



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA

RECHERCHE SCIENTIFIQUE

UNIVERSITE SAAD DAHLEB BLIDA 1

Faculté des Sciences de la Nature et de la vie

Département de Biologie Et Physiologie Cellulaire

Option : Biologie Moléculaire Et Cellulaire

Mémoire de projet de fin d'études en vue de l'obtention du diplôme

de Master

Thème :

*Identification des gènes et des voies de signalisations associés
maladies auto-immunes cas Diabète type1 par analyse
bio-informatique.*

Présenté par :

Soutenu le : Lundi 07 juillet 2025.

BRAHIMI YOUSRA

KERBADJ CHAIMA

Devant le jury composé de :

Présidente : HAMZI Wahiba

MCA-Université Saad Dahleb Blida1

Examineur : KHIER Mokrane

MCB-Université Saad Dahleb Blida1

Promoteur: SENINET Mohamed El-Hadi

Bioinformaticien d'étude, Recherche et
Développement.

Année universitaire : 2024/2025

Remerciements

Nous remercions en premier lieu le Dieu Tout-Puissant de nous avoir illuminé et ouvert les portes du savoir et de nous avoir donné la volonté et le courage d'élaborer ce travail.

*Nous tenons à remercier profondément et sincèrement tous ceux qui ont participé de près ou de loin à la réalisation de ce travail. Nous exprimons nos sincères remerciements à notre encadrant **Seninet Mohamed El-Hadi** pour sa disponibilité, son soutien constant, ses conseils précieux et sa confiance tout au long de ce projet.*

*Nous adressons également nos remerciements respectueux aux membres du jury, présidente Mme **HAMZI Wahiba** et l'examineur **Mr MOKRANE Khier** pour l'honneur qu'ils nous font en évaluant ce mémoire et pour leurs remarques constructives qui contribueront à son amélioration. Un merci tout particulier à notre Chef d'option Mme **SAADI Leila** et tous les enseignants qui ont enrichi notre parcours académique à travers leur savoir et leur dévouement. Leur engagement a été un pilier important de notre formation.*

Enfin, nous souhaitons exprimer toute notre reconnaissance à nos camarades de la promotion dont l'entraide, la collaboration et le soutien moral ont été essentiels pour traverser cette aventure universitaire.



Dédicaces

J'ai le grand plaisir de dédie ce modeste travail :

À ma maman, Sellah Nadia

Pour ton amour inconditionnel, tes sacrifices infinis et tes prières silencieuses.

Tu es ma force tranquille et mon premier soutien, je t'aime profondément.

À mon papa, Brahimi Maamer

Pour ta fatigue silencieuse, pour chaque effort que tu fais sans jamais te plaindre et pour ton amour discret qui s'est révélé sous forme de larmes dans nos moments de joie, que dieu te garde.

À ma sœur unique, Sarah

Ma complice de toujours, dans les rires comme dans les moments difficiles.

À mes deux chers frères, Yacine et M'hamed

Véritables piliers de ma vie, votre présence est un cadeau que la vie m'a offert deux fois.

À mes grands-parents

***Papa M'hamed**, tu as toujours été une source de courage, d'amour et la force solide qui soutient notre grande famille, tu n'as jamais cessé de m'encourager d'aller en avant. Tu es notre exemple. Fièvre d'être mon papa.*

***Muima**, merci pour ta douceur, ta patience et tous l'amour que tu nous donne. Merci d'être notre refuge chaleureux. Que dieu te protège*

*A toute ma famille **BRAHIMI Et SELLAH**, mes tentes, oncles, mes cousines et cousins.*

À ma collègue Chaima, qui a partagé ce travail avec moi. Merci pour ta bonne humeur, ta patience et ta collaboration.

À notre encadrant, Monsieur Seninet M^{ed} El-Hadi pour son accompagnement, sa patience et la qualité de ses conseils, qui ont contribué à la réalisation de ce mémoire.

Fousra

Dédicaces

*Ce travail est dédié avec tout mon amour à mes très chers parents, **Kerbadj Youcef et Berrouane Saïda**, pour leur amour infini, leurs sacrifices et leur soutien sans limite.*

*À mes sœurs, **Nidal et Meriem**.*

*À toute ma famille (**Kerbadj et Berrouane**).*

À ma douce nièce, Zineb, petite étoile de ma vie, dont le sourire m'a donné du courage même dans les moments les plus difficiles.

*À ma chère binôme, **Brahimi Yousra**, pour ta patience, ta confiance et ton bel esprit d'équipe.*

*À **Oukali Karim**, pour ton soutien discret, ta gentillesse et ta présence réconfortante.*

*Et enfin, à **Monsieur Seninet M^{ed} El-Hadi**, mon encadrant, pour ses précieux conseils et son accompagnement tout au long de ce travail.*

Chaima

Le diabète de type 1 (DT1) est une maladie auto-immune caractérisée par la destruction progressive des cellules bêta pancréatiques, entraînant une carence absolue en insuline.. Dans le cadre de cette étude, une approche bioinformatique in silico a été utilisée pour analyser le profil d'expression génique des monocytes issus de différents groupes : patients récemment diagnostiqués DT1 (ROT1D), leurs frères et sœurs à haut (HRS) et faible risque (LRS), ainsi que des témoins sains non apparentés (uHC). Les données transcriptomiques ont été extraites de la base GEO (série GSE232310) et traitées à l'aide de plusieurs outils tels que GEO2R, Limma, DESeq2 pour l'identification des gènes différentiellement exprimés, et STRING-db, DAVID, Enrichr-KG ou NetworkAnalyst pour l'analyse des voies biologiques et des réseaux protéine-protéine.

Les analyses ont permis d'identifier 293 gènes différentiellement exprimés dans le groupe ROT1D vs uHC et 184 gènes dans le groupe HRS vs LRS. Ces gènes sont majoritairement codants, bien que les ARN non codants (lncRNA) soient également présents, suggérant une régulation post-transcriptionnelle dans le DT1. L'enrichissement et l'analyse fonctionnel des de ces résultats a révélé une activation significative de voies liées à l'immunité, notamment l'activation des lymphocytes T CD4+/CD8+, la cytotoxicité médiée par les cellules NK, la signalisation via le récepteur TCR, et la voie de signalisation de l'insuline ainsi que des gènes clés tels que SOCS2, CBLB, AKT3 et les claudines jouant un rôle dans l'inflammation pour le groupe ROT1D vs uHC. Cependant chez le groupe HRS vs LRS, les gènes clés tels que les claudines (CLDN1, CLDN2) ont montré une balance entre une surexpression et sous-expression significative, indiquant des altérations importantes dans les processus inflammatoires et les jonctions cellulaires. L'analyse d'expression différentielle du groupe uHC vs HRS a permis d'identifier deux gènes clés dont la régulation est inversée le CHCHD5 et TRIT1 pointent vers une dysfonction mitochondriale potentiellement complexe et multifacette dans les monocytes des individus à haut risque de diabète de type I. Les réseaux d'interaction protéine-protéine ont mis en lumière des protéines centrales jouant un rôle crucial dans la modulation de la maladie, des interactions clés dans les voies immunitaires, métaboliques et de jonctions serrées, offrant de potentielles cibles thérapeutiques et biomarqueurs précoces. pourraient constituer des biomarqueurs précoces du DT1.

Mots clés : Diabète de type 1 – Monocytes – Expression génique – Jonctions serrées – Dysfonction mitochondriale

Type 1 diabetes (T1D) is an autoimmune disease characterised by the progressive destruction of pancreatic beta cells, leading to absolute insulin deficiency. In this study, an in silico bioinformatics approach was used to analyse the gene expression profile of monocytes from different groups: newly diagnosed T1D patients (ROT1D), their high-risk (HRS) and low-risk (LRS) siblings, and unrelated healthy controls (uHC). Transcriptomic data were extracted from the GEO database (series GSE232310) and processed using several tools such as GEO2R, Limma, DESeq2 for the identification of differentially expressed genes, and STRING-db, DAVID, Enrichr-KG, or NetworkAnalyst for the analysis of biological pathways and protein-protein networks.

The analyses identified 293 differentially expressed genes in the ROT1D vs uHC group and 184 genes in the HRS vs LRS group. These genes are predominantly coding, although non-coding RNAs (lncRNAs) are also present, suggesting post-transcriptional regulation in T1D. Enrichment and functional analysis of these results revealed significant activation of immunity-related pathways, including CD4+/CD8+ T cell activation, NK cell-mediated cytotoxicity, TCR signalling, and insulin signalling, as well as key genes such as SOCS2, CBLB, AKT3, and claudins playing a role in inflammation for the ROT1D vs. uHC group. However, in the HRS vs. LRS group, key genes such as claudins (CLDN1, CLDN2) showed a balance between significant overexpression and underexpression, indicating significant alterations in inflammatory processes and cell junctions. Differential expression analysis of the uHC vs HRS group identified two key genes whose regulation is reversed, CHCHD5 and TRIT1, pointing to potentially complex and multifaceted mitochondrial dysfunction in the monocytes of individuals at high risk of type I diabetes. Protein-protein interaction networks have highlighted central proteins that play a crucial role in modulating the disease, key interactions in immune, metabolic and tight junction pathways, offering potential therapeutic targets and early biomarkers. These could constitute early biomarkers for T1D.

Key words :

Type 1 diabetes - Monocytes - Gene expression - Tight junctions - Mitochondrial dysfunction

داء السكري من النوع 1 (DT1) هو مرض مناعي ذاتي يتميز بالتدمير التدريجي لخلايا بيتا في البنكرياس، مما يؤدي إلى نقص مطلق في الأنسولين. في إطار هذه الدراسة، تم استخدام نهج بيولوجي حاسوبي في السيليكون لتحليل ملف تعبير الجينات للخلايا الأحادية من مجموعات مختلفة: مرضى تم تشخيصهم مؤخراً بـ (DT1) (ROT1D)، وإخوتهم وأخواتهم المعرضين لخطر مرتفع (HRS) ومنخفض (LRS)، بالإضافة إلى شهود أصحاء غير أقارب (uHC). تم استخراج البيانات النسخية من قاعدة GEO (سلسلة GSE232310) ومعالجتها باستخدام عدة أدوات مثل GEO2R و Limma و DESeq2 لتحديد الجينات المعبرة بشكل مختلف، و STRING-db و DAVID و Enrichr-KG أو NetworkAnalyst لتحليل المسارات البيولوجية وشبكات البروتين-البروتين.

سمحت التحليلات بتحديد 293 جيناً يتم التعبير عنها بشكل مختلف في المجموعة ROT1D مقابل uHC و 184 جيناً في المجموعة HRS مقابل LRS. هذه الجينات هي في الغالب جينات مشفرة، على الرغم من وجود الحمض النووي الريبي غير المشفر (lncRNA) أيضاً، مما يشير إلى وجود تنظيم ما بعد النسخ في DT1. كشف التخصيب والتحليل الوظيفي لهذه النتائج عن تنشيط كبير للمسارات المرتبطة بالمناعة، ولا سيما تنشيط الخلايا التائية CD4+/CD8+، والسمية الخلوية التي تتوسطها الخلايا القاتلة الطبيعية، والإشارات عبر مستقبلات TCR، ومسار إشارات الأنسولين، بالإضافة إلى جينات رئيسية مثل SOCS2 و CBLB و AKT3 والكلودينات التي تلعب دوراً في الالتهاب للمجموعة ROT1D مقابل uHC. ومع ذلك، في المجموعة HRS مقابل LRS، أظهرت الجينات الرئيسية مثل الكلودينات (CLDN1، CLDN2) توازناً بين فرط التعبير والتعبير المنخفض بشكل كبير، مما يشير إلى تغييرات كبيرة في العمليات الالتهابية والوصلات الخلوية.

سمح تحليل التعبير التفاضلي لمجموعة uHC مقابل HRS بتحديد جينين رئيسيين يتم تنظيمهما بشكل معكوس، وهما CHCHD5 و TRIT1، مما يشير إلى خلل وظيفي ميتوكوندريا معقد ومتعدد الأوجه في خلايا الدم البيضاء لدى الأفراد المعرضين لخطر الإصابة بداء السكري من النوع الأول. سلطت شبكات التفاعل بين البروتينات الضوء على البروتينات المركزية التي تلعب دوراً حاسماً في تعديل المرض، والتفاعلات الرئيسية في المسارات المناعية والتمثيل الغذائي والوصلات الضيقة، مما يوفر أهدافاً علاجية محتملة وعلامات بيولوجية مبكرة. قد تشكل علامات بيولوجية مبكرة لمرض السكري من النوع 1.

الكلمات المفتاحية: داء السكري من النوع الأول - الخلايا الأحادية - التعبير الجيني - الوصلات الضيقة - الخلل الوظيفي للميتوكوندريا

Liste des abréviations :

CD4+ : Lymphocytes T Helper

CD8+ : T Cytotoxiques

CTLA4 : Protéine 4 associée aux lymphocytes T cytotoxiques.

DID : Diabète insulino-dépendant.

DLGAP5 : Discs Large-Associated Protein 5

GADA: Glutamic Acid Decarboxylase Antibodies.

HRS: Siblings at high genetic risk.

IA-2 : Anticorps contre l'antigène des îlots pancréatiques 2.

IAA : Auto-anticorps dirigés contre l'insuline.

ICA : Islet Cell Antibodies.

IDDM : Diabète sucré insulino-dépendant.

IL2RA : Récepteur alpha de l'interleukine 2.

lncRNA : Long Non-Coding RNA

LRS: Siblings at low genetic risk.

NK : Natural Killer

NKT: Natural Killer T cells.

ROT1D: Recently diagnosed type 1 diabetes patients.

TCR : T Cell Receptor

Th : Cellules T auxiliaires.

uHC : Unrelated healthy controls.

ZnT8 : Anticorps contre le transporteur de zinc 8

La neuropathie diabétique : est une complication invalidante du diabète sucré qui touche des millions de personnes dans le monde. Elle se caractérise par des lésions nerveuses résultant d'une exposition prolongée à une glycémie élevée. Elle peut provoquer divers symptômes, notamment des douleurs, des engourdissements, une faiblesse musculaire, un dysfonctionnement autonome et des ulcères du pied (**Strand et al.,2024**).

La rétinopathie diabétique (RD) : est une microangiopathie, c'est-à-dire une atteinte des petits vaisseaux sanguins, causée par les effets chroniques du diabète sucré. C'est la maladie vasculaire de la rétine la plus fréquente. La RD était la cinquième cause mondiale de cécité évitable et la cinquième cause de déficience visuelle modérée à sévère chez les personnes âgées de 50 ans et plus. (**Fung et al., 2022**).

La néphropathie diabétique : caractérisée par une altération progressive de la fonction rénale due à l'hyperglycémie prolongée. Elle se manifeste par une albuminurie persistante (présence d'albumine dans les urines), une élévation de la pression artérielle et une diminution progressive du débit de filtration glomérulaire (DFG), pouvant évoluer vers une insuffisance rénale terminale. (**Molitch et al.,2010**).

L'artériopathie oblitérante des membres inférieurs (AOMI) : se définit comme un rétrécissement du calibre des artères à destination des membres inférieurs, qui entraîne une perte de charge hémodynamique, avec ou sans traduction clinique. Elle est particulièrement fréquente chez les diabétiques, présente une double particularité, la sévérité de l'atteinte le plus souvent distale, plus fréquente ; et le risque élevé de progression vers l'ulcération et la gangrène. (**Ndambwe et al.,2021**).

La sarcocystose : est une maladie parasitaire rare causée par le parasite protozoaire *Sarcocystis*. Elle peut se manifester par une sarcocystose intestinale chez l'homme ou une sarcocystose musculaire, où des sarcocystes se développent dans différents types de muscles..

(<https://www.malacards.org/card/sarcocystosis>)

Liste des Figures

Figure 1 : Principale complications chroniques du diabète de type 1.	3.
Figure 2 : Pathogénèse de l'auto-immunité du diabète de type 1	7.
Figure 3: Représentation de l'outil NCBI GEO et son mode d'emploi.	15.
Figure 4: Présentation de l'outil STRING-db pour la construction et la visualisation des réseaux d'interaction protéine-protéine (PPI), l'annotation fonctionnelle et d'enrichissement.	17.
Figure 5: Présentation de l'outil DAVID pour l'annotation fonctionnelle et d'enrichissement.	18.
Figure 6: Présentation de l'outil Enrichr-KG pour l'analyse d'enrichissement de gènes.	19.
Figure 7 : Présentation de l'outil NetworkAnalyst pour l'analyse de réseaux omiques.	20.
Figure 8: Présentation de la base des données ENSEMBL pour l'annotations complètes de génomes.	20.
Figure 9 : Présentation de la base de données GeneCards pour la recherche de gènes et son mode d'utilisation.	21.
Figure 10: Présentation de l'outil Metascape pour l'analyse de listes de gènes.	22.
Figure 11 : Représentation du diagramme de Venn illustrant le chevauchement des gènes significativement exprimés entre les différents groupes étudiés.	23.
Figure 12: Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'étude « ROT1D vs uHC ».	24.
Figure 13 : Distribution des gènes identifiés par type dans le groupe des gènes « ROT1D vs uHC ».	25.
Figure 14: Heatmap des termes GO et des processus biologiques enrichis dans l'analyse transcriptomique du groupe des gènes « ROT1D vs uHC ».	27.
Figure 15: Représentation graphique des principales voies et processus biologiques enrichis, classés par signification et nombre de gènes du groupe des gènes « ROT1D vs uHC ».	29.
Figure 16: Réseau d'interaction protéine-protéine (PPI) des gènes différentiellement exprimés gènes du groupe des gènes « ROT1D vs uHC ».	32.
Figure 17 : Expression relative ($\log_2(\text{Fold Change})$) des gènes clés identifiés dans les voies immunitaires et de signalisation du groupe des gènes « ROT1D vs uHC ».	33.

Figure 18: Représentation schématique de la voie de signalisation de l'insuline avec indication des gènes régulés du groupe des gènes « ROT1D vs uHC »..	35.
Figure 19 : Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'étude « HRS vs LRS »..	37.
Figure 20: Distribution des gènes identifiés par type dans le groupe des gènes « HRS vs LRS »..	38.
Figure 21 : Heatmap des termes enrichis dans l'analyse transcriptomique du groupe des gènes « HRS vs LRS »..	40.
Figure 22: Représentation graphique des principales voies et processus biologiques enrichis, classés par signification et nombre de gènes du groupe des gènes « HRS vs LRS ».....	41.
Figure 23 : Réseau d'interaction protéine-protéine (PPI) des gènes différentiellement exprimés gènes du groupe des gènes « HRS vs LRS »..	44.
Figure 24: Expression relative (log2(Fold Change)) des gènes clés identifiés dans le processus biologique Tight junction interactions du groupe des gènes « HRS vs LRS »..	45.
Figure 25: Représentation schématique de la voie de signalisation des jonctions serrées (Tight Junction) avec indication des gènes régulés du groupe des gènes « HRS vs LRS».....	47.
Fifure 26 : Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'études « uHC vs HRS ».....	49.
Figure 27 : Représentation par histogramme du niveau d'expression du gène CHCHD5/TRIT1 dans les groupes d'étude « uHC vs HRS».....	50.
Figure 28 : Ensemble des voies métaboliques et des maladies associés (Reactome Disease Pathway),soulignant les liens avec la fonction mitochondriale.....	52.
Figure 29 : Des différentes étapes impliquées dans la sécrétion d'insuline.	Annexes.

Liste des Tableaux

Tableau I: Les effectifs des groupes étudiés.....	13.
Tableau II: Top 10 des processus biologiques (GO) et des voies de signalisation significativement enrichis.....	30.
Tableau III : Processus biologiques (GO) et appariement des gènes enrichis.....	42.

Remerciements	
Dédicaces	
Résumé	
Abstract ملخص	
Liste d'abréviations	
Glossaire	
Liste des figures	
Liste des tableaux	
Introduction.....	1.
Chapitre I : Recherche bibliographique	
I. Diabète Auto-immune (Diabète type 1) :	2.
I.1 Définition de DT1 :	2 .
I.2 Complication de DT1 :	2.
I.2.1 Complication aiguës :	2.
I.2.2 Complication chronique :	3.
I.3 Physiopathologie du diabète de type 1 :	4.
I.3.1 Destruction des cellules bêta :	4.
I.3.2 Perte de sécrétion d'insuline :	4.
I.3.3 Rôle des lymphocytes T :	5.
I.3.4 Rôle des auto-anticorps :	5.
I.4 Etiologie (Facteur génétique) :	7.
II. Approches génétiques :	9.
II.1 NGS (Next-Generation Sequencing):	9.
II.2 RNA-seq (RNA sequencing):	9.
II.3 Microarrays (puces à ADN) :	10.

III.	Bioinformatique et analyse des données transcriptomiques :.....	10.
------	---	-----

III.1	Définition et importance de la bioinformatique :.....	10.
-------	---	-----

III.2	Outils et bases de données bioinformatiques :.....	11.
-------	--	-----

Chapitre II : Matériel & Méthodes

I.	Matériel :	12.
----	------------------	-----

Population :		12.
--------------	-------	-----

II.	Méthodes :	12.
-----	------------------	-----

II.1	Identification des gènes différentiellement exprimés	13.
------	--	-----

II.1.1	NCBI GEO2R:	13.
--------	-------------------	-----

II.2	Analyse exploratoire, expression différentielle et analyses des voies de signalisation sur des données RNA-seq	15.
------	--	-----

II.2.1	RStudio :	15.
--------	-----------------	-----

II.2.2	Limma :	16.
--------	---------------	-----

II.2.3	DESeq2 :	16.
--------	----------------	-----

II.2.4	Biclustering (packages Biclust et QUBIC sur R) :	16.
--------	--	-----

II.3	Annotation d'enrichissement et d'Analyse de Réseaux	17.
------	---	-----

II.3.1	STRING-db	17.
--------	-----------------	-----

II.3.2	DAVID :	18.
--------	---------------	-----

II.3.3	Enrichr-KG :	18.
--------	--------------------	-----

II.3.4	NetworkAnalyst :	19.
--------	------------------------	-----

II.4	Annotation des gènes exprimées de manières différentielles	20.
------	--	-----

II.4.1	ENSEMBL :	20.
--------	-----------------	-----

II.4.2	GeneCards :	21.
--------	-------------------	-----

II.4.3 Metascape :	22.
Chapitre III : Résultats et Discussions	
I. Analyse profil d'expressions des gènes dans nos groupe d'étude	23.
I.1 ROT1D contre uHC :	24.
I.2 HRS contre LRS	36.
I.3 uHC vs HRS.....	48.
Conclusion.....	54.
Références bibliographiques.....	56.
Anenex.....	64
.	

Introduction

Introduction :

Le diabète de type I (DT1) encore appelé diabète insulino-dépendant, est une pathologie chronique résulte de la destruction progressive et irréversible des cellules bêta pancréatiques, conduisant à une carence absolue en insuline et à un déséquilibre durable de l'homéostasie glucidique. À long terme, le DT1 peut provoquer des complications microvasculaires et macrovasculaires. Le processus immunologique débute par la production d'autoanticorps dirigés contre l'insuline, GAD, IA-2, IA-2 β ou ZnT8. La présence de plusieurs autoanticorps définit le stade 1, suivi d'un stade 2 marqué par une dysglycémie, puis d'un stade 3 où apparaissent les symptômes cliniques. La susceptibilité génétique est liée à des polymorphismes dans plusieurs gènes, surtout dans le complexe HLA class 2 (comme HLA-DR3 et DR4) et d'autres gènes tels que INS, CTLA4, IL2RA, PTPN22, qui modifient la réponse immunitaire. (Zajec et al., 2022) (Mittal et al., 2024).

Cette étude *in silico* vise à approfondir la compréhension de la pathogenèse du DT1 en analysant les profils d'expression génique des monocytes, cellules immunitaires clés. Nous avons examiné des données transcriptomiques (GSE232310) issues de patients DT1 récents (ROT1D), de sujets à haut (HRS) et faible risque (LRS), et de contrôles sains (uHC). Utilisant une suite d'outils bioinformatiques : des bases comme GEO (Gene Expression Omnibus) et GeneCards fournissent un accès à des données transcriptomiques et des annotations génétiques de haute qualité. Les outils analytiques tels que Limma et DESeq2 sont largement utilisés pour l'analyse différentielle de l'expression génique, tandis que GEO2R offre une interface plus accessible pour explorer les données disponibles dans GEO. Une fois les gènes différentiellement exprimés identifiés, des outils d'enrichissement fonctionnel comme DAVID, STRINGEnrichr, KEGG et Reactome permettent d'interpréter les résultats en les reliant à des voies biologiques ou à des réseaux de protéines). Enfin, la visualisation des données est essentielle pour faciliter l'interprétation : des représentations comme les volcano plots, les heatmaps, les diagrammes de Venn et les réseaux PPI (protein-protein interaction) permettent de synthétiser visuellement des ensembles complexes de résultats.

Notre objectif est d'identifier les signatures moléculaires distinctes et les voies signalisations altérées. Ce travail vise à révéler de nouveaux mécanismes et biomarqueurs potentiels pour le DT1.

Chapitre □

Recherche Bibliographique

I. Diabète Auto-immune (Diabète type 1) :

I.1 Définition de DT1 :

Le diabète de type 1 ou insulino-dépendant (DID) est une maladie auto-immune rare (5 à 10 % des cas de diabète) et se développe le plus souvent chez les enfants et les adolescents, mais peut survenir à tout âge, caractérisée par une réaction auto-immune à médiation cellulaire ciblant les cellules β pancréatiques, entraînant une carence absolue en insuline (**Addissouky et al., 2024**).

Les symptômes classiques au début comprennent la polyurie (l'augmentation du volume d'eau bue par jour), la polydipsie (augmentation du volume d'urines émises par jour) et amaigrissement (la perte de poids involontaire), mais le tableau clinique varie individuellement. Les adultes atteints de DT1 d'apparition récente présentent généralement des symptômes similaires à ceux observés chez les enfants, mais peuvent avoir une progression plus progressive (**Lucier et Mathias, 2024**).

I.2 Complication de DT1 :

I.2.1 Complication aiguës :

a) L'acidocétose diabétique :

L'acidocétose diabétique est une complication grave du diabète, surtout du type 1, causée par une carence d'insuline. Elle entraîne une forte augmentation de la glycémie, la production de cétones, une acidité du sang et une déshydratation. C'est une urgence médicale qui nécessite un traitement rapide pour éviter des complications graves comme l'œdème cérébral ou la septicémie (**Elendu et al., 2023**), (**Lizzo et al., 2023**).

b) L'hypoglycémie diabétique :

L'hypoglycémie est fréquente chez les personnes utilisant de l'insuline ou certains comprimés pour réduire l'hyperglycémie. Cela s'explique par des facteurs tels qu'une activité physique imprévue, une alimentation insuffisante, qui peuvent entraîner une diminution des

besoins en insuline, entraînant une chute importante de la glycémie.

Les signes indiquant que la glycémie est trop basse peuvent inclure : Sueurs froides, visage pâle, Maux de tête, Se sentir incroyablement affamé, Frissons, sensation de faiblesse dans les genoux.

Le traitement de l'hypoglycémie repose sur l'administration de sucre sous plusieurs Formes: boissons sucrées, morceaux de sucre si le patient est conscient ; perfusion Intraveineuse de glucose si le patient est inconscient ; injection intramusculaire de Glucagon (InformedHealth.org,2021)

I .2.2 Complication chroniques :

Les complications liées au diabète, à la fois microvasculaires (rétinopathie, néphropathie et neuropathie) et macrovasculaires (maladies cardiovasculaires, artériopathie périphérique), (Figure 2), (Powers, 2021).

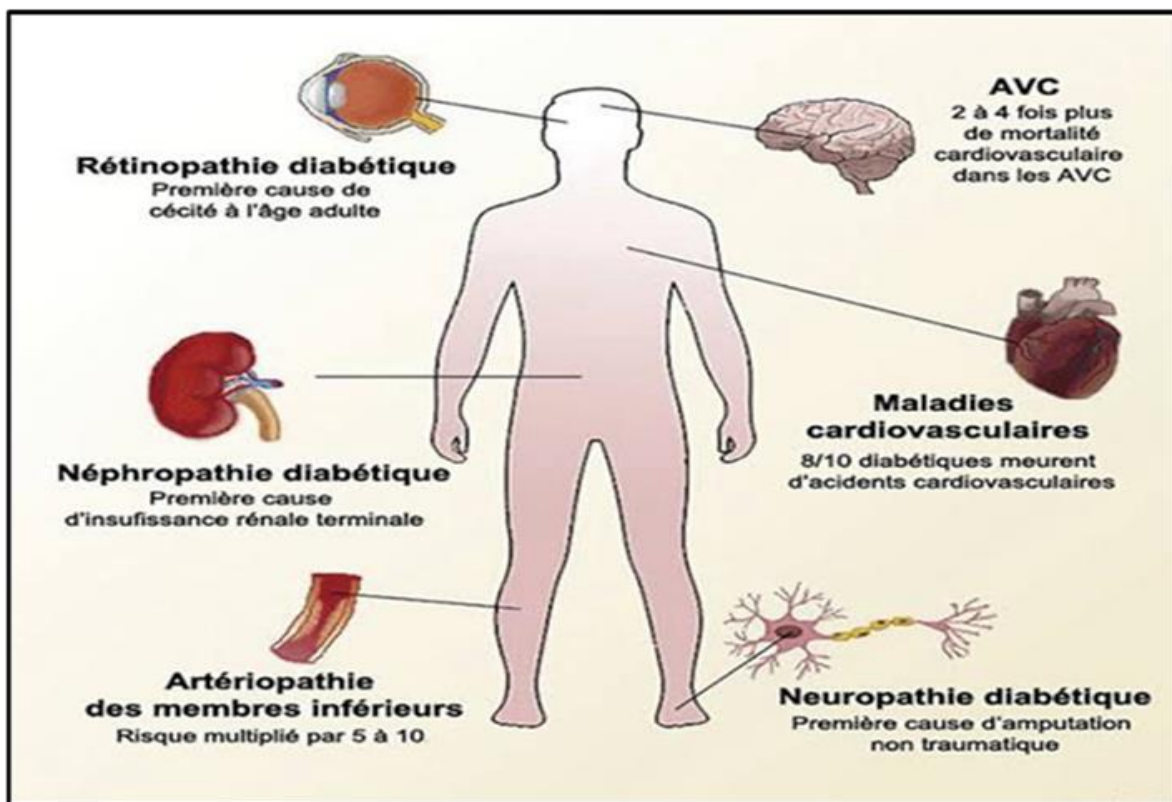


Figure 1 : Principales complications chroniques du diabète de type 1 (Foussier & Zergane, 2021).

I. 3 Physiopathologie du diabète de type 1 :

La pathogenèse du diabète de type 1 (DT1) est médiée par une attaque auto-immune ciblant les cellules bêta pancréatiques productrices d'insuline. Les lymphocytes T autoréactifs et les auto-anticorps contribuent à la destruction sélective des cellules bêta, provoquant une carence en insuline et une hyperglycémie (Addissouky *et al.*, 2024).

I .3.1 Destruction des cellules bêta :

La pathogenèse du diabète de type 1 (DT1) repose sur la perte progressive des cellules bêta pancréatiques, un processus qui débute bien avant l'apparition des symptômes cliniques. Cette destruction des cellules bêta se manifeste par une insulite, caractérisée par une infiltration de cellules immunitaires, notamment des lymphocytes T, des lymphocytes B, des macrophages et des cellules dendritiques dans les îlots de Langerhans. Les cytokines sécrétées par ces cellules inflammatoires induisent l'apoptose des cellules bêta. La destruction de 80 à 95 % des cellules bêta se produit lorsque l'hyperglycémie devient visible (Addissouky *et al.*, 2024).

Les lymphocytes T CD4⁺ et CD8⁺ attaquent directement les cellules bêta en reconnaissant des auto-antigènes comme l'insuline, le GAD65, l'IA-2 et le ZnT8, via les voies HLA. Des déclencheurs environnementaux, tels que les infections virales, endommagent les cellules bêta, libérant des auto-antigènes qui activent les cellules T autoréactives recrutent d'autres cellules immunitaires et maintiennent l'inflammation. Cette destruction progressive dépasse la capacité de régénération des cellules bêta, aboutissant à une sécrétion insuffisante d'insuline (Addissouky *et al.*, 2024).

I .3.2 Perte de sécrétion d'insuline :

Les cellules bêta normales ont une capacité à réguler la sécrétion d'insuline selon le taux de glucose. Dans le DT1, cette régulation est endommagée en raison du stress des cellules bêta qui perturbe la fonction sécrétoire. Les cytokines pro-inflammatoires (IL-1 β , TNF- α , IFN- γ) libérées par les cellules immunitaires dans les îlots pancréatiques endommagent les cellules bêta. Cela perturbe les voies de signalisation régulant la sécrétion d'insuline. Les cellules bêta tentent de compenser en augmentant leur production d'insuline et leur masse

mais cette réponse devient insuffisante face à la perte cellulaire. La sécrétion d'insuline diminue alors, menant à une hyperglycémie par défaut de régulation de la production hépatique et de l'absorption périphérique du glucose, provoquant les symptômes du diabète (Addissouky *et al.*, 2024).

I. 3.3 Rôle des lymphocytes T :

Les lymphocytes T CD4⁺ et CD8⁺ jouent un rôle clé dans l'attaque auto-immune contre les cellules bêta. Les CD4⁺ reconnaissent les peptides des cellules bêta via les molécules HLA de classe II et se différencient en Th1 et Th17, sécrétant des cytokines inflammatoires (IFN- γ , TNF- α , IL-17) qui amplifient l'inflammation et recrutent d'autres cellules immunitaires. Ils soutiennent aussi les cellules B dans la production d'auto-anticorps. Les CD8⁺ reconnaissent les antigènes via HLA de classe I et détruisent directement les cellules bêta par l'action de granzymes, de perforine et de cytokines pro-apoptotiques. Les lymphocytes T régulateurs qui suppriment normalement les réponses auto-immunes sont incapables de contrôler l'expansion des lymphocytes T pathogènes dans le DT1. En plus des lymphocytes T CD4⁺ et CD8⁺, d'autres lymphocytes T comme les cellules NKT et les lymphocytes T gamma-delta ($\gamma\delta$) participent à l'attaque auto-immune contre les cellules bêta dans le diabète de type 1 (Addissouky *et al.*, 2024).

I. 3.4 Rôle des auto-anticorps :

Les auto-anticorps dirigés contre des auto-antigènes des îlots (comme l'insuline, GAD65, IA-2 et ZnT8) sont des marqueurs sensibles et spécifiques du DT1. Produits par les lymphocytes B en collaboration avec les lymphocytes T, ils ne détruisent pas directement les cellules bêta, mais contribuent à l'inflammation en formant des complexes immunitaires et en facilitant l'activation des lymphocytes T auto réactifs. Leur présence aide à prédire le risque de DT1 et à identifier les personnes à risque avant l'apparition des symptômes

Les lymphocytes T proviennent de la moelle osseuse et mûrissent dans le thymus, où ils apprennent à distinguer les antigènes du soi et des antigènes étrangers (sélection négative). Les lymphocytes T régulateurs et les lymphocytes T pathogènes reconnaissent les antigènes

des cellules auto- ou bêta, mais avec des affinités différentes, ce qui explique leurs rôles opposés. Dans le diabète de type 1, ces lymphocytes ciblent les protéines des cellules bêta, comme l'insuline. Lorsqu'elles sont présentées par des cellules présentatrices d'antigènes sur le CMH/HLA, les lymphocytes T s'activent dans les ganglions lymphatiques et migrent vers les îlots de Langerhans et détruisent spécifiquement les cellules bêta. Les Tregs régulent cette réponse pour maintenir la tolérance périphérique. Si cette réponse immunitaire ne peut pas être régulée, cela entraîne une carence en insuline, une hyperglycémie et le développement du DT1, comme montre **(Figure 2)**. **(Addissouky et al., 2024)**

HLA de classe I et II, indispensables à la présentation des antigènes aux lymphocytes T. Précisément, les gènes HLA de classe II (DR3, DR4, DQ) sont associés à un risque accru : Plus de 90 % des patients DT1 sont porteurs de DR3 et/ou DR4.

Les individus hétérozygotes DR3/DR4 ont un risque 30 à 40 fois supérieur à celui d'un individu moyen (**Philips Radermecker, 2012**).

Les variations dans ces gènes augmentent le risque de diabète de type 1, mais elles ne suffisent pas à provoquer la maladie, et d'autres gènes sont impliqués.

Parmi les gènes en dehors du système HLA (non-HLA) :

- ✚ IDDM2, localisé sur le chromosome 11 correspond au gène de l'insuline .
- ✚ IDDM12, sur le chromosome 2, qui contient plusieurs gènes candidats possibles comme le gène CTLA-4 (**Dean & McEntyre, 2004**).
- ✚ **Le gène CTLA-4** : Code un récepteur qui s'exprime sur les cellules T après leur activation par les cellules présentatrices d'antigènes, via l'interaction CD28-B7. Lorsque les cellules T rencontrent des auto-antigènes, CTLA-4 agit comme un frein en se liant aux protéines B7 sur les CPA, limitant ainsi l'activation excessive des cellules T. Des mutations ou polymorphismes peuvent réduire son efficacité, perturbant la régulation de la réponse immunitaire et favorisant une attaque des cellules saines (**El Wafai et al., 2011**).
- ✚ **Le gène PTPN22**, localisé sur le chromosome 1p13, code pour une phosphatase tyrosine spécifique des lymphocytes appelée Lyp. Cette protéine aide à freiner leur activation et leur prolifération. Elle agit en se liant à une autre protéine, Csk, qui bloque la signalisation du récepteur T (TCR). Une mutation ponctuelle dans le gène PTPN22 entraîne l'empêchement de Lyp de bien se lier à Csk, ce qui entraîne une signalisation excessive du TCR et une activation anormale des lymphocytes T (**Gizeh et al., 2013**).
- ✚ **Le gène IL2Ra**, situé sur le chromosome 10p15, code pour un facteur de différenciation essentiel qui supprime les lymphocytes T auto-réactifs grâce à CD25 et

- ✚ régle également l'activité des cellules NK, des cellules B et des cellules Treg. Le polymorphisme de CD25, peuvent altérer la fonction des cellules Treg, qui sont responsables du maintien de l'équilibre immunitaire en inhibant les réponses autoimmunes. Lorsque cette fonction est perturbée, Cela peut entraîner une activation excessive des lymphocytes T (**Frommer et Kahaly,2021**).

II. Approches génétiques :

II. 1 NGS (Next-Generation Sequencing):

Le séquençage de nouvelle génération (NGS) est un outil puissant utilisé dans la recherche en génomique, qui peut séquencer des millions de fragments d'ADN à la fois, fournissant des informations détaillées sur la structure des génomes, les variations génétiques, l'activité des gènes et les changements dans le comportement des gènes (**Satam et al., 2023**).

Pour étudier le diabète, le NGS a aidé à identifier des gènes et des variations liés à la maladie, à mieux comprendre leur rôle dans son développement et à estimer comment elle se répand dans la population. Il a aussi permis de trouver de nouvelles cibles pour développer des traitements adaptés à chaque patient. Avec ses progrès récents, qui rendent le séquençage plus rapide, plus précis et moins coûteux, le NGS est devenu un outil clé pour améliorer le diagnostic, le suivi et les traitements du diabète (**Fareed et al.,2022**).

II. 2 RNA-seq (RNA sequencing):

Le séquençage de l'ARN est une méthode de profilage du transcriptome qui utilise la technologie de séquençage de nouvelle génération (**Fareed et al., 2022**).

RNA-seq permet une caractérisation précise des lymphocytes T et B dans divers compartiments (sang, ganglions pancréatiques, îlots de Langerhans) chez les patients atteints de diabète de type 1. Cette technique identifie des gènes exprimés de façon spécifique dans les cellules T des îlots, impliqués dans la pathogenèse de la maladie. Elle offre aussi la possibilité de détecter des signatures d'expression génique capables de prédire le

développement du diabète avant même l'apparition des auto-anticorps .Elle est plus sensible que les microarrays, mais son coût élevé et la complexité de l'analyse restent ses principales limites. (**Hanna et al., 2021**)

II .3 Microarrays (puces à ADN) :

Les micropuces à ADN (ou bio-puces) sont des supports solides, comme du verre ou du silicium, sur lesquels sont fixées des sondes d'ADN permettant de détecter et mesurer l'expression de milliers de gènes simultanément grâce au principe d'hybridation avec des séquences cibles marquées par fluorescence (**Martin Dufva, 2009**).

Les microarrays permettent d'analyser rapidement et à moindre coût l'expression de milliers de gènes en parallèle, avec des applications en génotypage, diagnostic et recherche biomédicale. Toutefois, leur sensibilité est limitée, ils dépendent de sondes connues, présentent des risques de cross-hybridation et nécessitent une analyse technique et bioinformatique complexe (**Timoth, 2001**).

III. Bioinformatique et analyse des données transcriptomiques :

III.1 Définition et importance de la bioinformatique :

La bioinformatique est une discipline interdisciplinaire qui combine l'informatique, les statistiques et la biologie afin d'analyser et d'interpréter les données biologiques massives. En transcriptomique, elle joue un rôle essentiel car les technologies modernes comme le **RNA-seq** et les **puces à ADN (microarrays)** génèrent des volumes considérables de données brutes, nécessitant des approches computationnelles robustes pour leur traitement et leur interprétation (**Conesa et al., 2016**)(**Mitra D et al,2022**).

L'application de méthodes bioinformatiques permet notamment d'identifier les gènes différentiellement exprimés, de détecter des signatures transcriptomiques et de comprendre les mécanismes moléculaires sous-jacents. Sans cette discipline, il serait pratiquement impossible d'extraire des informations biologiquement pertinentes, par exemple dans l'étude de maladies

kkcomplexes comme le diabète de type 1 (DT1), où des centaines de gènes peuvent être régulés simultanément (**Pant et al., 2024**).

III.2 Outils et bases de données bioinformatiques :

Pour mener des analyses transcriptomiques, plusieurs bases de données et outils bioinformatiques sont indispensables. Des bases comme GEO (Gene Expression Omnibus) <https://www.ncbi.nlm.nih.gov/geo/>, Ensembl (<http://www.ensembl.org/>) et GeneCards <https://www.genecards.org/> fournissent un accès à des données transcriptomiques publiques et des annotations génétiques de haute qualité. Les outils analytiques tels que Limma (**Ritchie ME et al., 2015**) et DESeq2 (**Love MI et al., 2014**) sont largement utilisés pour l'analyse différentielle de l'expression génique, tandis que GEO2R offre une interface plus accessible pour explorer les données disponibles dans GEO. Une fois les gènes différentiellement exprimés identifiés, des outils d'enrichissement fonctionnel comme DAVID (<https://davidbioinformatics.nih.gov/>), STRING, (<https://string-db.org/>) Enrichr, (<https://maayanlab.cloud/enrichr-kg>). KEGG et Reactome permettent d'interpréter les résultats en les reliant à des voies biologiques ou à des réseaux de protéines). (**Huang da W et al., 2009**) Enfin, la visualisation des données est essentielle pour faciliter l'interprétation : des représentations comme les volcano plots, les heatmaps, les diagrammes de Venn et les réseaux PPI (protein-protein interaction) permettent de synthétiser visuellement des ensembles complexes de résultats.

Chapitre □

Matériels et méthodes

Cette étude a pour but d'identifier les gènes à expression différentielle et les voies de signalisations impliquées grâce à une analyse *in silico* en utilisant divers outils bioinformatiques.

I. Matériel :

Les analyses de cette étude ont été effectuées sur un micro-ordinateur "Hp" sous Windows 10 et Microsoft Office 2016.

Population :

Selon l'étude menée **par Pant et al., 2024**, un total 62 participants répartis en quatre groupes : 14 sujets atteints de diabète de type 1 récent "ROT1D" (âge, $12,4 \pm 2,8$ ans), 15 frères et sœurs d'individus à haut risque "HRS" (âge, $12,9 \pm 3,4$ ans), 19 frères et sœurs d'individus à faible risque "LRS" (âge, $11,7 \pm 2,8$ ans), et 14 témoins sains non apparentés "uHC" (âge, $15,3 \pm 2,8$ ans).

II. Méthodes :

L'analyse transcriptomique de l'expression génique a été effectuée par des approches *in silico* en utilisant plusieurs outils bio-informatiques, notamment GEO2R,

- Les données de la puce à ADN (GSE232310) ont été extraites à partir de la base de données Gene expression.
(<https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE232310>)
- Cette étude vise à établir les profils d'expression génique des monocytes à l'échelle du génome, à partir de cellules mononuclées du sang périphérique (PBMC), chez des patients présentant un diabète de type 1 récemment diagnostiqué (ROT1D), des frères et sœurs non diabétiques partageant un risque génétique variable, ainsi que des témoins sains non apparentés.

- La puce Affymetrix Human Genome U133 Plus 2.0 a été utilisée pour obtenir l'analyse transcriptomique.

Tableau I : Les effectifs des groupes étudiés.

<u>Groupes</u>	<u>Nombre de patients</u>	<u>Tranche d'âge</u>
Patients atteints de diabète type 1 récemment diagnostiqué (ROT1D).	14	12,4 ± 2,8 ans
Frères et sœurs à haut risque génétique (HRS).	19	12,9 ± 3,4 ans
Frères et sœurs à faible risque génétique (LRS).	15	11,7 ± 2,8 ans
Témoins sains non apparentés (uHC).	14	15,3 ± 2,8 ans

II.1 Identification des gènes différentiellement exprimés

II.1.1 NCBI GEO2R:

NCBI GEO2R <https://www.ncbi.nlm.nih.gov/geo/>

Est un outil web interactif du NCBI qui permet de comparer deux ou plusieurs groupes d'échantillons d'une série GEO issus à partir de résultats bruts de puces à ADN, RNAseq, afin d'identifier les gènes différentiellement exprimés selon les conditions expérimentales. Les résultats sont présentés sous forme de tableau de gènes classés par : le nom des gènes,

l'identifiant, le P value, la localisation chromosomique du gène, le numéro d'accèsion du gène, ainsi que sous forme de graphiques (volcano plot, diagrammes, etc.) pour visualiser ces gènes et évaluer la qualité des données.

NCBI GEO2R utilise le package limma (Linear Models for Microarray Analysis) du projet Bioconductor, basé sur le langage R, pour analyser les données issues des puces à ADN.

Limma applique des modèles linéaires pour identifier les gènes exprimés différemment en fonction des conditions expérimentales.

Nous avons retenu quatre groupes : (patients atteints de diabète de type 1 récemment diagnostiqué (ROT1D), frères et sœurs à haut risque génétique (HRS), frères et sœurs à faible risque génétique (LRS) et témoins sains non apparentés (uHC). La sélection des gènes s'est appuyée sur les paramètres suivants : nom, identifiant, valeur p, localisation chromosomique et le numéro d'accèsion du gène.

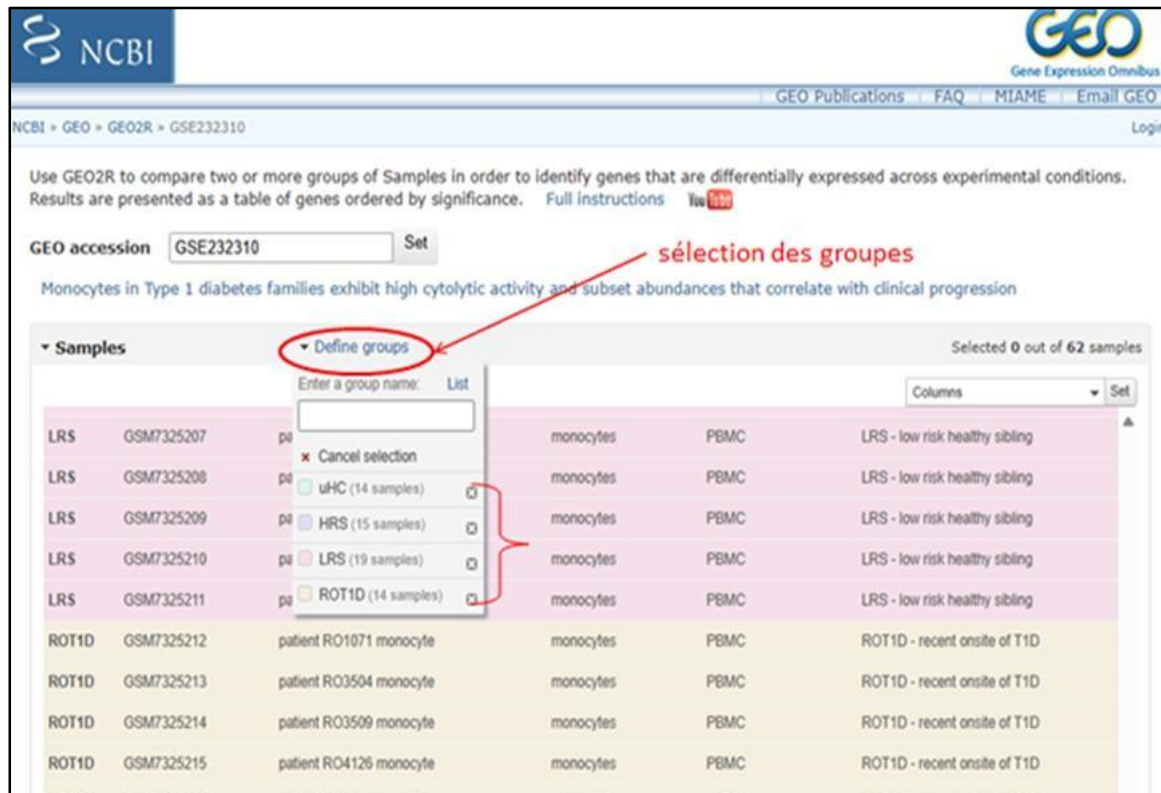


Figure 3 : Représentation de l'outil NCBI GEO et son mode d'emploi.

II.2 Analyse exploratoire, expression différentielle et analyses des voies de signalisation sur des données RNA-seq

Cette section décrit les principaux outils et packages bioinformatiques utilisés pour l'analyse de données d'expression génique, l'identification de gènes différentiellement exprimés, le clustering, l'annotation fonctionnelle et la construction de réseaux.

II.2.1 RStudio :

RStudio est un environnement de développement intégré (IDE) gratuit et open-source pour le langage de programmation statistique R <http://www.posit.co>. Il fournit une interface conviviale pour l'écriture de code R, la gestion de projets, la visualisation de données, le débogage et la génération de rapports. (Posit team, 2025). RStudio a été utilisé comme plateforme principale pour exécuter tous les scripts et packages d'analyse statistique des données d'expression génique.

II.2.2 Limma :

limma (Linear Models for Microarray Data) est un package R de Bioconductor largement utilisé pour l'analyse d'expression différentielle dans les expériences d'omics, y compris le RNA-seq. Il est basé sur des modèles linéaires généralisés et est réputé pour sa robustesse dans l'analyse de petits nombres d'échantillons (**Ritchie ME *et al.*, 2015**). Limma a été employé pour identifier les gènes dont l'expression est significativement modifiée entre les différentes conditions ou groupes étudiés, en intégrant des ajustements pour les effets de batch si nécessaire.

II.2.3 DESeq2 :

DESeq2 est un autre package R populaire de Bioconductor, spécifiquement conçu pour l'analyse d'expression différentielle de données de comptages RNA-seq. Il utilise des modèles statistiques basés sur la distribution binomiale négative pour modéliser les données de comptages brutes et calculer les fold changes et les p-values ajustées (**Love MI *et al.*, 2014**). DESeq2 a été utilisé en complément ou en alternative à limma pour l'identification des gènes différentiellement exprimés, en tenant compte de la variabilité des données de séquençage.

II.2.4 Biclustering (packages Biclust et QUBIC sur R) :

Le biclustering est une technique d'apprentissage automatique qui permet d'identifier des sous-ensembles de gènes qui présentent des profils d'expression similaires à travers des sousensembles d'échantillons. Contrairement au clustering traditionnel qui regroupe soit les gènes, soit les échantillons, le biclustering identifie des blocs rectangulaires dans la matrice d'expression.

- ✦ Biclust : Ce package R fournit une interface pour diverses algorithmes de biclustering, facilitant l'exploration et la visualisation des biclusters ([\(\[https://cran.rproject.org/web/packages/biclust/index.html\]\)](https://cran.rproject.org/web/packages/biclust/index.html)).
- ✦ QUBIC : Ce package R implémente un algorithme de biclustering basé sur des modèles quasi-bicliques, particulièrement efficace pour les données de RNA-seq, où il peut identifier des patterns d'expression cohérents même en présence de bruit ([\(\[http://bioconductor.org/packages/QUBIC\]\)](http://bioconductor.org/packages/QUBIC)). Le biclustering a été appliqué pour découvrir des gènes corrélés sur des sous-ensembles d'échantillons, permettant d'identifier des signatures moléculaires spécifiques à des sous-groupes phénotypiques.

II.3 Annotation d'enrichissement et d'Analyse de Réseaux

II.3.1 STRING-db

Search Tool for the Retrieval of Interacting Genes/Proteins est une base de données et un outil web permettant de construire et de visualiser des réseaux d'interaction protéine-protéine (PPI) (<https://string-db.org/>)

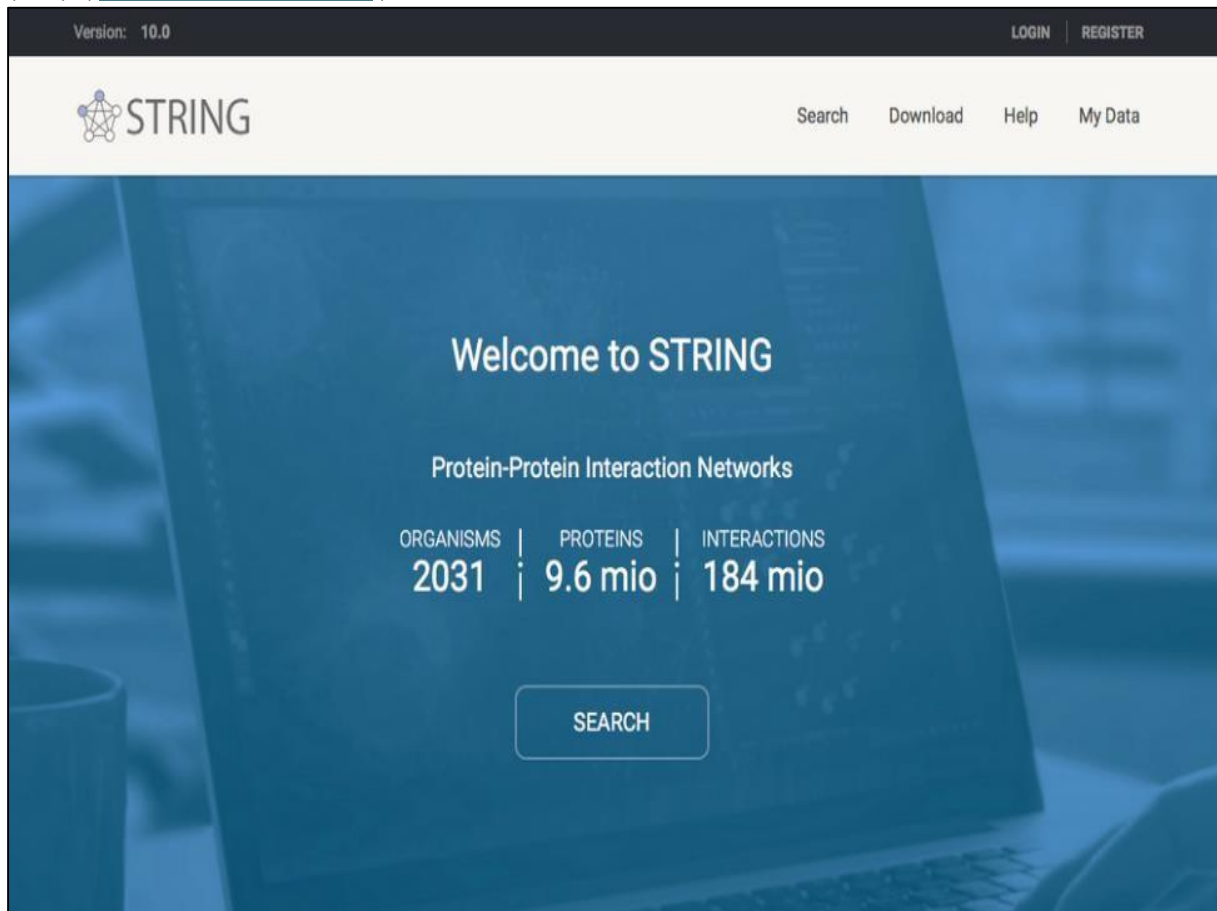


Figure 4 : Présentation de l'outil STRING-db pour la construction et la visualisation des réseaux d'interaction protéine-protéine (PPI), l'annotation fonctionnelle et d'enrichissement.

Il agrège les informations sur les interactions connues (expérimentales, prédictions, co-expression, etc.) provenant de diverses sources (Szklarczyk D *et al.*, 2021)

STRING-db a été utilisé pour explorer les relations fonctionnelles entre les protéines codées par les gènes différentiellement exprimés et pour identifier les protéines clés au sein de ces réseaux.

II.3.2 DAVID :

DAVID (Database for Annotation, Visualization and Integrated Discovery) est une suite d'outils bioinformatiques qui fournit des fonctionnalités d'annotation fonctionnelle et d'enrichissement (<https://davidbioinformatics.nih.gov/>).



Figure 5 : Présentation de l'outil DAVID pour l'annotation fonctionnelle et d'enrichissement.

Il permet d'identifier les processus biologiques (Gene Ontology - GO), les voies de signalisation (KEGG, Reactome) et les domaines protéiques qui sont surreprésentés dans une liste de gènes donnée (**Huang da W *et al.*, 2009**). DAVID a été employé pour réaliser des analyses d'enrichissement afin de comprendre les fonctions biologiques et les mécanismes sous-jacents aux ensembles de gènes régulés.

II.3.3 Enrichr-KG :

Enrichr-KG fait référence à une extension d'Enrichr, une plateforme web robuste pour l'analyse d'enrichissement de gènes(<https://maayanlab.cloud/enrichr-kg>).

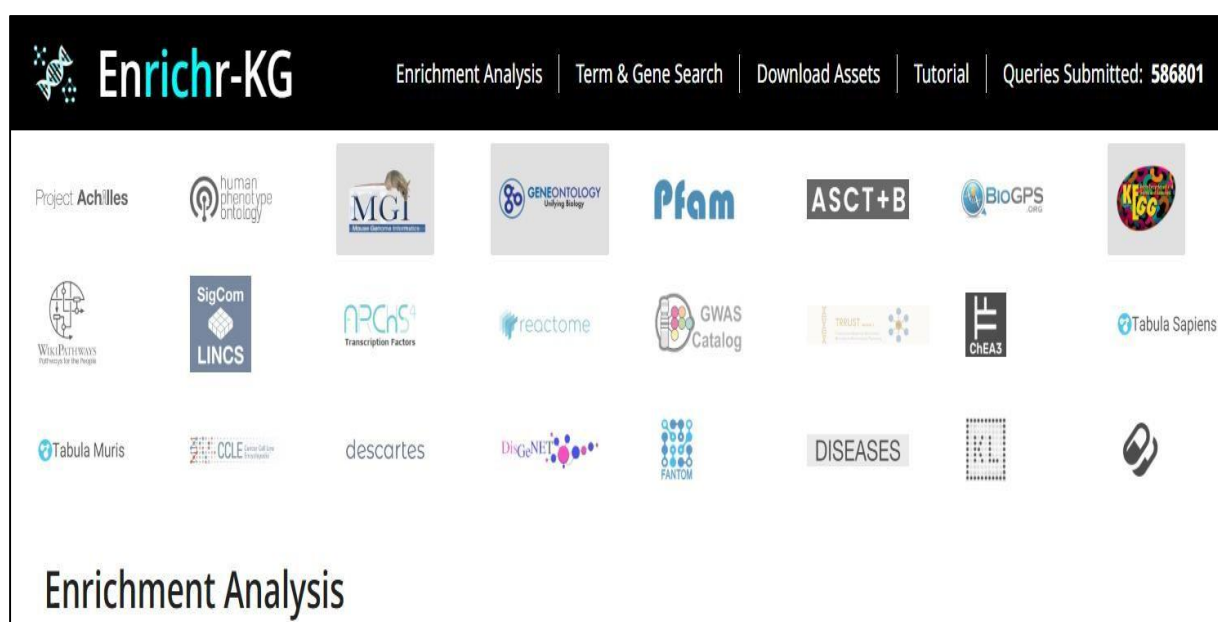


Figure 6: Présentation de l'outil Enrichr-KG pour l'analyse d'enrichissement de gènes.

Enrichr est connu pour sa rapidité et sa vaste collection de bibliothèques d'enrichissement, couvrant des termes GO, des voies, des régulations de miRNA, des prédictions de médicaments, et plus encore (Xie Z et al., 2021). Il a été utilisé pour des analyses d'enrichissement fonctionnel complémentaires à DAVID, permettant d'obtenir une perspective plus large sur les processus biologiques associés aux gènes d'intérêt.

II.3.4 NetworkAnalyst :

NetworkAnalyst est une plateforme web-based conviviale pour l'analyse de réseaux omiques (<https://www.expressanalyst.ca>), la visualisation et l'intégration de données. Elle permet la construction de réseaux PPI, l'analyse d'enrichissement de voies, l'identification de gènes hub et de modules (Xia J et al., 2015).

NetworkAnalyst a été utilisé pour des analyses de réseaux complexes et l'intégration des résultats d'expression différentielle et d'enrichissement, offrant une vue systémique des interactions moléculaires.

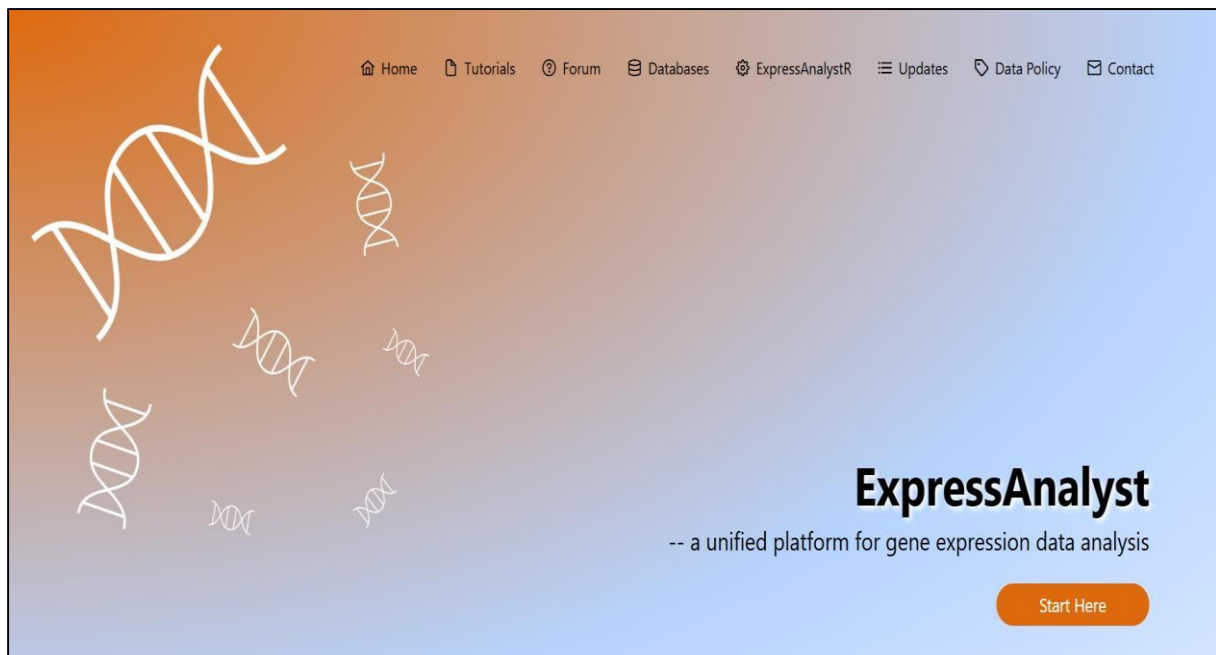


Figure 7 : Présentation de l’outil NetworkAnalyst pour l’analyse de réseaux omiques.

II.4 Annotation des gènes exprimées de manières différentielles

II.4.1 ENSEMBL :

ENSEMBL est une base de données génomique majeure qui fournit des annotations complètes de génomes (gènes, transcrits, variants, etc.) pour des dizaines d'espèces eucaryotes (<http://www.ensembl.org/>).

Name	Description	Online tool	Upload limit	Download script	Documentation
Ensembl Variant Effect Predictor	Analyse your own variants and predict the functional consequences of known and unknown variants via our Ensembl Variant Effect Predictor (VEP) tool.		50MB*		
Variant Recoder	Translate a variant identifier, HGVS notation or genomic SPDI notation to all possible variant IDs, HGVS, VCF format and genomic SPDI.		Maximum 1000 variants recommended		
BLAST/BLAT	Search our genomes for your DNA or protein sequence.		50MB		
File Chameleon	Convert Ensembl files for use with other analysis tools				
Assembly Converter	Map (liftover) your data's coordinates to the current assembly		50MB		
ID History Converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.		50MB		
Linkage Disequilibrium Calculator	Calculate LD between variants using genotypes from a selected population.				
VCF to PED converter	Parse a vcf file to create a linkage pedigree file (ped) and a marker information file, which together may be loaded into id visualization tools like Haploview.				
Data Slicer	Get a subset of data from a BAM or VCF file.				

* For larger datasets we provide an API script that can be downloaded (you will also need to install our Perl API, below, to run the script).

Figure 8 : Présentation de la base des données ENSEMBL pour l’annotations complètes de génomes.

Il sert de ressource fondamentale pour obtenir des informations détaillées sur les gènes (IDs, symboles, localisations chromosomiques, types de gènes comme codants ou lncRNA) et pour la cartographie des lectures de séquençage. ENSEMBL a été utilisé pour l'annotation et la caractérisation des gènes identifiés dans les analyses différentielles.

II.4.2 GeneCards :

GeneCards <https://www.genecards.org/> est une base de données génétique humaine intégrative et interrogeable qui fournit des informations complètes sur tous les gènes humains annotés ou prédits. Elle rassemble automatiquement des données provenant d'environ 200 sources, incluant des informations génomiques, transcriptomiques, protéomiques, génétiques, cliniques et fonctionnelles.

Son principe repose sur l'unification de ces données pour permettre une navigation efficace entre gènes, maladies, variants, protéines, cellules et voies biologiques. GeneCards propose aussi des outils associés comme MalaCards pour les maladies, PathCards pour les voies biologiques, et VarElect pour prioriser les phénotypes en lien avec des variants, utile dans les projets de séquençage de nouvelle génération (Stelzer *et al.*, 2016)



Figure 9 : Présentation de la base de données GeneCards pour la recherche de gènes et son mode d'utilisation.

II.4.3 Metascape :

Metascape <https://metascape.org/gp/index.html#/main/step1> est une plateforme en ligne gratuite dédiée à l'analyse bio-informatique des listes de gènes. Elle permet de découvrir les fonctions biologiques associées aux gènes, d'identifier les voies de signalisation dans lesquelles ils sont impliqués, et d'explorer les interactions entre protéines. En utilisant plusieurs bases de données, telles que Gene Ontology (GO), KEGG, et Reactome, Metascape offre une analyse complète qui aide à interpréter les résultats d'études génomiques (Zhou *et al.*, 2019).

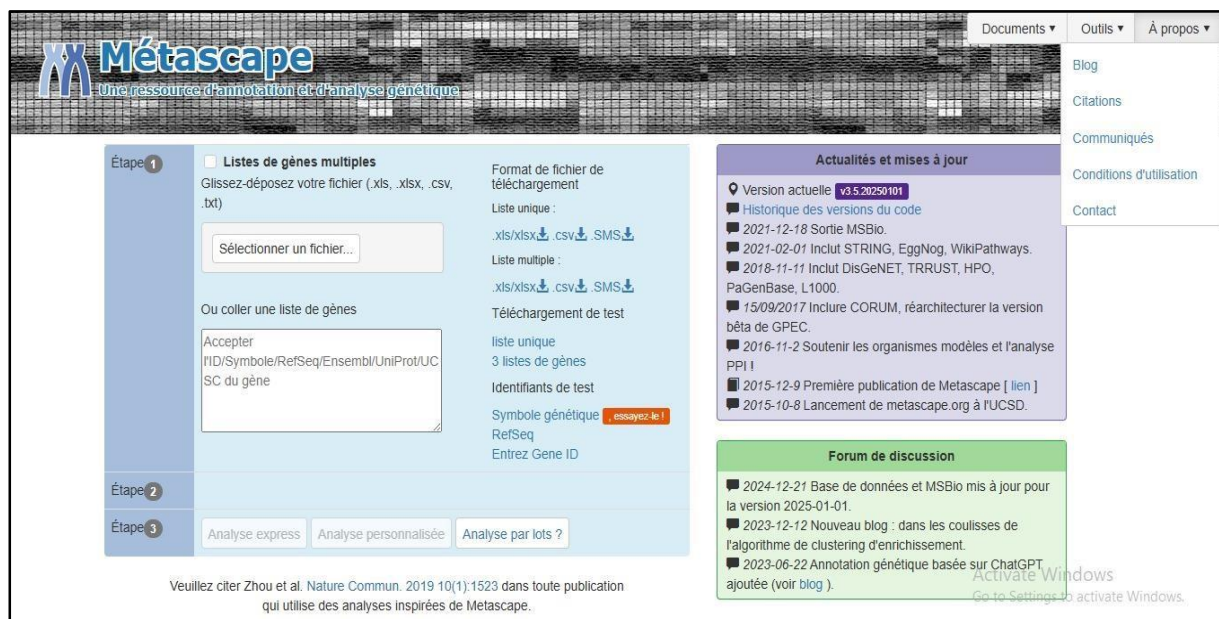


Figure 10 : Présentation de l'outil Metascape pour l'analyse de listes de gènes.

Chapitre □ I

Résultats et Discussions

I. Analyse profil d'expressions des gènes dans nos groupe d'étude

Les résultats obtenus À partir de l'analyse GEO2R réalisée avec l'outil limma, un total de **54196 gènes** ont été analysé dans la série **GSE232310**, en appliquant un seuil de significativité ajusté ($P_{adj} < 0,05$). Le diagramme de Venn **Figure 11** a permis de visualiser les chevauchements de gènes significativement exprimés entre les quatre groupes étudiés.

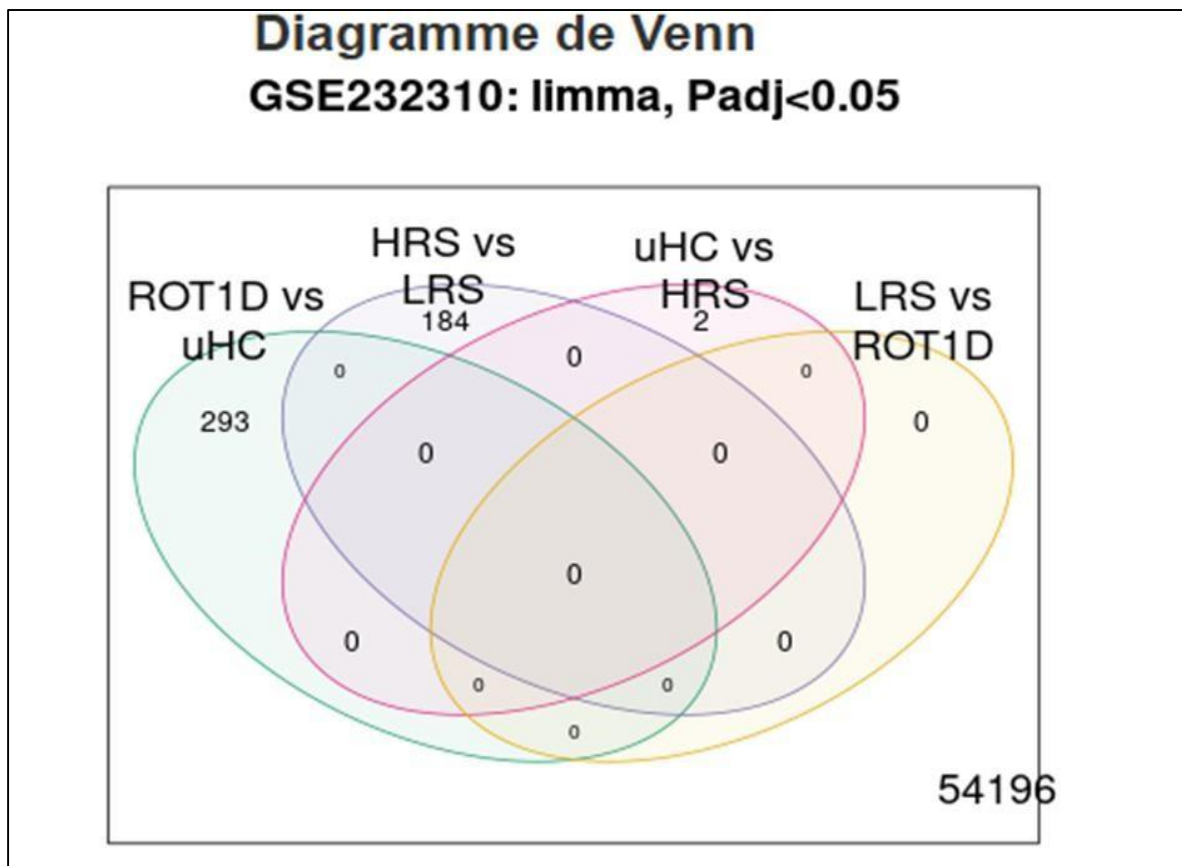


Figure 11: Représentation du diagramme de Venn illustrant le chevauchement des gènes significativement exprimés entre les différents groupes étudiés.

D'après la **figure 13**, les zones de chevauchement entre les cercles du diagramme de Venn montrent le nombre de gènes communs entre plusieurs comparaisons, quatre comparaisons ont été faites : **(ROT1D vs uHC), (HRS vs LRS), (uHC vs HRS) et (LRS vs ROT1D).**

On peut constater que le nombre total de gènes testés dans cette étude est de 54 196. Parmi

ceux-ci, 293 gènes sont exprimés de manière différentielle dans la comparaison entre ROT1D et uHC. En outre, 184 gènes présentent une expression différentielle chez HRS contre LRS. Enfin, 2 gènes sont spécifiquement exprimés de manière à l'intersection de "ROT1D vs uHC" et "HRS vs LRS" correspond à des gènes significativement exprimés dans ces deux comparaisons, le "0" dans la zone de chevauchement signifie qu'aucun gène n'est partagé entre ces groupes selon le seuil de significativité ($P_{adj} < 0,05$) différentielle entre les groupes uHC vs HRS.

I.1 ROT1D contre uHC :

Le diagramme volcan (volcanoplot), généré via l'outil Limma, permet de représenter la significativité statistique ($-\log_{10} P$) en fonction de l'ampleur du changement d'expression ($\log_2$ fold change) pour identifier visuellement les gènes différentiellement exprimés entre les groupes d'étude.

La **Figure 12** présente un graphe en volcan (volcano plot) illustrant les résultats de l'analyse d'expression différentielle des gènes dans les monocytes, dans le contexte de la base de données GSE232310 comparant les échantillons ROT1D apparentés à des patients avec un risque de diabète de type 1 aux échantillons uHC (contrôles sains non apparentés). L'axe des X représente le $\log_2$ (changement de pli) ($\log_2$ (fold change)), quantifiant la magnitude de l'expression des gènes, tandis que l'axe des Y représente la significativité statistique en $-\log_{10}$ (p-valeur).

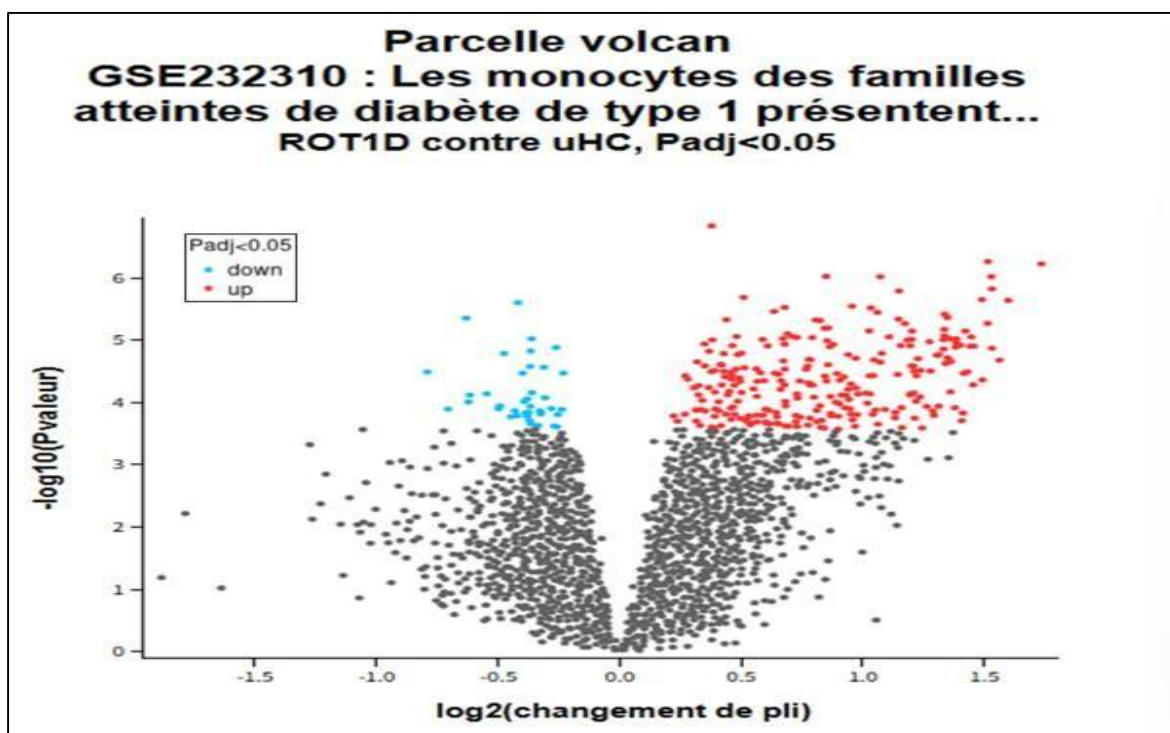


Figure 12 : Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'étude « ROT1D vs uHC ».

Les points rouges (à droite) indiquent les gènes significativement surexprimés chez ROT1D ($\log_2(\text{fold change})$ positif, et p-valeur très faible (≤ 0.0001). Les points bleus (à gauche) représentent les gènes significativement sous-exprimés chez ROT1D ($\log_2(\text{fold change})$ négatif, et p-valeur très faible < 0.0001). Les points gris correspondent aux gènes dont l'expression n'est pas significativement modifiée. Le graphe révèle un nombre plus important de gènes significativement surexprimés (points rouges) par rapport aux gènes sous-exprimés (points bleus), indiquant une asymétrie dans la réponse transcriptomique des monocytes.

L'analyse intégrée de l'expression différentielle et des voies via le package DESeq2, nous a permis de mettre en œuvre l'Analyse exploratoire, l'expression différentielle et les analyses de voies sur des 293 gènes trouvés dans le groupe de recherche ROT1D vs uHC.

Le résultat de cette analyse (**Figure 13**) démontre que 245 gènes ont passé le filtre, 221 ont été convertis en Ensembl_ID de gène. Les 24 gènes restants ont été conservés dans les données en utilisant les identifiants originaux.

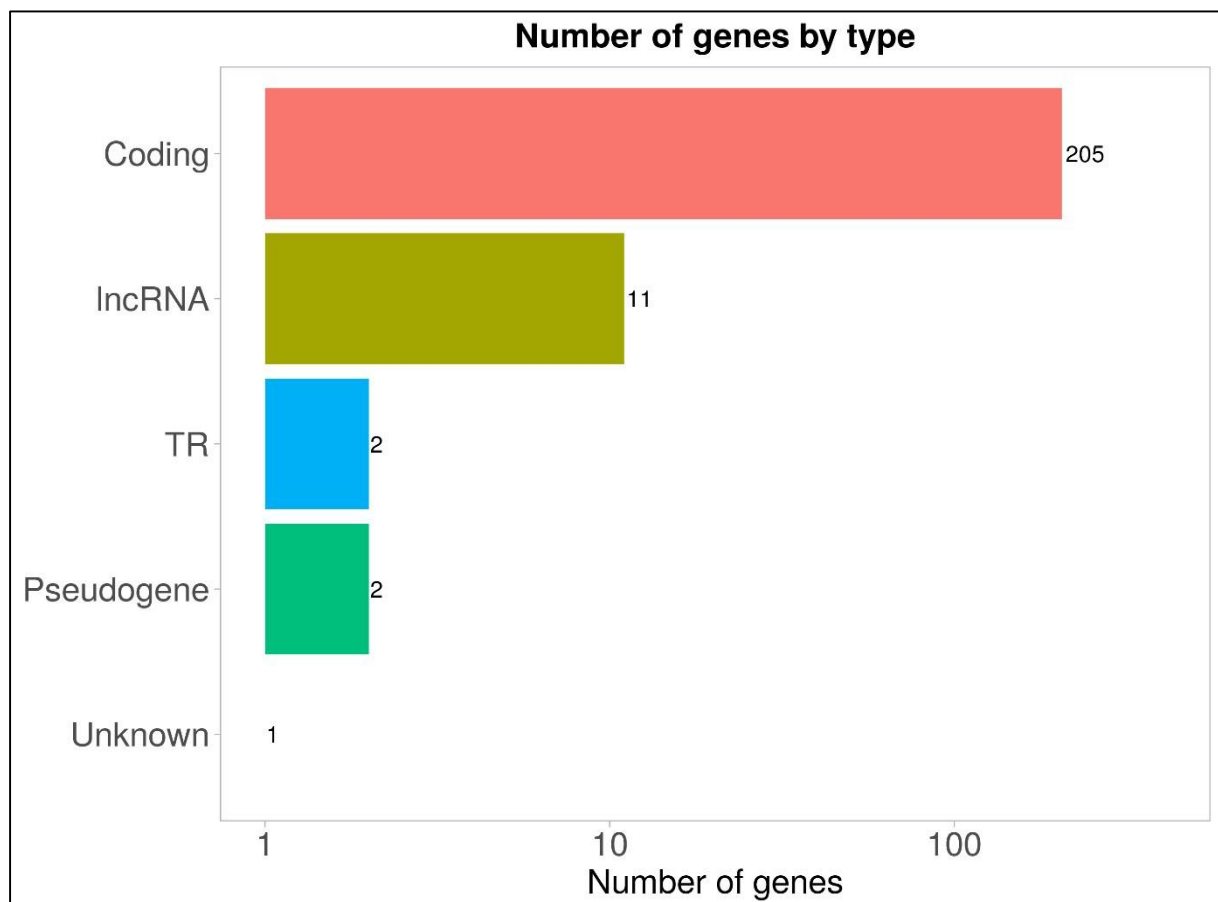


Figure 13 : Distribution des gènes identifiés par type dans le groupe des gènes « ROT1D vs uHC ».

La classification des gènes identifiés par type a révélé une prédominance marquée des gènes codants. Sur un total de 221 gènes, 205 (environ 92.7%) étaient des gènes codants. Les ARN non codants longs (lncRNA) représentaient le deuxième type le plus abondant avec 11 gènes (environ 5%). Les gènes codant pour les récepteurs (TR) et les pseudogènes étaient chacun représentés par 2 gènes (environ 0.9% chacun). Enfin, un seul gène est resté de type inconnu, soulignant la robustesse de l'annotation.

La distribution des gènes par type, dominée par les gènes codants, est cohérente avec la plupart des analyses transcriptomiques visant à identifier des processus biologiques ou des réponses immunitaires. La forte proportion de gènes codants suggère que les processus d'intérêt sont principalement médiatisés par des protéines. La présence de lncRNA, bien que minoritaire, est notable. Ces molécules jouent des rôles régulateurs de plus en plus reconnus dans diverses voies biologiques, y compris l'immunité, et leur identification pourrait indiquer des mécanismes de régulation transcriptionnelle ou post-transcriptionnelle des processus étudiés. Les faibles nombres de gènes TR et de pseudogènes sont attendus dans ce contexte, ces derniers étant souvent filtrés ou jouant un rôle moins direct dans l'expression des phénotypes.

L'identification d'un gène 'inconnu' justifierait une investigation plus approfondie pour améliorer notre compréhension fonctionnelle des éléments moins caractérisés.

En utilisant l'approche du biclustering avec les packages Biclust et QUBIC sur R nous avons pu découvrir les gènes corrélés sur le sous-ensemble d'échantillons de nos deux groupes. **(Figure 14)**

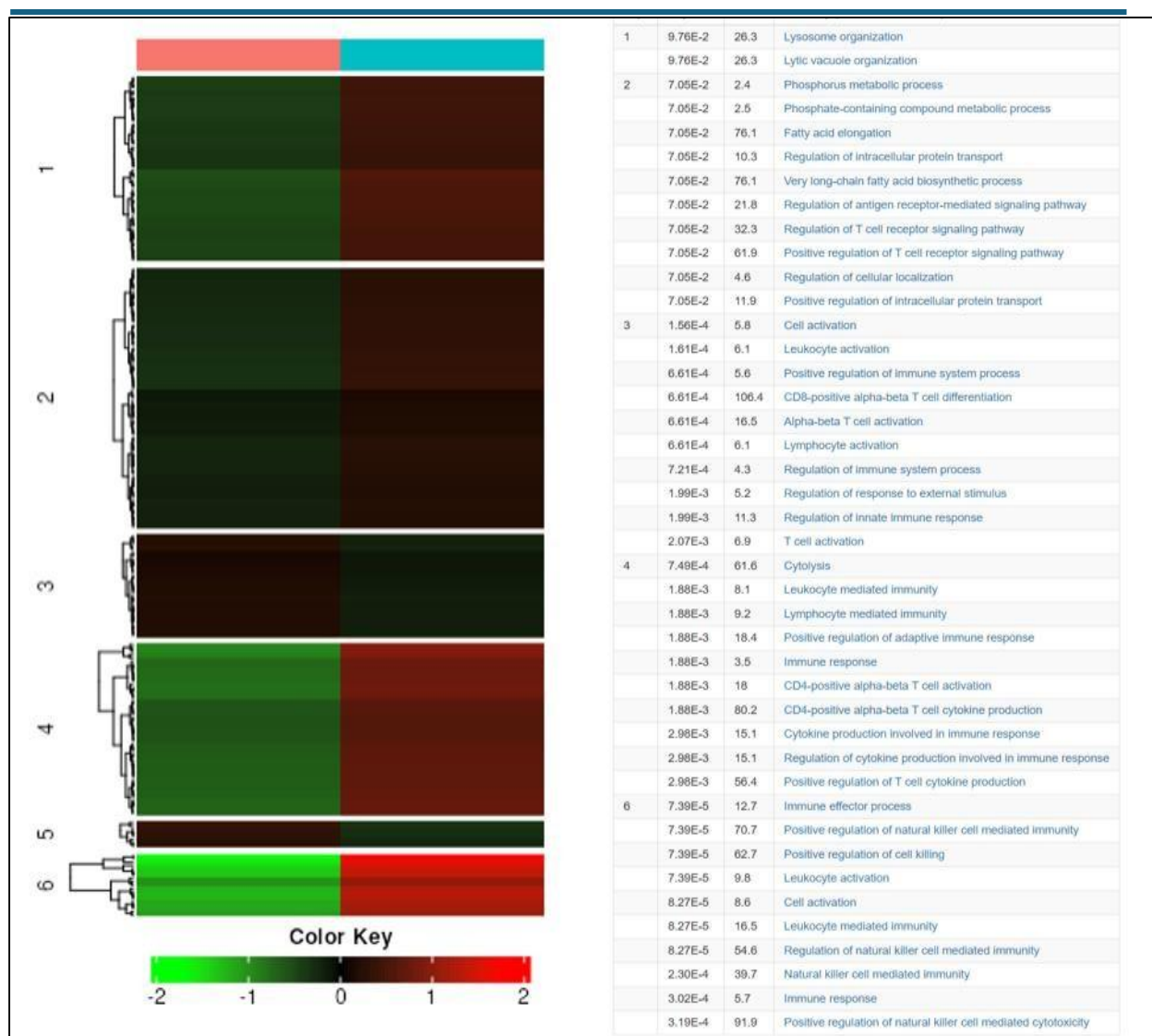


Figure 14 : Heatmap des termes GO et des processus biologiques enrichis dans l'analyse transcriptomique du groupe des gènes « ROT1D vs uHC ».

La **Figure 14** présente une heatmap des termes Gene Ontology (GO) et des processus biologiques significativement enrichis, regroupés par clustering hiérarchique. La partie gauche de la figure (heatmap) illustre le niveau d'expression des gènes au sein des différentes catégories, où le rouge indique une surexpression et le vert une sous-expression. La partie droite détaille les termes GO enrichis avec leurs p-values associées. On observe des clusters de gènes (numérotés de 1 à 6) présentant des profils d'expression distincts. Par exemple, le cluster 1 est fortement associé à des processus de 'Lysosome organization' et 'Vacuole organization', tandis que le cluster 6 regroupe des gènes impliqués dans 'Cell activation', 'Leukocyte mediated immunity' et 'Natural killer cell mediated immunity'

L'analyse des termes GO enrichis, visualisée dans la **Figure 14**, fournit des informations cruciales sur les processus biologiques altérés dans notre groupe d'étude. La surexpression des gènes du cluster 6, notamment ceux liés à l'activation cellulaire, l'immunité à médiation leucocytaire et l'immunité à médiation par les cellules Natural Killer (NK), est particulièrement pertinente. Cela indique une activation robuste et spécifique de la réponse immunitaire de l'hôte, potentiellement en réponse à la maladie auto-immune. Ces résultats sont cohérents avec le diabète type I.

Inversement, les gènes des clusters supérieurs (ex : cluster 1), associés à des fonctions métaboliques ou de transport intracellulaire, montrent des profils d'expression variables, suggérant des perturbations dans le métabolisme ou la machinerie cellulaire de ce groupe d'étude. La capacité à identifier des processus spécifiques comme la 'Positive regulation of T cell receptor signaling pathway' (cluster 1 et 2) et la 'CD4-positive alpha-beta T cell activation' (cluster 4) renforce l'idée d'une réponse immunitaire adaptative impliquée. La présence de 'Pseudogene' ou 'TR' dans la classification des gènes pourrait expliquer leur rôle potentiel dans ces processus enrichis.

La **Figure 15** présente une visualisation des processus biologiques et des voies de signalisation significativement enrichis, basée sur l'analyse des gènes différentiellement exprimés. L'axe des Y représente les différents termes ou voies identifiés, tandis que la taille des cercles indique le nombre de gènes impliqués dans chaque voie ('Gene count'). La coloration des cercles, allant du vert clair au bleu foncé, reflète la significativité de l'enrichissement (p-value ajustée, FDR), le bleu foncé indiquant les enrichissements les plus statistiquement significatifs ($FDR \leq 2.0 \times 10^{-5}$).

Les processus les plus fortement enrichis et impliquant un nombre important de gènes sont dominés par les thèmes liés à la réponse immunitaire et à l'activation cellulaire, comme l'"activation des leucocytes" et l'"activation des lymphocytes", qui montrent les p-values ajustées les plus faibles et un nombre élevé de gènes.

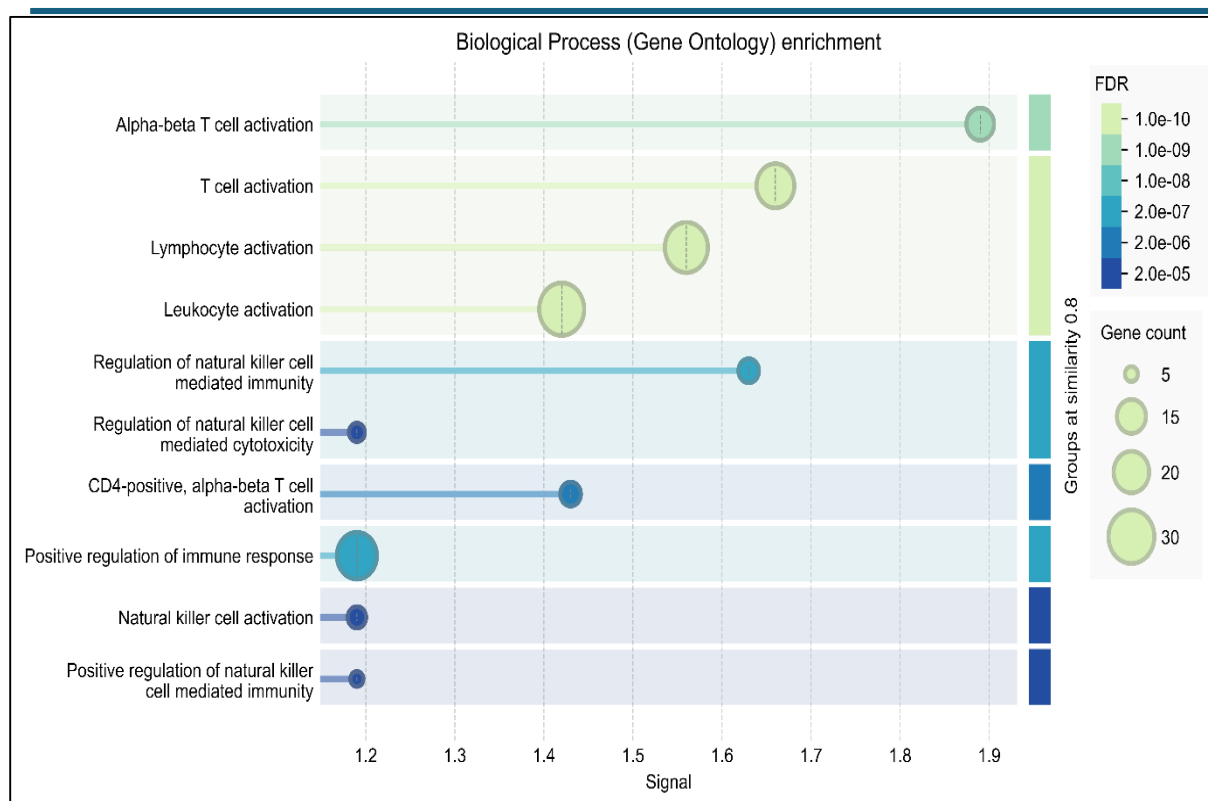


Figure 15 : Représentation graphique des principales voies et processus biologiques enrichis, classés par signification et nombre de gènes du groupe des gènes « ROT1D vs uHC ».

Le **Tableau II** présente les dix principaux processus biologiques (Gene Ontology) et voies de signalisation qui se sont avérés significativement enrichis parmi les gènes identifiés dans notre analyse. Ces enrichissements sont classés par p-value ajustée (Adj.Pval). Les processus immunologiques sont fortement représentés, avec l'"activation des leucocytes" (Adj.Pval = 1.02×10^{-7} , Fold = 11.2) en tête de liste, suivi par l'"activation des lymphocytes" (Adj.Pval = 2.83×10^{-6} , Fold = 7.9) et l'"activation cellulaire" (Adj.Pval = 2.84×10^{-5} , Fold = 8).

Des voies spécifiques telles que la 'différenciation des cellules Th1 et Th2' et la 'cytotoxicité médiatisée par les cellules Natural Killer' sont également mises en évidence. Des processus plus généraux comme la 'réponse immunitaire' (Adj.Pval = 2.94×10^{-3} , Fold = 6.8) et le 'processus du système immunitaire' (Adj.Pval = 3.07×10^{-2} , Fold = 6.6) confirment l'ampleur de l'engagement immunitaire."

Tableau II : Top 10 des processus biologiques (GO) et des voies de signalisation significativement enrichis.

N°	Adj.Pval	Fold	Go Biological Process	Pathways
1	1.02E-7	11.2	Leukocyte activation	Th1 and Th2 cell differentiation
2	2.83E-6	7.9	Lymphocyte activation	Natural killer cell mediated cytotoxicity
3	2.84E-5	8	Cell activation	Th17 cell differentiation
4	1.20E-3	6.7	Alpha-beta T cell activation	T cell receptor signaling pathway
5	2.94E-3	6.8	Immune response	PD-L1 expression and PD-1 checkpoint pathway in cancer
6	3.07E-2	6.6	Immune system process	Inflammatory bowel disease
7	3.14E-2	2.4	T cell activation	Pathways in cancer
8	3.98E-2	3.3	Immune response-activating cell surface receptor signaling pathway	Ras signaling pathway
9	4.92E-2	3.9	Immune response-activating signal transduction	Cellular senescence
10	7.04E-2	6.1	Positive regulation of immune response	Hedgehog signaling pathway

L'analyse d'enrichissement des termes GO et des voies de signalisation (**Tableau II**) corrobore et approfondit l'activation robuste de la réponse immunitaire de l'hôte, telle qu'observée précédemment. La prédominance de termes comme 'activation des leucocytes', 'activation des lymphocytes', et 'activation cellulaire' avec des valeurs de Fold Change élevées (>7) suggère une réponse immunitaire cellulaire intense et globalement activée en réponse au diabète type I.

La mise en évidence de la 'différenciation des cellules Th1 et Th2' est particulièrement significative, indiquant une polarisation potentielle de la réponse T helper, cruciale dans la régulation de l'immunité contre les pathogènes intracellulaires ou extracellulaires, et pertinente pour la leishmaniose. La détection de la 'cytotoxicité médiatisée par les cellules Natural Killer' et de l'activation des cellules T alpha-bêta' confirme l'implication des mécanismes de l'immunité innée et adaptative cellulaire.

L'enrichissement de voies comme 'PD-L1 expression and PD-1 checkpoint pathway in cancer' et 'Pathways in cancer' peut sembler inattendu dans un contexte non-cancéreux mais souligne l'interconnexion des voies de signalisation. Le point de contrôle PD-1/PD-L1 est un régulateur clé de la réponse immunitaire et son engagement peut moduler la persistance ou

l'élimination de l'agent pathogène. L'enrichissement de la 'voie de signalisation Ras' et de la 'sénescence cellulaire' suggère des processus de signalisation cellulaire et de vieillissement cellulaire qui pourraient être modulés pendant le diabète type I, méritant des investigations plus approfondies pour comprendre leur rôle précis dans la pathogénie ou la résolution.

Selon l'étude menée par **Irvine KM *et al.*, 2012**, Les chercheurs ont caractérisé l'expression génique des monocytes chez des patients diabétiques de type 1 sous insuline, environ 3 mois après le diagnostic. Ils ont ensuite analysé les caractéristiques cliniques de ces patients du diagnostic jusqu'à un an après. L'étude a identifié un profil d'expression génique spécifique dans les monocytes du sang périphérique qui permet de stratifier cliniquement le contrôle glycémique des patients au cours de la première année suivant le diagnostic de diabète de type 1. Ce qui suggère que les monocytes pourraient refléter l'état inflammatoire et la progression de la maladie.

La **Figure 16** présente le réseau d'interaction protéine-protéine (PPI) construit à partir des gènes différentiellement exprimés identifiés dans notre étude. Chaque nœud représente une protéine (produit d'un gène), et les arêtes (lignes) entre les nœuds indiquent des interactions connues entre ces protéines. La densité des interconnexions suggère un réseau complexe et fortement régulé. On observe des clusters de nœuds, notamment un grand cluster central densément connecté (en rouge), indiquant un rôle pivot de ces protéines dans les processus biologiques sous-jacents. Des nœuds périphériques (en vert, jaune, bleu) ou moins connectés sont également présents, certains pouvant interagir avec le cœur du réseau, tandis que d'autres pourraient avoir des rôles plus spécifiques ou marginaux dans les conditions étudiées.

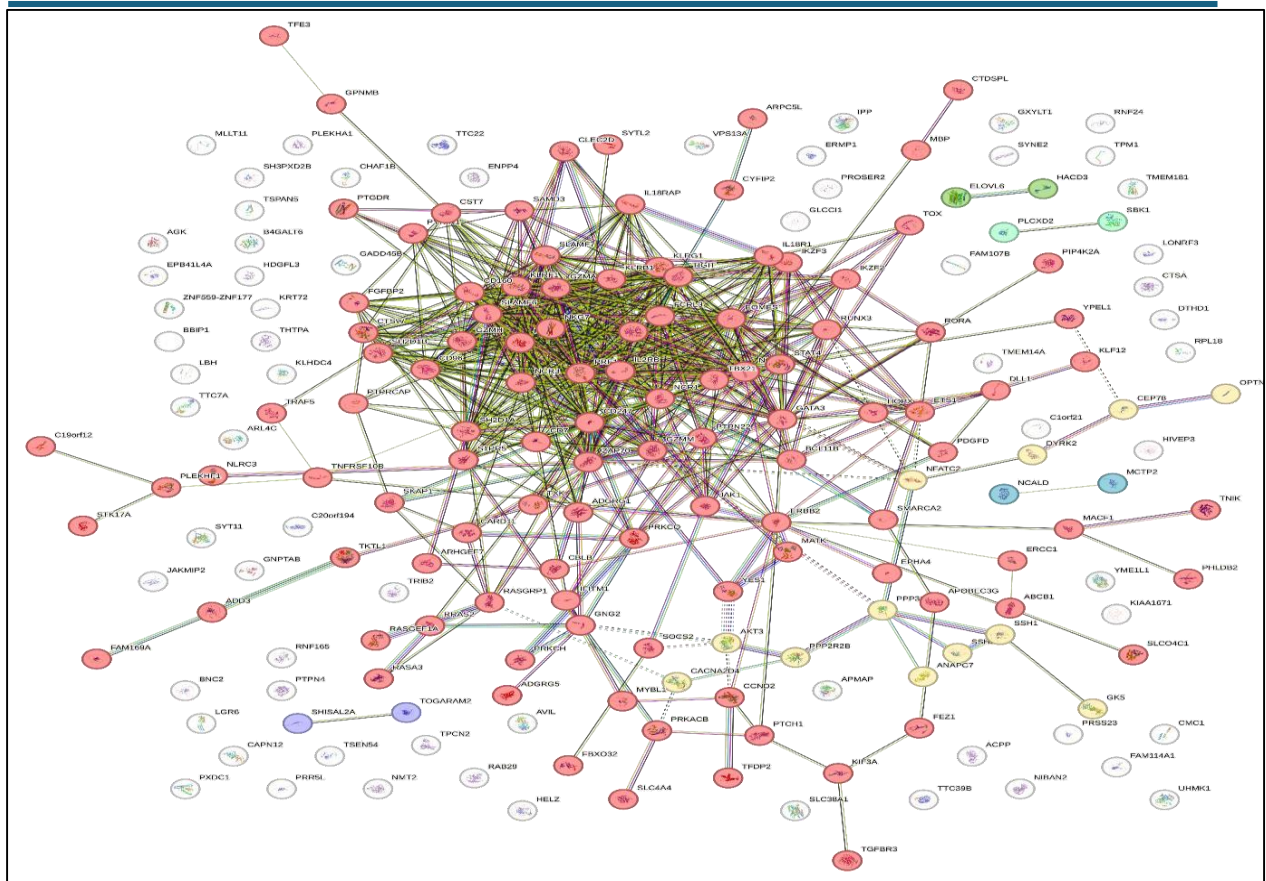


Figure 16 : Réseau d'interaction protéine-protéine (PPI) des gènes différentiellement exprimés gènes du groupe des gènes « ROT1D vs uHC ».

L'analyse du réseau d'interaction protéine-protéine (**Figure 16**) fournit des informations essentielles sur les relations fonctionnelles et la coordination entre les gènes dont l'expression est altérée. La structure du réseau, caractérisée par un cluster central très interconnecté, suggère l'existence de protéines clés (hub proteins) qui orchestrent et modulent des processus biologiques fondamentaux. Ces 'hubs' représentent des points d'intégration pour de multiples signaux, et leur identification est cruciale pour comprendre la pathogénèse du diabète type I.

La présence de gènes liés à l'activation immunitaire et à l'inflammation dans ce réseau central corrobore les résultats des analyses d'enrichissement de termes GO (**Figure 16** et **Tableau II**). Cela indique non seulement que ces gènes sont différentiellement exprimés, mais qu'ils agissent également de concert au sein de modules fonctionnels.

Les interactions complexes observées dans ce réseau peuvent également mettre en lumière des voies de compensation ou de régulation croisée qui pourraient ne pas être apparentes avec une simple analyse d'expression différentielle.

L'identification de sous-réseaux ou de modules au sein du PPI pourrait révéler des mécanismes biologiques distincts ou des phases différentes de la réponse à ROT1D vs uHC.

Ces résultats ouvrent des perspectives pour des études fonctionnelles ciblées visant à valider le rôle des protéines centrales et à explorer de nouvelles stratégies thérapeutiques agissant sur ces réseaux d'interactions."

La **Figure 17** présente l'expression différentielle (exprimée en $\log_2(\text{Fold Change})$) des gènes spécifiques identifiés comme étant des acteurs clés dans les processus biologiques mis en évidence par nos analyses précédentes (**Figure 15** et **Tableau II**). Les valeurs positives de $\log_2(\text{Fold Change})$ indiquent une surexpression de ces gènes dans la condition étudiée.

On observe que tous les gènes représentés sont surexprimés. Parmi eux, SBK1, SOCS2, et SBK2 montrent les plus fortes augmentations d'expression, avec des $\log_2(\text{Fold Change})$ supérieurs à 0.6. D'autres gènes comme CBLB, AKT3, et PRKACB sont également surexprimés de manière significative. Les gènes CALM3 / CALM2 / CALM1 présentent des niveaux de surexpression plus modérés, mais néanmoins présents.

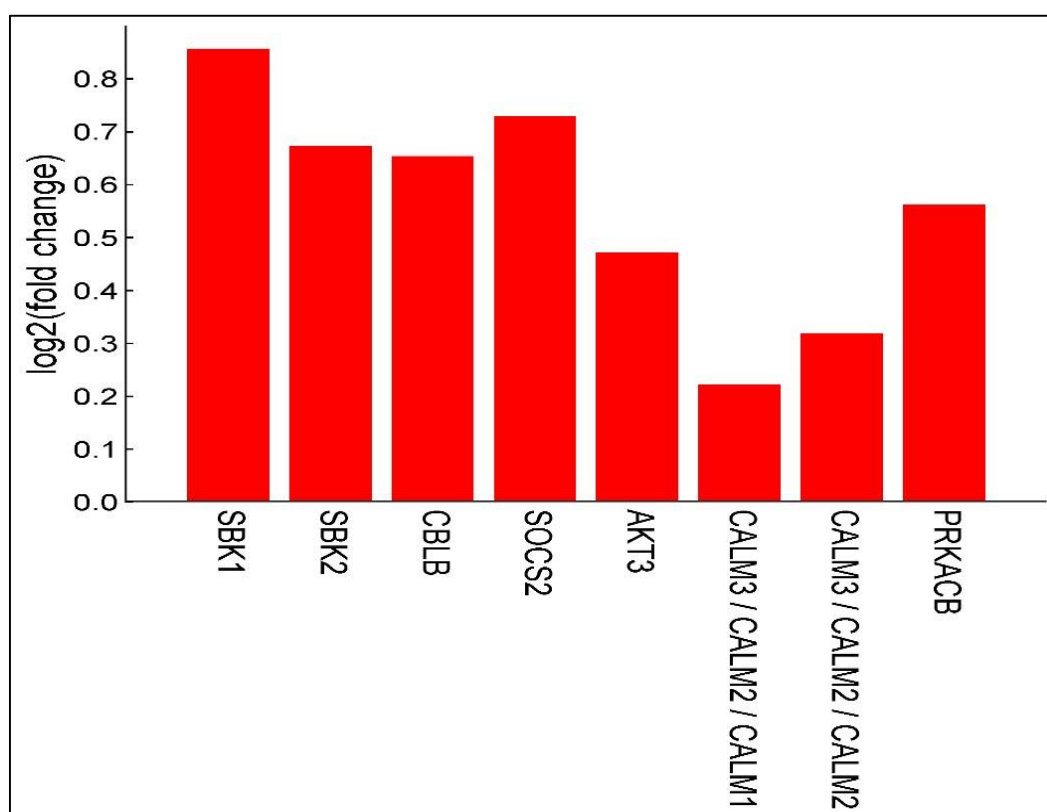


Figure 17 : Expression relative ($\log_2(\text{Fold Change})$) des gènes clés identifiés dans les voies immunitaires et de signalisation du groupe des gènes « ROT1D vs uHC ».

L'analyse ciblée de l'expression des gènes sélectionnés (**Figure 17**) vient confirmer et affiner les observations issues des analyses d'enrichissement et des réseaux d'interaction protéine-protéine (**Figure 15, Tableau II et Figure 17**).

La surexpression de gènes tels que SBK1 et SOCS2, ainsi que CBLB et AKT3, suggère une modulation complexe des voies de signalisation intracellulaire en réponse au diabète.

- ✦ SOCS2 (Suppressor of Cytokine Signaling 2) est un régulateur clé des voies de signalisation des cytokines, souvent impliqué dans la régulation négative de l'inflammation et de la réponse immunitaire. Sa surexpression pourrait indiquer un mécanisme de rétroaction négative visant à moduler une réponse immunitaire initialement intense. (**Carow B et al., 2014**)
- ✦ CBLB (Casitas B-cell lymphoma B) est un E3 ubiquitine ligase qui régule négativement l'activation des cellules T. Son augmentation d'expression pourrait participer à la modulation de l'immunité adaptative, un point crucial dans les infections chroniques comme la leishmaniose où la balance entre activation et régulation immunitaire est essentielle. (**Paolino M et al., 2010**)
- ✦ AKT3 est une isoforme de la kinase AKT, une voie majeure dans la survie cellulaire, la prolifération et le métabolisme, mais aussi impliquée dans la régulation immunitaire. Sa surexpression suggère une activation de cette voie dans le contexte de la réponse de l'hôte. (**Carow B et al., 2014**)
- ✦ La présence de gènes liés à la calmoduline (CALM1, CALM2, CALM3), bien que moins fortement surexprimés, est pertinente. La calmoduline est une protéine de liaison au calcium impliquée dans une multitude de processus cellulaires, y compris la signalisation des cellules immunitaires. (**Yang CF et al., 2021**)

Nous avons procédé à une analyse d'enrichissement fonctionnel. Cette étape a consisté à interroger des bases de données de connaissances établies, la base de données KEGG (Kyoto Encyclopedia of Genes and Genomes) et Gene Ontology (GO),

L'identification de la voie de signalisation de l'insuline comme significativement impactée découle directement d'une approche d'analyse systématique et non biaisée de nos données transcriptomiques. Après l'identification des gènes différentiellement exprimés (mise en évidence dans la **Figure 17** pour des gènes clés comme SOCS2, CBLB, AKT3),

La voie de signalisation de l'insuline a émergé comme l'une des voies statistiquement les plus enrichies parmi nos gènes régulés. Ceci signifie qu'un nombre significativement plus élevé de gènes de cette voie étaient différemment exprimés que ce qui serait attendu par simple hasard.

La localisation de gènes spécifiques tels que 'IRS', 'PI3K', 'AKT' et 'SOCS2' au sein de cette voie (**Figure 18**) gènes dont nous avons confirmé la surexpression (**Figure 17**) a renforcé la pertinence biologique de cette découverte.

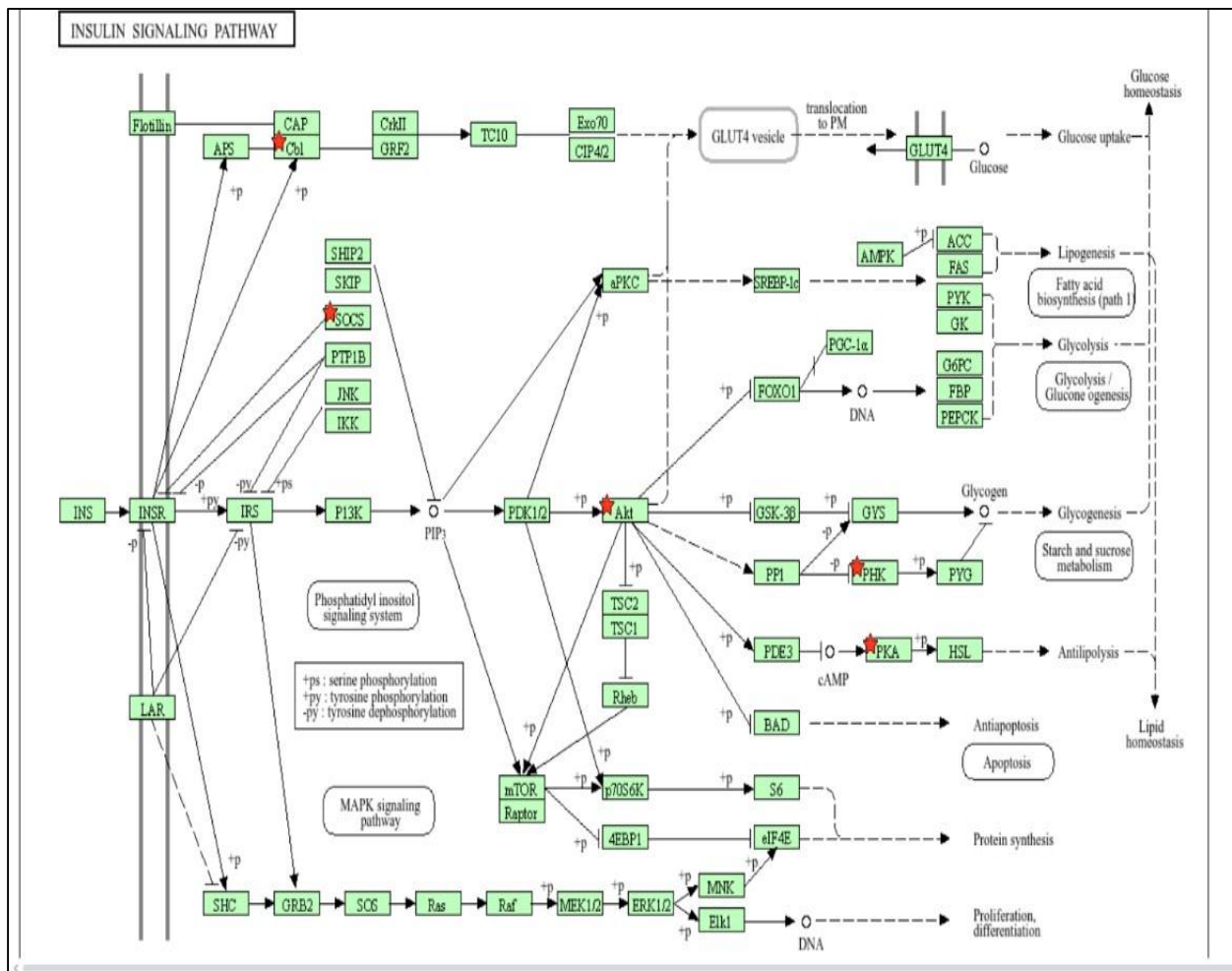


Figure 18 : Représentation schématique de la voie de signalisation de l'insuline avec indication des gènes régulés du groupe des gènes « ROT1D vs uHC ».

L'analyse du réseau d'interaction protéine-protéine (Figure 18) a par ailleurs montré que plusieurs de ces gènes de la voie de l'insuline sont interconnectés, suggérant une perturbation coordonnée de ce système.

La **Figure 18** illustre la voie de signalisation de l'insuline récupérée de la base de données KEGG, et met en évidence la localisation des gènes identifiés comme différemment

exprimés dans nos analyses. Des nœuds marqués d'un astérisque rouge indiquent des gènes et / ou des complexes protéiques qui ont été trouvés surexprimés dans le cadre de notre étude.

On observe la présence de gènes surexprimés à différents niveaux de cette cascade de signalisation, notamment en amont au niveau des récepteurs et des protéines adaptatrices (ex: 'IRS', 'PI3K', 'AKT'), ainsi qu'en aval, impactant potentiellement des processus tels que le transport du glucose (via 'GLUT4'), la glycogénèse, la lipogenèse et la synthèse protéique. La surexpression de certains gènes comme 'SOCS2', déjà identifié dans la **Figure 17**, est également localisée dans cette voie, confirmant son rôle potentiel dans la régulation de la signalisation.

Ainsi, l'obtention de cette voie de signalisation comme un résultat central de notre étude est une déduction basée sur des preuves statistiques et biologiques issues de l'ensemble de notre analyse transcriptomique, et non une supposition initiale. Elle met en lumière une facette inattendue ou accentuée de la réponse de l'hôte au diabète type I, ouvrant de nouvelles pistes de recherche sur les interactions entre le métabolisme et la réponse immunitaire.

I.2 HRS contre LRS

La **Figure 19** présente un graphe en volcan (volcano plot) illustrant les résultats de l'analyse d'expression différentielle des gènes dans les monocytes, dans le contexte de la base de données GSE232310.

Ce graphique compare spécifiquement les échantillons du groupe à haut risque de développer un diabète de type 1 (HRS - High-Risk Subjects) aux échantillons du groupe à faible risque (LRS - Low-Risk Subjects), avec une p-valeur ajustée (P_{adj}) inférieure à 0.05. L'axe des X représente le $\log_2(\text{changement de pli})$ ($\log_2(\text{fold change})$), quantifiant la magnitude de la régulation des gènes, tandis que l'axe des Y représente la significativité statistique en $-\log_{10}(\text{p-valeur})$.

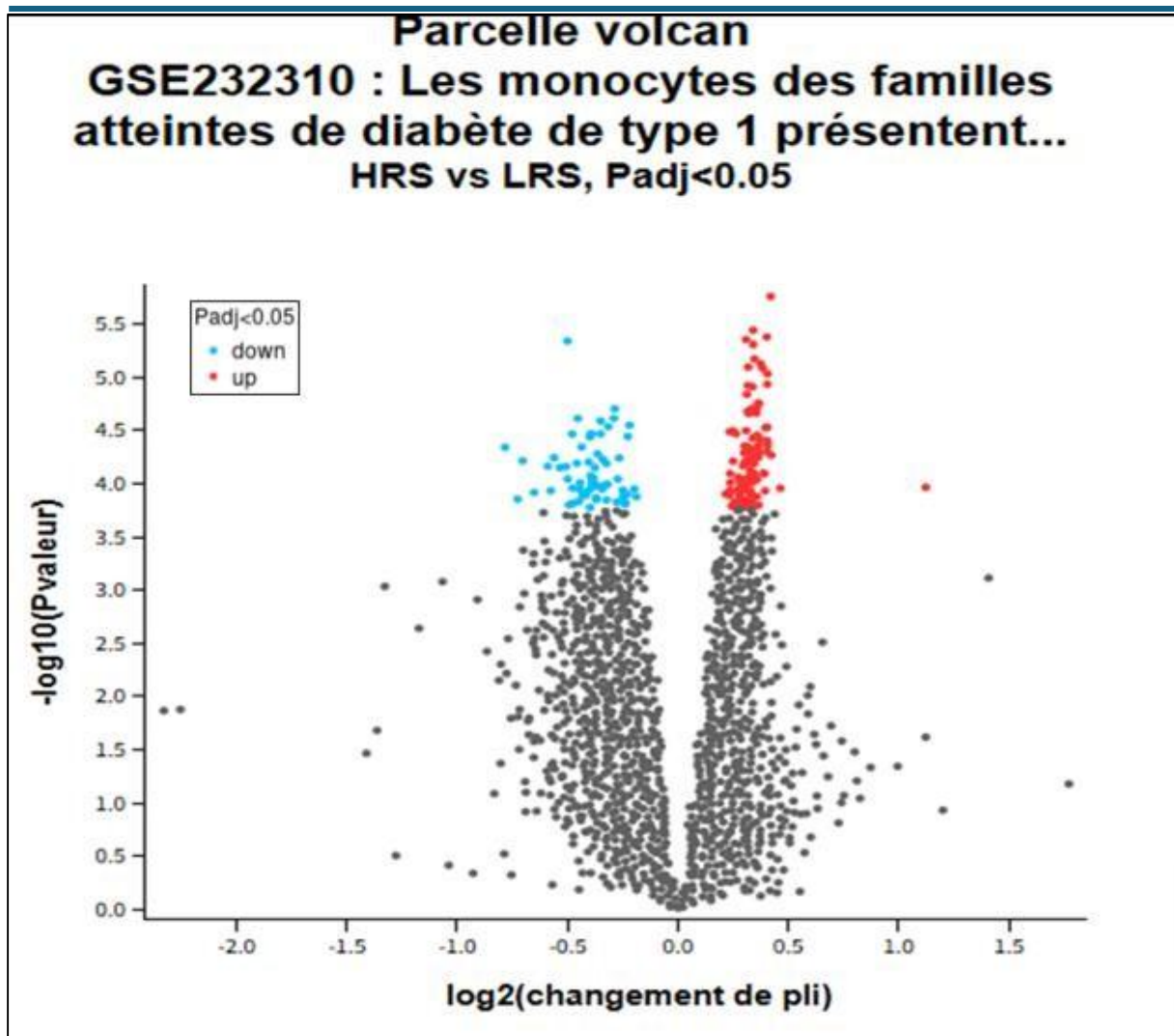


Figure 19 : Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'étude « HRS vs LRS ».

Les points rouges (situés à droite) indiquent les gènes significativement surexprimés ($\log_2(\text{fold change})$ positif, $\text{Padj} < 0.05$). Les points bleus (situés à gauche) représentent les gènes significativement sous-exprimés ($\log_2(\text{fold change})$ négatif, $\text{Padj} < 0.05$). Les points gris correspondent aux gènes dont l'expression n'est pas significativement modifiée. Le graphe révèle un nombre comparable, mais légèrement supérieur, de gènes significativement surexprimés (points rouges) par rapport aux gènes sous-exprimés (points bleus), indiquant une modulation bidirectionnelle, mais avec une légère prépondérance d'activation transcriptionnelle, dans la réponse des monocytes entre les groupes HRS et LRS."

Le graphe en volcan (**Figure 19**) complète l'analyse précédente (**Figure 12**) en affinant la comparaison au sein de la cohorte GSE232310, en se concentrant sur les différences entre les monocytes des sujets à haut risque (HRS) et à faible risque (LRS) dans le diabète de type I.

La présence de gènes à la fois significativement surexprimés et sous-exprimés dans cette comparaison suggère une dysrégulation transcriptionnelle complexe au niveau des monocytes, potentiellement liée à la progression ou à la susceptibilité au diabète de type 1.

Contrairement à la **Figure 12** qui montrait une asymétrie plus prononcée vers la surexpression, cette figure met en lumière une modulation plus équilibrée, avec un nombre important de gènes sous-exprimés, ce qui pourrait indiquer des mécanismes de suppression immunitaire ou des voies métaboliques altérées.

L'analyse intégrée de l'expression différentielle et des voies via le package DESeq2 nous a permis de mettre en œuvre l'Analyse exploratoire, expression différentielle et analyses de voies sur des 184 gènes trouvés dans le groupe de recherche **HRS vs LRS**.

Le résultat de cette analyse (**Figure 20**) démontre que 184 gènes ont passé le filtre, 147 ont été convertis en Ensembl_ID de gène. Les 37 gènes restants ont été conservés dans les données en utilisant les identifiants originaux.

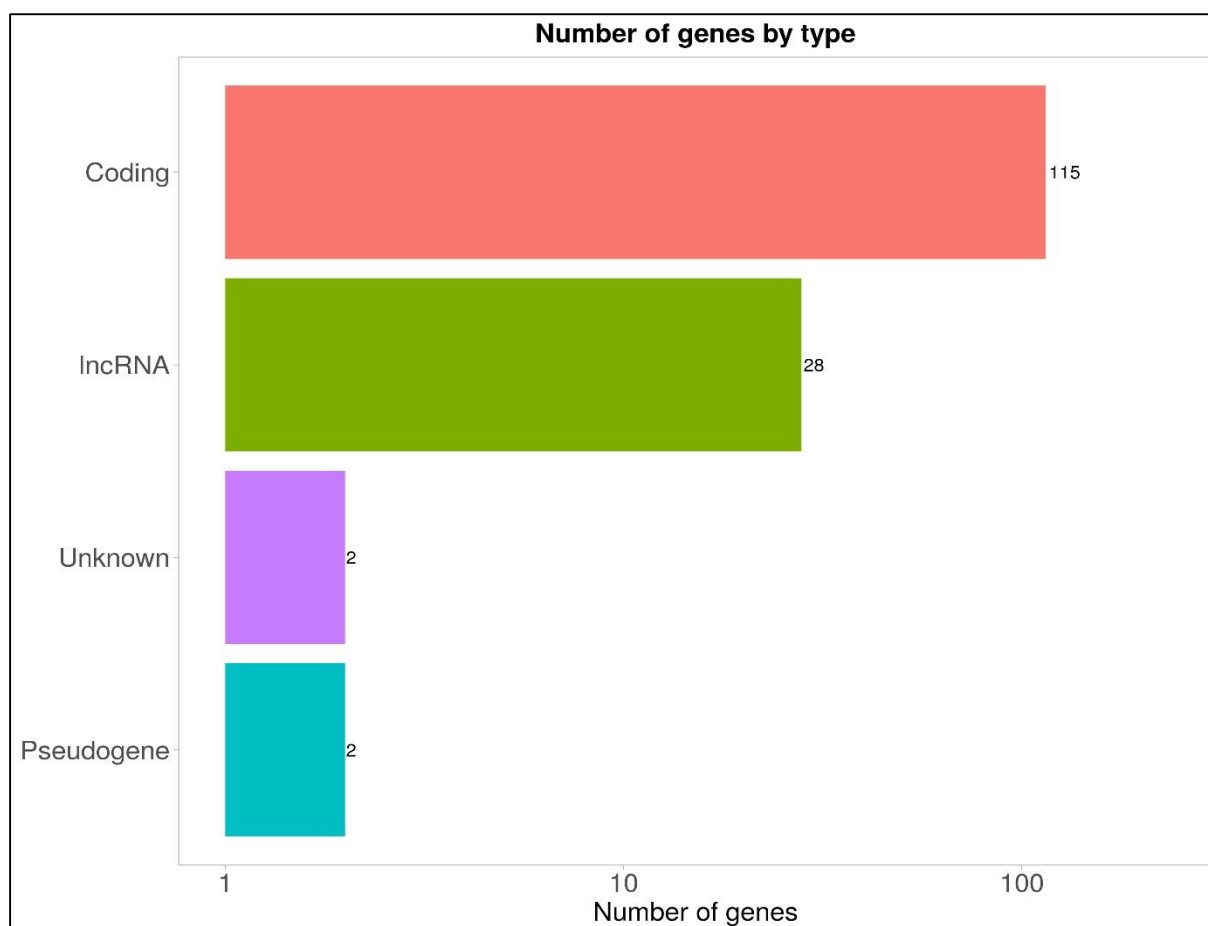


Figure 20 : Distribution des gènes identifiés par type dans le groupe des gènes « HRS vs LRS ».

La **Figure 20** présente la classification des gènes identifiés par type dans l'ensemble de données analysé. Sur un total de 147 gènes, la majorité, soit 115 gènes (environ 78.2%), sont des gènes codants. Les ARN non codants longs (lncRNA) constituent le deuxième groupe le plus important avec 28 gènes (environ 19%). On observe également la présence de 2 pseudogènes (environ 1.4%), tandis que 2 gènes (environ 1.4%) sont classés comme de type inconnu.

La répartition des gènes par type (**Figure 20**) apporte des nuances importantes par rapport aux observations initiales de notre analyse transcriptomique. Bien que les gènes codants restent majoritaires, la proportion de lncRNA (près de 19%) est notablement plus élevée dans cet ensemble de données par rapport à d'autres analyses où leur proportion est souvent plus faible par rapport à la **Figure 13** précédente qui montrait environ 5% de lncRNA. Cette augmentation suggère une implication potentiellement accrue des régulations non codantes dans les processus biologiques mis en jeu.

Les lncRNA sont de plus en plus reconnus pour leur rôle complexe dans la régulation de l'expression génique, incluant l'activation et la suppression transcriptionnelle, la modification de la chromatine et la séquestration de microARN. Leur prévalence accrue pourrait indiquer que les mécanismes régulatoires à l'œuvre sont plus sophistiqués ou qu'ils exploitent davantage les niveaux épigénétiques et post-transcriptionnels. L'identification de ces lncRNA comme une composante significative des gènes régulés ouvre des avenues de recherche pour comprendre leur rôle précis dans le groupe des gènes HRS vs LRS et leur potentiel comme biomarqueurs ou cibles thérapeutiques, au-delà des gènes codants traditionnels.

Nous avons exploré les corrélations géniques dans des sous-ensembles d'échantillons de nos deux groupes, en employant une méthode de biclustering mise en œuvre avec les packages R Biclust et QUBIC.

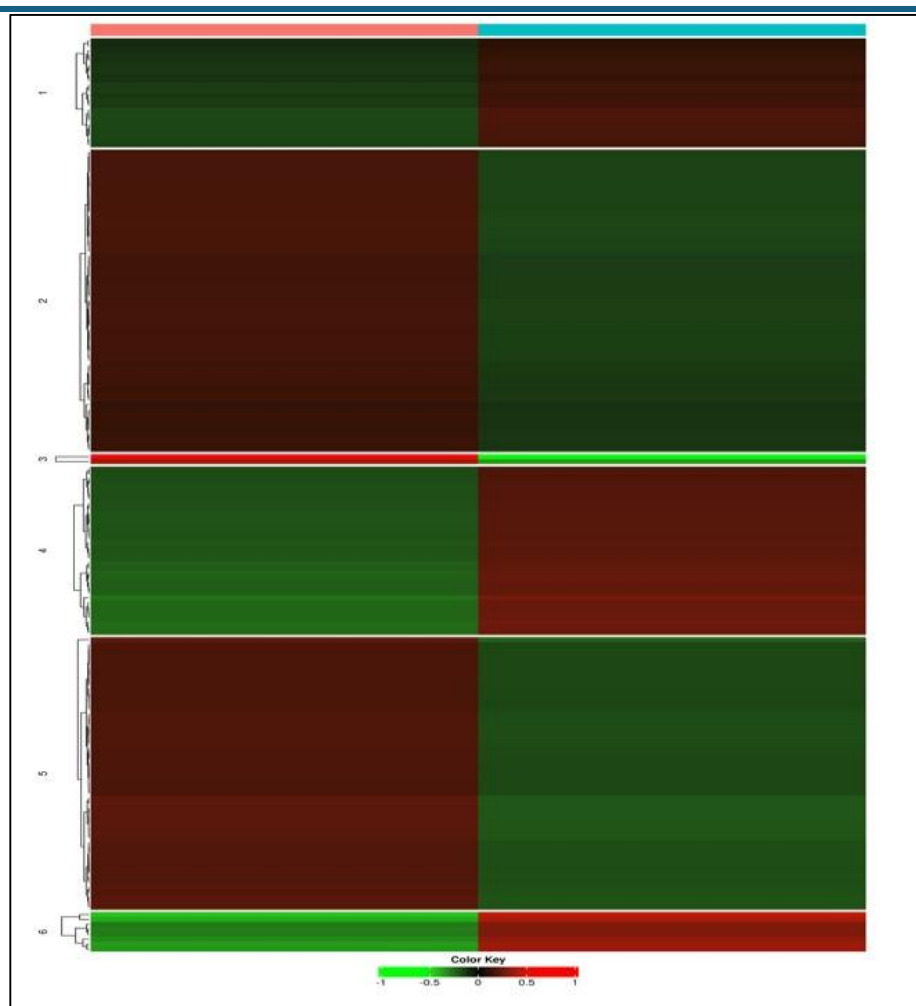


Figure 21 : Heatmap des termes enrichis dans l'analyse transcriptomique du groupe des gènes « HRS vs LRS ».

La Figure 21 présente une heatmap enrichis, regroupés par clustering hiérarchique et illustre le niveau d'expression des gènes au sein des différentes catégories. On observe des clusters de gènes (numérotés de 1 à 6) présentant des profils d'expression distincts.

Le cluster 3 le plus distingué montre un profil d'expression fortement sous-exprimé (vert vif) dans le groupe d'échantillons de gauche (celui avec la barre supérieure rouge) et une surexpression marquée (rouge vif) dans le groupe d'échantillons de droite (celui avec la barre supérieure cyan) est fortement associé au processus de 'Tight junction interactions'

L'analyse des termes GO enrichis, visualisée dans la **Figure 21**, fournit des informations cruciales sur les processus biologiques altérés dans notre modèle d'étude. Les jonctions serrées sont des complexes multiprotéiques essentiels à l'intégrité des barrières épithéliales et endothéliales dans l'organisme, contrôlant la perméabilité paracellulaire et régulant des

fonctions cruciales telles que l'homéostasie tissulaire et la signalisation cellulaire. (Zihni, C *et al.*, 2016).

La mise en évidence d'un cluster de gènes spécifiquement régulés et liés à ces jonctions suggère une altération ou une adaptation de la fonction de barrière dans le contexte de notre étude.

La concentration de gènes régulés démontré dans la **Figure 22** et leur lien avec les 'Tight junctions interactions' suggèrent une altération ou une adaptation significative de la fonction de barrière dans le contexte de notre étude. Cela pourrait indiquer :

- ✦ Une perturbation de la barrière physique : Une perméabilité accrue due à une inflammation, une infection, ou un stress métabolique (comme cela pourrait être le cas dans le diabète type I).
- ✦ Une tentative de réparation ou de renforcement : L'organisme pourrait tenter de compenser des dommages ou de modifier l'environnement tissulaire.
- ✦ Des changements dans la signalisation cellulaire : Les protéines des jonctions serrées ne sont pas seulement structurales, elles sont aussi impliquées dans des voies de signalisation qui influencent la prolifération, la différenciation et la survie cellulaire.

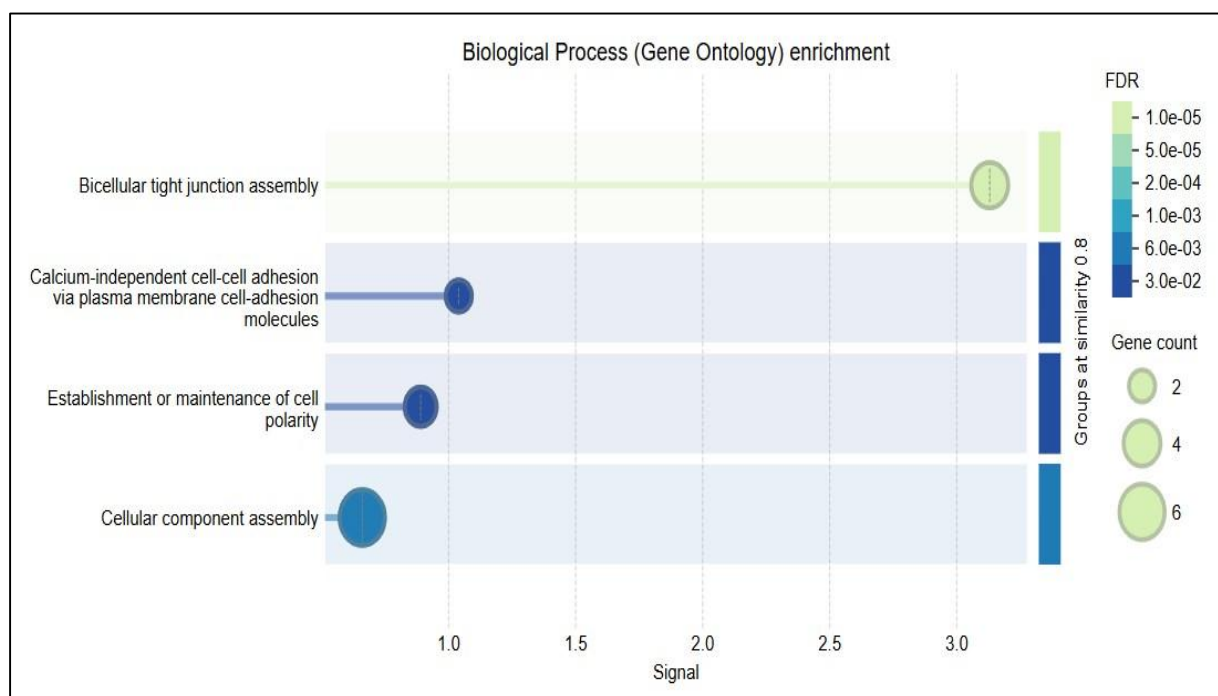


Figure 22: Représentation graphique des principales voies et processus biologiques enrichis, classés par signification et nombre de gènes du groupe des gènes « HRS vs LRS ».

Le **Tableau III** présente les 4 principaux processus biologiques (Gene Ontology) et appariement des gènes enrichis qui se sont avérés significativement enrichis parmi les gènes identifiés dans notre analyse. Ces enrichissements sont classés par p-value ajustée (Adj.Pval). Le processus le plus significativement enrichi est 'Bicellular tight junction assembly' (Adj.Pval = 1.09×10^{-5} , Fold = 2.43), impliquant des gènes clés tels que CLDN1, CLDN2, MPP7, et DLG1. Un processus plus général, 'Cellular component assembly', apparaît également (Adj.Pval = 0.0076, Fold = 0.9), partageant certains de ces gènes (DLGAP5, MPP5, CLDN1, CLDN2, MPP7, DLG1). De plus, l'"Calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules" (Adj.Pval = 0.0278, Fold = 2.5), avec CLDN1 et CLDN2, ainsi que l'"Establishment or maintenance of cell polarity" (Adj.Pval = 0.0327, Fold = 1.67), impliquant MPP5 et DLG1, sont également significativement enrichis.

Tableau III : Processus biologiques (GO) et appariement des gènes enrichis.

N	Adj.Pval	Fold	Go Biological Process	Matching gene
1	1.09e-05	2.43	Bicellular tight junction assembly	CLDN1,CLDN2,MPP7,DLG1
2	0.0076	0.9	Cellular component assembly	DLGAP5,MPP5,CLDN1,CLDN2,MPP7,DLG1
3	0.0278	2.5	Calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	CLDN1,CLDN2
4	0.0327	1.67	Establishment or maintenance of cell polarity	MPP5,MPP7,DLG1

L'analyse des termes GO enrichis présentée dans le **Tableau III** fournit un support fonctionnel solide à l'observation du Cluster 3 de la heatmap (**Figure 21**). La présence prédominante de processus comme 'Bicellular tight junction assembly' avec une très faible p-value ajustée confirme que la régulation des gènes dans ce cluster est directement liée à la formation et à la maintenance des jonctions serrées.

Les gènes identifiés dans ces processus, tels que CLDN1 et CLDN2 (codant pour des claudines, composants fondamentaux des jonctions serrées), ainsi que MPP7 et DLG1 (impliqués dans l'organisation des complexes de jonction et la polarité cellulaire), sont des acteurs cruciaux. Leur implication suggère que le Cluster 3 est sous-exprimé dans un groupe spécifique, comme discuté pour la heatmap.

Ces gènes sont sous-exprimés dans le groupe d'intérêt HRS vs LRS, indiquerait une compromission de l'intégrité des barrières cellulaires. Une perméabilité accrue des barrières (endothéliale ou épithéliale) est une caractéristique connue de nombreuses pathologies inflammatoires et métaboliques, y compris le diabète de type 1 où une dysfonction de la barrière intestinale ou vasculaire.

De plus, l'enrichissement des processus liés à l'adhésion cellulaire calcium-indépendante' et à l'établissement et / ou le maintien de la polarité cellulaire' renforce l'idée d'une modification des interactions cellule-cellule et de l'organisation tissulaire. Ces altérations peuvent avoir des conséquences profondes sur la fonction cellulaire, la migration des cellules immunitaires, ou la réponse aux signaux environnementaux. La corrélation entre les profils d'expression des gènes du Cluster 3 et leur association fonctionnelle avec les jonctions serrées et l'adhésion souligne l'importance de ces mécanismes dans la diabète.

La **Figure 23** présente un sous-réseau d'interaction protéine-protéine (PPI) centré sur les gènes identifiés comme cruciaux pour les processus de jonctions serrées et d'adhésion cellulaire (**Tableau III et Figure 23**). On observe un réseau interconnecté mettant en relation des protéines telles que CLDN1, CLDN2, MPP5, MPP7, DLG1 et DLGAP5. Les protéines CLDN1 et CLDN2 apparaissent comme des points centraux, interagissant directement entre elles et avec les protéines MPP et DLG. La présence de DLGAP5 en périphérie suggère un rôle de régulation ou d'interaction moins directe dans ce sous-réseau spécifique.

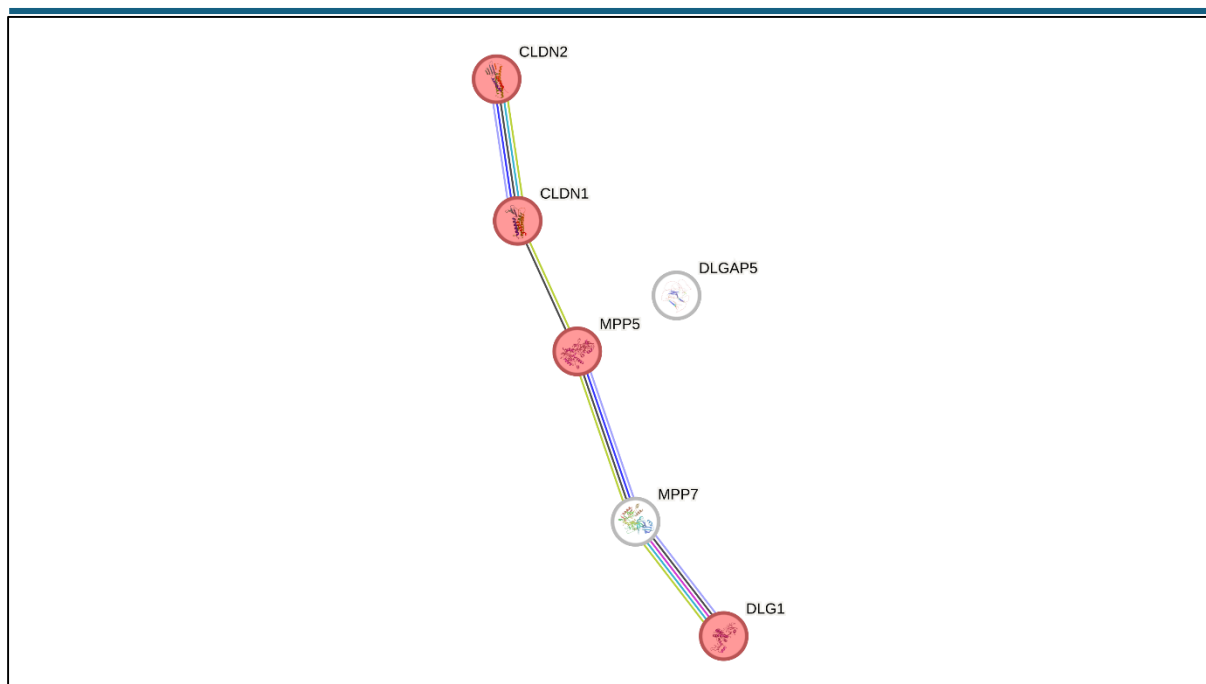


Figure 23 : Réseau d'interaction protéine-protéine (PPI) des gènes différentiellement exprimés gènes du groupe des gènes « HRS vs LRS ».

L'analyse du réseau d'interaction protéine-protéine (**Figure 23**) des gènes liés aux jonctions serrées offre une vision moléculaire détaillée des résultats obtenus précédemment. La forte interconnectivité observée entre CLDN1, CLDN2, MPP5, MPP7 et DLG1 n'est pas surprenante, car ces protéines sont des composants bien établis des complexes de jonctions serrées et des structures d'adhésion cellulaire. (**Zihni, C et al., 2016**)

La **Figure 24** présente l'expression différentielle de gènes spécifiques identifiés comme étant des acteurs clés dans le processus biologique mis en évidence par nos analyses précédentes

(**Figure 23 et Tableau III**) connus pour leur rôle dans les jonctions serrées : **MPP5, CLDN1, CLDN2 et DLG1**.

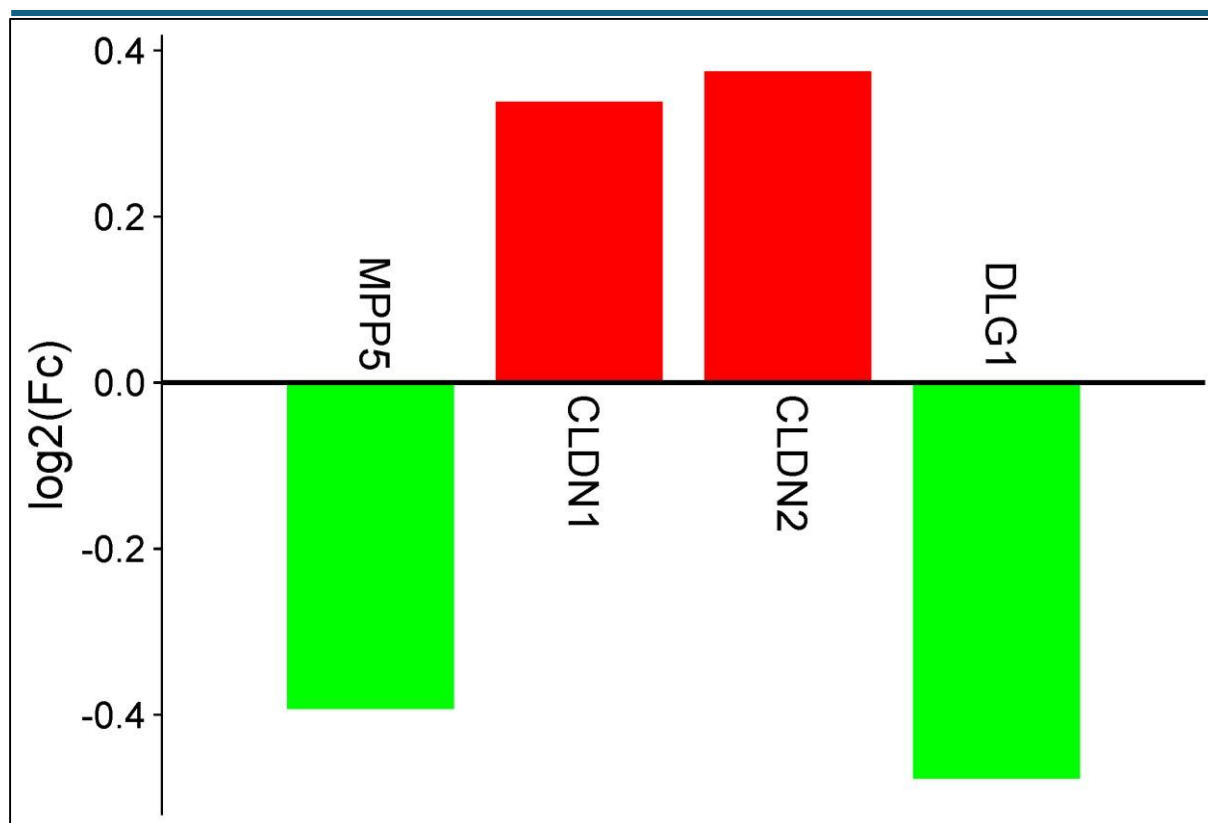


Figure 24 : Expression relative ($\log_2(\text{Fold Change})$) des gènes clés identifiés dans le processus biologique Tight junction interactions du groupe des gènes « HRS vs LRS ».

Les barres rouges indiquent une surexpression, tandis que les barres vertes représentent une sous-expression. On observe une surexpression significative de CLDN1 et CLDN2, avec des $\log_2(\text{Fold Change})$ positifs et importants (environ 0.35 pour les deux). En revanche, MPP5 et DLG1 montrent une sous-expression notable, avec des $\log_2(\text{Fold Change})$ négatifs (autour de -0.4 pour MPP5 et -0.5 pour DLG1)."

- ✦ CLDN1 et CLDN2 (Claudines) sont des protéines transmembranaires essentielles qui forment les brins des jonctions serrées et régulent directement la perméabilité paracellulaire. Leurs interactions directes soulignent leur rôle fondamental dans l'architecture de ces jonctions. **(Tsukita S *et al.*, 2000)**
- ✦ MPP5 (MAGUK p55 scaffolding protein 5, alias PALS1) et MPP7 (MAGUK p55 scaffolding protein 7) sont des protéines d'échafaudage MAGUK qui organisent les complexes protéiques aux jonctions cellulaires et sont impliquées dans la polarité cellulaire et la signalisation. **(Emeline Assémat *et al.*, 2007)**

- ✦ DLG1 (Discs Large MAGUK scaffold protein 1) est également une protéine d'échafaudage MAGUK jouant un rôle crucial dans l'organisation des jonctions cellule-cellule et la signalisation. (**Mantovani F et al., 2003**)
- ✦ DLGAP5 (DLG associated protein 5) interagit avec DLG1 et est impliqué dans l'organisation du cytosquelette et la régulation du cycle cellulaire, suggérant un lien avec la dynamique des jonctions. (**Dis Res, 2018**)

Ces résultats d'expression individuelle des gènes clés des jonctions serrées (Figure 24) fournit une validation directe et quantitative des observations tirées des analyses d'enrichissement (**Tableau III**) et du réseau d'interaction protéine-protéine (Figure 23), ainsi que des profils de clusters de la heatmap (**Figure 21**). La régulation opposée de ces gènes est particulièrement informative.

La surexpression des claudines (CLDN1 et CLDN2), qui sont des composants transmembranaires essentiels des brins des jonctions serrées, suggère une tentative de renforcement ou une modification qualitative de ces jonctions. Cela pourrait être une réponse adaptative visant à limiter la perméabilité paracellulaire en face d'un stress ou d'une inflammation. Cependant, une surexpression aberrante de certaines claudines peut aussi conduire à une dysfonction ou une perméabilité 'ouverte' ou 'fermée' anormale, en fonction de l'isoforme et du contexte cellulaire.

Inversement, la sous-expression marquée de MPP5 et DLG1 est cruciale. Ces protéines d'échafaudage MAGUK sont fondamentales pour l'assemblage, l'organisation et la signalisation des complexes de jonction serrée et d'adhésion. Leur diminution suggère une désorganisation structurelle ou une altération de la signalisation en aval des jonctions. Une déficience de MPP5 et DLG1 pourrait compromettre la polarité cellulaire, la régulation du cytosquelette et la transmission des signaux nécessaires au maintien de l'intégrité de la barrière.

En somme, ce résultat révèle une dysrégulation complexe et bidirectionnelle des composants des jonctions serrées. Alors que les éléments structuraux claudines pourraient être surexprimés pour compenser ou remodeler, les protéines clés d'organisation et de signalisation (MPP5, DLG1) sont sous-exprimées, indiquant potentiellement une barrière structurellement présente mais fonctionnellement compromise ou altérée dans le groupe étudié. Cette balance déséquilibrée pourrait contribuer à la physiopathologie du diabète type I.

Nous avons procédé à une analyse d'enrichissement fonctionnel. Cette étape a consisté à interroger des bases de données de connaissances établies, la base de données KEGG (Kyoto Encyclopedia of Genes and Genomes) et Gene Ontology (GO),

La **Figure 25** représente la voie de signalisation des jonctions serrées (Tight Junction), telle qu'annotée dans la base de données KEGG.

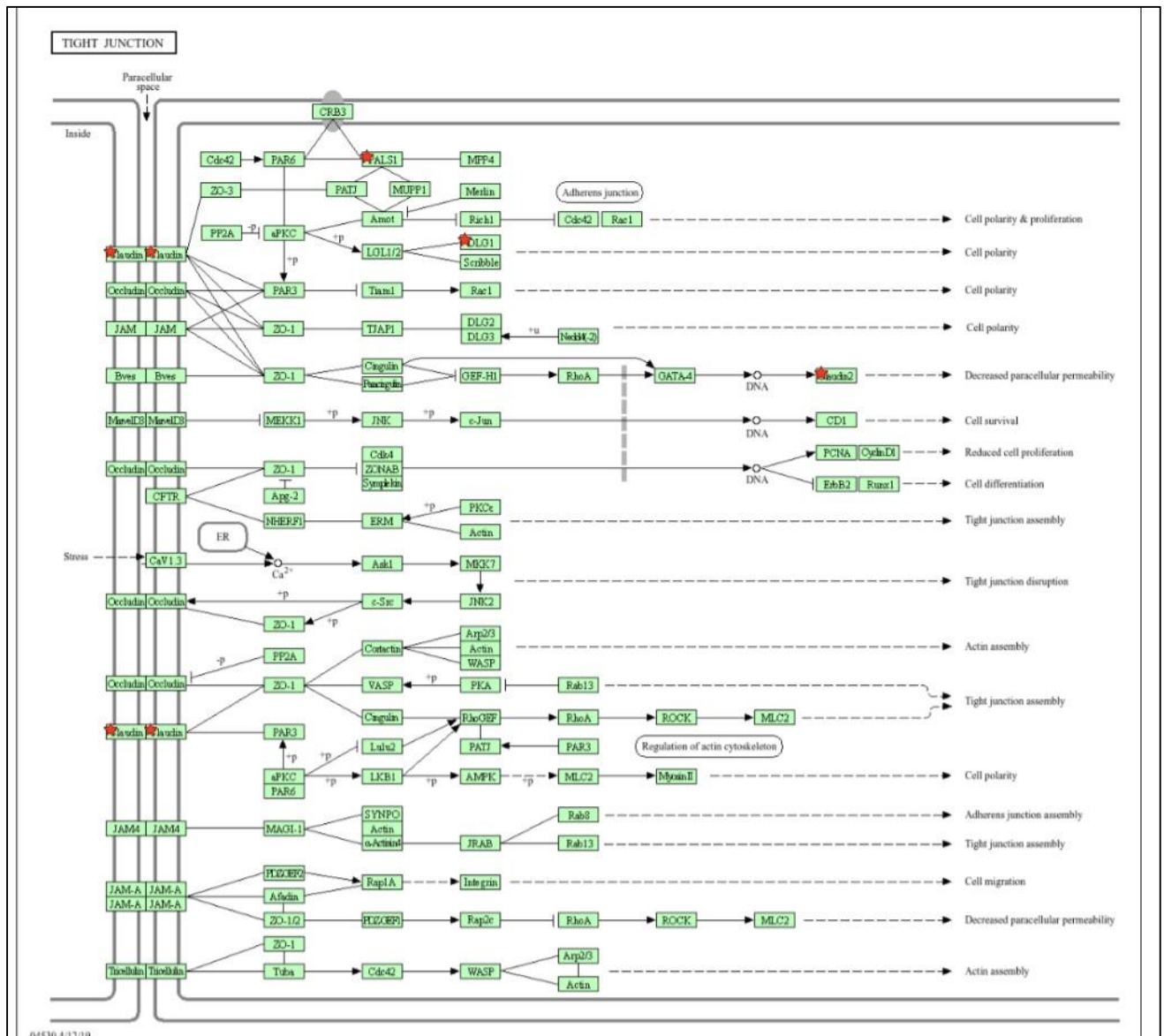


Figure 25: Représentation schématique de la voie de signalisation des jonctions serrées (Tight Junction) avec indication des gènes régulés du groupe des gènes « HRS vs LRS ».

Cette carte schématique illustre les interactions complexes entre les différentes protéines et molécules impliquées dans la formation, la régulation et la fonction des jonctions serrées. Les éléments marqués d'une étoile rouge indiquent des gènes et / ou des complexes protéiques qui ont été identifiés comme étant régulés dans nos analyses.

La cartographie des gènes régulés sur la voie de signalisation des jonctions serrées (**Figure 25**) fournit une explication mécanistique détaillée de l'enrichissement du 'Tight junction interactions' dans le **Tableau III**. La présence de gènes régulés à des points stratégiques de cette voie suggère une perturbation active des mécanismes régulant l'intégrité et la fonction des jonctions serrées.

L'impact sur des molécules clés comme les claudines (CLDN) et les molécules de JAM (Junctional Adhesion Molecules), qui sont des composants structuraux essentiels des jonctions, est particulièrement critique, ce qui corrobore avec l'étude de **Visser.J et al., 2009** qui met en évidence le rôle de l'augmentation de la perméabilité intestinale (via les jonctions serrées) dans la pathogenèse de diverses maladies auto-immunes.

D'autre part l'étude menée par **Robles-Osorio ML et al., 2023**, qui présente rôle fondamental des jonctions serrées et de leur perturbation dans le développement des complications chroniques du diabète, y compris la perméabilité accrue des barrières (sang-cerveau, sangrétine) tels que :

- ✦ Modifications de la perméabilité de la barrière : Des jonctions serrées dysfonctionnelles peuvent altérer la capacité d'une barrière (par exemple, intestinale, vasculaire) à contrôler le passage des substances, ce qui est un facteur connu dans plusieurs maladies inflammatoires et métaboliques, y compris les complications du diabète.
- ✦ Altérations de la polarité et de la signalisation cellulaire : Les jonctions serrées ne sont pas seulement physiques ; elles sont aussi des plateformes de signalisation qui influencent la croissance cellulaire, la différenciation et la réponse aux stimuli externes. La régulation des protéines de cette voie peut donc avoir des conséquences en cascade sur le comportement cellulaire.

I.3 uHC vs HRS

La **Figure 26** présente un graphe en volcan (volcano plot) illustrant les résultats de l'analyse d'expression différentielle des gènes dans les monocytes, dans le contexte de la base de données GSE232310.

Ce graphique compare spécifiquement les échantillons des contrôles sains non apparentés (uHC) aux échantillons des sujets à haut risque de développer un diabète de type I (HRS - HighRisk Subjects), avec une p-valeur ajustée (Padj) inférieure à 0.05.

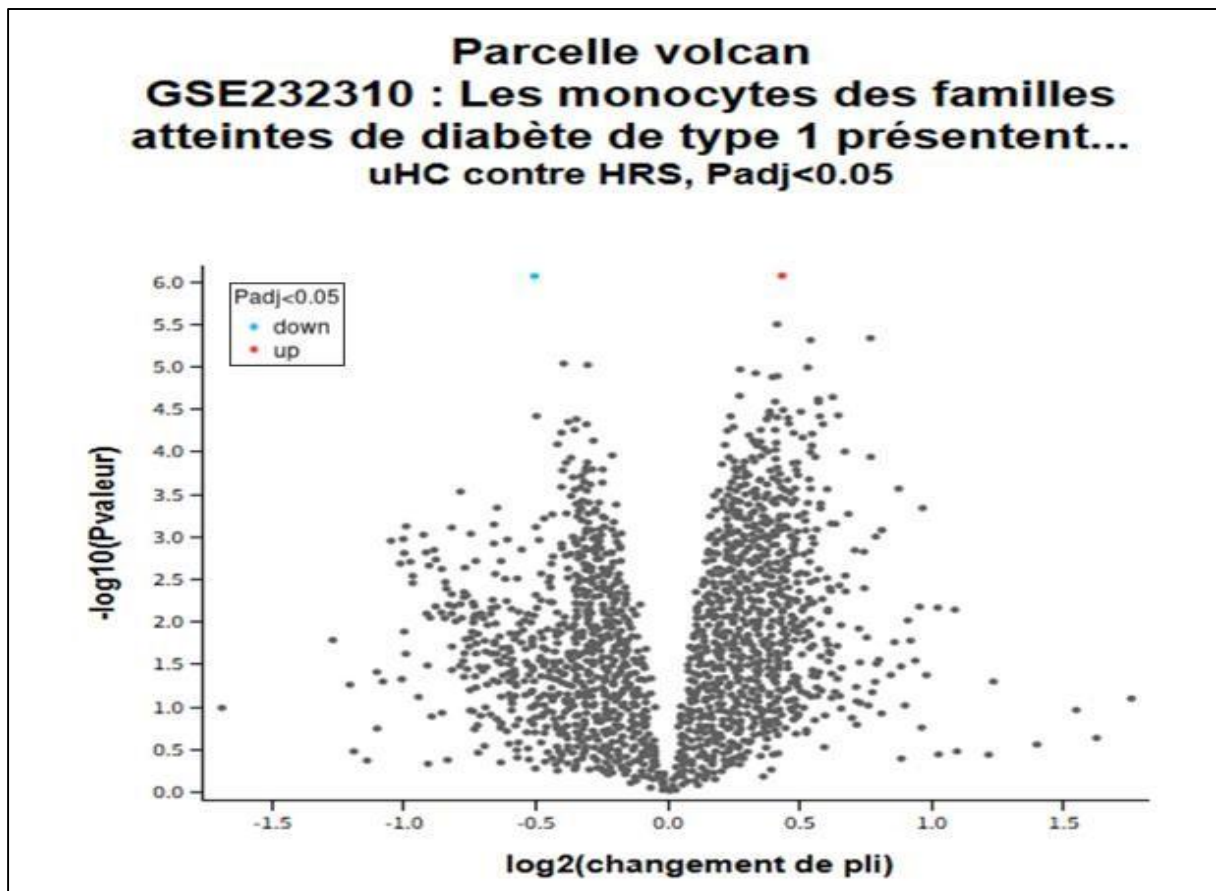


Figure 26 : Représentation volcano plot des gènes exprimés de manière différentielle chez les groupes d'étude « uHC vs HRS ».

Les points rouges (situés à droite) indiquent les gènes significativement surexprimés ($\log_2(\text{fold change})$ positif, $\text{Padj} < 0.05$). Les points bleus (situés à gauche) représentent les gènes significativement sous-exprimés ($\log_2(\text{fold change})$ négatif, $\text{Padj} < 0.05$). Les points gris correspondent aux gènes dont l'expression n'est pas significativement modifiée. Le graphe révèle un nombre beaucoup plus important de gènes significativement régulés à la hausse (points rouges) qu'à la baisse (points bleus) lorsque l'on compare les sujets à HRS. La large dispersion des points et la présence de nombreux gènes au-delà des seuils de significativité (particulièrement en surexpression) indiquent une dysrégulation transcriptionnelle significative et étendue dans les monocytes associés au risque de diabète de type I.

L'analyse d'expression différentielle a permis d'identifier des gènes clés dont la régulation est inversée entre les groupes (**Figure 27**).

Le gène **CHCHD5** (représenté par le point rouge en haut à droite) est significativement surexprimé chez le groupe à Haut Risque (HRS) par rapport aux Contrôles Sains Non Apparentés (uHC). Inversement, le gène **TRIT1** (représenté par le point bleu en haut à gauche) est significativement sous-exprimé chez le groupe HRS par rapport aux uHC."

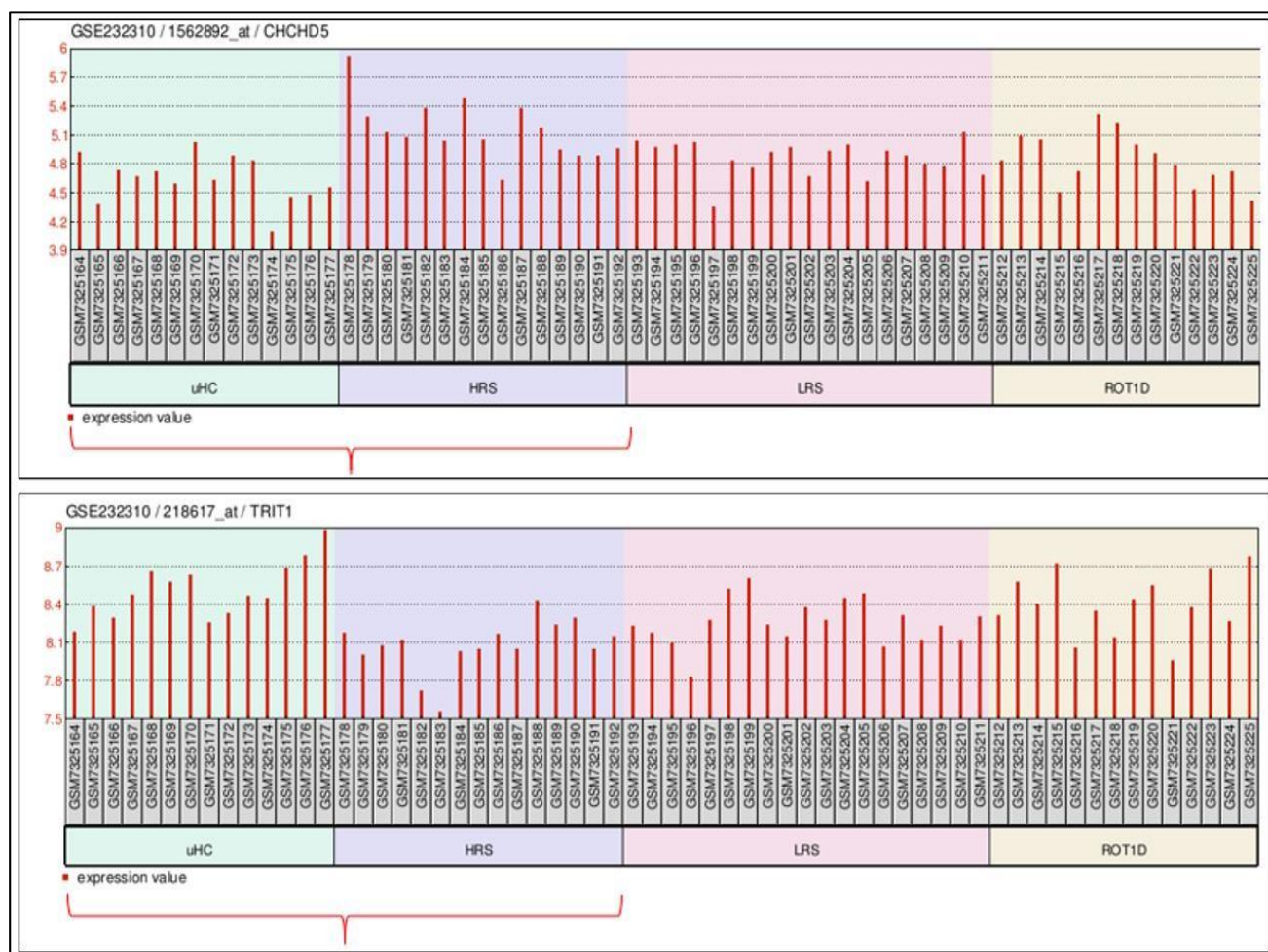


Figure 27 : Représentation par histogramme du niveau d'expression du gène CHCHD5 / TRIT1 dans les groupes d'étude « uHC vs HRS ».

- ✦ CHCHD5 (Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 5) est un gène situé sur le chromosome 2q14. Il code une protéine mitochondriale localisée dans l'espace intermembranaire de la mitochondrie (Figure 30), où elle participe activement au repliement oxydatif des protéines. Cette protéine joue un rôle crucial dans le processus de respiration cellulaire. Parmi les maladies associées à CHCHD5 la sarcocystose (**Wu et al., 2017**).
- ✦ TRIT1 (TRNA Isopentenyltransferase 1), est un gène situé sur le chromosome 1p34. Il code une protéine ciblant la mitochondrie et modifie les ARN de transfert (ARNt) en ajoutant un groupe diméthylallyle à l'adénine en position 37. Cette

modification est importante pour maintenir le cadre de lecture correct lors de la traduction des protéines. Ce gène est considéré comme un suppresseur de tumeur et son expression peut ralentir la croissance cellulaire. Les maladies associées à TRIT1 incluent le déficit combiné en phosphorylation oxydative 35 et le déficit combiné en phosphorylation oxydative. (Niu *et al.*, 2025).

La surexpression de CHCHD5 chez les sujets à haut risque (HRS) est particulièrement intéressante. Une altération des fonctions mitochondriales est souvent impliquée dans les maladies métaboliques et auto-immunes, incluant le diabète de type 1. Cette surexpression pourrait indiquer un stress mitochondrial accru, une tentative d'adaptation métabolique, ou un mécanisme pathologique spécifique au développement de la maladie dans les monocytes des individus à risque. Des études antérieures ont montré que les mitochondries des cellules immunitaires jouent un rôle actif dans la régulation de l'immunité et de l'inflammation.

À l'inverse, la sous-expression de TRIT1 chez les sujets à haut risque (HRS) est tout aussi pertinente. TRIT1 (Tyrosyl-tRNA Synthetase 1, Mitochondrial) est une enzyme mitochondriale impliquée dans la modification des ARNt mitochondriaux, essentielle à la traduction des protéines codées par le génome mitochondrial. Une sous-expression de TRIT1 pourrait suggérer une dysfonction de la traduction mitochondriale et, par extension, une altération de la respiration cellulaire et de la production d'ATP dans les monocytes du groupe HRS. Cela renforcerait l'hypothèse d'un déséquilibre métabolique ou énergétique sous-jacent, qui pourrait rendre ces cellules plus vulnérables ou contribuer à la pathogenèse de la maladie.

Ces régulations opposées de gènes mitochondriaux (CHCHD5 surexprimé et TRIT1 sous-exprimé) ne sont pas anecdotiques ; elles pointent vers une dysfonction mitochondriale potentiellement complexe et multifacette dans les monocytes des individus à haut risque de diabète de type 1. L'exploration plus approfondie de ces deux gènes et de leurs conséquences fonctionnelles au niveau cellulaire et mitochondrial est essentielle pour comprendre les mécanismes précoces de la maladie et développer de potentiels biomarqueurs ou cibles thérapeutiques." Ces résultats vont dans le même sens de la revue établie par Sivitz WI *et al.*, 2010, ou ils explorent en détail les perturbations qualitatives, quantitatives et fonctionnelles des mitochondries dans le diabète (type I et II), soulignant leur importance dans la cause et les complications, qui peut appuyer l'idée que les changements dans CHCHD5 et TRIT1 reflètent une dysfonction mitochondriale globale.

La **Figure 28** présente une cartographie générale des voies métaboliques et des maladies associées, telle que dérivée des bases de données de pathways Reactome Disease.

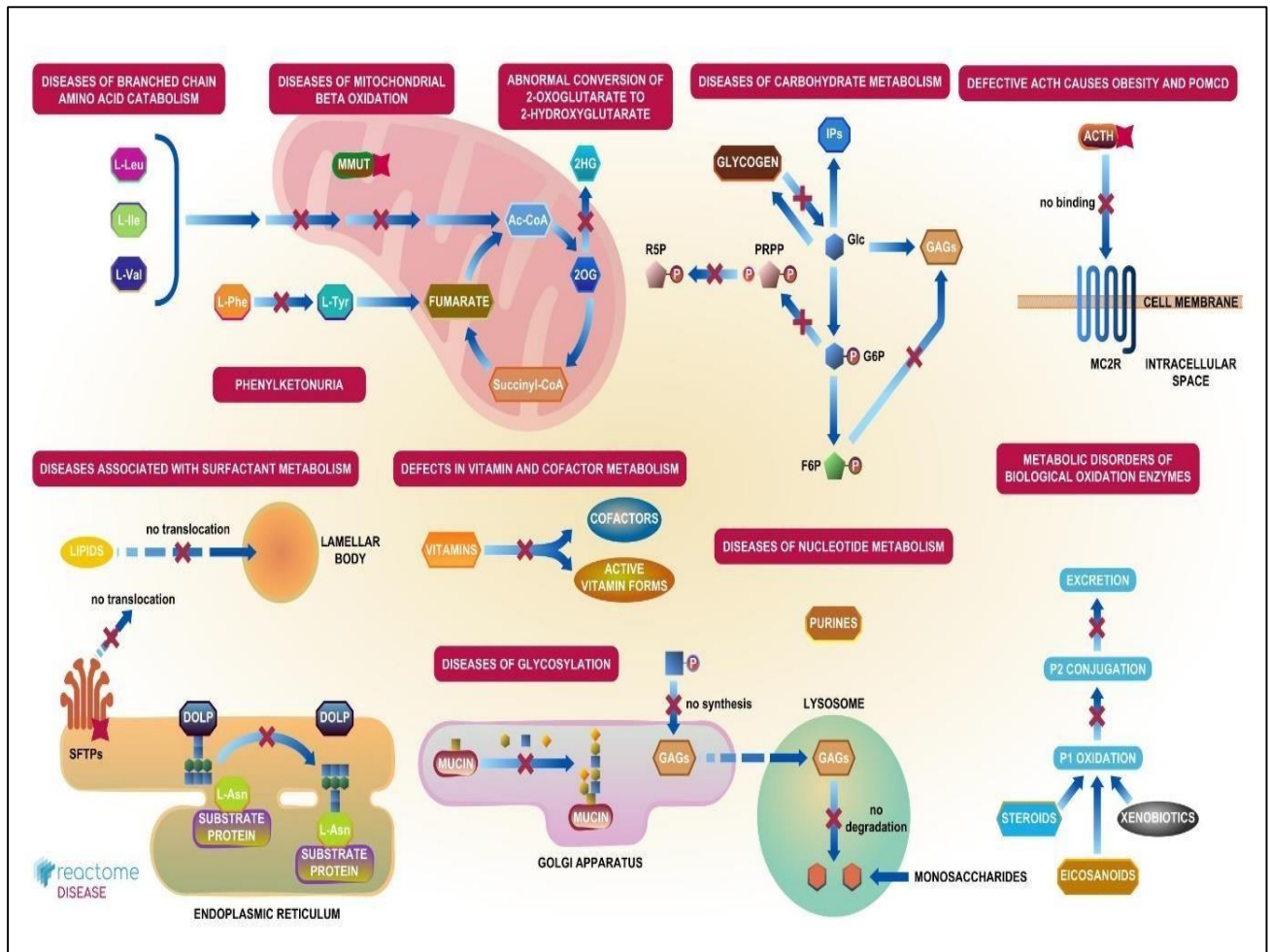


Figure 28 : Ensemble des voies métaboliques et des maladies associées (Reactome Disease Pathway), soulignant les liens avec la fonction mitochondriale.

Ce diagramme illustre diverses catégories de troubles métaboliques, incluant les maladies du catabolisme des acides aminés ramifiés, les maladies de la bêta-oxydation mitochondriale, les troubles du métabolisme des glucides, et les défauts du métabolisme des vitamines et cofacteurs. Chaque section de la carte détaille les interconnexions entre les molécules, les enzymes et les processus perturbés, et leurs conséquences sur l'homéostasie cellulaire et les fonctions physiologiques. La figure met en lumière la complexité et l'interdépendance des différentes voies métaboliques et leur implication dans la pathogenèse des maladies."

Cette vue d'ensemble des voies métaboliques et des maladies associées (Figure 30) enrichit la compréhension des implications systémiques des changements d'expression génique observés dans notre étude, notamment en ce qui concerne la fonction mitochondriale.

Les analyses précédentes ont révélé des dysrégulations de gènes comme CHCHD5 (surexprimé dans le groupe à haut risque) et TRIT1 (sous-exprimé), tous deux directement impliqués dans la biogenèse et la fonction mitochondriale.

La présence des 'Maladies de la bêta-oxydation mitochondriale' et d'autres processus métaboliques clés dans cette figure corrobore l'hypothèse d'une dysfonction mitochondriale centrale dans le contexte de notre étude sur le diabète de type I.

L'importance de CHCHD5 dans cette figure réside dans son rôle en tant que protéine de la famille CHCHD, essentielle à la formation et à la stabilité des crêtes internes mitochondriales, qui sont cruciales pour une respiration cellulaire efficace et la production d'ATP. Sa surexpression (comme observé précédemment) pourrait refléter une tentative de compensation face à un stress mitochondrial, ou au contraire, une perturbation du processus de remodelage mitochondrial qui conduit à une dysfonction.

De même, la régulation de gènes comme TRIT1 (impliqué dans la traduction des protéines mitochondriales) et leur impact sur des processus tels que la bêta-oxydation (dégradation des acides gras dans la mitochondrie) et le métabolisme des glucides, suggère des défauts dans la production d'énergie et l'utilisation des substrats.

Dans le diabète de type I, une dysfonction mitochondriale dans les cellules immunitaires (comme les monocytes dans notre cas d'étude) et les cellules bêta-pancréatiques est un domaine de recherche actif.

Selon **Cabrera SM *et al.*, 2015**, les altérations métaboliques visualisées dans la **figure 28**, couplées à la régulation de gènes mitochondriaux clés, indiquent que ces perturbations pourraient non seulement affecter l'efficacité énergétique des cellules, mais aussi moduler la production de ROS (espèces réactives de l'oxygène) et les signaux inflammatoires, contribuant ainsi à la pathogenèse auto-immune.

Conclusion

Conclusion :

Cette étude *in silico* a permis d'explorer de manière approfondie les profils d'expression génique des monocytes chez des individus à différentes étapes de la progression du diabète de type 1 (DT1), depuis le diagnostic récent jusqu'au statut à risque familial. En exploitant les données transcriptomiques nous avons pu démêler des signatures moléculaires complexes et identifier des processus biologiques et des voies de signalisation clés potentiellement altérés dans cette maladie auto-immune.

Nos analyses d'expression différentielle ont mis en lumière une dysrégulation génique significative. La comparaison entre les patients ROT1D et les uHC a révélé une prédominance des 293 gènes surexprimés, suggérant une activation transcriptionnelle asymétrique et une réponse pro-inflammatoire dans les monocytes. Alors que, la comparaison entre les HRS et LRS a montré 184 gènes différentiellement exprimés avec une modulation plus équilibrée, indiquant des mécanismes complexes de susceptibilité et de progression du DT1. La caractérisation des types de gènes a souligné la prédominance des gènes codants, mais aussi la présence notable de lncRNA, suggérant des niveaux de régulation post-transcriptionnelle potentiellement importants dans la pathogenèse.

L'enrichissement fonctionnel a été particulièrement révélateur, convergeant vers une activation robuste et spécifique de la réponse immunitaire cellulaire. Des processus tels que l'activation des leucocytes et des lymphocytes, la différenciation des cellules T CD4⁺/CD8⁺, la cytotoxicité médiée par les cellules NK, et la signalisation du récepteur TCR ont été fortement enrichis. Ces résultats renforcent l'idée que les monocytes et d'autres cellules immunitaires participent activement à la réponse auto-immune qui caractérise le DT1. L'identification de gènes clés surexprimés tels que SOCS2, CBLB et AKT3, reconnus pour leur rôle dans la modulation de l'inflammation et de l'immunité adaptative, corrobore ces observations et suggère des tentatives de l'organisme de réguler une réponse immunitaire intense ou dérégulée.

Un aspect particulier de notre étude est la mise en évidence d'une altération significative des voies liées aux jonctions serrées. Un cluster distinct de gènes fortement

associé à ce processus a montré une régulation bidirectionnelle, avec une surexpression des claudines (CLDN1, CLDN2), des composants structuraux, et une sous-expression des protéines d'échafaudage comme MPP5 et DLG1. Cette dysrégulation complexe suggère une perturbation de l'intégrité

des barrières cellulaires. Une perméabilité accrue des barrières, notamment intestinale ou vasculaire, est un facteur connu dans diverses maladies inflammatoires et métaboliques, y compris le DT1, et pourrait contribuer à l'inflammation systémique et à la destruction des îlots pancréatiques.

Nos résultats soulignent l'implication mitochondriale, avec la surexpression de CHCHD5 et la sous-expression de TRIT1 dans les monocytes du groupe HRS. Ces gènes, essentiels à la biogenèse et à la fonction mitochondriale, indiquent une dysfonction mitochondriale potentiellement complexe et multifacette. Une altération de la respiration cellulaire et du métabolisme énergétique dans les monocytes pourrait affecter leur fonction immunitaire et contribuer au développement du DT1. Ces observations concordent avec des études antérieures sur le rôle pivot de la dysfonction mitochondriale dans les maladies auto-immunes et métaboliques.

En conclusion, notre étude fournit des observations robustes que les monocytes du sang périphérique chez les individus à risque et ceux atteints de DT1 présentent des signatures transcriptomiques distinctes et des dysrégulations fonctionnelles clés. Les altérations dans les voies immunitaires, la fonction des jonctions serrées et le métabolisme mitochondrial représentent des mécanismes importants dans la pathogenèse du DT1. Ces constatations ouvrent des perspectives pour l'identification de nouveaux biomarqueurs précoces et le développement de stratégies thérapeutiques ciblées, potentiellement axées sur la restauration de l'intégrité des barrières et l'optimisation de la fonction mitochondriale afin de prévenir et / ou d'atténuer la progression du diabète de type I.

Références Bibliographiques

A

Aitman, T J. “DNA microarrays in medical practice.” *BMJ* (Clinical research ed.) vol. 323,7313 (2001): 611-5. doi:10.1136/bmj.323.7313.611

Addissouky T, Ali M, El Sayed I, Wang Y. Type 1 diabetes mellitus: retrospect and prospect. *Bull Natl Res Cent* .2024Apr .48,42.doi:10.1186/s42269-024-01197-z.

Antar S, Ashour N, Sharaky M, Khattab M ,Ashour N, Zaid R.T, Roh E.J, Elkamhawy A. Diabetes mellitus: Classification, mediators, and complications; A gate to identify potential targets for the development of new effective treatments.2023Dec ;168:115734. doi: 10.1016/j.biopha.2023.115734.PMID: 37857245

C

Cabrera SM, Chen YG, Hagopian WA, Hessner MJ. Blood-based signatures in type 1 diabetes. *Diabetologia*. 2016 Mar;59(3):414-25. doi: 10.1007/s00125-015-3843-x. Epub 2015 Dec 23. PMID: 26699650; PMCID: PMC4744128.)

Carow B, Rottenberg ME. SOCS3, a Major Regulator of Infection and Inflammation. *Front Immunol*. 2014 Feb 19;5:58. doi: 10.3389/fimmu.2014.00058. PMID: 24600449; PMCID: PMC3928676.

Chieffi P. Aurora B: A new promising therapeutic target in cancer. *Intractable Rare Dis Res*. 2018 May;7(2):141-144. doi: 10.5582/irdr.2018.01018. PMID: 29862159; PMCID: PMC5982624.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016 Jan 26;17:13. doi: 10.1186/s13059-016-0881-8. Erratum in: *Genome Biol*. 2016 Aug 26;17(1):181. doi: 10.1186/s13059-016-1047-4. PMID: 26813401; PMCID: PMC4728800.

Daziano G. Rôle du propeptide de la sortiline et de ses dérivés dans les mécanismes de survie de la cellule bêta pancréatique [thèse de doctorat]. Nice (FR): Université Côte d’Azur; 2021. Numéro National de Thèse : 2021COAZ6025.

Dean L, McEntyre J. Genetic Factors in Type 1 Diabetes. 2004 Jul 7. In: Dean L, McEntyre J, editors. The Genetic Landscape of Diabetes [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2004.

Dufva M. Introduction to microarray technology. *Methods Mol Biol.* 2009;529:1-22. doi: 10.1007/978-1-59745-538-1_1. PMID: 19381982.

E

Emeline Assémat, Elsa Bazellères, Emilie Pallesi-Pocachard, André Le Bivic, Dominique Massey-Harroche, Polarity complex proteins, *Biochimica et Biophysica Acta (BBA) – Biomembranes*, Volume 1778, Issue 3, 2008, Pages 614-630, ISSN 0005-2736, <https://doi.org/10.1016/j.bbamem.2007.08.029>.

Ei Wafai RJ, Chmaisse HN, Makki RF, Fakhoury H. Association of HLA class II alleles and CTLA-4 polymorphism with type 1 diabetes. *Saudi J Kidney Dis Transpl.* 2011 Mar;22(2):273-81. PMID: 21422625.

Elendu C, David JA, Udoyen AO, Egbunu EO, Ogbuiyi-Chima IC, Unakalamba LO, Temitope AI, Ibhiedu JO, Ibhiedu AO, Nwosu PU, Koroyin MO, Eze C, Boluwatife AI, Alabi O, Okabekwa OS, Fatoye JO, Ramon-Yusuf HI. Comprehensive review of diabetic ketoacidosis: an update. *Ann Med Surg (Lond).* 2023 May 23;85(6):2802-2807. doi: 10.1097/MS9.0000000000000894. PMID: 37363479; PMCID: PMC10289692

Elmore S. Apoptosis: a review of programmed cell death. *Toxicol Pathol.* 2007 Jun;35(4):495-516. doi: 10.1080/01926230701320337. PMID: 17562483; PMCID: PMC2117903.

F

Fareed M, Chauhan W, Fatma R, Din I, Afzal M, Ahmed Z. Next-generation sequencing technologies in diabetes research. *Diabetes Epidemiol Manag.* 2022;7:100097. doi:10.1016/j.deman.2022.100097.

Fous

sier L, Zergane L. Le diabète, une pathologie chronique. L'Aide-Soignante. 2021;35(224):10-12. doi:10.1016/j.aidsoi.2020.12.005.

Fung THM, Patel B, Wilmot EG, Amoaku WMK. Diabetic retinopathy for the nonophthalmologist. Clin Med (Lond). 2022 Apr;22(2):112–6. doi: 10.7861/clinmed.2021-0792.

Frommer L, Kahaly GJ. Type 1 Diabetes and Autoimmune Thyroid Disease-The Genetic Link. Front Endocrinol (Lausanne). 2021 Mar 10;12:618213. doi: 10.3389/fendo.2021.618213. PMID: 33776915; PMCID: PMC7988207.

G

Galli-Tsinopoulou A. The role of PTPN22 C1858T gene polymorphism in diabetes mellitus type 1: first evaluation in Greek children and adolescents. Biomed Res Int. 2013;2013:721604. doi: 10.1155/2013/721604. PMID: 23936838; PMCID: PMC3727122.

Giza S, Goulas A, Gbandi E, Effraimidou S, Papadopoulou-Alataki E, Eboriadou M, Galli-Tsinopoulou A. The role of PTPN22 C1858T gene polymorphism in diabetes mellitus type 1: first evaluation in Greek children and adolescents. Biomed Res Int. 2013;2013:721604. doi: 10.1155/2013/721604. Epub 2013 Jul 15. PMID: 23936838; PMCID: PMC3727122.

H

Hanna SJ, Tatovic D, Thayer TC, Dayan CM. Insights from single cell RNA sequencing into the immunology of type 1 diabetes: cell phenotypes and antigen specificity. Front Immunol. 2021 Oct 1;12:751701. doi:10.3389/fimmu.2021.751701.

Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009 Jan;37(1):1-13. doi: 10.1093/nar/gkn923. Epub 2008 Nov 25. PMID: 19033363; PMCID: PMC2615629.]

I

Irvine KM, Gallego P, An X, Best SE, Thomas G, Wells C, Harris M, Cotterill A, Thomas

R. Peripheral blood monocyte gene expression profile clinically stratifies patients with recent-onset type 1 diabetes. *Diabetes*. 2012 May;61(5):1281-90. doi: 10.2337/db11-1549. Epub 2012 Mar 8. PMID: 22403299; PMCID: PMC3331753.

L

Lizzo JM, Goyal A, Gupta V. Adult Diabetic Ketoacidosis. 2023 Jul 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 32809558.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNAseq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.

Lucier J, Mathias PM. Type 1 Diabetes. 2024 Oct 5. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 29939535.

M

Mantovani F, Banks L. Regulation of the discs large tumor suppressor by a phosphorylation-dependent interaction with the beta-TrCP ubiquitin ligase receptor. *J Biol Chem*. 2003 Oct 24;278(43):42477-86. doi: 10.1074/jbc.M302799200. Epub 2003 Aug 5. PMID: 12902344.

Molitch ME, Steffes M, Sun W, Rutledge B, Cleary P, de Boer IH, Zinman B, Lachin J, Epidemiology of Diabetes Interventions and Complications Study Group. Development and progression of renal insufficiency with and without albuminuria in adults with type 1 diabetes in the diabetes control and complications trial and the epidemiology of diabetes interventions and complications study. *Diabetes Care*. 2010 Jul;33(7):1536-43. doi: 10.2337/dc09-1098. PMID: 20413518; PMCID: PMC2890355.

Mitra D, Mitra D, Bensaad MS, Sinha S, Pant K, Pant M, Priyadarshini A, Singh P, Dassamiour S, Hambaba L, Panneerselvam P, Das Mohapatra PK. Evolution of bioinformatics and its impact on modern bio-science in the twenty-first century: special attention to pharmacology, plant science and drug discovery. *Comput Toxicol*. 2022;24:100234.

N

Ndambwe Moussio V, Jemea B, Bayiha JE, Epacka Ewane M, Ngo Nonga B. Prévalence de l'artériopathie oblitérante des membres inférieurs chez les patients diabétiques à l'Hôpital Général de Douala. *Health Sci Dis.* 2021 Apr;22(4):29–34.

Niu X, Pan X, He G, Xuan C, Tian Q, Yuan Y, Chen J, Song Y, Tang Y, Zhou T. To determine the role of TRIT1 in the diagnosis, prognosis and immunoinvasion of liver hepatocellular carcinoma. *Front Immunol.* 2025 May 5;16:1563442. doi: 10.3389/fimmu.2025.1563442. PMID: 40391218; PMCID: PMC12086158.

P

Pant T, Lin CW, Bedrat A, Jia S, Roethle MF, Truchan NA, Ciecko AE, Chen YG, Hessner MJ. Monocytes in type 1 diabetes families exhibit high cytolytic activity and subset abundances that correlate with clinical progression. *Sci Adv.* 2024 May 17;10(20):eadn2136. doi: 10.1126/sciadv.adn2136. PMID : 38758799; PMCID: PMC11100571.

Philips JC, Radermecker RP. Le diabète de type 1 : de la prédisposition génétique à un contexte environnemental hypothétique. *Rev Med Liège.* 2012;67(5-6):319-325.

Paolino M, Penninger JM. Cbl-b in T-cell activation. *Semin Immunopathol.* 2010 Jun;32(2):137-48. doi: 10.1007/s00281-010-0197-9. Epub 2010 Feb 21. PMID: 20458601.

Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA.

Powers AC. Type 1 diabetes mellitus: much progress, many opportunities. *J Clin Invest.* 2021 Apr 15;131(8):e142242. doi: 10.1172/JCI142242. PMID: 33759815; PMCID: PMC8262558.

R

Robles-Osorio ML, Sabath E. Tight junction disruption and the pathogenesis of the chronic complications of diabetes mellitus: A narrative review. *World J Diabetes.* 2023 Jul 15;14(7):1013-1026. doi: 10.4239/wjd.v14.i7.1013. PMID: 37547580; PMCID: PMC10401447.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*

Res. 2015 Apr 20;43(7):e47. doi: 10.1093/nar/gkv007. Epub 2015 Jan 20. PMID: 25605792; PMCID: PMC4402510.

S

Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, Malonia SK. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)*. 2023 Jul 13;12(7):997. doi: 10.3390/biology12070997. Erratum in: *Biology (Basel)*. 2024 Apr 24;13(5):286. doi: 10.3390/biology13050286. PMID: 37508427; PMCID: PMC10376292.

Sivitz WI, Yorek MA. Mitochondrial dysfunction in diabetes: from molecular mechanisms to functional significance and therapeutic opportunities. *Antioxid Redox Signal*. 2010 Apr;12(4):537-77. doi: 10.1089/ars.2009.2531. PMID: 19650713; PMCID: PMC2824521.

Solis-Herrera C, Triplitt C, Reasner C, DeFronzo RA, Cersosimo E. Classification of Diabetes Mellitus. 2018 Feb 24. In: Feingold KR, Ahmed SF, Anawalt B, Blackman MR, Boyce A, Chrousos G, Corpas E, de Herder WW, Dhatariya K, Dungan K, Hofland J, Kalra S, Kaltsas G, Kapoor N, Koch C, Kopp P, Korbonits M, Kovacs CS, Kuohung W, Laferrère B, Levy M, McGee EA, McLachlan R, Muzumdar R, Purnell J, Rey R, Sahay R, Shah AS, Singer F, Sperling MA, Stratakis CA, Trencle DL, Wilson DP, editors. *Endotext* [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000–. PMID: 25905343.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. 2016 Jun 20;54:1.30.1-1.30.33. doi: 10.1002/cpbi.5. PMID: 27322403.

Strand N, Anderson M.A, Attanti S, Gill B, Wie C, Dawodu A, Pagan-Rosado R, Harbell MW, Maloney JA. Diabetic Neuropathy: Pathophysiology Review. *Curr Pain Headache Rep*. 2024 Jun;28(6):481-487. doi: 10.1007/s11916-024-01243-5. PMID: 38558164.

Strand N, Anderson MA, Attanti S, Gill B, Wie C, Dawodu A, Pagan-Rosado R, Harbell MW, Maloney JA. Neuropathie diabétique: revue de physiopathologie. *Curr Pain Headache Rep*.

2024 Jun;28:481–7. doi: 10.1007/s11916-024-01243-5.

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021 : customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D605-D612. doi: 10.1093/nar/gkaa1074. Erratum in: *Nucleic Acids Res.* 2021 Oct 11;49(18):10800. doi: 10.1093/nar/gkab835. PMID: 33237311; PMCID: PMC7779004.

T

Tsukita S, Furuse M. The structure and function of claudins, cell adhesion molecules at tight junctions. *Ann N Y Acad Sci.* 2000;915:129-35. doi: 10.1111/j.1749-6632.2000.tb05235.x. PMID: 11193568.

V

Visser J, Rozing J, Sapone A, Lammers K, Fasano A. Tight junctions, intestinal permeability, and autoimmunity: celiac disease and type 1 diabetes paradigms. *Ann N Y Acad Sci.* 2009 May;1165:195-205. doi: 10.1111/j.1749-6632.2009.04037.x. PMID: 19538307; PMCID: PMC2886850.

W

Wu L, Gao L, Zhao X, Zhang M, Wu J, Mi J. A new risk locus in CHCHD5 for hypertension and obesity in a Chinese child population: a cohort study. *BMJ Open.* 2017 Sep 11;7(9):e016241. doi: 10.1136/bmjopen-2017-016241. PMID: 28893745; PMCID: PMC5722095.

X

Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc.* 2021 Mar;1(3):e90. doi: 10.1002/cpz1.90.

PMID: 33780170; PMCID: PMC8152575

Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based metaanalysis of gene expression data. *Nat Protoc.* 2015 Jun;10(6):823-44. doi: 10.1038/nprot.2015.052. Epub 2015 May 7. PMID: 25950236.

Y

Yang CF, Tsai WC. Calmodulin: The switch button of calcium signaling. *Tzu Chi Med J.* 2021 Aug 23;34(1):15-22. doi: 10.4103/tcmj.tcmj_285_20. PMID: 35233351; PMCID: PMC8830543.

Z

Zhou Q, Melton DA. Pancreas regeneration. *Nature.* 2018 May;557(7705):351-358. doi: 10.1038/s41586-018-0088-0. Erratum in: *Nature.* 2018 Aug;560(7720):E34. doi: 10.1038/s41586-018-0294-9. PMID: 29769672; PMCID: PMC6168194.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019 Apr 3;10(1):1523. doi: 10.1038/s41467-019-09234-6. PMID: 30944313; PMCID: PMC6447622.

Zihni, C., Mills, C., Matter, K. et al. Tight junctions: from simple barriers to multifunctional molecular gates. *Nat Rev Mol Cell Biol* **17**, 564–580 (2016).
<https://doi.org/10.1038/nrm.2016.80>

Annexes

-Diabète type 2 :

Le diabète de type 2, également appelé diabète non insulino-dépendant (DNID) ou diabète de l'adulte représente la majorité des cas (90 à 95 %), englobe les personnes qui ont une résistance à l'insuline et qui ont généralement une carence en insuline relative (l'insuline est toujours présente, mais pas en quantité suffisante par rapport aux besoins de l'organisme) (**American Diabetes Association, 2010**)

La majorité des patients atteints de diabète de type 2 présentent une obésité abdominale (graisse autour des organes). Ce type de graisse est une forme de graisse ectopique (graisse stockée dans des endroits inhabituels) qui est étroitement liée à la résistance à l'insuline (**Solis-Herrera *et al.*, 2018**).

-Physiologie du Pancréas :

Le pancréas joue un rôle crucial dans la digestion et la régulation du métabolisme énergétique. Il est constitué de deux parties morphologiquement et fonctionnellement : le pancréas exocrine, formé par les cellules acineuses et canalaire, et le pancréas endocrinien, constitué des îlots de Langerhans. (**Zhou *et Melton*, 2018**).

Les cellules acineuses exocrines produisent un ensemble d'enzymes digestives, notamment des lipases, des protéinases et des amylases, qui sont sécrétées dans les canaux pancréatiques et s'écoulent dans l'intestin grêle pour décomposer les graisses, les protéines et les glucides en vue de leur absorption. Les îlots endocrines représentent moins de 5 % de la masse pancréatique totale, mais comptent néanmoins plus d'un milliard de cellules chez l'homme. Chaque type cellulaire des îlots de Langerhans sécrète une hormone spécifique : (les cellules β) produisent l'insuline, (les cellules α) le glucagon, (les cellules δ) la somatostatine, (les cellules PP) le polypeptide pancréatique et (les cellules ϵ) la ghréline (**Zhou *et Melton*, 2018**).

-Insuline :

L'insuline est une hormone peptidique convertie à partir de la proinsuline dans les cellules bêta du pancréas. La proinsuline est composée d'une chaîne A (21 acides aminés), d'une chaîne B (30 acides aminés) et d'un peptide C (31 acides aminés). L'insuline mature est produite après l'excision du peptide C par une série de clivages protéolytiques. L'insuline a pour fonction de réguler la glycémie et d'induire son stockage dans le foie, les muscles et le tissu adipeux.

(Kawasaki, 2023)

-La sécrétion de l'insuline :

Lorsque la glycémie augmente, le glucose entre dans les cellules bêta du pancréas grâce à des transporteurs spécifiques appelés GLUT2. Une fois dans la cellule, le glucose subit la glycolyse qui le transforme en pyruvate. Ce pyruvate est ensuite utilisé par la mitochondrie pour produire de l'ATP (Daziano, 2021).

L'augmentation de la concentration en ATP provoque la fermeture des canaux potassiques KATP. En empêchant la sortie des ions potassium (K^+), cela entraîne leur accumulation dans le cytoplasme, modifiant la charge électrique de la cellule et provoquant une dépolarisation de la membrane plasmique (Daziano, 2021).

Cette dépolarisation ouvre des canaux calciques voltage-dépendants, permettant une entrée massive d'ions calcium (Ca^{2+}) dans la cellule. Cette augmentation de calcium intracellulaire déclenche alors l'exocytose des granules d'insuline, libérant l'hormone dans le sang pour réguler la glycémie (Daziano, 2021).

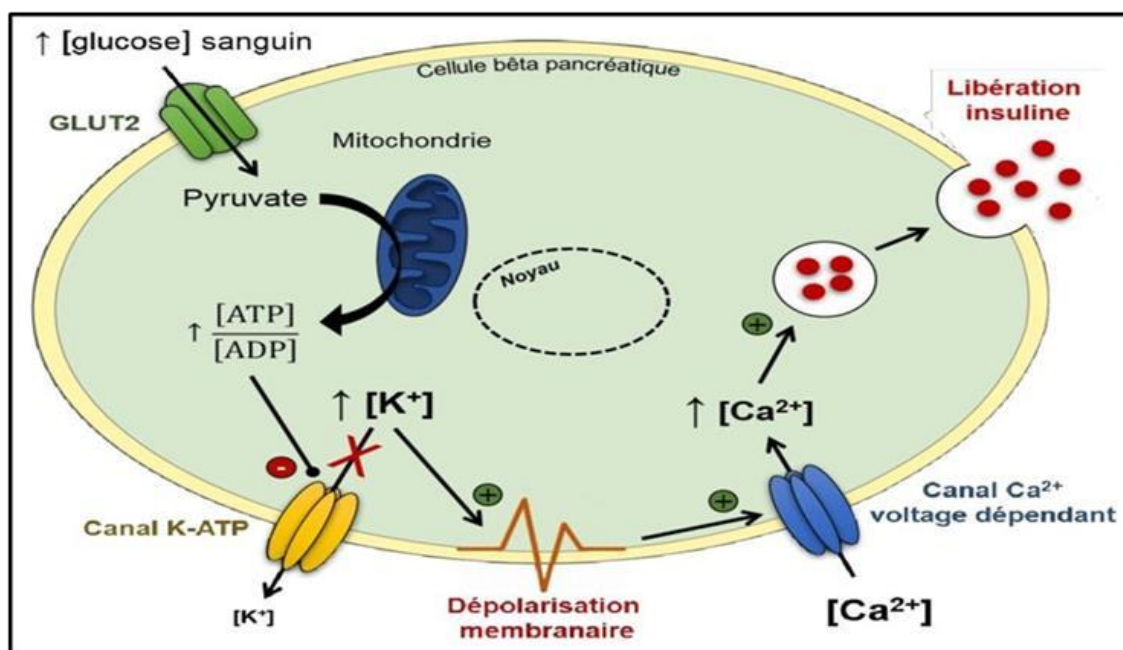


Figure 29 : Des différentes étapes impliquées dans la sécrétion d'insuline (Daziano, 2021)
 Liste des gènes identifier pour les groupes : « uHC vs ROTD1 » ; « HRS vs LRS » ; « HRS vs uHC » :

<https://we.tl/t-jXYdHDCW2>

[illegible]

```

# Build histogram of P-values for all genes. Normal test #
assumption is that most genes are not differentially expressed. tT2
<- topTable(fit2, adjust="fdr", sort.by="B", number=Inf)
hist(tT2$adj.P.Val, col = "grey", border = "white", xlab = "P-adj",
ylab = "Number of genes", main = "P-adj value distribution")

# summarize test results as "up", "down" or "not expressed" dT <-
decideTests(fit2, adjust.method="fdr", p.value=0.05, lfc=0)

# Venn diagram of results
vennDiagram(dT, circle.col=palette())

# create Q-Q plot for t-statistic
t.good <- which(!is.na(fit2$F)) # filter out bad probes
qqt(fit2$t[t.good], fit2$df.total[t.good], main="Moderated t statistic")
# volcano plot (log P-value vs log fold change)
colnames(fit2) # list contrast names ct <- 1
# choose contrast of interest
# Please note that the code provided to generate graphs serves as a
guidance to
# the users. It does not replicate the exact GEO2R web display due to
multitude
# of graphical options.
#
# The following will produce basic volcano plot using limma function:
volcanoplot(fit2, coef=ct, main=colnames(fit2)[ct], pch=20,
highlight=length(which(dT[,ct]!=0)), names=rep('+', nrow(fit2)))

# MD plot (log fold change vs mean log expression) # highlight
statistically significant (p-adj < 0.05) probes plotMD(fit2,
column=ct, status=dT[,ct], legend=F, pch=20, cex=1) abline(h=0)

#####
# General expression data analysis ex
<- exprs(gset)

# box-and-whisker plot
dev.new(width=3+ncol(gset)/6, height=5) ord
<- order(gs) # order samples by group
palette(c("#1B9E77", "#7570B3", "#E7298A", "#E6AB02", "#D95F02",
"#66A61E", "#A6761D", "#B32424", "#B324B3", "#666666")) par(mar=c(7,4,2,1))
title <- paste ("GSE232310", "/", annotation(gset), sep = "")
boxplot(ex[,ord], boxwex=0.6, notch=T, main=title, outline=FALSE, las=2,
col=gs[ord])
legend("topleft", groups, fill=palette(), bty="n") dev.off()

# expression value distribution par(mar=c(4,4,2,1))
title <- paste ("GSE232310", "/", annotation(gset), " value distribution",
sep = "")
plotDensities(ex, group=gs, main=title, legend = "topright")

# UMAP plot (dimensionality reduction) ex <- na.omit(ex)
# eliminate rows with NAs ex <- ex[!duplicated(ex), ] #
remove duplicates ump <- umap(t(ex), n_neighbors = 15,
random_state = 123) par(mar=c(3,3,2,6), xpd=TRUE)
plot(ump$layout, main="UMAP plot, nbrs=15", xlab="", ylab="", col=gs,
pch=20, cex=1.5)

```

```

legend("topright", inset=c(-0.15,0), legend=levels(gs), pch=20,
col=1:nlevels(gs), title="Group", pt.cex=1.5) library("maptools")
# point labels without overlaps
pointLabel(ump$layout, labels = rownames(ump$layout), method="SANN",
cex=0.6)

# mean-variance trend, helps to see if precision weights are needed
plotSA(fit2, main="Mean variance trend, GSE232310")

```

R script pour l'analyse d'expression différentielle de données de comptages RNA-seq

```

# --- 1. Installation et chargement des packages nécessaires ---
# Il est recommandé d'installer BiocManager d'abord si ce n'est pas déjà fait
# if (!requireNamespace("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")

# Installation des packages (décommenter et exécuter si non installés)
# BiocManager::install("DESeq2")
# BiocManager::install("GEOquery") # Pour télécharger les données GEO
# BiocManager::install("Biobase") # Dépendance de GEOquery pour ExpressionSet
# install.packages("ggplot2")
# install.packages("EnhancedVolcano")
# install.packages("pheatmap")

# Chargement des packages
library(DESeq2)
library(GEOquery)
library(Biobase)
library(ggplot2)
library(EnhancedVolcano)
library(pheatmap)

# --- 2. Préparation des données d'entrée : Téléchargement et chargement de
GSE232310 ---

# 2.1. Téléchargement des données GEO
# Le paramètre 'GSEMatrix = TRUE' tente de télécharger les données sous forme
de matrice d'expression
# Cela peut prendre un certain temps en fonction de la taille du
dataset cat("Téléchargement des données GSE232310 depuis GEO...\n") gse
<- getGEO("GSE232310", GSEMatrix = TRUE, AnnotGPL = FALSE)

```

```

# Si getGEO retourne une liste (plusieurs plateformes), sélectionnez
la première if (length(gse) > 1) { gse <- gse[[1]]
} cat("Données GSE232310 téléchargées et
chargées.\n")
# 2.2. Extraction de la matrice de comptages (expression data) et des données
phénotypiques (colData)
# Pour les microarrays, 'exprs' contient les données d'expression normalisées.
# Pour RNA-seq, on chercherait des raw counts, mais GEO peut fournir des
données déjà traitées.
# Il est crucial de vérifier la nature exacte des données dans 'exprs(gse)'. #
Si ce sont des valeurs normalisées (ex: log2 RMA), DESeq2 attend des comptages
bruts.
# Dans ce cas, ce script serait adapté pour des données déjà normalisées ou il
faudrait trouver les raw counts.
# Pour l'exemple, nous allons utiliser les valeurs d'expression directement si
elles sont suffisantes,
# mais idéalement, DESeq2 nécessite des comptages bruts.
# Si vous avez un fichier de comptages bruts séparément, utilisez-le ici.
# Par exemple: counts_data <- read.csv("your_raw_counts.csv", row.names=1)
# Pour les données de microarray comme GSE232310, exprs(gse) contient souvent
des valeurs log-transformées.
# DESeq2 est conçu pour les comptages bruts. Si vous n'avez pas accès aux
comptages bruts de GSE232310,
# vous devrez adapter votre analyse (par exemple, utiliser limma pour données
normalisées,
# ou trouver les comptages bruts si disponibles sur GEO).
# Pour des raisons d'illustration, nous allons simuler des "comptages" à
partir des données normalisées si elles ne sont pas brutes.
# Dans un cas réel, ASSUREZ-VOUS d'avoir les RAW COUNTS.
counts_data_raw <- Biobase::exprs(gse)
# Si les données sont log-transformées, les "délog-transformer" peut les
rendre non entières,
# ce qui n'est pas idéal pour DESeq2. Nous allons simuler des comptages bruts
ici pour que DESeq2 fonctionne.
# Dans une VRAIE analyse RNA-seq, vous auriez un fichier de comptages bruts
(ex: de featureCounts, Salmon, Kallisto). counts_data <-
round(2^counts_data_raw) # Conversion simplifiée pour l'exemple
counts_data[counts_data == 0] <- 1 # Assurer qu'il n'y a pas de zéros après
arrondi pour log
col_data <-
Biobase::pData(gse)

# 2.3. Nettoyage et préparation de col_data pour les conditions

```



```

# Vérifiez les noms de colonnes dans col_data pour identifier la colonne des
conditions.
# Ex: col_data$"disease:ch1" ou col_data$"characteristics_ch1.1" etc.
# Utilisez View(col_data) pour inspecter.

# Assumons que la colonne de condition est nommée 'disease_state' ou
similaire.
# Pour GSE232310, d'après les informations, les groupes sont ROT1D, uHC, HRS,
LRS.
# Il faut trouver la colonne correspondante dans pData(gse).
# Pour l'exemple, je vais créer une colonne 'condition' basée sur les
descriptions connues.
# L'utilisateur devra adapter ceci en inspectant 'col_data' de son côté.
# Exemple d'adaptation:
# col_data$condition <- factor(col_data$disease_state:ch1) # Remplacez par
le nom réel de votre colonne
# Ou construisez-la manuellement si les informations sont dans plusieurs
colonnes:

# Simuler la colonne de condition basée sur les noms d'échantillons ou des
heuristiques
# Pour l'exemple, nous allons créer une colonne 'condition' simplifiée.
# L'utilisateur DOIT adapter cela avec les vraies données de col_data
de GSE232310 sample_conditions <- rep(NA, nrow(col_data))
sample_conditions[grepl("ROT1D", col_data$title, ignore.case = TRUE)] <-
"ROT1D" sample_conditions[grepl("uHC", col_data$title, ignore.case = TRUE)]
<- "uHC" sample_conditions[grepl("HRS", col_data$title, ignore.case =
TRUE)] <- "HRS" sample_conditions[grepl("LRS", col_data$title, ignore.case
= TRUE)] <- "LRS"
# Si les titres ne sont pas suffisants, explorez d'autres colonnes de col_data
# View(col_data) # Utilisez cette commande pour explorer col_data dans RStudio

col_data$condition <- factor(sample_conditions)
# Retirer les échantillons sans condition définie si NA sont présents col_data
<- col_data[!is.na(col_data$condition), ]
counts_data <- counts_data[, rownames(col_data)] # Assurer que la matrice de
comptages correspond

# Vérification que les noms d'échantillons correspondent
stopifnot(all(rownames(col_data) ==
colnames(counts_data))) cat("Données de comptages et
phénotypiques préparées.\n")
# --- 3. Création de l'objet DESeqDataSet ---
# C'est l'objet principal qui contient les données de comptages et les
métadonnées.
dds <- DESeqDataSetFromMatrix(

```

```

countData = counts_data, colData = col_data, design = ~ condition #
Spécifie le design expérimental, ici nous comparons les conditions
)

# Filtrage des gènes faiblement exprimés (facultatif mais recommandé) #
Garder uniquement les gènes avec au moins 10 comptages sur au moins 3
échantillons
# Pour de grandes études, vous pouvez ajuster les seuils. keep <-
rowSums(counts(dds) >= 10) >= 3 dds <- dds[keep,] cat(paste0("Nombre
de gènes après filtrage des faibles expressions : ", nrow(dds), "\n"))

# Référence de la condition pour la comparaison
# Pour les comparaisons spécifiques de votre étude (ROT1D vs uHC ou HRS vs
LRS)
# Choisissez la référence appropriée.
# Exemple pour ROT1D vs uHC:
dds$condition <- releval(dds$condition, ref = "uHC")
# Exemple pour HRS vs LRS:
# dds$condition <- releval(dds$condition, ref = "LRS")
# --- 4. Exécution de l'analyse d'expression différentielle ---
# C'est l'étape principale où DESeq2 effectue la normalisation,
l'estimation de la dispersion et le test statistique. cat("Exécution de
l'analyse DESeq2...\n") dds <- DESeq(dds) cat("Analyse DESeq2 terminée.\n")

# --- 5. Récupération des résultats de l'expression différentielle ---
# Spécifiez la comparaison désirée :
# Pour ROT1D vs uHC : res <- results(dds, contrast =
c("condition", "ROT1D", "uHC"))
# Pour HRS vs LRS (décommentez et remplacez la ligne ci-dessus si c'est la
comparaison souhaitée) :
# res <- results(dds, contrast = c("condition", "HRS", "LRS"))
# Trier les résultats par p-value ajustée
res_ordered <- res[order(res$padj),]

# Afficher un résumé des résultats
cat("\nRésumé des résultats DESeq2:\n")
summary(res)

# Afficher les 10 premiers gènes différentiellement exprimés les plus
significatifs

```

```

cat("\nTop 10 des gènes différentiellement exprimés:\n")
print(head(res_ordered, 10))

# Filtrer les gènes significatifs (par exemple, padj < 0.05 et
# |log2FoldChange| > 0.5) res_sig <- subset(res_ordered, padj < 0.05 &
# abs(log2FoldChange) > 0.5) cat(paste0("\nNombre de gènes significatifs
# (padj < 0.05 & |log2FC| > 0.5) : ", nrow(res_sig), "\n"))

# Écrire les résultats dans un fichier CSV (décommenter pour sauvegarder)
# write.csv(as.data.frame(res_ordered), file =
# "resultats_deseq2_GSE232310_complets.csv") #
write.csv(as.data.frame(res_sig), file =
"genes_differentially_expressed_GSE232310_significatifs.csv")
# --- 6. Visualisation des résultats ---

# 6.1. MA plot (visualisation de la moyenne d'expression vs le fold
# change) cat("\nGénération du MA Plot...\n") plotMA(res, main="MA Plot -
# DESeq2 (GSE232310)", ylim=c(-2,2))
# 6.2. Volcano Plot (visualisation du log2FC vs la significativité)
cat("Génération du Volcano Plot...\n") EnhancedVolcano(res,
lab = rownames(res), x = 'log2FoldChange', y
= 'padj', title = 'Volcano Plot : Comparaison (GSE232310)', #
Adaptez le titre à votre comparaison subtitle = 'Gènes
différentiellement exprimés', pCutoff = 0.05, # Seuil pour la
p-valeur ajustée FCcutoff = 0.5, # Seuil pour le log2(Fold
Change) pointSize = 1.5, labSize = 3.0,
colAlpha = 0.8,
legendLabels=c('Non régulé', 'Log2FC', 'Padj', 'Padj
et Log2FC'), drawConnectors = TRUE,
widthConnectors = 0.75, colConnectors = 'grey30')

# 6.3. Heatmap des gènes les plus significatifs cat("Génération
# de la Heatmap...\n")
# Transformation des données pour la visualisation (variance
# stabilizing transformation) vsd <- vst(dds, blind=FALSE)

# Sélection des 50 gènes les plus significatifs pour la heatmap (ou moins si
# res_sig est petit)

```

```
# Assurez-vous d'avoir suffisamment de gènes significatifs if
(nrow(res_sig) > 0) {  num_top_genes_for_heatmap <- min(50,
nrow(res_sig))  top_genes <- head(rownames(res_sig),
num_top_genes_for_heatmap) # Utilise les gènes significatifs  mat <-
assay(vsd)[top_genes, ]  mat <- mat - rowMeans(mat) # Centrer les données
par gène pour mieux voir les différences relatives

  # Ajouter des annotations pour les colonnes (conditions)
annotation_col = data.frame(
  Condition = col_data[colnames(mat), "condition"], # S'assurer que les
noms correspondent      row.names = colnames(mat)
)  pheatmap(mat,          cluster_rows = TRUE,
show_rownames = TRUE, # Afficher les noms des gènes
cluster_cols = TRUE,      annotation_col = annotation_col,
main = paste0("Heatmap des ", num_top_genes_for_heatmap, " gènes les plus
significatifs"))
} else {  cat("Aucun gène significatif trouvé pour générer une
heatmap.\n") }

# --- Fin du script d'analyse DESeq2 ---
```



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA

RECHERCHE SCIENTIFIQUE

UNIVERSITE SAAD DAHLEB BLIDA 1

Faculté des Sciences de la Nature et de la vie

Département de Biologie Et Physiologie Cellulaire

Option : Biologie Moléculaire Et Cellulaire

Mémoire de projet de fin d'études en vue de l'obtention du diplôme

de Master

Thème :

*Identification des gènes et des voies de signalisations associés
maladies auto-immunes cas Diabète type1 par analyse
bio-informatique.*

Présenté par :

BRAHIMI YOUSRA

KERBADJ CHAIMA



Soutenu le : Lundi 07 juillet 2025.

Devant le jury composé de :

Présidente : HAMZI Wahiba

MCA-Université Saad Dahleb Blida1

Examineur : KHIER Mokrane

MCB-Université Saad Dahleb Blida1

Promoteur: SENINET Mohamed El-Hadi

Bioinformaticien d'étude, Recherche et

Developpement.

M. KHIER.
[Signature]

Année universitaire : 2024/2025