifications post-traductionnelles se fait par l'ajout de groupes fonctionnels, de groupes peptidiques ou par des modifications structurales ou chimiques ; par exemple : phosphorylation ou Protéolyse, la protéolyse est l'hydrolyse des protéines sous l'action des enzymes : la protéine se décompose en plusieurs morceaux (**Couée et al., 2010**).

III. ARNs non codants

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 (**costa, 2007**). 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, Despois, 2003**).

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 codant considéré comme des séquences d'ADN « poubelle » accumulée au cours de l'évolution (**Feng et al., 2014**).

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, Rinn, 2012**).

IV. Une panoplie d'ARN longs non codants

Les lncRNA sont des transcrits de plus de 200 nucléotides qui ne codent pas pour des protéines (**Gibb et al., 2011**). Bien que cette définition soit en apparence arbitraire, elle permet de distinguer les lncRNA des petits ARN régulateurs, tels que les miARN, les piARN et les autres petits ARN nucléaires. Les lncRNA, transcrits généralement par l'ARN polymérase II (**Amit I et al., 2009**), constituent un groupe de taille très hétérogène, certains pouvant s'étendre sur plusieurs dizaines de kilobases (kb). Les gènes de lncRNA partagent plusieurs caractéristiques avec les gènes codant pour des protéines, comme des profils épigénétiques

similaires, la présence de signaux d'épissage et de polyadénylation, ainsi que la taille des exons et des introns (Ulitsky,Bartel,2013). Toutefois, et comparativement aux ARNm, les lncRNA sont généralement plus enrichis dans le noyau (Kapranov et al.,2007), et montrent une plus faible conservation de séquence, bien que certains d'entre eux soient fortement conservés (Ulitsky,Bartel,2013).

De plus, les gènes de lncRNA s'expriment plus faiblement que les gènes codants et leur expression est remarquablement spécifique de certains tissus (Kung et al.,2013). En fonction de leur position par rapport aux gènes codants, les gènes de lncRNA peuvent être divisés en deux grandes catégories : les lncRNA intergéniques et les lncRNA intragéniques (Ponting et al.,2009 ; Rinn, chang,2012) (Figure11).

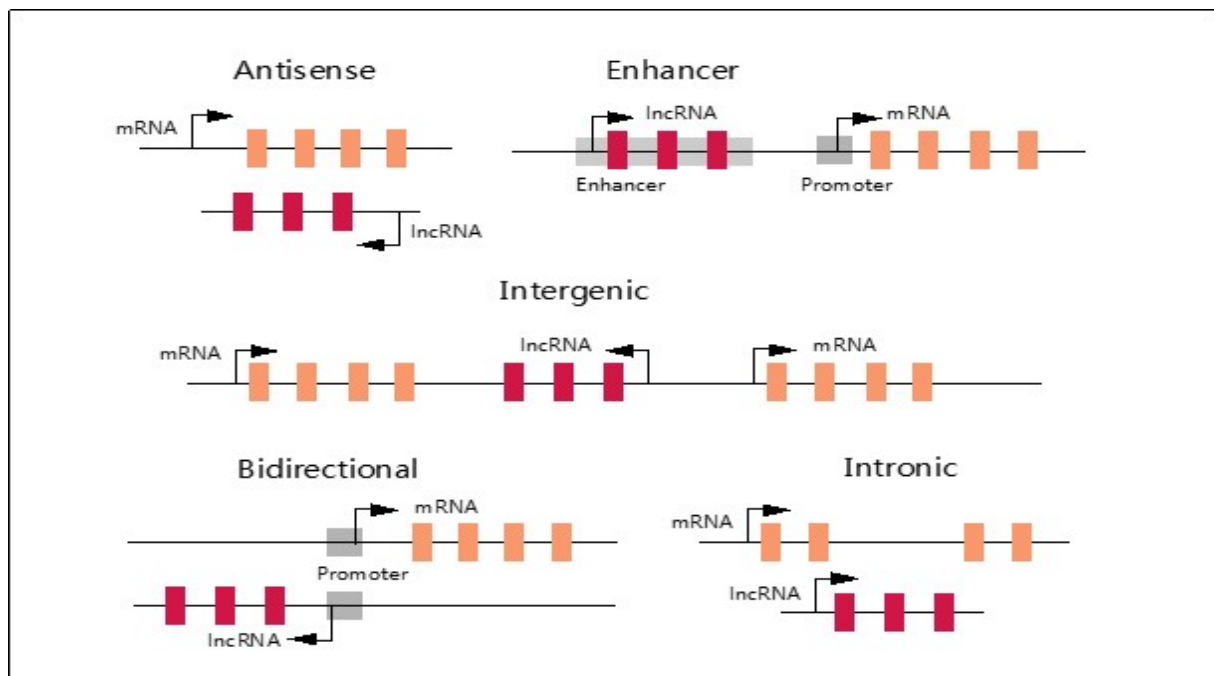


Figure11 : représentation schématique des principales catégories de lncRNA. (Ulitsky,Bartel, 2013).

Les lncRNA intergéniques, localisés par définition dans les régions non annotées du génome, sont généralement dénommés lincRNA. Ils représentent, à l'heure actuelle, la classe de lncRNA la mieux étudiée (Ulitsky,Bartel,2013). Les lncRNA intragéniques, quant à eux, peuvent être subdivisés en fonction de la façon dont ils chevauchent les gènes codants ou de leur orientation par rapport à ceux-ci (antisens, intronique, etc.). Il est à noter que de

nombreux gènes de lncRNA intergéniques ont un site d'initiation de la transcription proche de celui d'un gène codant, la transcription se faisant sur le brin opposé (transcription divergente). Nous avons récemment démontré que les gènes associés à ces transcrits divergents codent fréquemment pour des régulateurs transcriptionnels impliqués dans le développement et la différenciation cellulaire (Lepoivre et al.,2013).Enfin, certains lncRNA chevauchent de petits ARN, tels que des petits ARN nucléaires (snoRNA) (Abel et al., 2014),ou des miARN, avec des liens fonctionnels potentiels, comme dans le cas de régions soumises à l'empreinte génomique (Koerner et al.,2009).

De multiples lncRNA contiennent des éléments répétés, comme des LINE ou des SINE, avec des implications fonctionnelles potentielles (Kapusta et al.,2013 ; Kelly,2012). Cependant, pris dans leur ensemble, les lncRNA n'ont pas de séquence ou de structure conservée qui pourrait être indicative d'une fonction particulière (Ulitsky,Bartel,2013).De ce fait, la plupart des études visant à identifier des lncRNA potentiellement pertinents dans un contexte physiologique ou pathologique donné sont basées sur des analyses de coexpression ou de corégulation. On transpose alors la fonction des gènes voisins ayant un profil similaire aux lncRNA dont la fonction est à définir (Guttman, Rinn,2012). Ces approches sont aussi connues sous le terme « culpabilité par association » (Figure12).

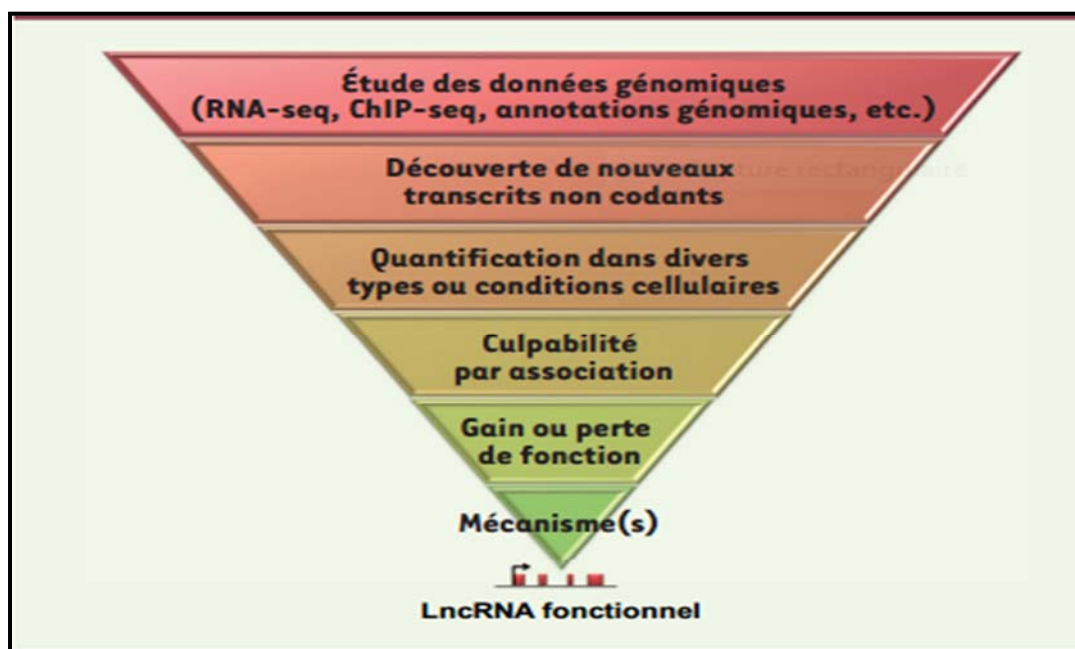


Figure12: Approches expérimentales menant à la caractérisation fonctionnelle des lncRNA. (Guttman, Rinn,2012).

V. Fonctions des lncRNA dans les processus physiologiques

Les longs ARN non codants jouent un rôle important dans plusieurs processus physiologiques, telles que la régulation épigénétique de l'expression génique (Wang et al,2011) ou la maturation des ARN messagers(Yang et al.,2011). De plus, plusieurs travaux récents indiquent qu'ils seraient impliqués dans le développement de pathologies humaines.

En effet, de nombreux longs ARN non codants sont impliqués dans le remodelage de la chromatine.

Les lncRNA sont aussi largement impliqués dans les processus post-transcriptionnels liés à la biogenèse des ARNm, tels que l'épissage, le transport, la traduction et la dégradation des ARNm.(figure13)

Les lncRNA peuvent potentiellement lier l'ADN, des protéines ou d'autres ARN, formant des plateformes d'échafaudage et permettant ainsi l'interaction entre diverses molécules fonctionnelles, Les LncRNAs peuvent se lier avec les microARNs (miRNAs), cette liaison empêche les miRNA d'effectuer leurs fonctions, participent ainsi à la régulation de la synthèse protéique. On appelle cette fonction « éponge à miRNA »(Anko,Neugebauer, 2010).(figure13)

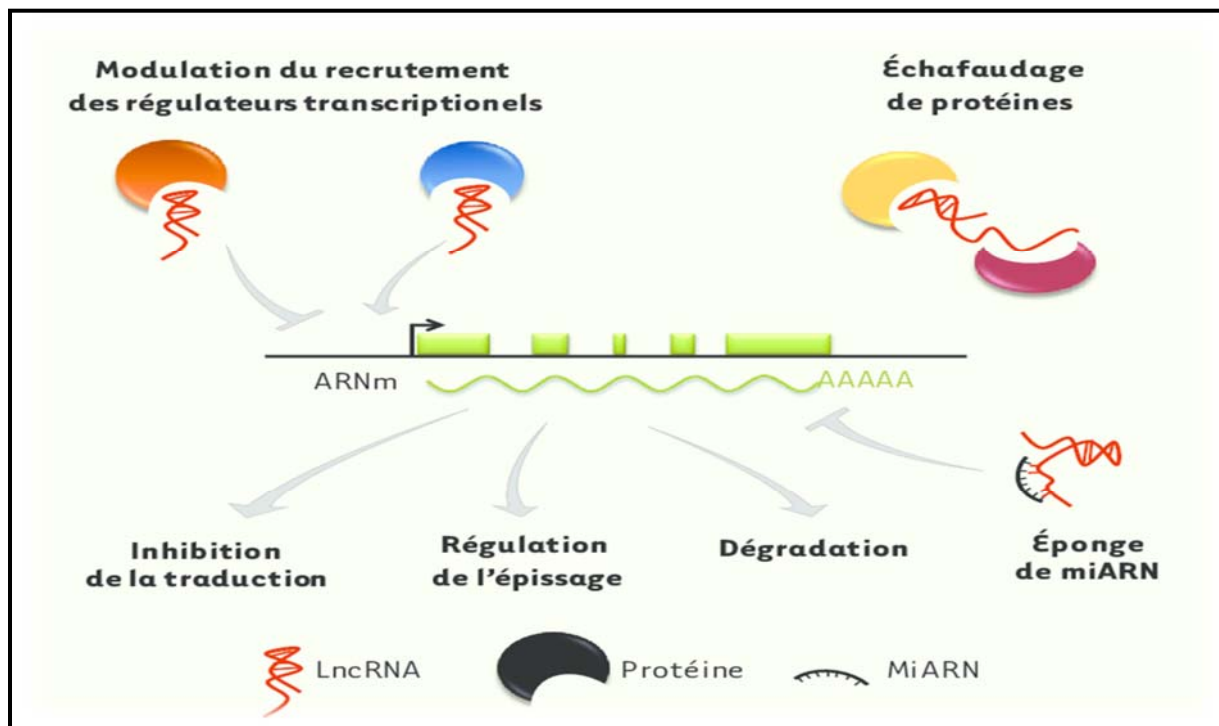


Figure13 : Exemples de mécanismes d'action des LncRNA. (Anko, Neugebauer,2010)

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+ T cells, es CD4+ TC, CD11c+... cependant leurs fonctions exactes restent encore peu explorées (**Fitzgerald, caffrey,2014**).

VI. LncRNA - cellules β

Des études récentes ont révélé des milliers lncRNAs dans le pancréas humain. Cellules β lncRNAs sont souvent de type cellulaire spécifique et présentent une régulation dynamique lors de la différenciation ou en changeant les concentrations de glucose. Bien que ces caractéristiques suggèrent un rôle de lncRNAs dans les cellules β , dans la régulation des gènes et dans le diabète ; la fonction de la cellule LncRNA reste largement inconnu.

Une grande fraction Des lncRNAs sont connus pour être activé lors de la différenciation cellulaire (**Moran et al., 2012**).cette spécificité cellulaire a également été notée pour les LncRNAs dans d'autres types de cellules (**Cabili et al., 2011; Derrien et al., 2012**) et souligne la possibilité que les LncRNAspeuvent réguler les programmes génétiques important pour la différenciation.

En outre, plusieurs cellules β LncRNAs ont été présentés pour être régleménté par des concentrations extracellulaires de glucose, suggérant un rôle potentiel des LncRNAs dans l'adaptation fonctionnelle des cellules β pour augmenter les demandes de sécrétion d'insuline (**Moran et al.,2012**).

Certaines cartes d'islotes sont mises en évidence dans des loci contenant des propriétés polygéniques ou défauts mendéliens associés au diabète humain, alors que les lncRNA sélectionnés montrent une déréglementation dans les îlots des donneurs d'organes avec le diabète de type 1 (DT1) (**Moran et al., 2012**). Collectivement, ces propriétés définissent une classe nouvellement identifiée de régulateurs candidats de la différenciation des cellules β , avec des implications potentielles pour le diabète sucré humain. Cependant, la vraie pertinence des lncRNA dépend de qu'ils produisent une fonction physiologique dans les cellules β humaines, qui reste à traiter systématiquement.

VII. Expression des lncRNAs dans les cellules pancréatiques humaines

En intégrant des données transcriptomiques et des cartes de chromatine, 1128 lncRNA ont été identifiés de manière fiable dans le pancréas humain et cellules β purifiés (**Moran et al., 2012**). Beaucoup de lncRNAs ont été clairement exprimés dans les cellules β , ce qui suggère leur importance rôles dans la programmation de développement, le bon fonctionnement et / ou l'entretien du tissu endocrinien pancréatique.

Dans une profondeur distincte l'étude de l'ARN-séq des cellules β humaines purifiées, 148 lincRNAs étaient surexprimé dans les cellules β par rapport à les Cellules non β (**Nica et al., 2013**), Tandis qu'une autre étude a découvert 12 cellules β et 5 cellules α spécifiques à une cellule lncRNAs (**Bramswig et al., 2013**). Ces résultats suggèrent l'importance de l'expression des lncRNA dans l'îlot pancréatique humaine.

VIII. Rôles potentiels des lncRNAs dans le développement, la fonction et maintenance des cellules β pancréatiques

L'implication des lncRNAs dans le développement d'un programme d'îlots pancréatiques est possible et peut avoir des effets directs à l'origine des différences cytoarchitecturales entre îlots dans les différentes espèces (**figure 14A**), et contribuer à l'expression des différentes protéines essentielles dans la sécrétion cellulaire des îlots (**Figure 14B**).

Dans la même ligne de raisonnement, un certain nombre de lncRNA spécifiques à un type d'îlot récemment signalés (**Bramswig et al., 2013**) pourraient jouer des rôles essentiels dans la confection de fonctions spécifiques au type de cellule dans le pancréas.

La majorité des lncRNAs associés à divers facteurs de modification de la chromatine sont des candidats tentant des facteurs qui pourraient fournir une spécificité moléculaire supplémentaire dans le ciblage des marqueurs épigénétiques dans les cellules des îlots pancréatiques. C'est tentant faire l'hypothèse que des lncRNAs spécifiques aux primates pourraient être impliqués dans l'expression de certains composants du couplage stimulus-

sécrétion dans les cellules β , qui à leur tour potentiellement contribue à la réponse spécifique aux espèces des cellules β aux éléments nutritifs stimuli (**figure 15B**).

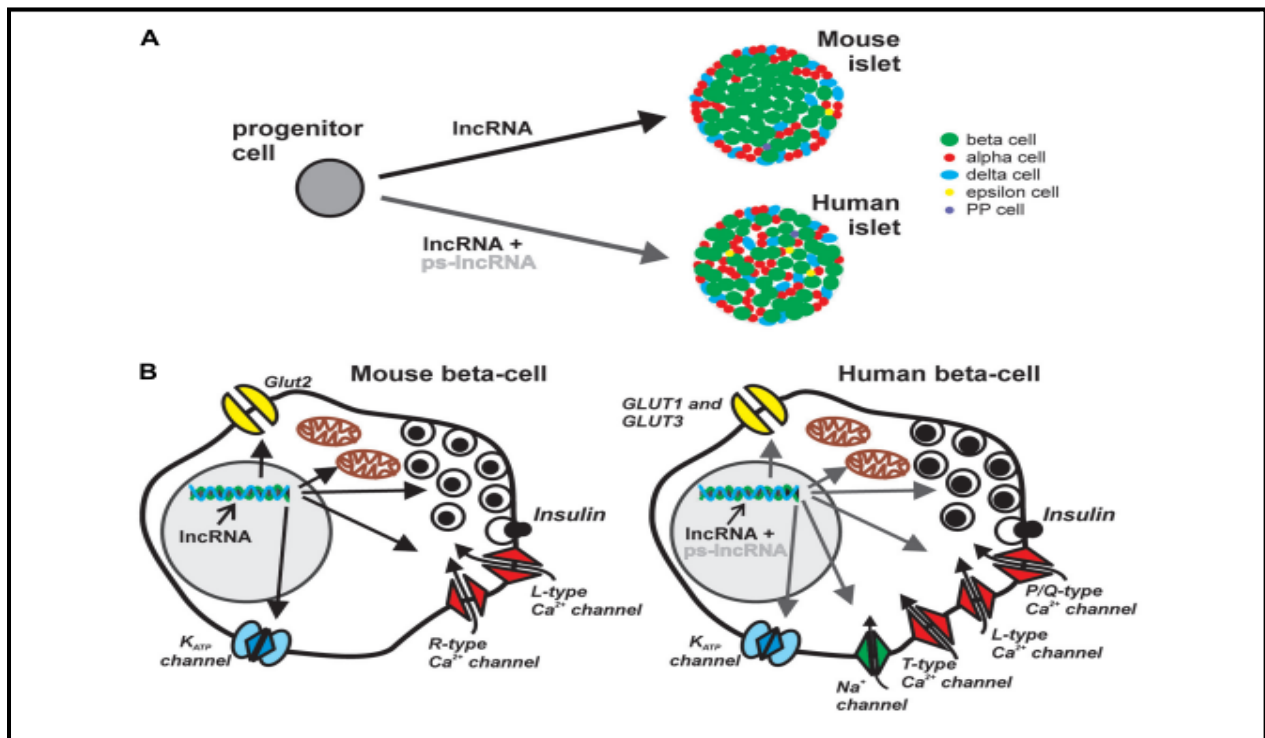


Figure 14 : Rôles potentiels des lncRNAs dans le développement d'îlots pancréatique et la fonction de cellules endocrines. (Bramswig et al., 2013)

Le présent travail a été réalisé au département de Biologie et physiologie cellulaire (BPC) de la faculté des sciences de la nature et de la vie (SNV) de l'université BLIDA1, pendant une période de 7 mois : du Février au Septembre 2017.

N'ayant pas de plateforme de séquençage à haut débit en Algérie, nous nous sommes focalisés dans notre travail à une étude *in silico* en utilisant les données des bases de données internationales et stockées et publiées dans plusieurs bases de données.

Rappelons que l'objectif de notre travail est de faire une analyse bioinformatique des LncRNAs afin d'identifier leur contribution dans le diabète type 1.

II.1. Matériel

II.1.1. Matériel non biologique

Rappelons que notre étude est une étude *in Silico*, qui nécessite l'utilisation du matériel informatique (ordinateur portable) et d'une connexion internet à haut débit.

II.1.1.1. Outils bio-informatiques

Les progrès de la biologie moléculaire et du grand projet du génome humain ont élargi nos outils tant que biologiste pour étudier les mécanismes pathogéniques des maladies humaines et la prédiction de certaines fonctions biologiques. Ainsi, nous avons utilisé des outils bio-informatiques (**ANNEXE I**) pour le traitement de nos échantillons moléculaires stockés dans des bases de données à accès libre (Open Source) nécessitant des interfaces web, des logiciels, des logiciels et des algorithmes et qui sont répertoriés dans le **tableau I** :

Tableau I : Description des outils bioinformatiques

Base de données	Commentaires	URL
NCBI	C'est une base de données qui fait progresser la science et la santé en fournissant l'accès à l'information biomédicale et génétique.	https://www.ncbi.nlm.nih.gov
UniProt	C'est une base de données ouverte, stable et accessible en ligne, elle est issue de la consolidation de l'ensemble des données produites par la communauté scientifique.	www.uniprot.org/
Galaxy	est une plateforme de travail en ligne Contenant divers outils pour effectuer un grand nombre d'opération sur des données génomiques.	https://usegalaxy.org/
Blast	outil développé par Genbank, qui permet de d'aligner votre séquence contre l'ensemble des séquences de la banque.	https://www.ncbi.nlm.nih.gov/BLAST/
Genmania	Représente multiples Association Réseau algorithme d'intégration.	https://genmania.org/

STRING	La base de données STRING vise à fournir une évaluation critique et une intégration des interactions protéines-protéines, y compris les associations directes (physiques) et indirectes (fonctionnelles).	http://string-db.org
---------------	---	---

II.1.2. Matériel biologique

Pour répondre à la problématique de notre étude in Silico, nous avons utilisé des échantillons stockés dans des bases de données à accès libre (Open Source) comme NCBI (National Center for Biotechnology Information). Ces échantillons ont été obtenus après une étude moléculaire qui avait nécessité une expérimentation à partir desquelles des molécules biologiques ont été obtenus puis traités par des plateformes à l'échelle internationale utilisant la technologie de séquençage à haut débit comme les NGS (Next-Generation-Sequencing). En effet, cette dernière, nous permet d'obtenir des séquences d'ADN ou de transcriptome et ceci pour répondre à plusieurs problématiques ou thématiques dans le souci de comprendre les mécanismes d'interactions moléculaires, les voies cellulaires et biologiques associées à des pathologies qui dans notre cas est le diabète type 1, ou lier aux diversités des espèces vivantes ou à des phénomènes biologiques.

Après une sélection des molécules biologiques que nous avons utilisées pour répondre à notre problématique, nous avons retenu les molécules qui sont répertoriés dans le tableau suivant :

Tableau II : Description des molécules biologiques

Molécules	Nature des séquences	Numéro d'accèsion	ID : identifiant	Nature de l'expérience	Nature du fichier	Format	Source Référence
ZnT8	Protéine	NC_000008	169026	séquençage	txt	FASTA	NCBI (GenBank)
INS	ADN	NC_000011	3630			FASTA	NCBI (GenBank)
PDX1	ARN	NC_000013	3651			FASTA	NCBI (GenBank)

Molécules	Nature des séquences	Numéro d'accèsion	ID : identifiant	Nature de l'expérience	Nature du fichier	Format	Source Reference
Lnc RNA (β cells)	ARN	GSM1093238 SRX248493	342774	Illumina HiSeq 2000	SRA	FASTQ	NCBI (SRA)

II.2. Etude Moléculaire

II.2.1. Traitement des données transcriptomiques

II.2.1.1. Outil Galaxy

➤ Principe

Galaxy est une plateforme web qui regroupe plusieurs outils bioinformatiques pour l'analyse des données NGS (Next-Generation-sequencing) sur une interface graphique. En effet, la plateforme Galaxy, permet de manipuler des séquences génomiques, de les aligner et enfin les analyser (<https://usegalaxy.org/>).

➤ Méthode

❖ Récupération des données brutes

- Pour télécharger les données sur la plateforme galaxy, nous cliquons sur « **Get Data** » puis sur " EBI SRA »
- Grâce au numéro d'accession (SRA) du fichier d'intérêt " séquence RNA-seq , nous téléchargeons nos échantillons « les SRA-RNA-seq » sous format Fast-Q.

❖ Control de qualité

La qualité des séquences générés sous format FASTQ vont être analysé par FASTQC
Ce logiciel va convertir les fichiers fastq en un format que Galaxy peut reconnaître il permet de générer un rapport de la qualité moyenne sous la forme HTML.

- Dans le panneau d'outils à gauche on click sur "NGS:QC and manipulation » ensuite sur l'outil "FastQC Read Quality reports".
- Nous pouvons visualiser ainsi, la qualité de la séquence.

❖ Suppression des adaptateurs/primers :

Cette étape est une étape obligatoire qui nécessite de Clipper et adapter les séquences pour cela ce Logiciel va permettre de supprimer les séquences d'adaptateurs, les amorces, les queues poly-A et d'autres types de séquences indésirables de notre séquençage à haut débit. Cela s'avère nécessaire lorsque la longueur du read est supérieure à celle de la molécule à séquencer, comme pour les LNC ARNs. Nous allons donc en premier lieu,

copiez notre séquence contaminée, aller à l'outil "Clip adapter séquence" copier et collez la séquence dans le champ puis appuyez sur exécuter pour nettoyer.

❖ Alignement et analyse de données de RNA-Seq:

"TopHat" est un logiciel d'alignement de séquences conçu spécialement pour le RNA-Seq. il permet de réaliser des alignements sur les jonctions d'épissage et sur le génome dans ce cas, un alignement sur un transcriptome virtuel extrait à partir des positions des exons est d'abord effectué, puis les reads non alignés sur ce transcriptome seront alignés sur le génome. Pour cette partie, on va suivre les étapes suivantes :

- Cliquez sur "NGS: Analyse ARN" après sur "TopHat Grapped-read mapper for Rna-seq Data" .
- Sélectionnez les fichiers FastQ, puis appuyez sur Exécuter puis sur le fichier "accepted Hits Data" puis visualiser les données par l'outil "IGB"

❖ Assemblage de transcrits ou Analyse d'ARN :

Cette étape nécessite l'utilisation de l'outil "Cufflinks" qui permet d'identifier un ensemble minimum de transcriptions indépendantes qui peuvent expliquer les lectures observées dans l'expérience RNA-Seq. Ainsi, l'abondance des transcrits en FPKM (**F**ragments **P**er **K**ilobase of exon model per **M**illion mapped fragments) qui sont des valeurs normalisées par la taille des transcrits et par la profondeur de séquençage. L'assemblage se fait le Protocol suivant :

- Appuyer sur "**NGS : RNA Analysis – Cufflinks**" sélectionner le fichier correspondant qui est au format "Bam" produit par "Tophat".
- Puis ajuster les paramètres selon notre séquence et exécuter pour assembler les lectures des transcrits.
- La plupart de ces fichiers sont décrits sous forme de tableaux en détails.

❖ Cuffcompare:

"CuffCompare" compare les transcrits assemblés à une annotation de référence et donne une information en détail sur les identités et les différences entre elles. Dans notre cas nous pouvons utiliser l'ensemble de données de transcription combiné pour comparer

les niveaux d'expression relatifs de chaque exon et chaque transcription dans nos deux ensembles de données d'entrée d'origine, en utilisant "CuffDiff" avec les étapes suivantes :

- On clique sur "NGS:RNA Analysis" puis sur "Cuffcompare".
- Sur le champ sélectionner le fichier "combined transcripts" de CuffCompare.
- Pour le fichier "BAM" choisissez "accepted-Hits" de "TopHat" puis exécuter.

❖ CuffDiff:

"CuffDiff" permet de tester les niveaux d'expression des gènes et des transcriptions dans plus d'une seule condition.

- Ouvrez l'outil "CuffDiff".
- Choisissez le fichier "transcript differential expression testing" généré par CuffDiff.
- Le deuxième fichier "accepted-Hits".
- Puis exécuter.

❖ Création de Workflows:

Comme vous avez pu le constater, enchaîner toutes ces étapes est fastidieux. C'est pourquoi il est intéressant d'automatiser leur enchaînement. Pour cela nous allons créer un workflow.

- Aller dans l'onglet "Workflow" et créer un nouveau workflow.
- Cliquer sur "input dataset" pour le fichier source.
- Dans la partie "NGS" sélectionnez l'outil "FastQ Groomer".
- Ajouter aussi "FastQC".
- Ensuite sur "Clip adapter sequence".
- Finalement sur "TopHat".
- Après l'ajout de ces fenêtres, mettez-la en ordre.

- Puis lier ces fenêtres entre eux.
- Sauvegarder votre "workflow".

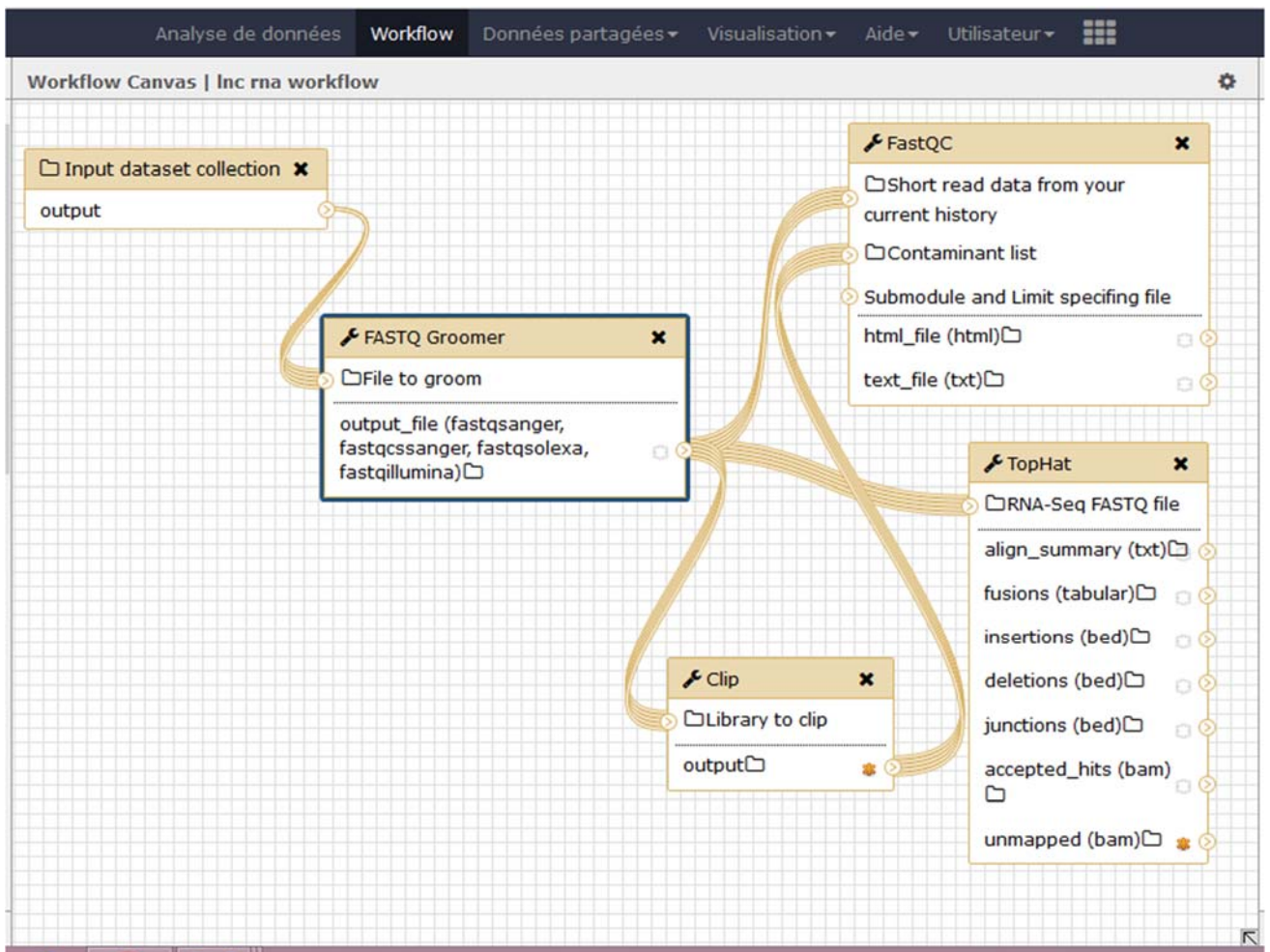


Figure 15 : L'organigramme de notre travail sur Galaxy.

II.2.2. Etude d'interactions des données moléculaires

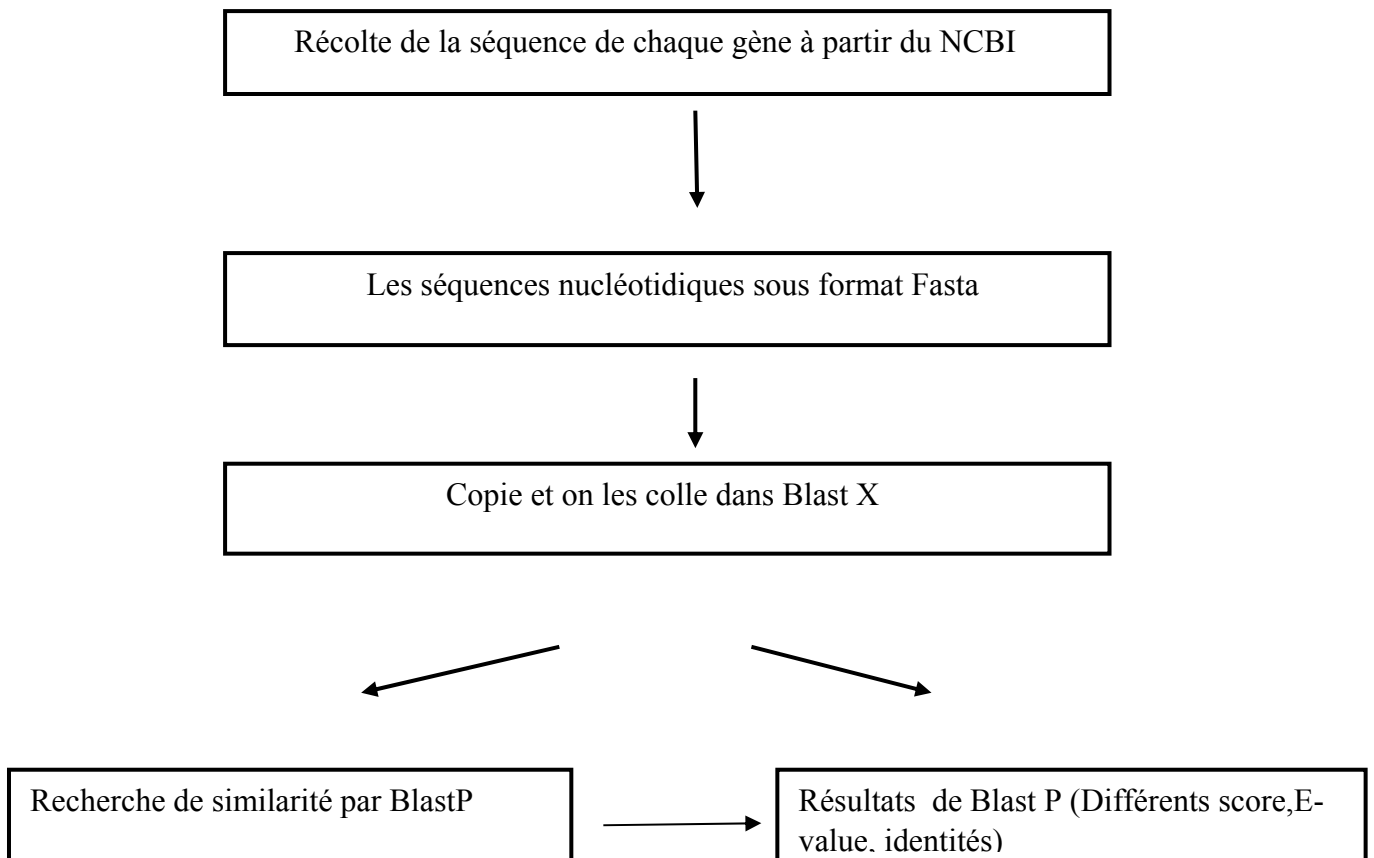
Différents outils s'offrent à nous, afin de traiter les données moléculaires et d'identifier les régions de similarité. Nous allons présenter ces outils ci-dessous.

II.2.2.1. Outil Blast

➤ Principe

L'utilisation de l'outil Blast de NCBI permet d'aligner une séquence contre l'ensemble des séquences de la banque. L'alignement se fait par sous séquences de 3 éléments minimum.

➤ Méthode



II.2.2.2. Outil Phyre2

Après l'obtention des différents résultats de Blast P, on a orienté notre étude sur ZnT8 qui est présent uniquement dans les cellules pancréatiques et qui est un rôle majeur dans la fixation de la molécule d'insuline (pas Zinc pas d'insuline).

Pour la visualisation 3D de ZnT8, nous avons utilisé l'outil phyre2 et Rasmol ; dont les étapes sont les suivantes. Premièrement, ouvrir le site Phyre2 et insérer le symbole ZnT8 puis écrire votre adress mail et appuyer sur search ; patientez quelque minutes, votre fichier est dans la boîte mail, par la suite, il faut convertir dans Rasmol pour obtenir la structure 3D de Znt8.

II.2.2.3. Outil UNIPROT

➤ Principe

UniProt est une base de données généraliste « open source », protéique qui permet d'aligner 2 ou plusieurs protéines est donc de prédire la similarité entre les séquences.

➤ Méthode

- Rendez-vous sur UniProt (<https://UniProt.org/>).
- Puis insérer le symbole de chaque protéine (INS-ZnT8- PDX).
- On clique sur search.
- On a un petit tableau des résultats de chaque protéine on les met dans add to basket.
- On selectionne dans l'onglet Basket les 3 protéines avec leur numéro d'accession, puis on clique sur align.

II.2.2.4. Outil GeneMANIA

GeneMANIA permet aux bio-informatiens de saisir un seul gène et de voir les réseaux d'interactions associées avec le gène, suivant la méthode ci-dessous :

- Rendez-vous sur la plate forme « Genmania » et Insérer les symboles des 3 gènes **INS** – **SLC30A8** – **PDX1**, puis lancer la recherche.
- Les résultats de Genmania vont apparaitre.

II.2.2.5. Etude des interactions protéiques par String

String vise à fournir une intégration des interactions proteine-proteine, d'où son utilisation pour l'étude d'interactions de 3 protéines **INS** – **ZnT8** et **PDX1**, suivant la méthode ci-dessous :

- Aller sur la plateformeSTRING(<http://string-db.org>)
- Cliquer sur multiple protéines

- Puis saisir les protéines cibles
- Choisir l'organisme (*Homo sapiens*)
- Puis appuyer sur search.

Rappelons que notre étude est une étude in silico qui utilise des outils bioinformatique pour comprendre et étudier le rôle des LncRNA dans les cellules β pancréatique et leurs relations avec le diabète type1

III.1. Résultats

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

III.1. 1. Alignement multiple par la base de données UniProt

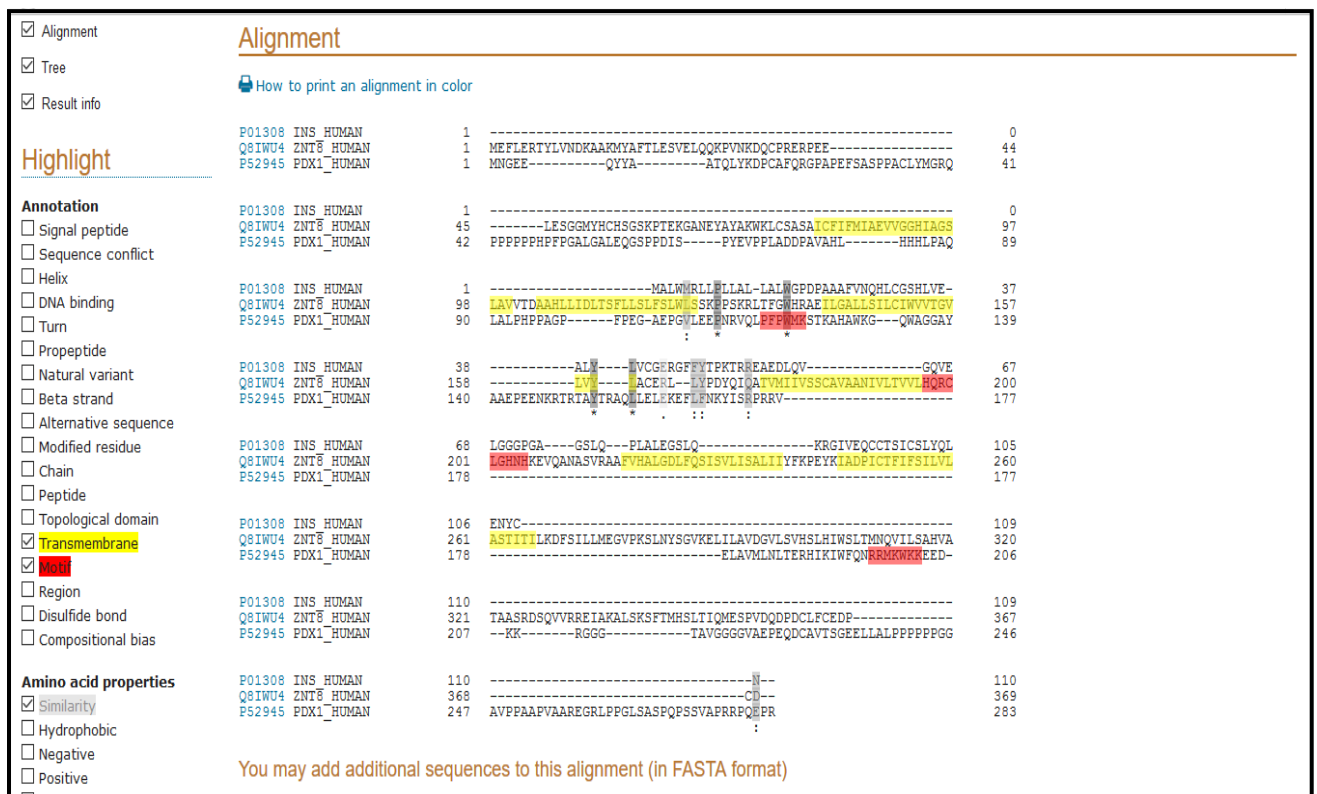


Figure 16 : Les pré- résultats d’alignement multiple.

La figure ci-dessus représente un alignement multiple de trois protéines INS, ZnT8 et PDX1 Humaine dont la description est la suivante

- Les régions de similarité sont visibles en gris.
- Les rouges qui représentent les motifs, sont présents que dans ZnT8 et PDX1 et absent dans INS.
- Les jaunes représentent les sites transmembranaires présentés que dans ZnT8 et PDX1.

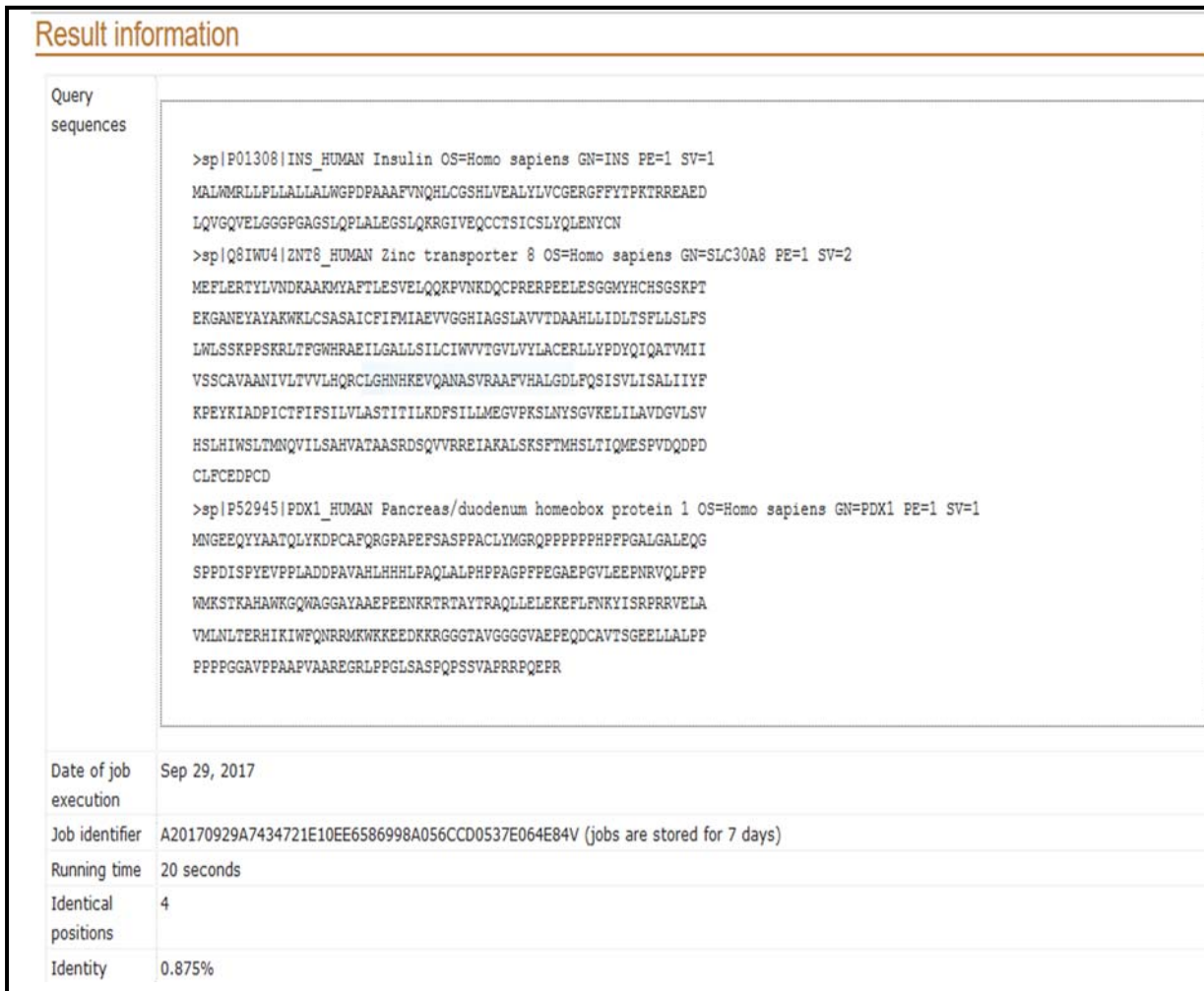


Figure 17 : les résultats d’alignement multiple.

Cet alignement se fait dans un temps très court 20 secondes. Ces résultats donnent les positions identiques des trois protéines (= 4), ainsi que les positions similaires (=6) avec un très faible pourcentage d’identités (0.875%) ce qui montre une grande différence entre les 3 protéines malgré qu’elles accomplissent des fonctions assez similaires.

III.1.2. Etude d'interaction entre les gènes

Cette étude se fait par une base de données qui nécessite des algorithmes d'interaction « GeneMANIA », et les résultats se présentent comme suit :

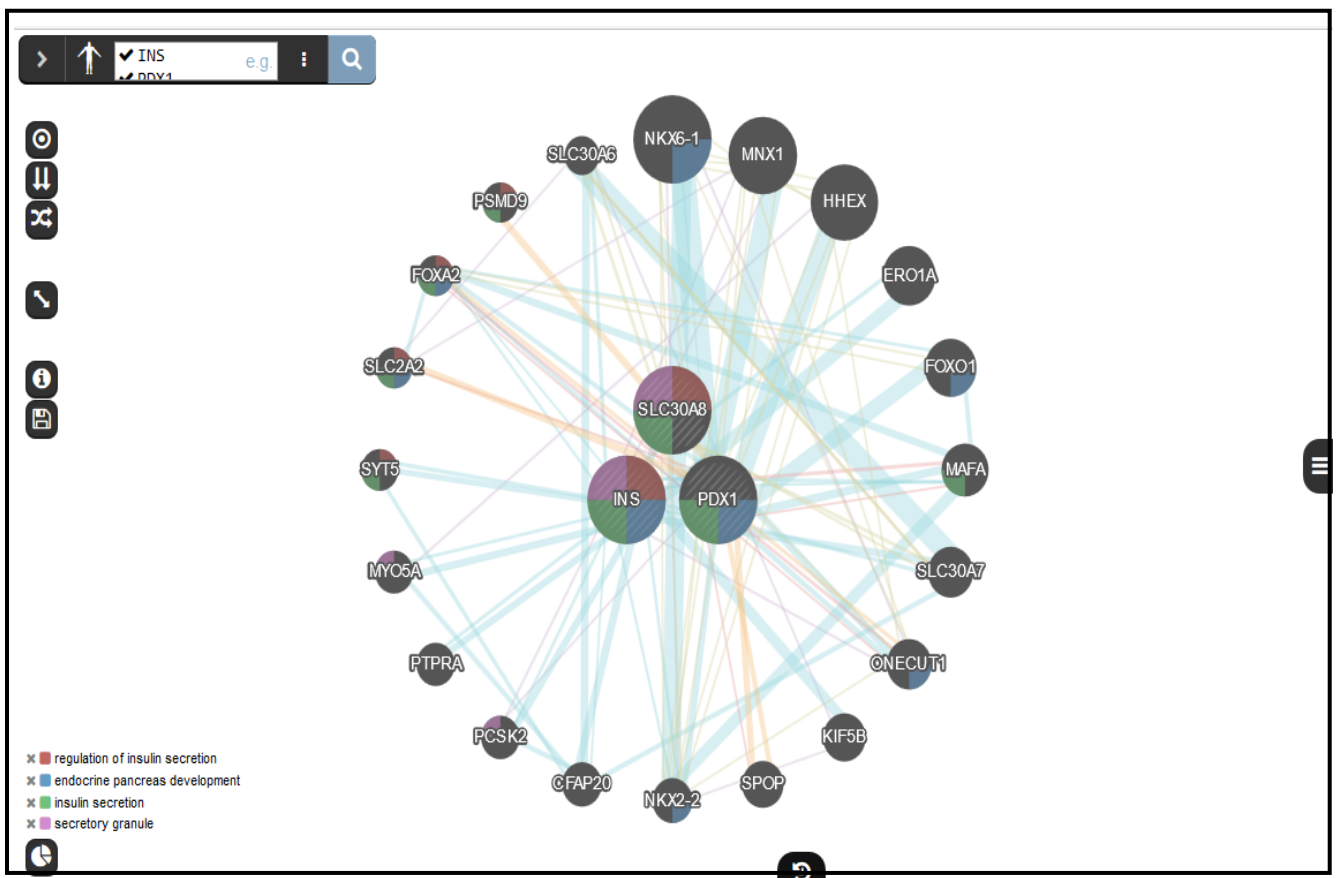


Figure 18 : Schéma représentatif des interactions géniques.

Les résultats montrent les différentes interactions entre nos gènes cibles (INS – PDX1 – SLC30A8 (ZnT8)).

Ces résultats touchent certains niveaux d'interactions :

- En premier, au niveau de la régulation de la sécrétion d'insuline. Il existe des interactions entre INS, SLC30A8 et d'autres gènes FOXA2, SLC2A2, SYT5 et le PSMD9 qui sont représentés en rouge.

- Les lignes bleus représentent les interactions au niveau du développement du pancréas endocrine. Les interactions sont visibles entre PDX1, INS, NKX6-1, FOXO1 et le FOXA2.
- Les lignes vertes représentent les interactions au niveau de la sécrétion d'insuline montrant des interactions avec multitude de gènes ; principalement avec nos gènes cibles (PDX1, INS, SLC30A8) et d'autres comme Wafa, SYT5, SLC2A2, FOXA2 et PSMO9.
- La couleur rose représente les interactions au niveau des granules sécrétoires. Ces interactions touchent INS, SLC30A8, PCSK2 et MYO5A.

III.1. 3.Annotation génomique par Blast

L'annotation génomique pour les 3 séquences INS – SLC30A8 – PDX1 débutant par la récolte de la séquence fasta de chacun d'eux ; passant par l'outil Blast X. Ce dernier, nous donne un champ vaste des résultats bruts. Que nous allons présenter dans le tableau ci-dessous.

Tableau III : Résultats du blast

Nom de la séquence	Identifiants	E-value	Score totale	Numéro d'accension	Longueur	Matches
insulin prepro protein [Homo sapiens]	99%	7e-54	174	NP-000198.1	110	1
insulin [Homo sapiens]	99%	2e-53	172	AEG1952.1	94	1
pancreas/duodenumhomeobox protein 1 [Homo sapiens]	100%	9e-44	331	NP_000200.1	283	2
insulinupstream factor 1 [Homo sapiens]	99%	6e-43	329	AAC05157.1	283	2
zinc transporter 8 isoform a [Homo sapiens]	99%	0.0	588	NP_776250.2	369	1
solute carrier family 30 member 8 [Homo sapiens]	99%	0.0	588	AAM80562.1	369	1
zinc transporter 8 isoform b [Homo sapiens]	99%	0.0	585	NP_001166282.1	320	1
SLC30A8 protein [Homo sapiens]	99%	0.0	583	ABQ59023.1	320	1
hypothetical protein[H. sapiens]	99%	0.0	578	CAD28545.1	320	1

III.1. 4. La modélisation 3D de ZnT8 :

Nous avons fait une recherche de domaine protéique de ZNT8 sur Blast P, cela pour prédire la structure 3D par un service web Phyre2 et qui est visualiser après par RasMol.

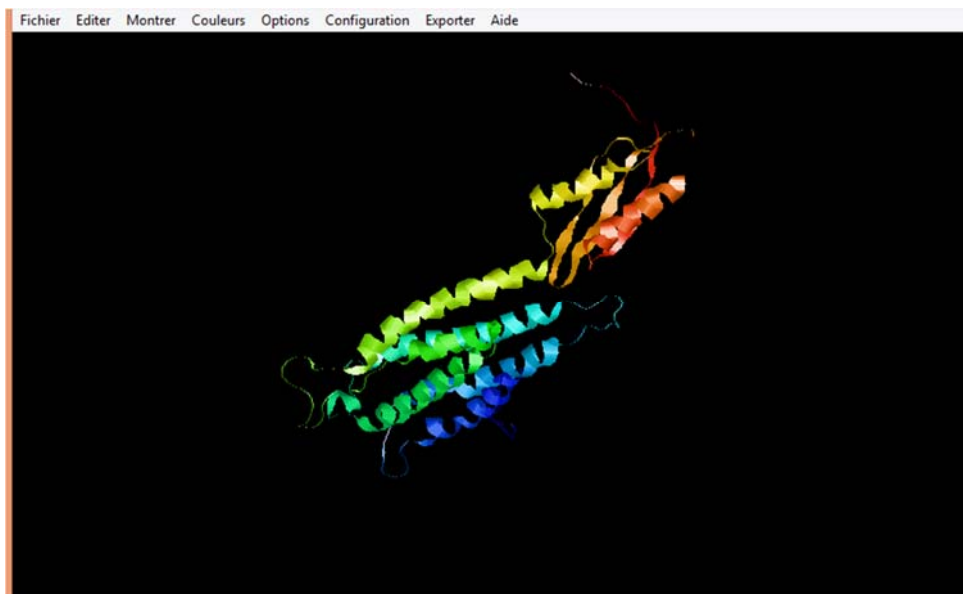


Figure 19: la structure 3D de ZnT8 visualisé par RasMol.

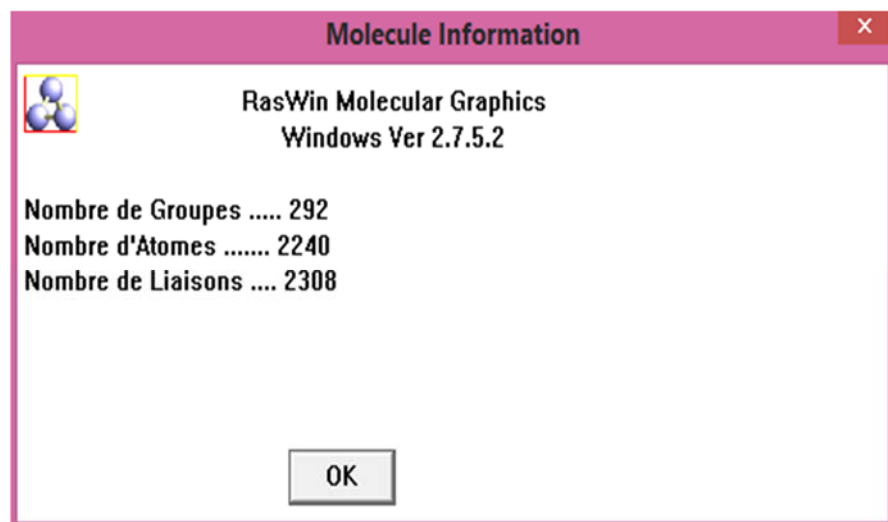


Figure 20 : le nombre de résidus de la protéine ZnT8.

Les résultats montrent différentes compositions de la molécule ZnT8 tel que le nombre de groupes qui est 292, le nombre d'atomes 2240, en plus le nombre de liaisons qui est égale 2308.

III.1. 5. Etude d'interaction protéique (String)

Les interactions qui se produisent entre les protéines sont au cœur du traitement cellulaire et leur caractérisation systématique contribuant à fournir un contexte dans la biologie des systèmes moléculaires. Cependant, les interactions connues et prédites sont dispersées sur de multiples ressources, et les données disponibles présentent des différences notables en termes de qualité et d'exhaustivité.

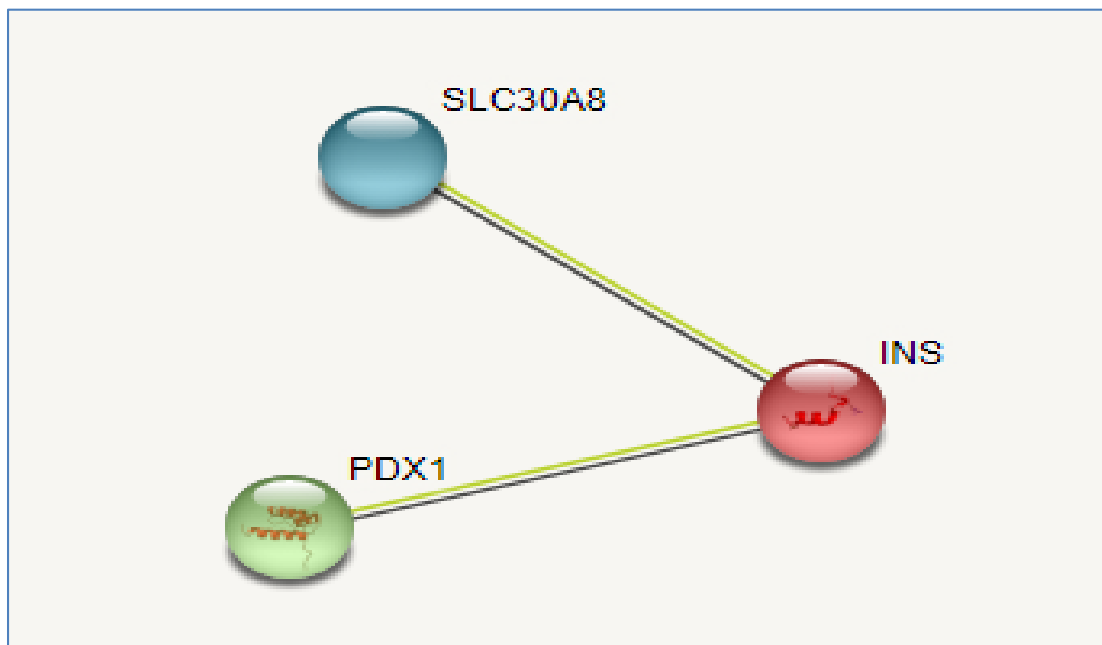


Figure 21 : Résultats d'interaction protéique par String.

Les résultats de STRING signifient que ces protéines ont plus d'interactions entre elles que ce qui serait attendu pour un ensemble aléatoire de protéines de taille similaire, tirées du génome. Un tel enrichissement indique que les protéines sont au moins partiellement liées biologiquement, en tant que groupe.

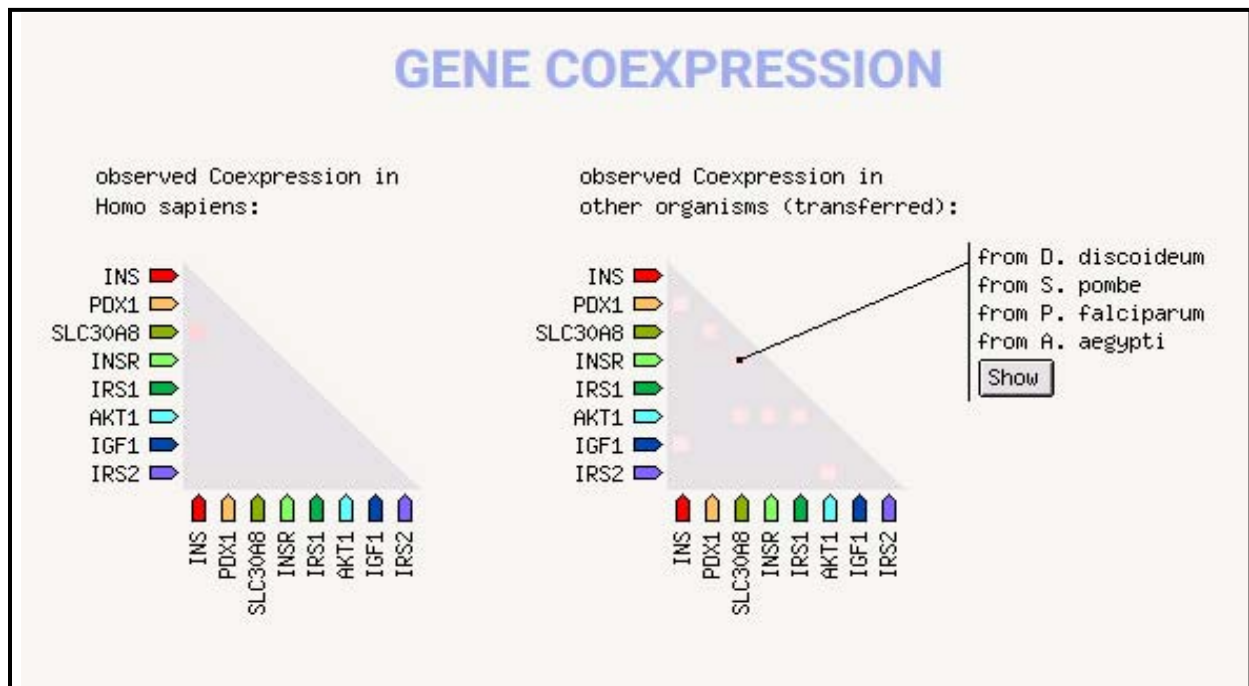


Figure 22 : les gènes de coexpression d'interaction.

La comparaison de la coexpression de nos gènes INS – SLC30A8 – PDX1(*Homo Sapiens*) avec d'autres organismes va donner les résultats suivants :

- Les gènes plus exprimés (SLC30A8 – INS) ont le score « 0 ,92 »
- La coexpression de gènes INS – PDX1 est de score « 0.46 »
- INS – IGF1 sont exprimés avec un score de « 0.51 »
- Le récepteur de l'insuline INSR et AKT1 ont le score « 0.54 ». (**ANNEXE III**)

RESULTATS

III.1.6-Outil Glaxy

1. Analyse FastQC

Nous avons récupéré les deux fichiers SRA RNA-seq des cellules β pancréatiques et avons pu traiter sous Galaxy.

Dans cette partie nous verrons comment contrôler la qualité des données de séquençage afin d'identifier les transcrits. Nous traitons ces résultats de manière comparative avec deux jeux de données.

Tableau IV : Les données statistique de RNA-seq

Measure	Avant le traitement	Après le traitement
File name	EBI_SRA__SRX248493_File_ SRR771476_1_fastq_gz.gz	Clip_on_data_10
File type	Conventional base calls	Conventional base calls
Encoding	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Total Sequences	16756990	976592
Sequences flagged as poor quality	0	0
Sequence length	101	15-96
%GC	44	45

File name	EBI_SRA__SRX248493_File_SRR771476_2_fastq_gz_gz	Trim_sequences_on_data_11
File type	Conventional base calls	Conventional base calls
Encoding	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Total Sequences	16756990	3997540
Sequences flagged as poor quality	0	0
Sequencelength	101	15-91
%GC	43	45

Cette graphe représente le score de qualité par base nous voyons trois zone de déférents couleur :

La zone vert : correspond ou bonne qualité.

La zone orange claire : correspond ou moyenne qualité.

La zone rose : correspond ou mauvaise qualité.

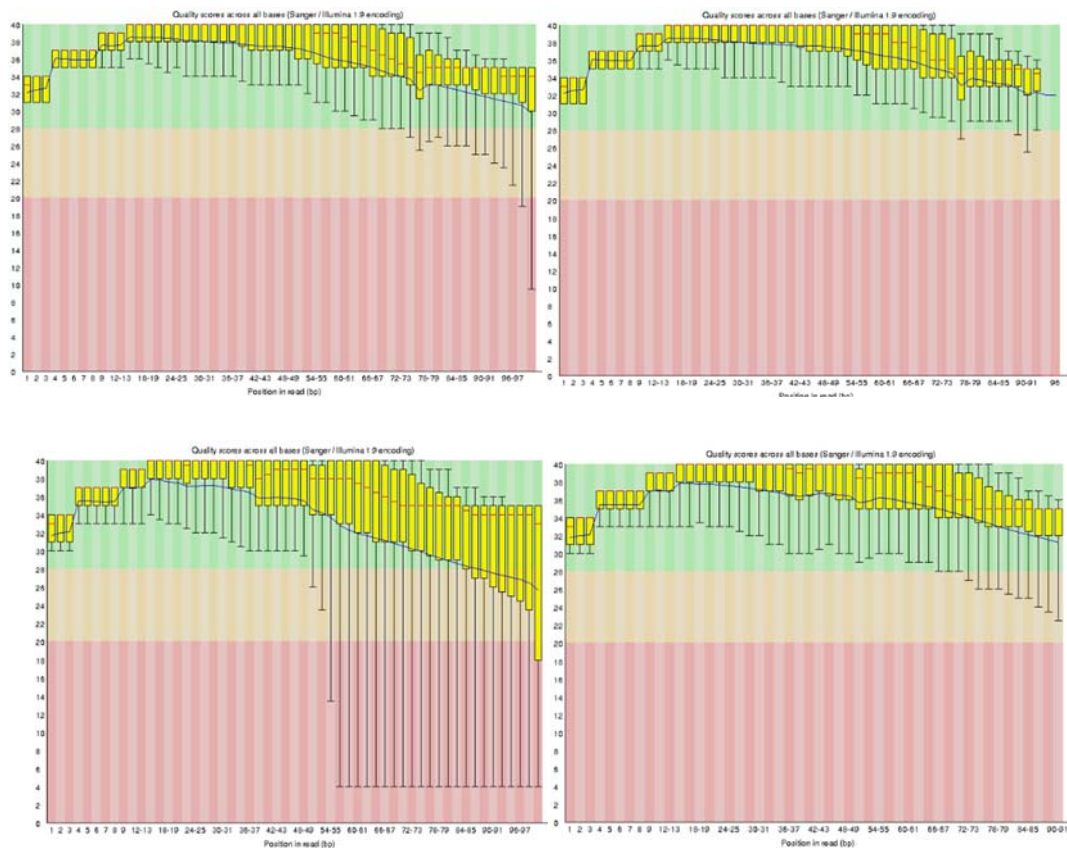


Figure 23: comparaison de la qualité des séquences.

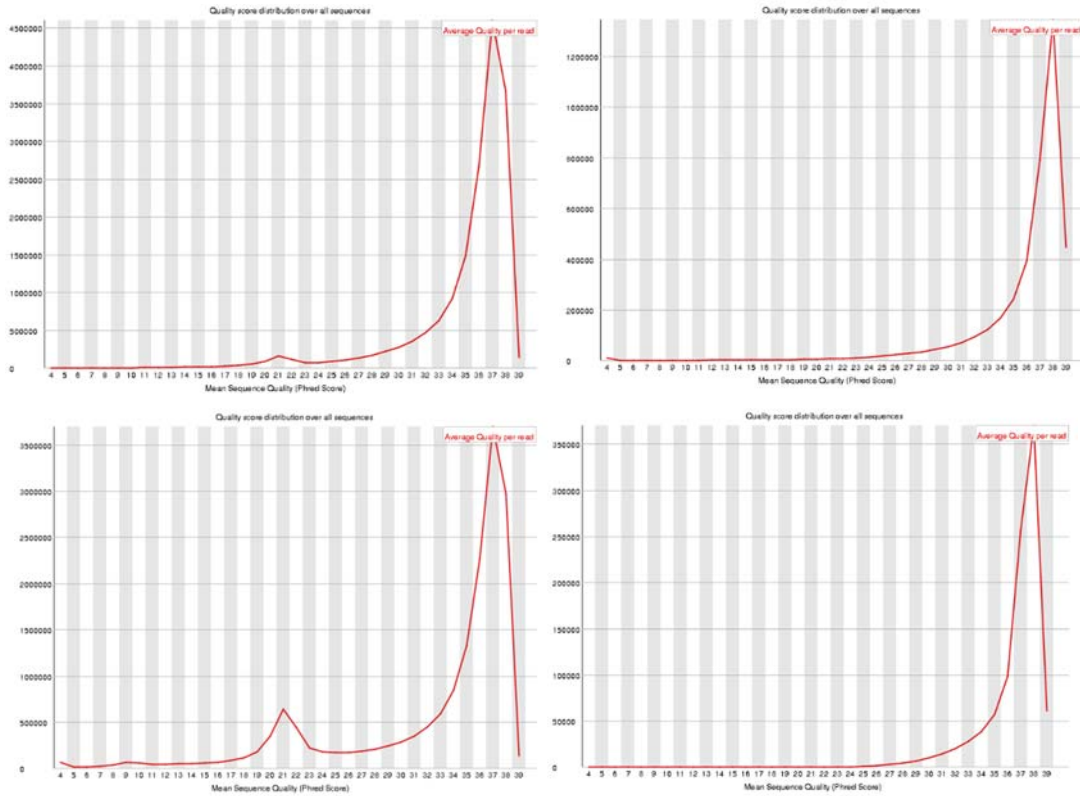


Figure 24 : Comparaison de la distribution de score de qualité par séquences.

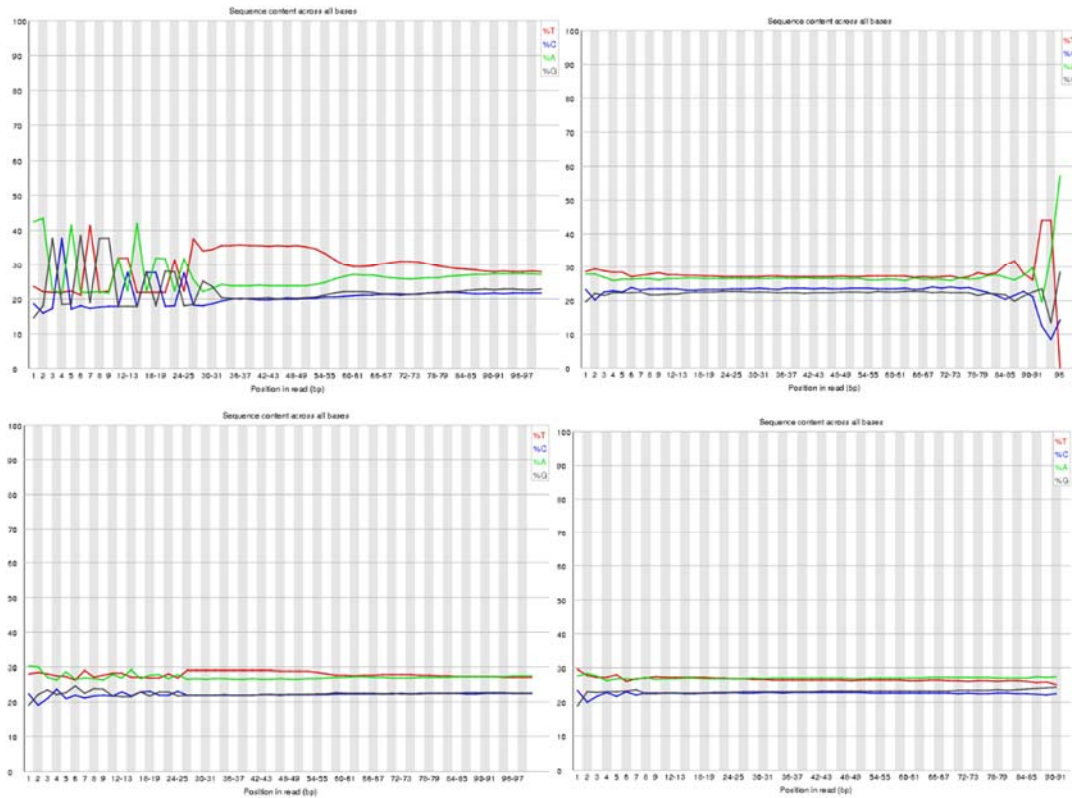


Figure 25 : Représentation graphique le contenu de séquence par base.

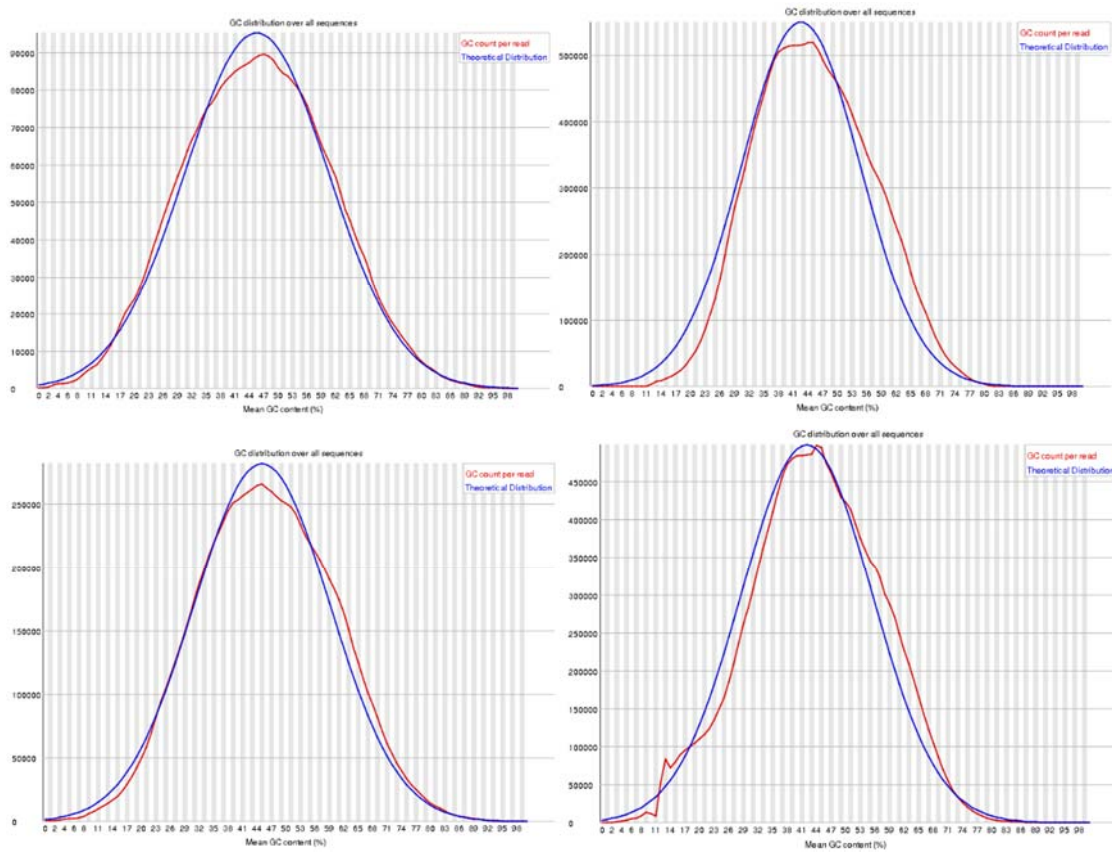


Figure 26 : Le nombre de GC contenu par séquence.

On remarque que le contenu en N << 5% à une position indique une erreur.

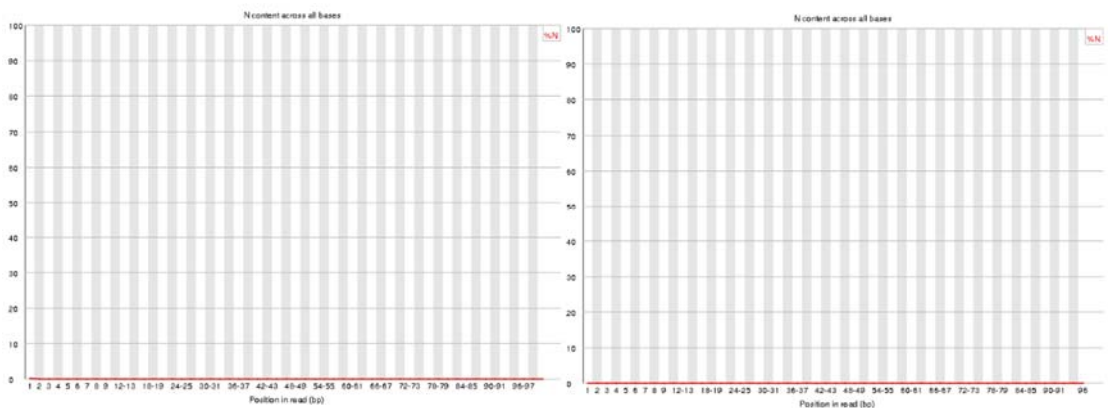


Figure 27 : Le pourcentage des reads.

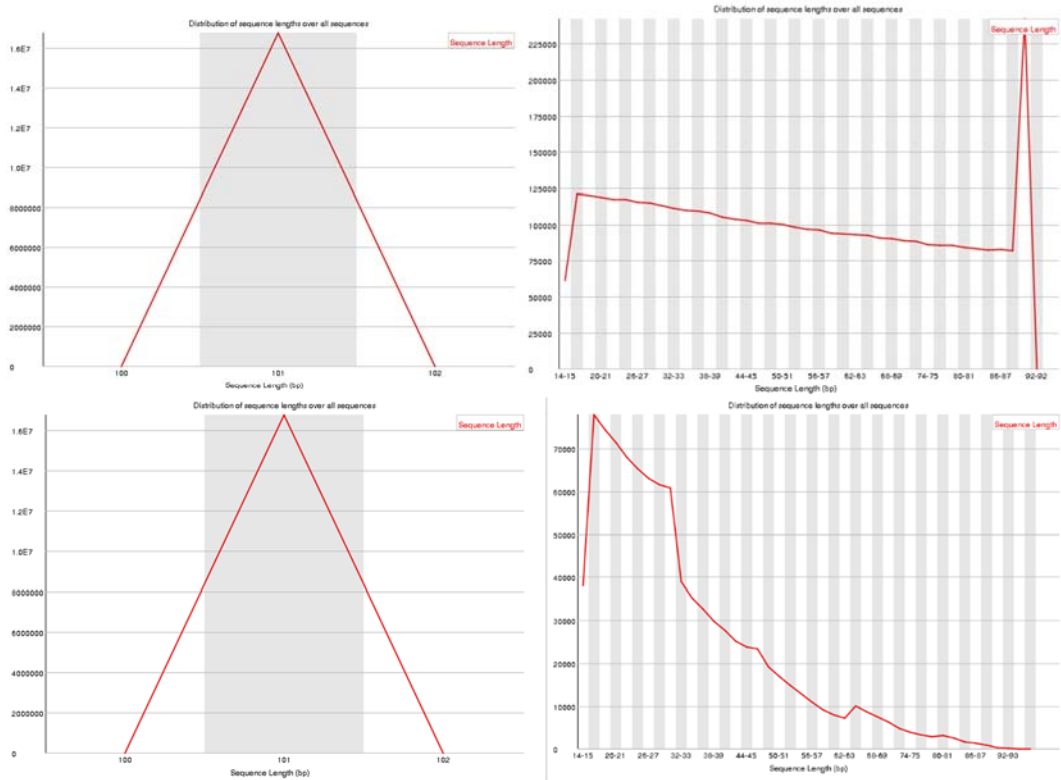


Figure 28 : La distribution de la taille des séquences.

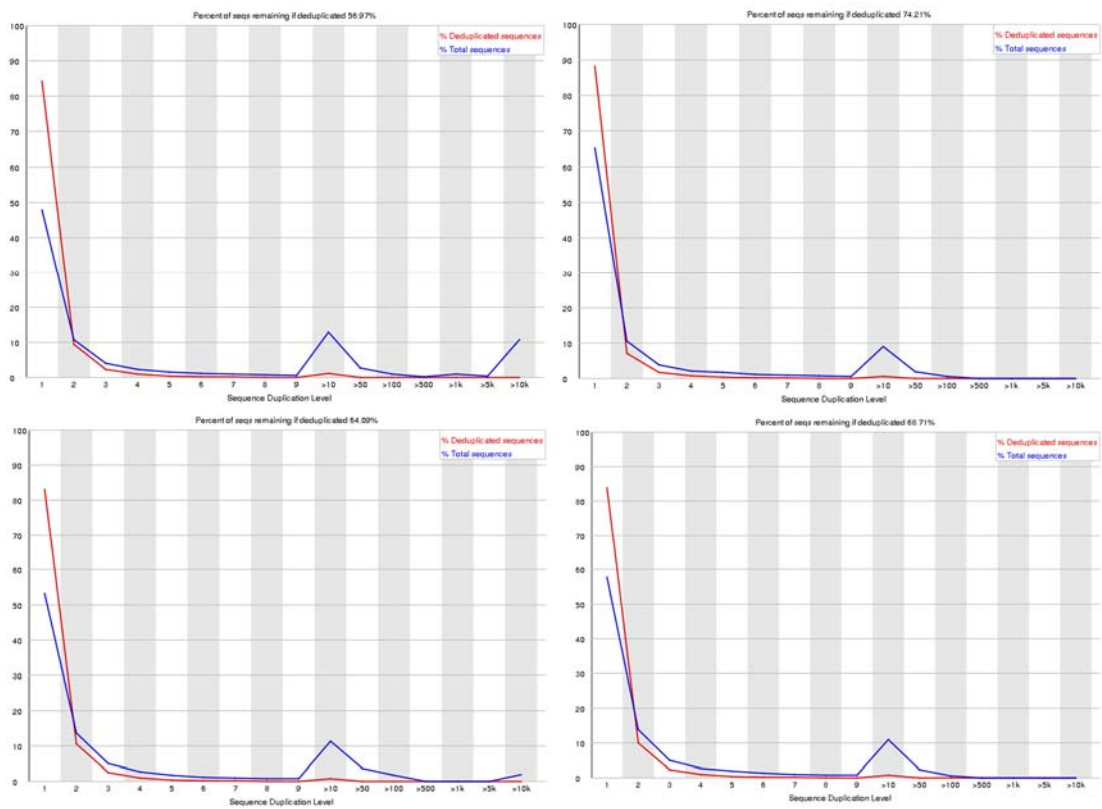


Figure 29 : Le degré de duplication des RNA-seq.

✖ Overrepresented sequences

Sequence	Count	Percentage	Possible Source
AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTT	332776	1.9858936479642226	No Hit

✖ Overrepresented sequences

Sequence	Count	Percentage	Possible Source
AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTT	1762736	10.51940712502663	No Hit

✔ Overrepresented sequences

No overrepresented sequences

Figure 30 : Les séquences contaminer des reads.

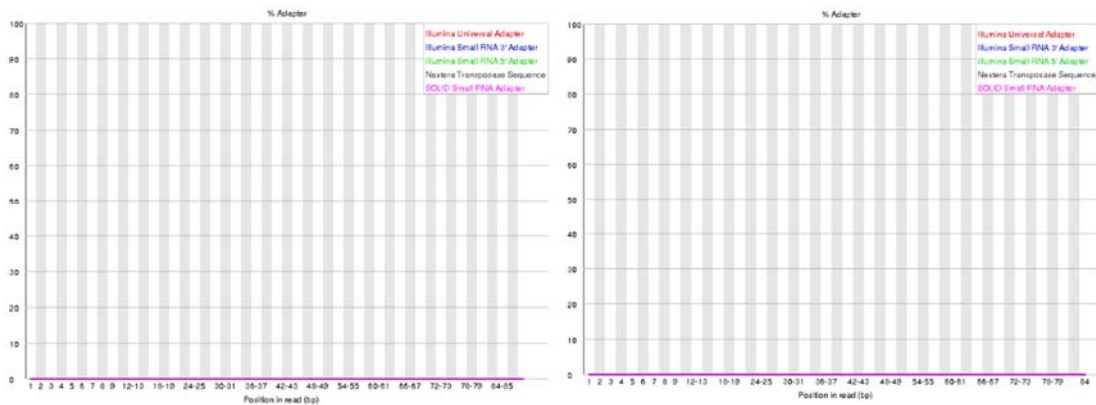


Figure 31 : Le nombre des adaptateurs dans les séquences.

2. Alignement avec TopHat

```

Left reads:
  Input      :    976592
  Mapped     :   145244 (14.9% of input)
  of these:   54986 (37.9%) have multiple alignments (2955 have >20)
Right reads:
  Input      :    976592
  Mapped     :   240791 (24.7% of input)
  of these:  105734 (43.9%) have multiple alignments (2966 have >20)
Unpaired reads:
  Input      :   3020948
  Mapped     :   447860 (14.8% of input)
  of these:  138453 (30.9%) have multiple alignments (28 have >20)
16.8% overall read mapping rate.

Aligned pairs:    35830
  of these:       21276 (59.4%) have multiple alignments
                  35468 (99.0%) are discordant alignments
0.0% concordant pair alignment rate.
    
```

Figure 32 : Résultats d'alignement.

Nous avons obtenu quatre tableaux :

Tableau V : Résultats des insertions des séquences.

Chrom	Start	End	Name	Score
trackname=insertions description="TopHat insertions"				
NC_000011.10	80179	80179	A	1
NC_000011.10	81983	81983	AA	1
NC_000011.10	157042	157042	AA	3
NC_000011.10	161399	161399	A	1
NC_000011.10	168966	168966	C	1
NC_000011.10	214419	214419	G	1
NC_000011.10	230979	230979	CA	1
NC_000011.10	245073	245073	C	1
NC_000011.10	264438	264438	T	21
NC_000011.10	277849	277849	C	2
NC_000011.10	301598	301598	A	2
NC_000011.10	403615	403615	G	1
NC_000011.10	463667	463667	C	1
NC_000011.10	465911	465911	G	1
NC_000011.10	482079	482079	G	1
NC_000011.10	502182	502182	C	1
NC_000011.10	592047	592047	G	1
NC_000011.10	600962	600962	T	1
NC_000011.10	643327	643327	A	5
NC_000011.10	643695	643695	G	1
NC_000011.10	670800	670800	A	1
NC_000011.10	671103	671103	C	1
NC_000011.10	715795	715795	T	1
NC_000011.10	731001	731001	C	1
NC_000011.10	731875	731875	G	1

Tableau VI : Les délétions des séquences.

Chrom	Start	End	Score
track name=deletions description="TopHat deletions"			
NC_000011.10	62344	62345	1
NC_000011.10	163856	163858	1
NC_000011.10	243162	243163	10
NC_000011.10	301594	301595	2
NC_000011.10	304880	304881	2

NC_000011.10	335234	335236	1
NC_000011.10	361493	361494	3
NC_000011.10	361586	361588	4
NC_000011.10	389089	389090	1
NC_000011.10	389091	389092	1
NC_000011.10	403261	403262	1
NC_000011.10	422831	422832	1
NC_000011.10	482261	482262	1
NC_000011.10	516288	516289	1
NC_000011.10	543872	543874	2
NC_000011.10	545112	545114	8
NC_000011.10	563670	563671	5
NC_000011.10	565246	565247	1
NC_000011.10	582655	582656	2
NC_000011.10	585820	585821	1
NC_000011.10	600857	600859	1
NC_000011.10	601069	601070	1
NC_000011.10	603763	603764	23

Tableau VII : les jonctions d'épissage.

Chrom	Start	End	Name	Score	Strand	Thick Start	Thick End	Item RGB	Block Count	Block Sizes	Block Starts
track name=junctions description="TopHat junctions"											
NC_000011.10	127853	129479	JUNC0000001	1	-	127853	129479	255,0,0	2	41,20	0,1606
NC_000011.10	290574	290698	JUNC0000002	1	+	290574	290698	255,0,0	2	26,21	0,103
NC_000011.10	293667	293853	JUNC0000003	22	+	293667	293853	255,0,0	2	60,24	0,162
NC_000011.10	314310	314969	JUNC0000004	7	+	314310	314969	255,0,0	2	46,48	0,611
NC_000011.10	399172	400028	JUNC0000005	6	+	399172	400028	255,0,0	2	24,62	0,794
NC_000011.10	404085	404281	JUNC0000006	8	+	404085	404281	255,0,0	2	50,46	0,150

NC_00001 1.10	406516	406900	JUNC0 000000 7	11	-	406516	406900	255,0,0	2	22,58	0,326
NC_00001 1.10	495029	498022	JUNC0 000000 8	13	-	495029	498022	255,0,0	2	24,52	0,2941
NC_00001 1.10	560998	561798	JUNC0 000000 9	1	+	560998	561798	255,0,0	2	35,37	0,763

Tableau VIII : score d'alignements.

NAME	FLAG	RNAME	POS	MAPQ	CIGAR	MRNM	MPOS	SIZE	SEQ
SRR771476. 1491337	329	NC_000 011.10	60100	0	15M	*	0	0	CTTGCTCATAGC AAA
SRR771476. 2772002	329	NC_000 011.10	60343	0	16M	*	0	0	AAAATGGTGCA TGAGT
SRR771476. 3171920	393	NC_000 011.10	60366	0	16M	*	0	0	GTGTGAAAAAA AACAA
SRR771476. 4399502	393	NC_000 011.10	60409	0	16M	*	0	0	ATCTAACATTCA AACAA
SRR771476. 1408518	329	NC_000 011.10	60409	0	16M	*	0	0	ATCTAACATTCA AACAA
SRR771476. 4178600	16	NC_000 011.10	60534	3	17M	*	0	0	GTGATTCATTTT TATTT
SRR771476. 10127239	409	NC_000 011.10	60550	0	16M	*	0	0	TTTCACCAAAT CTTTT
SRR771476. 379105	409	NC_000 011.10	60645	0	15M	*	0	0	TATTCCCTGGGC ATT
SRR771476. 863756	409	NC_000 011.10	60645	0	15M	*	0	0	TATTCCCTGGGC ATT
SRR771476. 2862962	433	NC_000 011.10	60645	0	15M	=	619646 78	61904 074	TATTCCCTGGGC ATT
SRR771476. 3412637	401	NC_000 011.10	60645	0	15M	=	728192 31	72758 602	TATTCCCTGGGC ATT
SRR771476. 3412637	401	NC_000 011.10	60645	0	15M	=	403542 25	40293 596	TATTCCCTGGGC ATT
SRR771476. 3602437	409	NC_000 011.10	60645	0	15M	*	0	0	TATTCCCTGGGC ATT

3. Cufflinks (voir l'annexe IV)

4. Cuffdiff (voir l'annexe IV)

5. Cuffcompare (voir l'annexe IV)

6. IGB (Integrated Genome Browser)

Dans notre cas l'IGB n'a trouvé aucune reads.

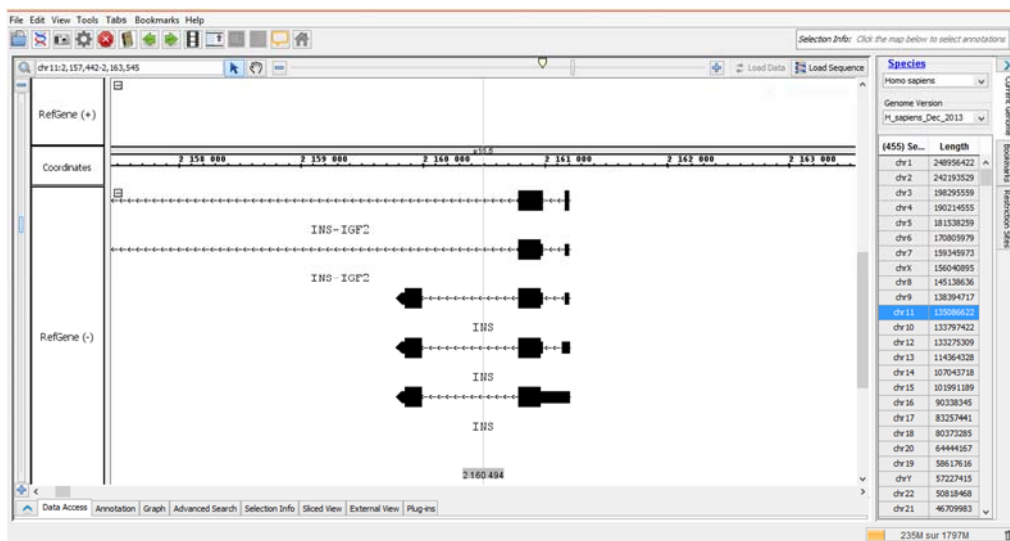
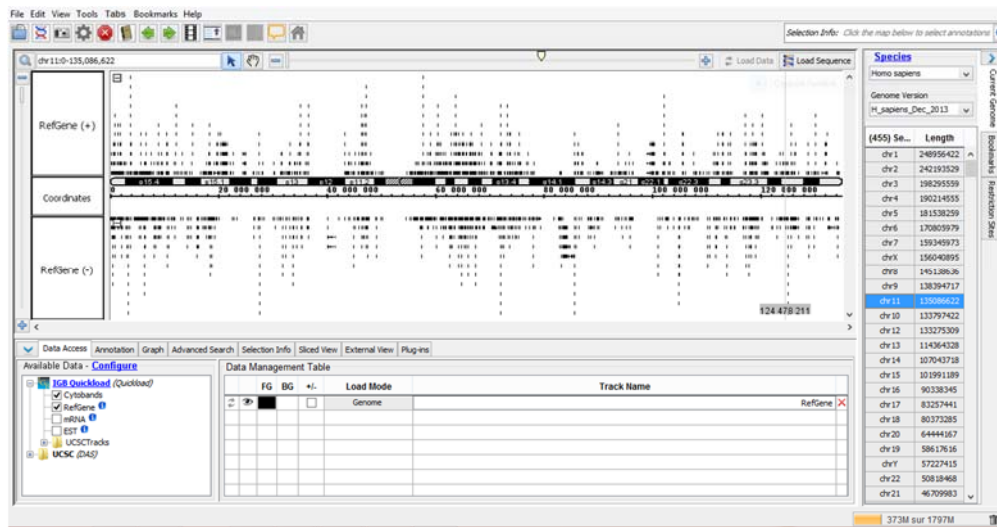


Figure 33 : visualisation des transcrits analysés.

Discussion

Dans l'évolution du diabète type 1 (DT1), les cellules β pancréatiques sont soumises à un processus auto-immun par infiltration de cellules immunitaires et la production d'autoanticorps. L'infiltration par les lymphocytes T au niveau des îlots de Langerhans va conduire à une apoptose puis une insulite, et à des complications du DT1 (**Vasumathi, Klaus (2014)**).

Lnc RNA est une nouvelle classe de transcrits qui participe au développement de plusieurs maladies. Seulement quelques unes sont connues pour leur rôle dans la sécrétion de l'insuline par les cellules β pancréatiques.

Notre étude a pour but d'évaluer la contribution des LncRNAs dans le dysfonctionnement des cellules β pancréatiques par le niveau d'expression des gènes impliqués dans la synthèse de l'insuline par les cellules β pancréatique.

D'après l'étude de (**Mothorie et al.,2015**), les MIN6 contribuent à l'expression d'un nombre très large de Lnc RNAs exprimés dans une lignée cellulaire MIN6 traitée avec des cytokines. Dans la même étude, les changements des taux des LncRNAs sont confirmés chez les îlots de souris et sont augmentés chez les souris NOD pré diabétiques.

L'insuffisance cellulaire β pancréatique conduit au développement du diabète. Au cours de la progression de la maladie, les cellules β adaptent la capacité de sécrétion pour compenser la glycémie élevée et la résistance à l'insuline périphérique. Ce mécanisme compensatoire implique une régulation fine pour moduler la capacité du réticulum endoplasmique (ER) et le contrôle de la qualité pour empêcher l'accumulation de proinsuline dépliée, une protéine majeure synthétisée dans la cellule β . Ces voies de signalisation sont

collectivement appelées réponse à la protéine dépliée (EPU). Les machines UPR sont nécessaires pour préserver l'homéostasie ER et l'intégrité des cellules β . En outre, les acteurs de l'EPU jouent un rôle clé en régulant la capacité de repliement des ER, en augmentant la dégradation des protéines mal repliées et en limitant le taux de traduction de l'ARNm.

Les études génétiques et biochimiques récentes sur les modèles de souris et les mutations des capteurs UPR humains démontrent une exigence claire des mécanismes UPR pour prévenir l'échec des cellules β et augmenter la masse cellulaire β et l'adaptation tout au long de la progression du diabète. Les cellules β pancréatiques ont une capacité marquée à s'adapter aux changements de l'environnement en augmentant sa masse et sa fonction. Les signes diabétiques se produisent lorsque ce mécanisme adaptatif ne parvient pas à compenser l'augmentation de la demande d'insuline. L'activation des acteurs de l'UPR est déclenchée au stade précoce du mécanisme de compensation et peut jouer un rôle central dans l'adaptation des cellules β et les fonctions ultérieures.

Des études menées au cours du développement du diabète sont nécessaires pour mieux comprendre le mécanisme de la capacité compensatoire et la perte de cellules β subséquentes chez les humains. Ceci est particulièrement intéressant, car il pourrait avoir un impact bénéfique pour le traitement des maladies métaboliques telles que le diabète (**Rabhi et al., 2014**).

Le facteur de transcription PDX1 joue un rôle important dans le développement des cellules β pancréatiques et dans la fonction sécrétoire. Il est important de noter que Pdx-1 est un régulateur transcriptionnel du gène de l'insuline ainsi que d'autres gènes clés impliqués dans le couplage stimulation-sécrétion (**Kulkarni et al., 2004**). Pdx-1 est également reconnu pour jouer un rôle dans l'adaptation de la cellule β au stress métabolique, et la haplo insuffisance Pdx-1 superposée à un fond de résistance à l'insuline sévère entraîne une diminution de l'expansion de la masse de cellules β compensatoires, du diabète et de la mortalité prématurée. **Sachdeva et al., 2009** ont récemment examiné les effets de l'obésité induite par le régime chez les souris Pdx-1 +/- et ont également constaté que la déficience en Pdx-1 limite de manière significative l'expansion de la masse cellulaire β dans les conditions de la HFD. Il est intéressant de noter que ces effets n'étaient pas secondaires à une diminution de la prolifération, mais plutôt à une diminution de la survie des cellules β . De même, l'apoptose des cellules β augmentée dans les îlots Pdx-1 +/- a également été observée par **Johnson et al., 2003**.

Les données récentes impliquent également le Pdx-1 dans le maintien de la santé du réticulum endoplasmique (ER). La pompe Ca^{2+} + ATPase 2b (SERCA2b) du réticulum sarcoplasmique maintient un gradient de Ca^{2+} + raide entre le cytosol et la lumière de l'urée. Dans les modèles de diabète, nos données ont démontré la perte de la Pdx-1 de la cellule β pancréatique qui se produit parallèlement à l'expression SERCA2b altérée, alors que l'analyse *in silico* du promoteur SERCA2b a révélé de multiples sites de liaison Pdx-1 putatifs. Nous avons émis l'hypothèse que la perte de Pdx-1 dans des conditions inflammatoires et diabétiques entraîne une diminution des niveaux de SERCA2b.

Le transporteur de zinc 8 (ZnT8) est un auto anticorps spécifiquement exprimé dans des granules sécrétoires contenant l'insuline des cellules β . Les études génétiques montrent que les génotypes de SLC30A8 peuvent déterminer une réponse protectrice ou diabétogène selon les facteurs environnementaux et de style de vie. L'expression de la protéine ZnT8, ainsi que la teneur en zinc dans les cellules β , est diminuée chez les souris diabétiques. Ainsi, le ZnT8 pourrait participer à la biosynthèse et à la libération d'insuline, et ensuite impliqué une détérioration de la fonction des cellules β par des mécanismes directs ou indirects dans le diabète de type 1. D'un point de vue clinique, la prévalence du ZnT8 est graduellement accrue le diabète auto-immune latent chez les adultes et le diabète sucré de type 1. La fréquence et les épitopes des lymphocytes T spécifiques de ZnT8 et la libération de cytokines par les lymphocytes T sont également différents chez les patients diabétiques et les témoins sains. En outre, la réponse à l'administration de ZnT8 est également différente dans le diabète sucré de type 1 et le diabète sucré de type 2 (Bo Yi et al.,2016).

Les différences sont principalement concentrées dans le rôle différent des génotypes de SLC30A8 dans la susceptibilité au diabète de type 1, aux différents résultats des interactions génétique-environnement et à la réponse différente à l'administration de ZnT8. Cependant, l'auto-immunité contre le diabète contre le ZnT8 et le rôle multifonctionnel de ZnT8 dans la fonction des cellules β montrent le lien mécanique potentiel entre le diabète sucré de type 1 et 2, les deux formes majeures du spectre du diabète. Des efforts sont nécessaires pour examiner le rôle et les fréquences des lymphocytes T spécifiques de ZnT8 dans LADA et le diabète sucré de type 2. L'association entre le génotype SLC30A8 et le risque de LADA doit également être démontrée. Bien que ZnT8 soit un antigène diabétogène qui peut participer au diabète, les mécanismes détaillés de la destruction des cellules β induites par le ZnT8 n'est

pas claire. Il est nécessaire de confirmer la fonction de régulation du ZnT8 dans l'homéostasie de l'insuline chez les souris NOD et chez les humains. Il est également nécessaire d'explorer les profils d'expression de cytokines produites par les lymphocytes T spécifiques de ZnT8 parmi différents types de diabète.

Conclusion

Le diabète type 1 constitue l'une des formes de diabète le plus fréquent chez les jeunes (diabète juvénile). Il se caractérise par la destruction des cellules β pancréatiques celles qui sont responsable de la production d'insuline.

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. Le défi majeur à l'heure actuelle reste la caractérisation fonctionnelle de la grande majorité des lncRNAs. Toutefois, l'identification et la cartographie de la totalité des lncRNAs présents dans le génome humain devraient accélérer l'atteinte de cet objectif. Cela implique des efforts conjoints entre la génomique fonctionnelle, l'épigénomique et la bioinformatique. Nous sommes à présent sûrs de leur importance dans de nombreux processus cellulaires et des pathologies, parmi eux le diabète.

En effet, de nombreuses études ont prouvé la corrélation entre l'expression des Lnc RNAs et le diabète type I. Dans notre travail, nous avons démontré sur le profil moléculaire les interactions entre PDX1 et ZnT8 par rapport à la molécule d'insuline et on a trouvé que le ZnT8 fait intervenir dans la fixation de la molécule d'insuline dans les cellules β pancréatiques (Sans Zinc pas d'insuline), et que le PDX1 joue un rôle important dans le développement des cellules β pancréatiques ainsi que dans la régulation transcriptionnelle du gène de l'insuline. Aussi, l'identification des LncRNAs qui sont exprimés différemment pour le but de rechercher les mutations dans leurs séquences qu'ils sont représentés des variantes génétiques. Nous avons utilisé différents logiciels d'annotation bioinformatique pour peu expliquer leurs conséquences sur les fonctions et la régulation clés de l'expression génique. Ces molécules LncRNAs restent encore non caractérisées jusqu'à ce jour.

L'importance de l'identification et l'analyse fonctionnelle de ces Lnc RNAs est de contribuer à une meilleure définition de leur rôle durant la destruction de cellules β pancréatiques. De plus, Ces molécules pourraient avoir des applications : comme être utilisées comme biomarqueurs pour le diagnostic et/ou le pronostic des pathologies avec le changement de leur expression associées aux différents types de diabète, ainsi que le développement de nouvelles stratégies thérapeutiques.

Références bibliographiques :

- Abel Y, Clerget G, Bourguignon-Igel V (2014). Les petits ARN nucléolaires nous surprennent encore ! *Med Sci (Paris)* ; 30 : 297-302.
- Amit I , Guttman M, 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.
- Anko M-L, Neugebauer KM (2010). Long noncoding RNAs Add Another Layer to pre-mRNA Splicing Regulation. *Molecular Cell* 39 :833-834.
- Batista P, Chang H (2013). Long non coding RNAs : Cellular address codes in development and disease. *Cell* 152,1298-1307.
- Boitard C, Efendic S, Ferrannini E, Henquin J, Donald F (2002). Steiner, and Erol Cerasi "A tale of two cousins: type 1 and type 2 diabetes." *Diabetes*, p. S1+.
- Bramswig, N. C., Everett, L. J., Schug, J., Dorrell, C., Liu, C., Luo, Y (2013). Epigenomic plasticity enables human pancreatic alpha to β cell reprogramming. *J. Clin. Invest.* 123,1275–1284. doi: 10.1172/JCI66514.
- Buysschaert m., vandeleene b. parus l., Hermans MP (1999). Le diabète sucré d'une réalité d'aujourd'hui à un défi de demain lonvain *med.* 118 : s 195.
- Cabili M, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611.
- Costa FF (2007) Non-coding RNAs : Lost in translation ? *Gene* 386 :1-10.
- Couée I, Fontaine-Poitou L, Guillaume V (2010). *Biologie et physiologie cellulaires et moléculaires.* Editions De Boeck Université, Bruxelles.

- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *GenomeRes.*22,1775-1789.doi:10.1101/gr.13215911.
- Djebali S, Davis C. A, Merkel, A, Dobin, A, Lassmann, T, Mortazavi, A, (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature1123
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat.Rev.Genet.*15,721.doi:10.1038/nrg3606.
- Feng Y, Fan Y, Huiqing C, Zicai L, Quan D (2014). The emerging Landscape of long non-coding RNAs.
- Fitzgerald KA, Caffrey DR (2014) Long non coding RNAs in innate and adaptive immunity. *Current opinion in immunology* 26 :140-146.
- Gargaud M, Despois D (2003). *Les traces du vivant: Presses universitaires de Bordeaux.*
- Gibb EA, Brown CJ, Lam WL (2011) the functional role of long non-coding RNA in human carcinomas 10 :38.
- Guttman M, Amit I, Garber M, French C, Lin M (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458,223–227.doi: 10.1038/nature07672.
- Guttman M, Rinn JL (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482 : 339-46.
- Hochheimer A, Tjian R (2003) Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes & Development* 17:1309 –1320. doi:10.1101/gad.1099903.
- Hyttinen V, Kaprio J, Kinnunen, L, Koskenvuo, M., Tuomilehto, J. (2003). Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 52, 1052–1055. doi: 10.2337/diabetes.52.4.1052.

- Johnson D, Ahmed T, Luciani S, Tran H, Fujita J, Misler S, Edlund H, Polonsky S (2003). Increased islet apoptosis in Pdx1+/- mice. *J. Clin. Invest.* 111, 1147–1160.
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta 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.
- Kapusta A, Kronenberg Z, Lynch VJ (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 9 : e1003470.
- Kelley DR, Rinn JL (2012). Transposable elements reveal a stem cell specific class of long noncoding RNAs. *Genome Biol* 13 : R107.
- Kevin S (1999). Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell*, Vol. 98, 1–4.
- Khalil A, Guttman M, Huarte M, Garber M, Raj A (2009). Many human large intergenic noncoding RNAs associate with chromatin modifying complexes and affect gene expression.
- Koerner M, Pauler F, Huang R, Barlow D (2009). The function of non-coding RNAs in genomic imprinting. *Development*; 136 : 1771-83.
- Krebs J, Kuo M., Allis C, and Peterson C (1999). Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes *GenesDev.* 13, 1412–1421.
- Kulkarni N, Jhala U, Winnay J, Krajewski S., Montminy M., Kahn C (2004) PDX-1 haploinsufficiency limits the compensatory islet hyperplasia that occurs in response to insulin resistance. *J. Clin. Invest.* 114, 828–836.
- Kung J, Colognori D, Lee J (2013) long non-coding RNAs : Past, Present, and Future. *Genetics* 193 :651-669.
- Lepoivre C, Belhocine M, Bergon A (2013). Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* 14 : 914.
- Loe-Mie, Y (2012). Contribution bioinformatique à l'analyse du transcriptome humain. Thèse. Université D'AIX Marseille.

- McCarthy M (2010). Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* 363,2339–2350. doi: 10.1056/NEJMra0906948.
- Medici F, Hawa M, Ianari A, Pyke D, Leslie R (1999). Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia* 42,146–150. doi: 10.1007/s001250051132.
- Meslier N, Gagnadoux F, Giraud P(2003). Impaired glucose-insulin metabolism in males with obstructive sleep apnoea syndrome. *Eur Respir J*; 22 (1): 156-160.
- Moran I, Akerman I, van de Bunt M, Xie R., Benazra M, Nammo T, (2012). Human β Cell transcriptome analysis uncovers lncRNAs that are tissue specific,dynamically regulated,and abnormally expressed in type 1 diabetes. *Cell Metab.* 16,435–448. doi: 10.1016/j.cmet.2012.08.01.
- Nica A,Ongen H,Irminger J,Bosco D,Berney T, Antonarakis S (2013). Cell-type, allelic, and genetic signatures in the human pancreatic β cell transcriptome. *Genome Res.* 23,1554–1562. doi: 10.1101/gr.150706.112.
- Ponting C,Oliver P, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136 :629-641.
- Racch D, (2004). Epidémiologie et physiopathologie des complications dégénératives du diabète sucré. *EMC – Endocrinologie.*1(1) : 29-42.
- Raven P, Johnson G, Mason K, Losos J, 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 J, Chang H(2012). Genome regulation by long noncoding RNAs. *Annual review of biochemistry* 81 :145-166.
- Sachdeva M, Claiborn K, Khoo C, Yang J, Groff D, Mirmira R, Stoffers D (2009). Pdx1 (MODY4) regulates pancreatic β cell susceptibility to ER stress. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19090–19095.
- Saunders A, Core L, Lis J (2006). Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7: 557–567. doi:10.1038/nrm1981.

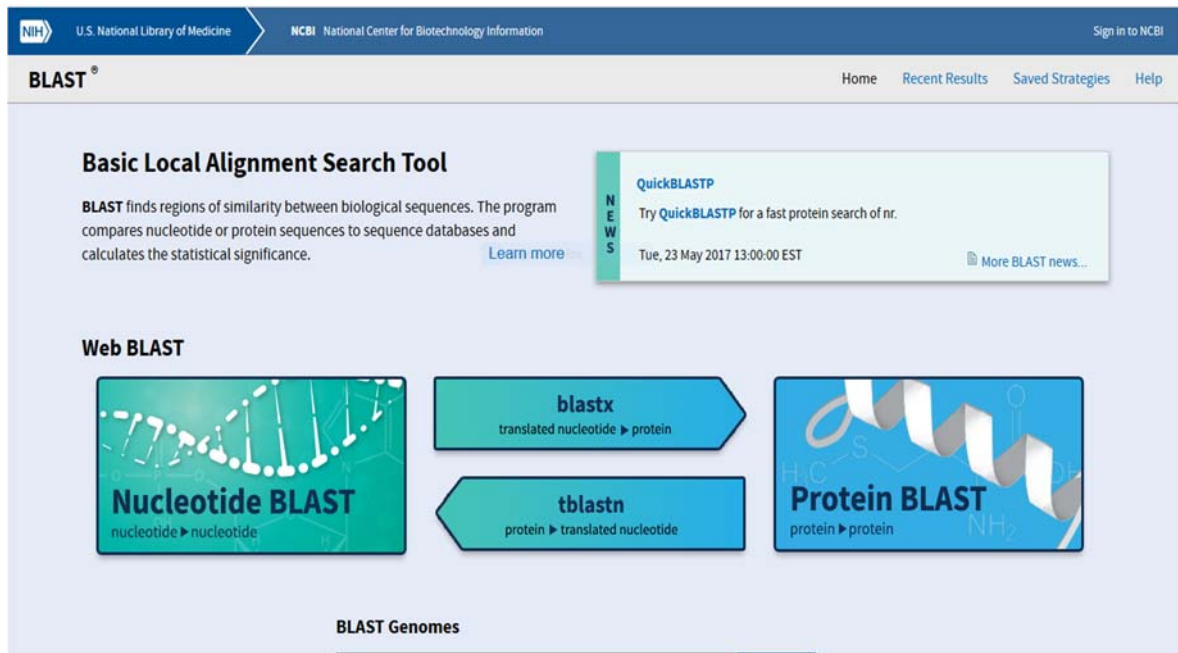
- Scharfmann, R., Pechberty, S., Hazhouz, Y., von Bulow, M., Bricout-Neveu, E., Grenier-Godard, M., et al. (2014). Development of a conditionally immortalized human pancreatic β cell line. *J. Clin. Invest.* 124, 2087–2098. doi: 10.1172/JCI72674.
- Ulitsky I, Bartel D (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* ; 154 : 26-46.
- Vasumathi K, Klaus H (2014). The Missing Lnc(RNA) between the pancreatic β -cell and diabetes .
- Wang X, Song X, Glass C, Rosenfeld M (2011). The long arm of long noncoding RNAs : roles as sensors regulating gene transcriptional programs. *Cold Spring Harbor perspectives in biology* 3:a003756.
- Yang Z, Zhou L, Wu L, Xie H, Zhang F, Zheng S (2011). Over expression of long non- coding RNA Hotair predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation . *Annals of surgical oncology* 18 :1243-1250.
- YI B, Huang G, Zhou Z (2016). Different role of Zinc transporter 8 between type1 diabetes mellitus and type 2 diabetes mellitus 7(4): 459–465.

ANNEXE I

BLAST

C'est une base de données qui permet d'aligner une séquence et la recherche de similarité entre les séquences.

<https://blast.ncbi.nlm.nih.gov/>



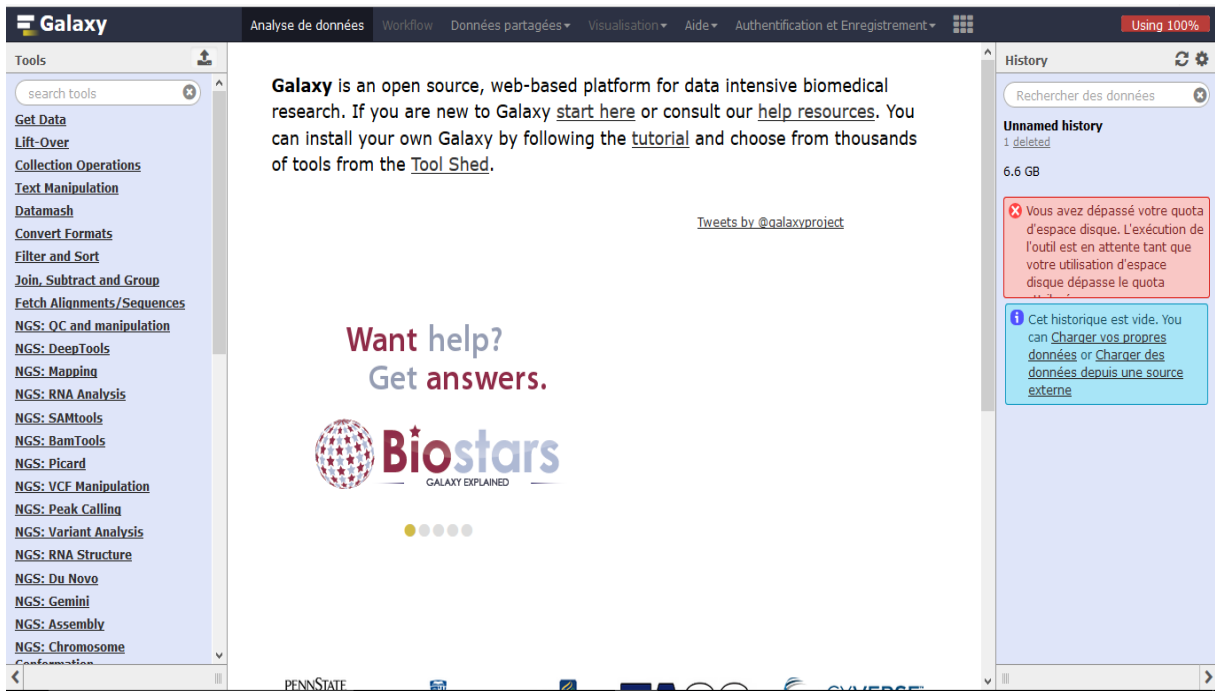
The image shows the homepage of the BLAST (Basic Local Alignment Search Tool) website. At the top, there is a navigation bar with the NIH logo, the text "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information", and a "Sign in to NCBI" link. Below the navigation bar, the word "BLAST" is prominently displayed. To the right of "BLAST" are links for "Home", "Recent Results", "Saved Strategies", and "Help". The main content area features a "Basic Local Alignment Search Tool" section with a brief description of BLAST's function and a "Learn more" link. To the right of this section is a "QuickBLAST" announcement box with a "NEWS" icon, stating "Try QuickBLASTP for a fast protein search of nr." and dated "Tue, 23 May 2017 13:00:00 EST". Below the "Basic Local Alignment Search Tool" section is a "Web BLAST" section with three main options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "tblastn" (protein to translated nucleotide). To the right of these options is a "Protein BLAST" section (protein to protein). At the bottom of the page, there is a "BLAST Genomes" section.

page d'accueil de BLAST.

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 telles 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 public ou notre propre instance.

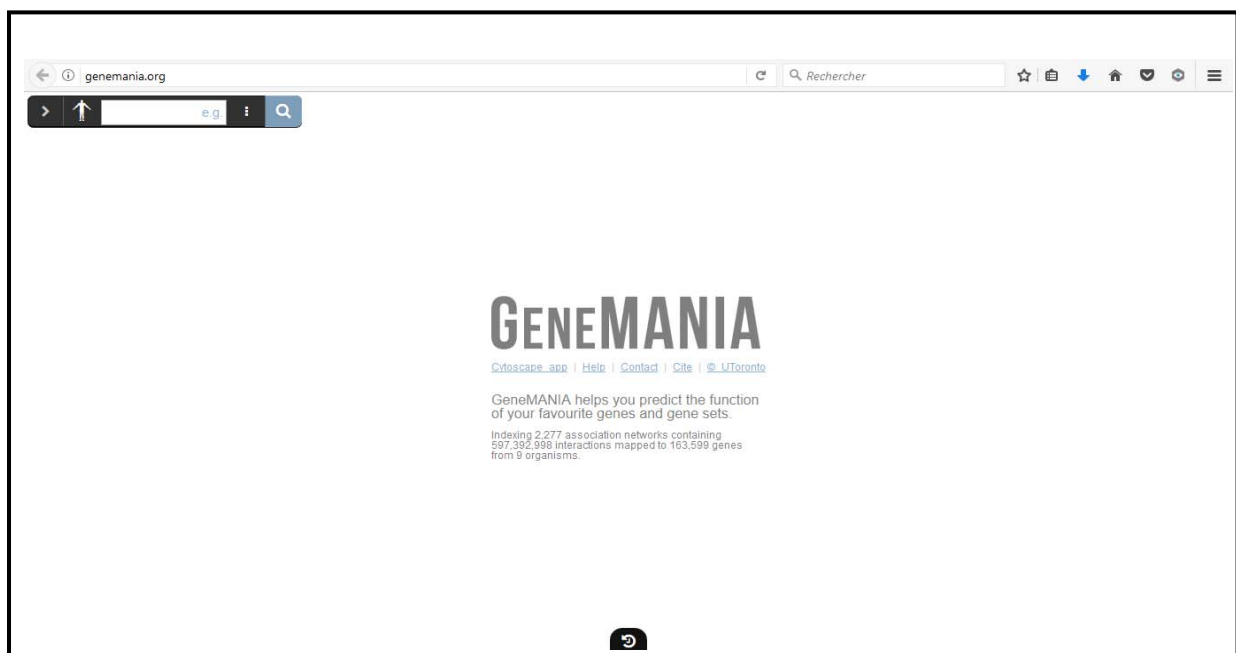
Il existe différentes approches pour commencer à manipuler avec GALAXY, son interface est assez simple ce qui facilite son utilisation, mais il existe aussi des pages spéciales contenant des tutoriels pour chaque outil (Wikigalaxy : <https://wiki.galaxyproject.org/>).



plateforme de travail GALAXY.

GeneMANIA

GeneMANIA représente multiples Association Réseau algorithme d'intégration, permet un scientifique sans aucune formation préalable en informatique bio-informatique de saisir un seul gène et de voir les réseaux d'interaction associée avec le gène.



page d'accueil de GeneMANIA.

NCBI (National Center for Biotechnology Information)

NCBI conduit des recherches dans la biologie informatique, développe des logiciels pour analyser des données de génome et fournir des informations biomédicales.

<https://www.ncbi.nlm.nih.gov/>

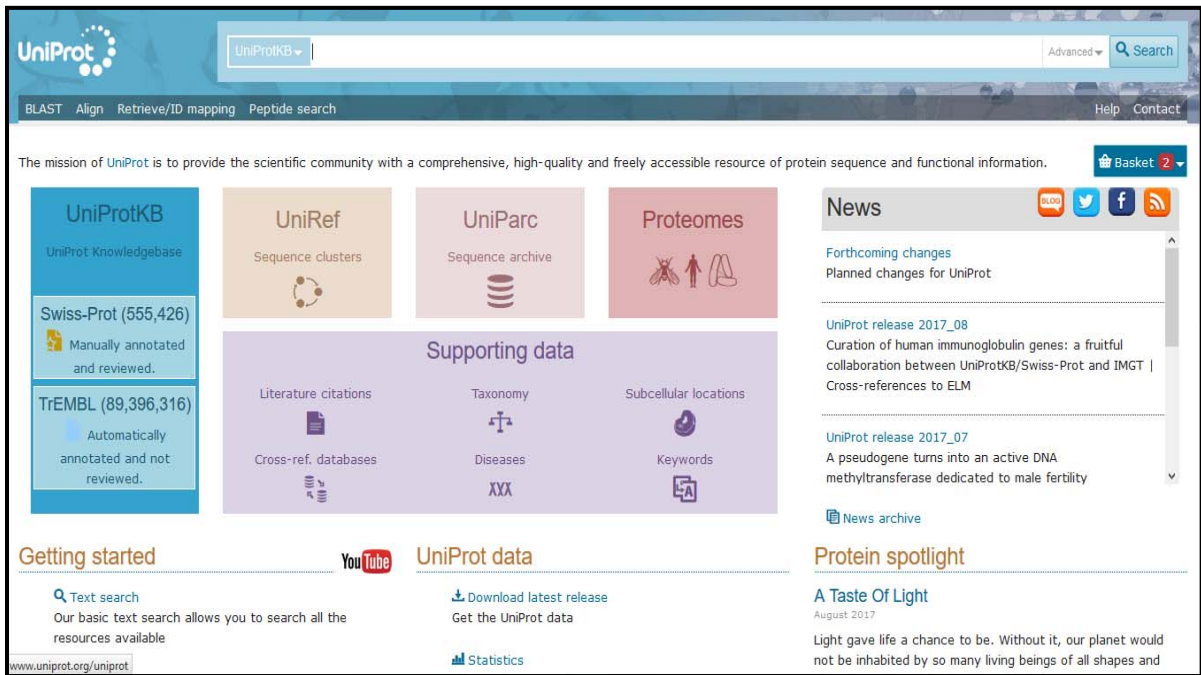
The screenshot shows the NCBI homepage with a blue header containing the logo, navigation links (Resources, How To), and a search bar. A left sidebar lists various biological categories. The main content area features a 'Welcome to NCBI' message and six service tiles: Submit (Deposit data or manuscripts into NCBI databases), Download (Transfer NCBI data to your computer), Learn (Find help documents, attend a class or watch a tutorial), Develop (Use NCBI APIs and code libraries to build applications), Analyze (Identify an NCBI tool for your data analysis task), and Research (Explore NCBI research and collaborative projects). A right sidebar lists 'Popular Resources' (PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and 'NCBI News & Blog' with recent updates.

Page d'accueil de NCBI.

UniProt (Universal Protein Ressource)

UniProt est une base de données généraliste protéique a 3 composants **UniProtKB** (*UniProt Knowledge Base*), UniRef (*UniProt Reference Clusters*), UniParc (*UniProt Archive*) C'est une base de données ouverte, stable et accessible en ligne, elle est issue de la consolidation de l'ensemble des données produites par la communauté scientifique.

[\(https://uniprot.org/\)](https://uniprot.org/)



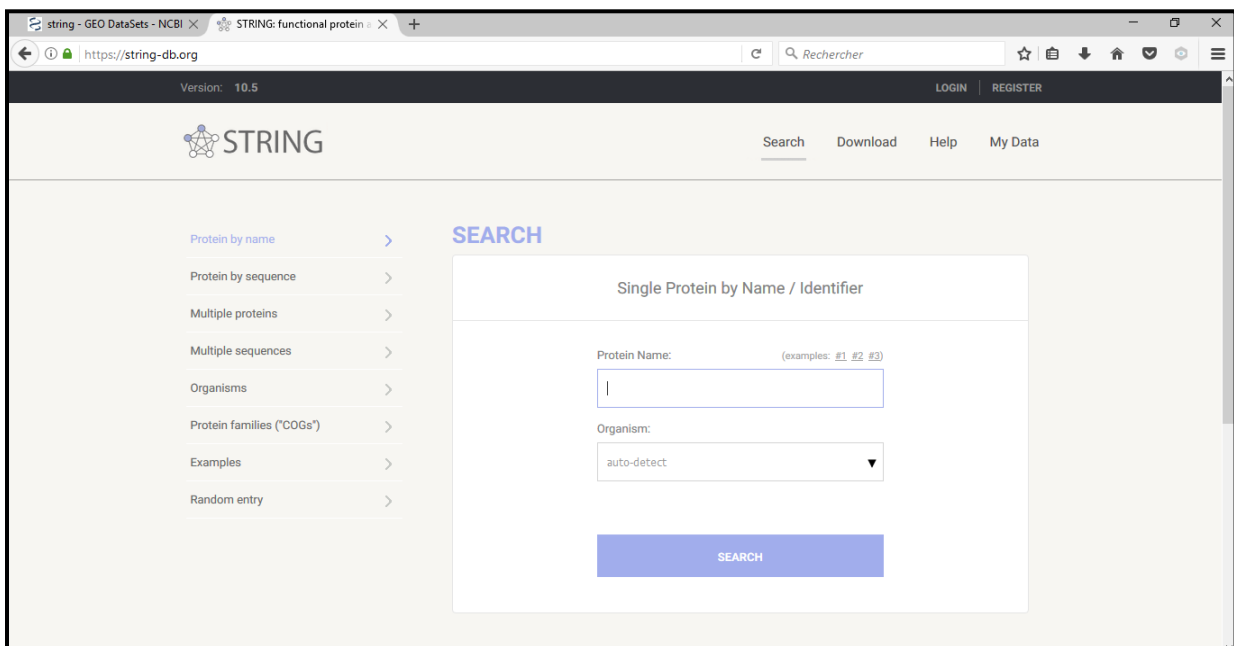
The screenshot shows the UniProt homepage. At the top, there is a navigation bar with the UniProt logo, a search bar containing 'UniProtKB', and a search button. Below the navigation bar, there are several main sections: UniProtKB (Swiss-Prot and TrEMBL), UniRef (Sequence clusters), UniParc (Sequence archive), and Proteomes. A 'Supporting data' section includes links for Literature citations, Taxonomy, Subcellular locations, Cross-ref. databases, Diseases, and Keywords. A 'News' section on the right features articles about forthcoming changes and UniProt releases. At the bottom, there are sections for 'Getting started' (Text search), 'UniProt data' (Download latest release, Get the UniProt data, Statistics), and 'Protein spotlight' (A Taste Of Light).

page d'accueil d'UniProt.

STRING

La base de données **STRING** vise à fournir une évaluation critique et une intégration des interactions protéines-protéines.

<http://string-db.org>



The screenshot shows the STRING database search page. The browser address bar displays 'https://string-db.org'. The page header includes the STRING logo, a search bar, and navigation links for 'Search', 'Download', 'Help', and 'My Data'. On the left side, there is a sidebar menu with options: 'Protein by name', 'Protein by sequence', 'Multiple proteins', 'Multiple sequences', 'Organisms', 'Protein families ("COGs")', 'Examples', and 'Random entry'. The main content area is titled 'SEARCH' and features a 'Single Protein by Name / Identifier' form. The form has a 'Protein Name:' input field with a placeholder '(examples: #1 #2 #3)', an 'Organism:' dropdown menu set to 'auto-detect', and a 'SEARCH' button.

page d'accueil de STRING

ANNEXE II

Les séquences Fasta :

```
>KR711678.1 Synthetic construct Homo sapiens clone CCSBHm_00028446
SLC30A8 (SLC30A8) mRNA, encodes complete protein
GTTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTTATAATGCCAACTTTGTACAAAAAAGTTGGCACCAT
GTACCACTGCCACAGTGGCTCCAAGCCCACAGAAAAGGGGGCGAATGAGTACGCCTATGCCAAGTGGAAA
CTCTGTTCTGCTTCAGCAATATGCTTCATTTTTCATGATTGCAGAGGTCGTGGGTGGGCACATTGCTGGGA
GTCTTGCTGTTGTACAGATGCTGCCACCTCTTAATTGACCTGACCAGTTTCCTGCTCAGTCTCTTCTC
CCTGTGGTTGTCATCGAAGCCTCCCTCTAAGCGGCTGACATTTGGATGGCACCAGCAGAGATCCTTGGT
GCCCTGCTCTCCATCCTGTGCATCTGGGTGGTACTGGCGTGCTAGTGTACCTGGCATGTGAGCGCCTGC
TGTATCCTGATTACCAGATCCAGGCGACTGTGATGATCATCGTTTTCCAGCTGCGCAGTGGCGGCCAACAT
TGTAATACTGTGGTTTTGCACCAGAGATGCCTTGGCCACAATCACAAGGAAGTACAAGCCAATGCCAGC
GTCAGAGCTGCTTTTGTGCATGCCCTTGGAGATCTATTTTCAGAGTATCAGTGTGCTAATTAGTGCACTTA
TTATCTACTTTAAGCCAGAGTATAAAATAGCCGACCCAATCTGCACATTCATCTTTTCCATCCTGGTCTT
GGCCAGCACCATCACTATCTTAAAGGACTTCTCCATCTTACTCATGGAAGGTGTGCCAAAGAGCCTGAAT
TACAGTGGTGTGAAAGAGCTTATTTTTAGCAGTCGACGGGGTGCTGTCTGTGCACAGCCTGCACATCTGGT
CTCTAACAATGAATCAAGTAATTCTCTCAGCTCATGTTGCTACAGCAGCCAGCTGGGACAGCCAAGTGGT
TCGGAGAGAAATTGCTAAAGCCCTTAGCAAAAGCTTTACGATGCACTCACTCACCATTTCAGATGGAATCT
CCAGTTGACCAGGACCCCGACTGCCTTTTCTGTGAAGACCCCTGTGACTTGCCAACTTTCTTGTACAAAG
TTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAAC
```

```
>KR710185.1 Synthetic construct Homo sapiens clone CCSBHm_00010262 INS
(INS) mRNA, encodes complete protein
GTTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTTATAATGCCAACTTTGTACAAAAAAGTTGGCATGGC
CCTGTGGATGCGCCTCCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCAGCCGAGCCTTT
GTGAACCAACACCTGTGCGGCTCACACCTGGTGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGCTTCT
TCTACACACCCAAGACCCGCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCC
TGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTTGCATTGTGGAACAATGC
TGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTACCCAACCTTCTTGTACAAAGTTG
GCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAAC
```

Séquence Fasta	Numéro d'accion
PDX1	NG_008183.1
Chromosome 11	NC_000011.10
Chromosome 8	NC_000008.11

ANNEXE III

Browse coexpression interactions in tabular form:

<u>node1</u>	<u>node2</u>	<u>node1 accession</u>	<u>node2 accession</u>	<u>node1 annotation</u>	<u>node2 annotation</u>	<u>coexpr. score</u>
AKT1	IRS1	ENSP00000270202	ENSP00000304895	<i>V-akt murine thymoma viral oncoge...</i>	<i>Insulin receptor substrate 1; May m...</i>	0.53
AKT1	IRS2	ENSP00000270202	ENSP00000365016	<i>V-akt murine thymoma viral oncoge...</i>	<i>Insulin receptor substrate 2; May m...</i>	0.53
AKT1	INSR	ENSP00000270202	ENSP00000303830	<i>V-akt murine thymoma viral oncoge...</i>	<i>Insulin receptor; Receptor tyrosine ...</i>	0.54
INS	IGF1	ENSP00000250971	ENSP00000302665	<i>Insulin; Insulin decreases blood glu...</i>	<i>Insulin-like growth factor 1 (somato...</i>	0.51
IGF1	INS	ENSP00000302665	ENSP00000250971	<i>Insulin-like growth factor 1 (somato...</i>	<i>Insulin; Insulin decreases blood glu...</i>	0.51
PDX1	INS	ENSP00000370421	ENSP00000250971	<i>Pancreatic and duodenal homeobo...</i>	<i>Insulin; Insulin decreases blood glu...</i>	0.46
SLC30A8	INS	ENSP00000415011	ENSP00000250971	<i>Solute carrier family 30 (zinc transp...</i>	<i>Insulin; Insulin decreases blood glu...</i>	0.92
INS	PDX1	ENSP00000250971	ENSP00000370421	<i>Insulin; Insulin decreases blood glu...</i>	<i>Pancreatic and duodenal homeobo...</i>	0.46
INS	SLC30A8	ENSP00000250971	ENSP00000415011	<i>Insulin; Insulin decreases blood glu...</i>	<i>Solute carrier family 30 (zinc transp...</i>	0.92
IRS1	AKT1	ENSP00000304895	ENSP00000270202	<i>Insulin receptor substrate 1; May m...</i>	<i>V-akt murine thymoma viral oncoge...</i>	0.53
IRS2	AKT1	ENSP00000365016	ENSP00000270202	<i>Insulin receptor substrate 2; May m...</i>	<i>V-akt murine thymoma viral oncoge...</i>	0.53
INSR	AKT1	ENSP00000303830	ENSP00000270202	<i>Insulin receptor; Receptor tyrosine ...</i>	<i>V-akt murine thymoma viral oncoge...</i>	0.54

Les résultats de STRING

ANNEXE IV

Cufflinks

Tableau représente résultats de cufflinks de l'expression des gènes.

1	2	3	4	5	7	8	9	10	11	12
tracking_id	class_code	nearest_ref_id	gene_id	gene_short_name	Locus	length	coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi
CUFF.1.1	-	-	CUFF.1	-	NC_000011.10:108639-108730	91	254.786	6505.64	2305.03	10706.2
CUFF.2.1	-	-	CUFF.2	-	NC_000011.10:10887-109032	145	141.489	4185.08	2635.62	5734.55
CUFF.3.1	-	-	CUFF.3	-	NC_000011.10:109296-109623	327	16.7772	692.972	511.298	874.645
CUFF.4.1	-	-	CUFF.4	-	NC_000011.10:109710-109903	193	108.562	3012.43	2202.45	3822.41
CUFF.5.1	-	-	CUFF.5	-	NC_000011.10:110234-110341	107	702.573	1381.11	9121.5	18500.7
CUFF.6.1	-	-	CUFF.6	-	NC_000011.10:110841-111116	275	26.5194	742.851	510.848	974.855

Tableau de l'expression des transcrits

1	2	3	4	5	6	7	8	9	10	11	12	13
tracking_id	class_code	nearest_ref_id	gene_id	gene_short_name	tss_id	locus	Length	Coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
CUFF.1.1	-	-	CUFF.1	-	-	NC_000011.10:108639-108730	91	254.786	6505.64	2305.03	10706.2	OK
CUFF.2.1	-	-	CUFF.2	-	-	NC_000011.10:10887-109032	145	141.489	4185.08	2635.62	5734.55	OK

CUFF.8.	CUFF.7.	CUFF.6.	CUFF.5.	CUFF.4.	CUFF.3.
1	1	1	1	1	1
-	-	-	-	-	-
-	-	-	-	-	-
CUFF.8	CUFF.7	CUFF.6	CUFF.5	CUFF.4	CUFF.3
-	-	-	-	-	-
-	-	-	-	-	-
NC_000011.10:1	NC_000011.10:1	NC_000011.10:1	NC_000011.10:1	NC_000011.10:1	NC_000011.10:1
106	902	275	107	193	327
181.26	4.85678	26.5194	702.573	108.562	16.7772
5638.28	184.208	742.851	13811.1	3012.43	692.972
2573.44	144.547	510.848	9121.5	2202.45	511.298
8703.12	223.87	974.855	18500.7	3822.41	874.645
OK	OK	OK	OK	OK	OK

Tableau de l'assemblage des transcrits

Seqname	Source	Feature	Start	End	Score	Strand	Attributes
NC_000011.10	Cufflinks	Transcript	108640	108730	1000	.	gene_id "CUFF.1";transcript_id "CUFF.1.1"; FPKM "6505.6368470949"; frac "1.000000"; conf_lo "2305.029046"; conf_hi "10706.244648"; cov "254.786007";
NC_000011.10	Cufflinks	Exon	108640	108730	1000	.	gene_id "CUFF.1";transcript_id "CUFF.1.1"; exon_number "1"; FPKM "6505.6368470949"; frac "1.000000"; conf_lo "2305.029046"; conf_hi "10706.244648"; cov

							"254.786007";
NC_000011.10	Cufflinks	transcript	108888	109032	1000	.	gene_id "CUFF.2";transcript_id "CUFF.2.1"; FPKM "4185.0830099996"; frac "1.000000"; conf_lo "2635.618926"; conf_hi "5734.547094"; cov "141.489232";
NC_000011.10	Cufflinks	Exon	108888	109032	1000	.	gene_id "CUFF.2";transcript_id "CUFF.2.1"; exon_number "1"; FPKM "4185.0830099996"; frac "1.000000"; conf_lo "2635.618926"; conf_hi "5734.547094"; cov "141.489232";
NC_000011.10	Cufflinks	Transcript	109297	109623	1000	.	gene_id "CUFF.3";transcript_id "CUFF.3.1"; FPKM "692.9716352508"; frac "1.000000"; conf_lo "511.298133"; conf_hi "874.645137"; cov "16.777205";
NC_000011.10	Cufflinks	Exon	109297	109623	1000	.	gene_id "CUFF.3";transcript_id "CUFF.3.1"; exon_number "1"; FPKM "692.9716352508"; frac "1.000000"; conf_lo "511.298133"; conf_hi "874.645137"; cov "16.777205";

Tableau de l'expression différentielle des transcrits

	CUFF.1001.1	CUFF.1000.1	CUFF.100.1	CUFF.10.1	CUFF.1.1	test_id	1
	CUFF.1001	CUFF.1000	CUFF.100	CUFF.10	CUFF.1	gene_id	2
-	-	-	-	-	-	Gene	3
NC_000011.10 :19526833- 19526902	NC_000011.10 :19441577- 19442020	NC_000011.10 :791797- 792011	NC_000011.10 :156980- 157183	NC_000011.10 :108639- 108730	NC_000011.10	locus	4
q1	q1	q1	q1	q1	q1	sample_1	5
q2	q2	q2	q2	q2	q2	sample_2	6
OK	OK	OK	OK	OK	OK	Status	7
28898	89.7726	4692.6	1278.64	10834.9	10834.9	value_1	8
28898	89.7726	4692.6	1278.64	10834.9	10834.9	value_2	9
0	0	0	0	0	0	log2(fold_chan ge)	10
0	0	0	0	0	0	test_stat	11
1	1	1	1	1	1	p_value	12
1	1	1	1	1	1	q_value	13
no	no	no	no	no	no	significant	14

Cuffcompare

Tableau de La comparaison des transcrits combinées

NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	NC_000011.10	Seqname
Cufflinks	Cufflinks	Cufflinks	Cufflinks	Cufflinks	Cufflinks	Cufflinks	Cufflinks	Cufflinks	Source
Exon	Exon	Transcript	Exon	Transcript	Exon	Exon	Exon	transcript	Feature
399967	399112	399112	314922	314311	314311	314311	314311	293639	Start
400108	399196	400108	315211	315211	314356	315211	315211	293922	End
.	Score
+	+	+	+	+	+	+	+	+	Strand
.	Frame
transcript_id "TCONS_000000_03";gene_id "XLOC_000002".	transcript_id "TCONS_000000_03";gene_id "XLOC_000002".	transcript_id "TCONS_000000_03";gene_id "XLOC_000002".	transcript_id "TCONS_000000_02";gene_id "XLOC_000002".	transcript_id "TCONS_000000_02";gene_id "XLOC_000002".	transcript_id "TCONS_000000_02";gene_id "XLOC_000002".	transcript_id "TCONS_000000_02";gene_id "XLOC_000002".	transcript_id "TCONS_000000_01";gene_id "XLOC_000001".	transcript_id "TCONS_000000_01";gene_id "XLOC_000001".	Attributs