

Popular Democratic Republic of Algeria
Ministry of High Education and
Scientific Research
Abbes Laghrour University- Khenchela-
Natural and life sciences Faculty
Molecular and Cellular Biology
Department



MEMOIRE DE FIN D'ETUDES DE MASTER ACADEMIQUE

Domaine : **Sciences de la nature et de la vie**

Filière : **Sciences Biologiques**

Spécialité : **Génétique**

Présenté par :

SAADI Ferial

SOUALHI HibatALLAH

Thème

Identification des variants génétique associés à
l'asthme à l'aide d'approches bioinformatique et
génomique

Mémoire soutenu publiquement le 18/06 2025 Devant le jury composé de :

Dr. BOUAZZA	L.	Président
Dr. RAHAL	K.	Examineur
Pr. HAMIDECHI	M.A.	Encadrant

Année Universitaire 2024/2025

Dédicace

Avec tout le respect que je vous dois, et l'expression sincère de ma reconnaissance, je dédie l'obtention de mon diplôme et ma joie qui l'accompagne.

*Tout d'abord, à l'âme de **ma chère maman** qui a quitté ce monde trop tôt ... à mon paradi., Tu n'es plus à mes côtés, mais ton souvenir continue de m'accompagner à chaque instant. C'est dans ton amour inconditionnel et dans les sacrifices que tu as faits pour moi que j'ai puisé la force d'avancer. Tu vivras à jamais dans mon cœur, avec tendresse et gratitude infinies*

***À mon père,** Merci papa, pour tout ce que tu as fait pour moi, pour tes sacrifices et ton amour inconditionnel.*

*À mes frères, **Adam** et **Hicham**, pour leur soutien inébranlable et leur inspiration. Merci d'avoir été mes anges gardiens, cette réussite et aussi la votre*

À la femme de mon frère,

Pour sa douceur, et son soutien discret qui ont marqué ce parcours.

*À **Tamim**, notre petit neveu*

La lumière de ma vie.

*À mon mari, **Imad***

Tu as été ma lumière dans les moments sombres, ma force quand je faiblissais, ton amour et ta confiance en moi ont été les moteurs silencieux de mon courage. Sans toi, ce rêve n'aurait pas eu la même saveur.

*A mon **Petit Ange**, avant même de te voir, je t'aime déjà plus que tout.*

*Sans oublier mon binôme **HibatAllah** pour son soutien moral sa patience et sa compréhension tout au long de ce projet.*

À toute ma famille, à tous mes amis.

À tous ceux que j'aime et ceux qui m'aiment.

FERIEL SAADI

Dédicace

Avant tout, louange et gratitude à **Dieu**, par Sa grâce les bonnes choses s'accomplissent.

C'est avec Son aide que j'ai pu franchir cette étape importante de mon parcours académique.

À **mon père**, mon repère et sources de soutien et de motivation, qui m'a appris que l'ambition n'a pas de limites.

À **ma mère**, douce et précieuse, la source d'amour dont les prières ont toujours été ma lumière dans les moments sombres.

Merci d'avoir été ma force dans chaque étape.

À **Hadil et Rihab**, mes sœurs adorées, votre tendresse et vos mots réconfortants ont toujours fait la différence.

À **Haitham**, mon petit frère, toujours prêt à répondre à mes demandes sans jamais se plaindre.

À **Hanane**, ma sœur de cœur, celle que la vie m'a offerte, ma confidente, celle qui a tout partagé avec moi – les rêves, la fatigue, les larmes et les fous rires. Tu as été mon véritable soutien et rien ne suffira jamais à exprimer ma reconnaissance. Ta présence est un soutien unique.

À **Hoyam**, rencontrée à la résidence universitaire et devenue une sœur. Notre rencontre était une belle surprise, et ton amitié un cadeau inestimable.

À **Feriel**, mon binôme, ta compagnie un trésor. Tu as ajouté une saveur particulière à chaque instant.

À tous ceux et celles qui ont marqué mon chemin, même pour un instant, je vous dédie ce travail avec tout mon amour et ma reconnaissance.

SOUALHI HibatAllah

Remerciements

Nous tenons tout d'abord à exprimer notre profonde gratitude à Monsieur le professeur HAMIDECHI M. A. notre encadrant pour sa précieuse aide, ses conseils avisés, sa disponibilité constante et sa bienveillance tout au long de ce travail. Son accompagnement a été essentiel à l'aboutissement de ce mémoire.

Je tiens à remercier sincèrement les membres du jury, en particulier Monsieur BOUAZZA L., président du jury, et Monsieur RAHAL K., examinateur, pour avoir accepté d'évaluer ce travail et pour le temps précieux qu'ils y ont consacré.

Nous remercions également l'ensemble des enseignants et enseignantes du Master Génétique, ainsi que tous les membres du département de Biologie Moléculaire, pour la qualité de l'enseignement dispensé au cours de ces deux années.

Un grand merci à nos collègues et camarades de promotion pour leur soutien, leur esprit d'équipe et les échanges enrichissants qui ont marqué cette belle aventure universitaire.

Enfin, nous dédions une pensée toute particulière à nos familles, notamment à nos parents, pour leur amour, leur soutien moral inconditionnel et leur patience. Sans eux, rien n'aurait été possible.

Résumé

Ce travail s'inscrit dans une approche combinée de bioinformatique et de génétique visant à identifier les polymorphismes mononucléotidiques (SNPs) associés à l'asthme. En s'appuyant sur les bases de données comme NCBI/dbSNP, MutationT@ster et gnomAD, nous avons analysé des SNPs liés à la maladie en évaluant leur fréquence, leur impact fonctionnel potentiel et leur répartition populationnelle. Une analyse phylogénétique a été réalisée à partir des séquences nucléiques contenant ces variants, dans le but d'étudier leur évolution et leur conservation. Ce mémoire met ainsi en lumière l'importance de la caractérisation génétique des variants dans la compréhension des mécanismes de susceptibilité à l'asthme.

Mots clés : *Variants génétiques – Asthme – Génomique - Bioinformatique*

ملخص

يتناول هذا العمل دراسة مشتركة بين المعلوماتية الحيوية وعلم الوراثة بهدف تحديد الطفرات النقطية المرتبطة (SNPs) بالربو.

تم اعتماد قواعد بيانات مثل لتحليل هذه الطفرات من حيث NCBI/dbSNP وMutationT@ster وgnomAD تكرارها، وتأثيرها الوظيفي المحتمل، وتوزيعها بين السكان. كما أُجري تحليل تطوري (فيلوجيني) باستخدام تسلسلات الحمض النووي التي تحتوي على هذه المتغيرات، وذلك لفهم مدى تطورها والمحافظة عليها عبر الأنواع. يسلط هذا البحث الضوء على أهمية تحليل المتغيرات الجينية لفهم آليات الاستعداد الوراثي للإصابة بالربو.

الكلمات المفتاحية: *المتغيرات الجينية – الربو – علم الجينوم – المعلوماتية الحيوية*

Abstract

This study is part of a combined bioinformatics and genetics approach aimed at identifying single nucleotide polymorphisms (SNPs) associated with asthma. Utilizing databases such as NCBI/dbSNP, MutationT@ster, and gnomAD, we examined disease-related SNPs by evaluating their allele frequency, predicted functional impact, and distribution across different populations. Additionally, a phylogenetic analysis was conducted using nucleotide sequences containing these variants to investigate their evolutionary conservation. This work underscores the importance of genetic variant characterization in advancing our understanding of the genetic predisposition to asthma.

Keywords: *Genetic variants – Asthma – Genomics – Bioinformatics.*

Table des Matières

INTRODUCTION.....	1
PARTIE 1 : ÉTUDE BIBLIOGRAPHIQUE	2
1- LES VARIANTS GÉNÉTIQUES ET LEUR RÔLE DANS LES MALADIES GÉNÉTIQUES	2
1-1/ Définition et types de variants génétiques :	2
1-2/ Les variants génétiques peuvent être associés à des maladies selon leur localisation et leur effet sur l'expression ou la fonction des gènes.....	3
2- IMPLICATION DES VARIANTS GÉNÉTIQUES DANS LA SUSCEPTIBILITÉ À L'ASTHME	4
2-1/ Exemples d'implication	4
2-2/ Les variants liés à l'asthme les plus communs	5
3- DISTRIBUTIONS STATISTIQUES DES VARIANTS ASSOCIÉS À L'ASTHME	7
1. Variants identifiés par des études pangénomiques	7
2. Variants spécifiques à certaines populations	7
3. Variants liés à l'atopie.....	7
4. Variants identifiés dans des populations spécifiques.....	8
5. Importance des études d'association pangénomique	8
4- MÉTHODES D'ÉTUDE DES ASSOCIATIONS GÈNES/ASTHME	11
1. Études d'Association Pangénomique (GWAS).....	11
2. Statistiques de l'Asthme	12
5. LES PRINCIPAUX TYPES DE PUCES ADN	13
5-1/ Les types de microarrays	15
5-2/ Les types d'études conduites avec une puce à ADN	16

5-3/ Le pré-traitement des données	17
PARTIE 2 : MATÉRIEL & MÉTHODES.....	18
1. PRÉTRAITEMENT ET VALIDATION DES DONNÉES.....	18
2. ANALYSE DES SNPs.....	22
i- MutationT@ster	22
ii- gnomAD	22
2-1/ Cas du variant rs2070901	22
3. IDENTIFICATION DES VARIANTS ASSOCIÉS À L'ASTHME	24
4. ANALYSE PHYLOGÉNÉTIQUE DES VARIANTS GÉNÉTIQUES.....	26
PARTIE 3 : RÉSULTATS & DISCUSSION	27
1- ANALYSE DES SNPs	27
1-1/ Annotation NCBI/SNP	27
1-2/ Analyse du SNP (rs2070901) sur MutationT@ster	27
1-3/ Analyse sur gnomAD	27
1-4/ Autres variants	28
2- IDENTIFICATION DES VARIANTS ASSOCIÉS À L'ASTHME	34
3- ANALYSE PHYLOGÉNÉTIQUE	35
CONCLUSION.....	40

Liste des figures

Figure 1 :Graphique de Manhattan des résultats des méta analyses effets aléatoires d'ascendance européenne et multi-ascendance du risque d'asthme.....	9
Figure 2 : Taux de prévalence de l'asthme standardisés selon l'âge et pourcentage moyen de personnes.....	14
Figure 3 : Représentation schématique des étapes de la réalisation de microarrays.....	16
Figure 4 : Étapes d'identification des variants associés à l'asthme à partir d'une liste de SNPs.....	27
Figure 5 : Cartographie du variant rs4286272 sur le chromosome 2 via le Genome Data Viewer (NCBI).....	31
Figure 6 : Vue globale du résultat de l'alignement multiple des variants génétique.....	38
Figure 7 : Vue agrandie d'une partie centrale de l'alignement multiple	39
Figure 8 : Organisation phylogénétique des variants génétiques liés à l'asthme.....	41

Liste des tableaux

Tableau 1 : Statistiques des associations observées	11
Tableau 2 : Caractéristiques des variants génétiques (SNPs) liés à l’asthme : localisation, fréquence et type de mutation.....	21_22
Tableau 3 : Résumé des paramètres clés de l’analyse de rs2070901 sue GenomeAD.....	30
Tableau 4 : Caractéristiques fonctionnelles et évolutives des variants génétiques.....	32
Tableau 5 : Fréquences alléliques du variant rs4286272 en fonction de la géographie et du sexe.....	34
Tableau 6 : Identification des SNPs en lien avec l’asthme à partir d’une analyse croisée de bases de données génétiques.....	36
Tableau 7 : Variants sans association significative avec l’asthme.....	37

INTRODUCTION

L'asthme est une maladie respiratoire chronique complexe qui a touché, selon l'OMS¹, près de 262 millions de personnes en 2019. D'autres études estiment ce nombre à environ 300 millions de personnes dans le monde, avec une tendance à la hausse de 50% tous les dix ans. Ces dernières années, les progrès en génomique ont permis de mettre en évidence de nombreux variants génétiques qui peuvent influencer l'expression ou la fonction de gènes clés dans les réponses immunitaires, inflammatoires ou encore dans la régulation de la fonction pulmonaire (Braman, 2006 ; Palikhe, 2024).

Dans ce contexte, l'utilisation d'approches bioinformatiques s'avère indispensable pour cribler efficacement de grands volumes de données génétiques et pour identifier les variants potentiellement impliqués dans la pathogénie de l'asthme. Des bases de données comme NCBI/dbSNP, MutationT@ster et gnomAD permettent de caractériser les SNPs d'intérêt en fournissant des informations sur leur localisation, leur fréquence allélique dans différentes populations, ainsi que leur impact fonctionnel présumé.

Par ailleurs, l'analyse phylogénétique des séquences nucléiques contenant ces SNPs offre une perspective complémentaire et enrichissante. Elle permet d'identifier des variants conservés ou spécifiques, et peut également révéler des relations structurales ou fonctionnelles entre différents loci associés à l'asthme.

Pour répondre à la problématique corrélée à ce travail stipulant que *'l'asthme, étant une maladie à forte composante génétique, comment peut-on, à l'aide d'approches bioinformatiques et génétiques, identifier et caractériser les SNPs associés à cette pathologie, et que peut nous apprendre l'analyse phylogénétique de ces variants sur leur origine, leur évolution et leur rôle potentiel dans la susceptibilité à l'asthme ?'* ; nous avons alors fixé un objectif principal de ce mémoire afin d'identifier et caractériser les variants génétiques associés à l'asthme à l'aide d'outils bioinformatiques, et de réaliser une analyse phylogénétique sur la base des séquences nucléiques correspondantes, dans le but de mieux comprendre leur rôle dans la susceptibilité à cette maladie.

¹ https://www.who.int/news-room/fact-sheets/detail/asthma%20%20/?utm_source=chatgpt.com

PARTIE 1 : ÉTUDE BIBLIOGRAPHIQUE

1- LES VARIANTS GÉNÉTIQUES ET LEUR RÔLE DANS LES MALADIES GÉNÉTIQUES

Les variants génétiques sont des modifications de la séquence d'ADN qui influencent la diversité biologique et les prédispositions aux maladies. Leur impact peut être neutre, bénéfique ou pathogène, selon leur effet sur la fonction des gènes (Bouzigon, 2010).

1-1/ Définition et types de variants génétiques : Les variants génétiques jouent un rôle essentiel dans la diversité génétique entre les individus et peuvent influencer des traits physiologiques, des prédispositions aux maladies et la réponse aux traitements médicaux. Ces variations peuvent être bénignes, pathogènes ou avoir un effet incertain, selon leur impact sur la fonction des gènes (Vercelli, 2008).

Ainsi, un variant génétique correspond à toute différence dans la séquence d'ADN par rapport à la séquence de référence du génome humain. Ces variations peuvent être classées en plusieurs catégories :

a-Variants ponctuels (SNPs - Single Nucléotide Polymorphismes)

- Ce sont des modifications d'un seul nucléotide (A, T, C ou G).
- Exemple : Un changement de C en T dans un gène peut modifier ou non l'acide aminé de la protéine codée.

b- Insertions et Délétions (Indels)

- Ajout ou suppression d'une ou plusieurs bases dans la séquence d'ADN.
- Peut provoquer un décalage du cadre de lecture (frameshift mutation) si la taille du changement n'est pas un multiple de trois.

c- Duplications et Expansions Répétées

- Certains gènes contiennent des séquences répétées qui peuvent s'étendre anormalement.

- Exemple : Maladies comme la maladie de Huntington, causée par une expansion excessive de triplets CAG dans le gène HTT.

D- Réarrangements Chromosomiques (Inversions, Translocations)

- Changements dans l'organisation de grandes régions du génome.
- Ces réarrangements peuvent perturber l'expression des gènes et entraîner des maladies.

1-2/ Les variants génétiques peuvent être associés à des maladies selon leur localisation et leur effet sur l'expression ou la fonction des gènes. On distingue :

a- Variants pathogènes dans les maladies monogéniques

- Ces maladies sont causées par des mutations dans un seul gène.
- Exemples :
 - Mutation du gène CFTR → Mucoviscidose
 - Mutation du gène HBB → Drépanocytose

b- Variants associés aux maladies multifactorielles

- Ces maladies résultent de l'interaction entre plusieurs gènes et des facteurs environnementaux.
- Exemples :
 - Asthme (associé aux SNPs dans ORMDL3, GSDMB)
 - Diabète de type 2 (associé à des variantes dans le gène TCF7L2)

c- Polymorphismes neutres et prédisposition génétique

- Certains variants n'ont pas d'effet immédiat mais peuvent augmenter le risque de développer une maladie sous certaines conditions.
- Exemple : Mutations BRCA1 et BRCA2 augmentent le risque de cancer du sein et de l'ovaire, mais ne causent pas systématiquement la maladie.

2- IMPLICATION DES VARIANTS GÉNÉTIQUES DANS LA SUSCEPTIBILITÉ À L'ASTHME

L'asthme est une maladie inflammatoire chronique respiratoire (ou des voies aériennes) résultant d'une interaction complexe entre des facteurs environnementaux et des prédispositions génétiques. Des études ont identifié plusieurs variants génétiques associés à une susceptibilité accrue à l'asthme, offrant ainsi des perspectives pour une meilleure compréhension et gestion de cette pathologie considérée comme hétérogène regroupant de nombreux phénotypes (Villeneuve et Guillemainault, 2020).

2-1/ Exemples d'implication : Parmi les gènes impliqués, ADAM33 a été l'un des premiers identifiés. Ce gène code pour une métalloprotéase impliquée dans le remodelage des voies respiratoires. Des polymorphismes au sein d'ADAM33 ont été associés à une hyperréactivité bronchique et à une inflammation des voies aériennes, suggérant un rôle clé dans la pathogenèse de l'asthme (Vercelli, 2008).

Un autre gène d'intérêt est GPRA (G protein-coupled receptor for asthma susceptibility), également connu sous le nom de GPR154. Des études ont montré que certaines variations de ce gène sont associées à une augmentation des taux d'IgE sériques et à une prédisposition à l'asthme, par exemple, dans les populations finlandaise et québécoise. L'expression accrue de GPRA dans les tissus bronchiques des asthmatiques suggère son implication dans les mécanismes inflammatoires de la maladie.

Les cytokines jouent également un rôle central dans l'asthme. L'interleukine 13 (IL-13), sécrétée principalement par les lymphocytes Th2, est impliquée dans la maturation des cellules B et la production d'IgE. Des polymorphismes du gène IL13 ont été associés à une susceptibilité accrue à l'asthme et à d'autres maladies atopiques. L'IL-13 contribue à l'hyperréactivité des voies respiratoires, à la sécrétion excessive de mucus et au remodelage bronchique, caractéristiques de l'asthme.

Des études d'association à l'échelle du génome ont également mis en évidence des variants situés dans des régions régulatrices, influençant l'expression de gènes liés à l'asthme. Par exemple, une étude internationale a identifié de nouvelles régions du génome associées à l'asthme, situées à proximité de marques épigénétiques dans les

cellules immunitaires, suggérant une interaction entre facteurs génétiques et épigénétiques dans la susceptibilité à l'asthme.

Enfin, le rôle des gènes du complexe majeur d'histocompatibilité de classe II (HLA II) a été exploré. Des associations significatives ont été trouvées entre certains allèles HLA II et la susceptibilité à l'asthme, suggérant que ces gènes pourraient influencer la réponse immunitaire aux allergènes environnementaux.

Ces découvertes soulignent l'importance des facteurs génétiques dans la prédisposition à l'asthme et ouvrent la voie à des approches thérapeutiques ciblées basées sur le profil génétique individuel des patients.

2-2/ Les variants liés à l'asthme les plus communs : Les avancées récentes en génomique ont permis d'identifier plusieurs variants génétiques associés à une susceptibilité accrue à l'asthme. Parmi ces variants, certains sont particulièrement récurrents dans la littérature scientifique contemporaine (Holloway, 2010).

1. Gène ADAM33 : a été l'un des premiers identifiés comme associé à l'asthme. Il code pour une métalloprotéase impliquée dans le remodelage des voies respiratoires. Des polymorphismes au sein d'ADAM33 ont été corrélés à une hyperréactivité bronchique et à une inflammation des voies aériennes, suggérant un rôle clé dans la pathogenèse de l'asthme.

2. Gène GPRA (GPR154) : codant pour un récepteur couplé aux protéines G, a été associé à l'asthme et à des niveaux élevés d'IgE sériques. Des études ont montré que certaines variations de ce gène sont liées à une augmentation des taux d'IgE et à une prédisposition à l'asthme, notamment dans les populations finlandaise et québécoise. L'expression accrue de GPRA dans les tissus bronchiques des asthmatiques suggère son implication dans les mécanismes inflammatoires de la maladie.

3. Interleukine 13 (IL-13) : Une cytokine sécrétée principalement par les lymphocytes Th2, jouant un rôle central dans la réponse immunitaire de type allergique. Des polymorphismes du gène IL13 ont été associés à une susceptibilité accrue à l'asthme, à l'hyperréactivité bronchique et à des niveaux élevés d'IgE. L'IL-13 contribue à

l'hyperréactivité des voies respiratoires, à la sécrétion excessive de mucus et au remodelage bronchique, caractéristiques de l'asthme.

4. Complexe Majeur d'Histocompatibilité de Classe II (HLA II) : Les gènes du Complexe Majeur d'Histocompatibilité de Classe II (HLA II) jouent un rôle crucial dans la présentation des antigènes et la modulation de la réponse immunitaire. Des associations significatives ont été trouvées entre certains allèles HLA II et la susceptibilité à l'asthme, suggérant que ces gènes pourraient influencer la réponse immunitaire aux allergènes environnementaux.

5. Variants Épigénétiques et Régions Régulatrices : Des études d'association à l'échelle du génome ont mis en évidence des variants situés dans des régions régulatrices, influençant l'expression de gènes liés à l'asthme. Par exemple, une étude internationale a identifié de nouvelles régions du génome associées à l'asthme, enrichies en marques épigénétiques, suggérant une interaction entre facteurs génétiques et épigénétiques dans la susceptibilité à l'asthme.

6. Gènes GSTM1, GSTP1 et GSTT1 : Codant pour des glutathion-S-transférases, sont impliqués dans la détoxification des produits oxydants. Des variants nuls (Null Variants : mutation génétique qui entraîne une perte complète de fonction du gène affecté) pour la GSTM1 sont associés à un effet délétère de l'exposition tabagique in utero, augmentant le risque de développer un asthme précoce et persistant.

7. Variants du Gène CACNA1H : Code pour la sous-unité $\alpha 1H$ des canaux calciques de type T, impliqués dans l'excitabilité neuronale. Des haplotypes spécifiques de CACNA1H, cohérités avec des duplications du gène TPSAB1, ont été identifiés chez des individus présentant une alpha-tryptasémie héréditaire, condition associée à des symptômes similaires à l'asthme.

8. Gène PHF11 : A été identifié comme un gène de susceptibilité à l'asthme et aux allergies. Des études ont montré que des variations de PHF11 sont associées à des niveaux d'IgE et à une prédisposition à l'asthme, bien que les mécanismes précis restent à élucider.

Ces découvertes mettent en lumière la complexité de la génétique de l'asthme, impliquant une multitude de gènes et de mécanismes. Une meilleure compréhension de ces variants génétiques ouvre la voie à des approches thérapeutiques personnalisées et à des stratégies de prévention ciblées (Bouziguon, 2010).

3- DISTRIBUTIONS STATISTIQUES DES VARIANTS ASSOCIÉS À L'ASTHME

Les études génétiques ont mis en évidence plusieurs variants associés à l'asthme, avec des fréquences variant selon les populations étudiées. Par exemple, des études ont identifié environ 680 polymorphismes nucléotidiques (SNPs) associés à l'asthme, représentant 90% de l'ensemble des variants génétiques liés à cette maladie (Dos Santos, 2015). La fréquence de ces variants peut varier selon les populations et les méthodologies utilisées, mais leur identification souligne l'importance de la composante génétique dans la susceptibilité à l'asthme (Bouzigon, 2005).

Les avancées en génétique ont permis d'identifier plusieurs variants associés à une susceptibilité accrue à l'asthme, avec des fréquences variant selon les populations (Demenais, 2018) :

1. Variants identifiés par des études pangénomiques : Les études d'association pangénomique (GWAS) ont permis d'identifier de nombreux variants associés à l'asthme. Par exemple, une méta-analyse a mis en évidence 878 variants génétiques situés dans 18 régions du génome, associé au risque d'asthme.

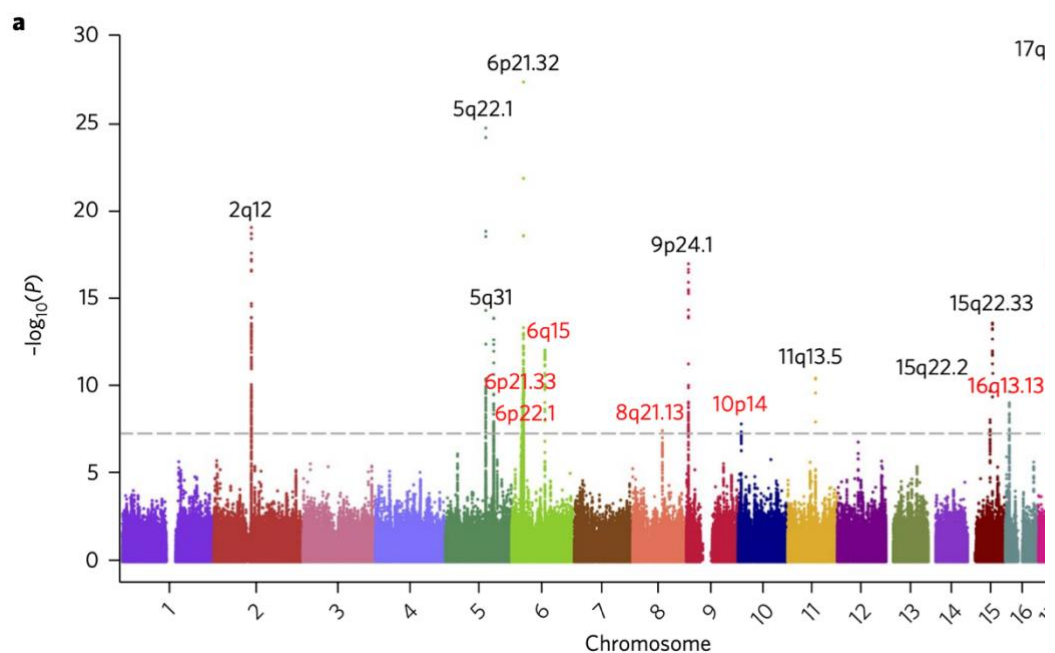
2. Variants spécifiques à certaines populations : Certains variants présentent des fréquences particulières selon les populations étudiées. Par exemple, le gène GPRA (GPR154) a été associé à l'asthme et à des niveaux élevés d'IgE sériques, notamment dans les populations finlandaise et québécoise. Ce gène contient sept haplotypes différents, dont quatre sont associés à l'asthme et à une augmentation des IgE.

3. Variants liés à l'atopie : Environ 75 % des cas d'asthme sont associés à l'atopie, une prédisposition génétique à produire des IgE en réponse à des allergènes environnementaux. Cette association suggère que les variants impliqués dans l'atopie peuvent également influencer la susceptibilité à l'asthme.

4. Variants identifiés dans des populations spécifiques : Des études ont identifié des variants spécifiques à certaines populations. Par exemple, des études ont montré que certaines variations du gène GPRA (GPR154) sont associées à l'asthme et à des niveaux élevés d'IgE sériques, notamment dans les populations finlandaise et québécoise.

5. Importance des études d'association pangénomique : Les études d'association pangénomique (GWAS) ont permis d'identifier de nombreux variants associés à l'asthme. Par exemple, une méta-analyse a mis en évidence 878 variants génétiques situés dans 18 régions du génome, associés au risque d'asthme.

Ces données illustrent la complexité génétique de l'asthme, avec des variants dont les fréquences et les impacts varient selon les populations (Demenais, 2018). Dans leur étude internationale, Demenais et al. (2018), ont identifié de nouvelles régions du génome associées à l'asthme et enrichies en marqueurs épigénétiques (figure 1).



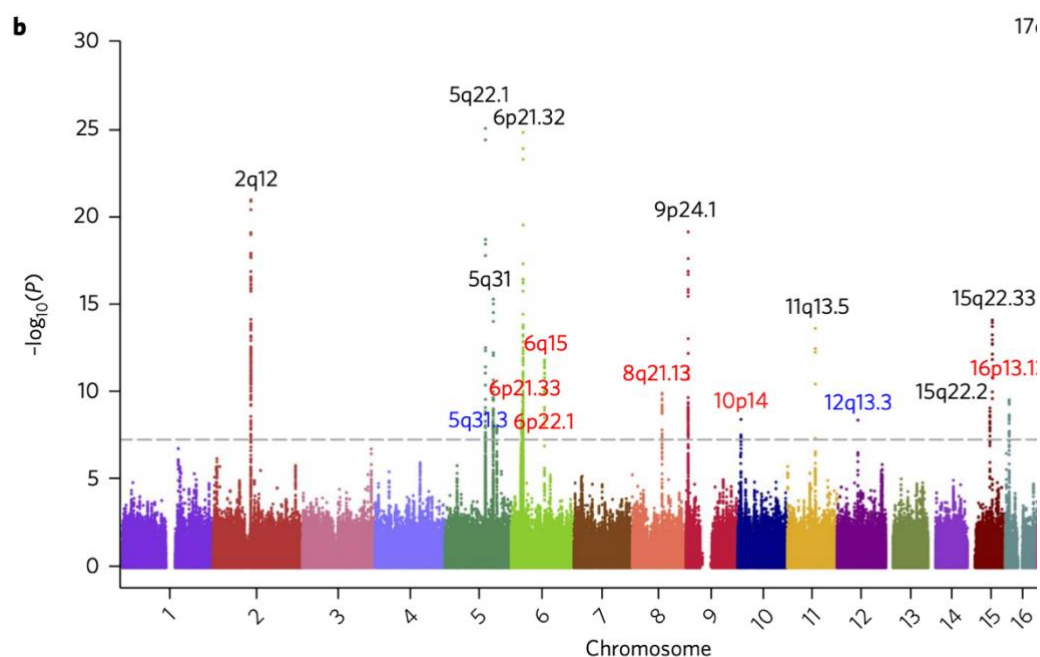


Figure 1 : Graphiques de Manhattan des résultats des méta-analyses à effets aléatoires d'ascendance européenne et multi-ascendance du risque d'asthme. **a :** La méta-analyse d'ascendance européenne concerne 19 954 cas d'asthme et 107 715 témoins. **b :** La méta-analyse multi-ascendance concerne 23 948 cas d'asthme et 118 538 témoins. Chaque locus est annoté en fonction de sa localisation dans la bande cytogénétique. L'axe des x représente la localisation chromosomique et l'axe des y représente $-\log_{10}(p)$ pour les tests d'association entre les SNP et l'asthme. Noir, loci déjà connus ; rouge, nouveaux loci identifiés dans la méta-analyse d'ascendance européenne ; bleu, nouveaux loci supplémentaires identifiés dans la méta-analyse multi-ascendance. La ligne horizontale en pointillés indique $p = 5 \times 10^{-8}$.

C'est ainsi qu'ils ont pu noter que 18 régions chromosomiques comprenaient cinq nouveaux loci associés à l'asthme à 5q31.3, 6p22.1, 6q15, 12q13.3 et 17q21.33 ; deux nouvelles associations à 6p21.33 et 10p14, indépendantes des signaux précédemment rapportés à ces loci dans des populations d'ascendance spécifique (respectivement d'ascendance latino et japonaise) ; deux associations à 8q21.13 et 16p13.13, précédemment rapportées pour l'asthme associé au rhume des foins, mais pas pour l'asthme seul, dans une étude portant sur des populations d'ascendance européenne ; et neuf loci d'asthme précédemment identifiés.

Dans leur étude, Stikker et al. (2023) confirment que compte tenu de sa prévalence élevée et de notre compréhension incomplète des mécanismes sous-jacents à la susceptibilité à la maladie, l'asthme est fréquemment étudié dans le cadre d'études

d'association pangénomique (GWAS), qui ont identifié des milliers de variantes génétiques associées au développement de l'asthme. Ils finissent par confirmer que leurs études GWAS sur l'asthme ont identifié 3 129 associations SNP-traités liées à plus de 140 loci, et l'EWAS a identifié collectivement des milliers de variantes épigénétiques associées à la maladie, mais concluent par le fait que, très peu d'associations ont été étudiées en profondeur pour déterminer quels variants sont la cause directe de l'asthme, quels mécanismes moléculaires sont à la base de l'impact des variantes génétiques et quelles voies biologiques pertinentes pour le risque de développer un asthme sont affectées.

La compréhension de l'association entre les gènes liés à l'asthme et les différents facteurs de risque est essentielle pour développer des stratégies de prévention et de traitement efficaces. Cette association est généralement évaluée à l'aide d'études d'association pangénomique (GWAS), qui permettent d'identifier des polymorphismes nucléotidiques simples (SNPs) associés à la maladie. Par exemple, des études ont identifié environ 680 SNPs associés à l'asthme, représentant 90% de l'ensemble des variants génétiques liés à cette maladie. Le Trans-National Asthma Genetic Consortium (TAGC) a identifié 878 SNPs associés à l'asthme répartis sur 18 loci différents. Ces SNPs influencent le risque d'asthme en régulant l'expression génique dans l'épithélium des voies respiratoires (Kim, 2020).

Il est important de noter que ces SNPs sont situés dans des régions non codantes du génome et ont généralement un effet limité individuellement. Cependant, leur identification contribue à une meilleure compréhension des mécanismes génétiques impliqués dans l'asthme. Par ailleurs, d'autres travaux ont rapporté que 64 gènes ont été associés à l'asthme ou à des phénotypes intermédiaires, avec 33 gènes mis en évidence par au moins deux études indépendantes (Dos Santos, 2015).

Il est à noter que la littérature consultée révèle que les évaluations clés de l'asthme par groupes d'étude à travers le monde ne montrent pas les mêmes résultats d'association en termes de fréquences et de loci (Tableau 1).

Tableau 1 : Statistiques des associations observées.

Cas	Contrôle	Association	Résultats	Auteur
994	1243	1	Locus ORM DL3/GSDMB en tête (principalement associé à l'asthme infantile).	Moffatt (2007)
84	263	3	Asthme professionnel (toluène diisocyanate) ; associations les plus significatives avec le gène CTNNA3	KIM (2009)
19954	107715	69	Grande méta-analyse des GWAS sur l'asthme à l'échelle mondiale, issues de populations ethniquement diverses. Première association du locus GATA3.	Demenais (2018)
64538	329321	167	Très vaste méta-GWAS détectant un grand nombre d'associations, y compris une étude de suivi in vivo du locus CD52	Han (2020)

4- MÉTHODES D'ÉTUDE DES ASSOCIATIONS GÈNES/ASTHME

L'identification des déterminants génétiques de l'asthme repose sur des études d'épidémiologie génétique, qui incluent principalement les analyses de liaison et les études d'association. Ces approches permettent d'identifier des régions chromosomiques impliquées dans la prédisposition à l'asthme ainsi que des gènes spécifiquement associés à son développement. Parmi ces gènes, certains appelés gènes de susceptibilité, contribueraient à la pathogénèse de l'asthme en augmentant la probabilité d'apparition des symptômes et en influençant la réponse immunitaire aux déclencheurs environnementaux (Dahmani, 2020). De plus, la compréhension des associations entre les gènes et l'asthme est essentielle pour développer des stratégies de prévention et de traitement efficaces. Plusieurs méthodes ont été développées pour étudier ces associations, chacune ayant ses avantages et ses limites :

1. Études d'Association Pangénomique (GWAS) : Les études d'association pangénomique sont des analyses à grande échelle visant à identifier des variations génétiques associées à des traits phénotypiques, tels que l'asthme. Ces études se concentrent généralement sur les polymorphismes nucléotidiques simples (SNP) et comparent leur fréquence entre des individus atteints d'asthme et des témoins sains. L'objectif est de repérer des SNPs dont la fréquence diffère significativement entre ces

deux groupes, suggérant une association avec la maladie. La démarche de GWAS nécessite, en plus des séquences SNPs, des informations phénotypiques telles que la présence de l'asthme, le niveau d'IgE, l'âge, le sexe, le tabagisme, etc. Une fois les données traitées, une régression logistique sur chaque SNP peut être réalisée pour apprécier des éventuelles associations significatives avec l'asthme.

2. Statistiques de l'Asthme : L'asthme demeure l'une des maladies respiratoires chroniques les plus répandues à l'échelle mondiale (figure 2). En 2021, on estimait à environ 260 millions le nombre de personnes vivant avec cette affection, soit une prévalence normalisée par âge de 3 340 cas pour 100 000 habitants (Yuan et al, 2021). Cette prévalence varie significativement selon les régions et les groupes démographiques. Par exemple, dans la région de la Méditerranée orientale, les adolescents présentent une prévalence légèrement plus élevée (10,10 %) comparativement aux enfants (9,70 %), avec une prédominance chez les garçons (11,48 %) par rapport aux filles (9,75 %). De plus, les zones urbaines affichent une prévalence supérieure (11,27 %) à celle des zones rurales (8,29 %), ce qui pourrait être attribué à des facteurs environnementaux tels que la pollution de l'air. En Afrique, une étude récente a révélé que des millions d'adolescents souffrent d'asthme non diagnostiqué, principalement en raison de l'urbanisation rapide et du manque d'accès aux soins de santé appropriés. Ces disparités soulignent l'importance d'une approche génétique pour comprendre les mécanismes sous-jacents de l'asthme et développer des interventions ciblées (Wachinou, 2019).

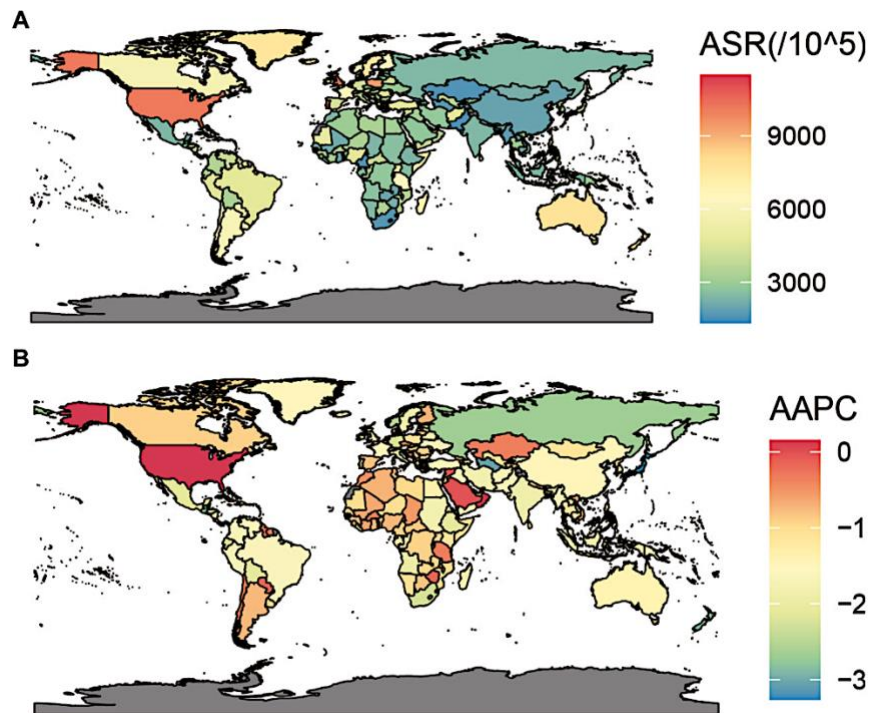


Figure 2 : Taux de prévalence de l'asthme standardisés selon l'âge et pourcentage moyen de personnes atteintes d'asthme selon l'âge dans 204 pays et territoires. Remarques : **(A)** Taux de prévalence standardisé selon l'âge en 2021. **(B)** Taux moyen de personnes atteintes d'asthme standardisés selon l'âge de 1990 à 2021. Variation annuelle moyenne en pourcentage.

5. LES PRINCIPAUX TYPES DE PUCES ADN

L'étude du profil d'expression génique des cellules et des tissus est devenue un outil majeur de découverte en médecine. Les expériences sur puces à ADN permettent de décrire les variations d'expression à l'échelle du génome, tant en santé qu'en maladie. De plus, une étude impartiale et systématique du profil d'expression génique devrait permettre d'établir une nouvelle taxonomie des syndromes obstétricaux et gynécologiques. Ainsi, une nouvelle ère s'ouvre où les processus et troubles de la reproduction pourraient être caractérisés à l'aide d'outils moléculaires et d'empreintes génétiques.

La conception, l'analyse et l'interprétation des expériences sur puces à ADN requièrent des connaissances spécialisées qui ne font pas partie du cursus standard de notre discipline. Les auteurs ont décrit les types d'études réalisables avec des expériences

sur puces à ADN (comparaison de classes, prédiction de classes, découverte de classes) et ont abordé les questions clés relatives à la conception expérimentale, au prétraitement des données et aux méthodes de sélection des gènes. Les types courants de représentation des données sont illustrés. Les pièges potentiels dans l'interprétation des expériences de microarray, ainsi que les atouts et les limites de cette technologie, sont mis en évidence (Tung, 2006).

Les puces à ADN peuvent mesurer simultanément le niveau d'expression de milliers de gènes dans un échantillon d'ARNm particulier. Un tel profilage d'expression à haut débit peut être utilisé pour comparer le niveau de transcription des gènes dans des conditions cliniques afin de : 1) identifier des biomarqueurs diagnostiques ou pronostiques ; 2) classer les maladies (par exemple, les tumeurs avec un pronostic différent qui sont indiscernables par examen microscopique) ; 3) surveiller la réponse au traitement ; et 4) comprendre les mécanismes impliqués dans la genèse des processus pathologiques. Pour ces raisons, les puces à ADN sont considérées comme des outils importants pour la découverte en médecine clinique (Schena, 2000 ; Zhang, 2005).

Une puce à ADN peut être considérée comme une analyse de Southern ou de Northern blot parallèle à grande échelle (au lieu d'un gel, les sondes sont fixées sur une surface inerte, qui deviendra la puce à ADN). L'ARNm est extrait des tissus ou des cellules, rétrotranscrit et marqué avec un colorant (généralement fluorescent), puis hybridé sur la puce, comme illustré en figure 3. L'hybridation et les lavages sont réalisés dans des conditions de stringence élevée afin de minimiser le risque d'hybridation croisée entre gènes similaires. L'étape suivante consiste à générer une image par imagerie par fluorescence induite par laser. Le principe de la quantification des niveaux d'expression est que la quantité de fluorescence mesurée à chaque emplacement spécifique de la séquence est directement proportionnelle à la quantité d'ARNm de séquence complémentaire présente dans l'échantillon analysé. Ces expériences ne fournissent pas de données sur le niveau d'expression absolu d'un gène particulier (concentrations réelles d'ARNm), mais sont utiles pour comparer le niveau d'expression entre différentes conditions et gènes sain ou malade (Knudsen, 2004).

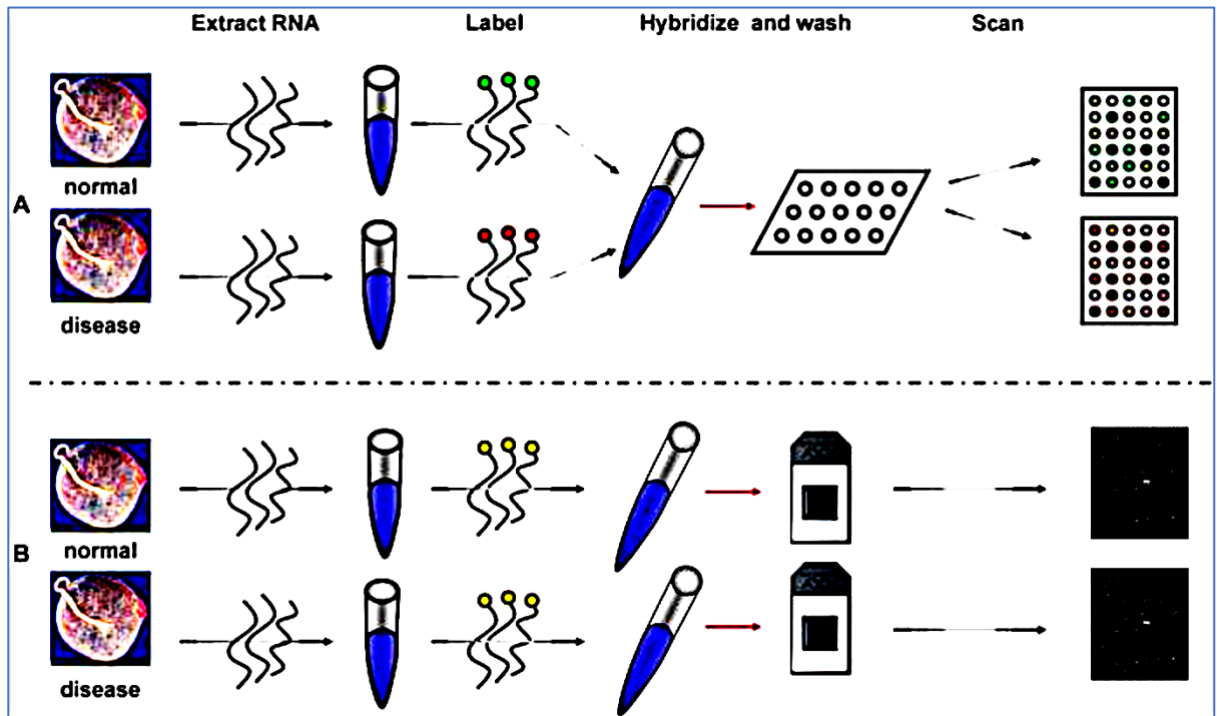


Figure 3 : Représentation schématique des étapes de la réalisation de microarrays. A (le panneau supérieur) illustre la technologie à deux canaux, tandis que B (le panneau inférieur) illustre la technologie à canal unique. Dans le panneau A, les ARNm normaux et pathologiques sont marqués avec deux colorants différents, mélangés puis hybridés sur le même microarray. Après lavage, le microarray est balayé à deux longueurs d'onde différentes pour produire deux images : une pour l'échantillon normal et une pour celui malade. Dans le panneau B (monocanal), chaque échantillon est marqué avec le même colorant fluorescent, mais hybridé indépendamment sur des microarrays différents.

5-1/ Les types de microarrays : les micropuces à ADN (ou puces à ADN) peuvent être classées selon trois critères :

i) La longueur des sondes : en fonction de leur longueur, les puces à ADN peuvent être classées en “puces à ADNc”, qui utilisent de longues sondes de plusieurs centaines ou milliers de paires de bases, et en “puces à oligonucléotides”, qui utilisent des sondes courtes (généralement de 50 pb ou moins).

ii) Les méthodes de fabrication comprennent : le « dépôt ou spotting » de séquences précédemment synthétisées et la « synthèse *in situ* ». Généralement, les puces à ADNc sont fabriquées par dépôt, tandis que les puces à oligonucléotides sont fabriquées par des technologies *in situ*. Ces technologies comprennent : la « photolithographie » (par exemple, Affymetrix, Santa Clara, Californie), l'« impression par jet d'encre » (par

exemple, Agilent, Palo Alto, Californie) et la « synthèse électrochimique » (par exemple, Combimatrix, Mukilteo, Washington).

iii) Les « puces monocanal » analysent un seul échantillon à la fois, tandis que les « puces multicanaux » peuvent analyser deux échantillons ou plus simultanément. La puce Affymetrix GeneChip est un exemple de puce oligonucléotidique monocanal.

En pratique, le terme « sonde » désigne la séquence nucléotidique fixée à la surface du microarray. Dans les expériences sur microarray, le terme « cible » désigne ce qui est hybridé aux sondes.

5-2/ Les types d'études conduites avec une puce à ADN : il existe trois principaux types d'applications des puces à ADN en médecine (Khatri, 2005 ; Tarca, 2006) :

1- La première consiste à identifier les différences d'expression entre des groupes prédéfinis d'échantillons. Il s'agit d'une expérience de « comparaison de classes », par exemple, l'identification de gènes différentiellement exprimés dans les placentas de femmes enceintes normales et de femmes atteintes de prééclampsie.

2- Une deuxième application, la « prédiction de classe », consiste à identifier l'appartenance à une classe d'un échantillon en fonction de son profil d'expression génétique (Intelligence Artificielle). Par exemple, il serait possible de prédire si une patiente présente (ou développera) une prééclampsie en fonction de son profil d'expression sanguine. Cela nécessite la construction d'un classificateur (un modèle mathématique) capable d'analyser le profil d'expression génétique d'un échantillon et de prédire son appartenance à une classe. Ce classificateur est construit à partir d'un ensemble représentatif d'échantillons dont l'appartenance à une classe est connue (par exemple, des femmes ayant une grossesse normale et celles développant ultérieurement une prééclampsie). Ce classificateur sera ensuite utilisé pour évaluer la probabilité de développer une prééclampsie chez les patientes non incluses dans sa construction.

3- Le troisième type d'application consiste à analyser un ensemble donné de profils d'expression génétique afin de découvrir des sous-groupes partageant des caractéristiques communes. Cette application est appelée « découverte de classes ». Par exemple, les profils d'expression d'un grand nombre de femmes atteintes de prééclampsie seront

mesurés afin d'identifier des sous-groupes de patientes présentant un profil d'expression génétique similaire. Cet effort vise à générer une taxonomie moléculaire de la maladie. Autrement dit, combien de types moléculaires de prééclampsie (sous-groupes) existent dans un échantillon de femmes atteintes de la maladie ?

5-3/ Le pré-traitement des données : Une fois les puces à ADN hybridées, les images obtenues servent à générer un ensemble de données. Cet ensemble de données doit être prétraité avant l'analyse et l'interprétation des résultats.

Le prétraitement est une étape qui extrait ou améliore les caractéristiques significatives des données et les prépare à l'application des méthodes d'analyse. Un exemple typique de prétraitement consiste à calculer le logarithme des valeurs d'intensité brutes. La « normalisation » est un type particulier de prétraitement effectué pour tenir compte des différences systématiques entre les ensembles de données. Un exemple de normalisation consiste à modifier les valeurs d'intensité brutes afin de compenser les différences d'efficacité des colorants lors d'expériences sur puces à ADN à deux canaux utilisant Cy3 (vert) et Cy5 (rouge).

La correction de bruit de fond (Background) est conçue pour ajuster l'hybridation non spécifique, c'est-à-dire l'hybridation de transcrits d'échantillons (cibles) dont les séquences ne correspondent pas parfaitement à celles des sondes sur la puce. Sur les puces à points, l'hybridation non spécifique incluse dans les valeurs d'intensité brutes peut être estimée à partir du niveau de fluorescence à proximité immédiate de la sonde (Yang, 2002).

PARTIE 2 : MATÉRIEL & MÉTHODES

L'analyse des variants génétiques associés à l'asthme repose sur l'exploitation d'un dataset contenant des informations sur les SNPs (Single Nucleotide Polymorphisms), leur localisation chromosomique et les gènes mappés. L'objectif est d'identifier les SNPs significatifs impliqués dans cette pathologie et d'évaluer leur impact fonctionnel à l'aide d'outils bioinformatiques et statistiques. L'ensemble des étapes proposées permet d'identifier des SNPs associés à l'asthme et d'évaluer leur rôle fonctionnel dans cette pathologie.

1. PRÉTRAITEMENT ET VALIDATION DES DONNÉES

Avant toute analyse, une étape de nettoyage et d'organisation des données est essentielle. Il convient de vérifier la présence de valeurs manquantes, de standardiser les noms des colonnes et d'harmoniser les formats des données.

Nos données relatives aux SNPs associés à l'asthme proviennent typiquement d'une étude d'association pangénomique (GWAS² : Genome-Wide Association Study) pour identifier les variations génétiques associées à une pathologie ou un trait particulier à l'échelle du génome, mais certaines annotations ont été également enrichies à l'aide de la base de données bioinformatique : Ensembl³.

Le programme GWAS génère des informations issues d'un traitement bioinformatique des programmes tels que : PLINK, SNPTEST ou GEMMA). Ces informations sont :

- P-value : calculée à partir des tests d'association statistiques entre SNPs et un phénotype (l'asthme) et mesure la significativité statistique de cette association SNP-Asthme. Il s'agit d'une valeur de probabilité liée à l'association entre le SNP et le phénotype étudié (ici, l'asthme), indiquant que cette association est due au hasard. Plus sa valeur est faible, plus l'association est statistiquement

² <https://www.ebi.ac.uk/gwas/>

³ <https://www.ensembl.org/index.html>

significative. En GWAS, un seuil de signification typique est $p < 5 \times 10^{-8}$ est utilisé en raison du grand nombre de tests simultanés.

- P-value MLOG : transformation logarithmique ($\log_{10}$) appliquée à la p-value pour la visualisation des résultats. Le calcul de cette transformation suit la formule : $MLOG = -\log_{10}(p\text{-value})$. Plus cette valeur est grande, plus l'association est fortement significative.
- SNPs Passing QC : Ce paramètre (binaire : 0 ou 1) indique si le SNP a passé les critères de contrôle de qualité ; un filtrage préalable des données basé sur plusieurs critères tels que le taux d'appel (call rate), respect des lois de Hardy-Weinberg, fréquence allélique et taux de génotypage. Dans notre analyse, ce paramètre fournit plusieurs informations importantes sur la provenance, le traitement et la qualité des données génétiques. En effet, Affymetrix indique que les données de génotypage proviennent d'une puce à ADN de la société Affymetrix à haute densité de type Axiom® et CytoScan HD. Ces puces contiennent des millions de sondes permettant de détecter des variants génétiques (SNPs) sur l'ensemble du génome. La valeur '(~ 8270000)' indique qu'environ 8,27 millions de SNPs ont été pris en compte après contrôle qualité (QC), y compris les SNPs imputés. Le QC est un processus qui exclut les SNPs de mauvaise qualité (taux de génotypage faible, biais, MAF trop bas : fréquence de l'allèle mineur du SNP). Un SNP avec une $MAF < 1\%$ est exclu car il est très rare et pose des problèmes statistiques. Cela permet de sélectionner les variants les plus fréquents pour avoir une puissance statistique suffisante. En fin, la valeur 'imputed' signifie que les données incluent des génotypes imputés. Cette imputation est une méthode statistique pour prédire les génotypes manquants à partir de ceux observés, en se basant sur un panel de référence tel que 1000 Genomes (ou également HRC : Haplotype Reference Consortium).

Nous signalons que les SNPs utilisés dans cette analyse proviennent de puces à ADN Affymetrix haute densité. Après contrôle qualité, environ 8,27 millions de variants (SNPs) ont été retenus, incluant ceux directement génotypés et ceux imputés à l'aide d'un

panel de référence. Les données brutes ont été téléchargées à partir de la base de données dbSNP⁴ du NCBI (National Center for Biotechnology Information).

La matrice de données des SNPs comprend les informations brutes suivantes en colonnes :

Date Added to Catalog, PubMedID, First author, Date, Journal, Link, Study, Disease /trait, Initial sample size, Replication sample size, Region, CHR_ID, Reported Gene(s), Mapped gene, Upstream_Gene_ID, Downstream_Gene_ID, SNP_Gene_IDS, Upstream_Gene_Distance, Downstream_Gene_Distance, StrongestSNP-Risk Allele, SNPs, Merged, SNP_ID_Current, Context, Intergenic, Risk Allele Frequency, p-value, pvalue_MLOG, Platform [SNPS Passing QC], CNV, Mapped_Trait, Mapped_Trait_URL, Study Accession et Genotyping Technology. Les lignes de cette matrice contiennent les SNPs.

La matrice finale, pour cette étude, est un tableau de 38 caractères annotés sur une liste de 199 SNPs associés à l'asthme.

Le traitement de cette matrice a permis d'éliminer les caractères (colonnes) dont les valeurs sont manquantes et les caractères qui donnent des informations globales non cliniques telles que l'URL ou encore le nom des auteurs et leurs journaux de publications respectifs. Ce tri a conduit à une matrice (Tableau 2) plus légère mais plus adéquate avec les attentes génétiques de ce travail.

⁴ <https://www.ncbi.nlm.nih.gov/snp/>

Tableau 2 : Caractéristiques des variants génétiques (SNPs) liés à l’asthme : localisation, fréquence et type de mutation

Région	SNPs	Pos_Chrom	Chr	Frq /1000 Génomes	Mutation	Séquence
1q23.3	rs2070901	161215268	1	0,389914	G>C/ G>T	GCATGGGGGA AGGCGTGGCA GGAAGAGGGG GACTCTGTGG TCAGGGAAGT GCTCGCTGAG CACAGCTGCA CAGTGCTGTC AGAACGGCCG ATCTCCAGCC CAAGATGATT CCAGC
1q21.3	rs61816761	152313385	1	NA	G>A/ G>C/ G>T	ACGAATGGTG TCCTGACCCT CTTGGGACGC TGAATGCCTG GAGCTGTCTC GTGCCTGCTC GTGGTGCGAT CCTTGCTTTC CTCCAGTGCT GGTCCCGGTC CGTCCATGGG CAGAG
1q21.3	rs72702900	152799487	1	0,0111852	T>A	CTTCCAGTTC ATGAGCCTGA GACATCTTTC CATTTGTTTG TGTCTTCTTC TATTTCTTCA TCAATGCTTT TAGCTTTCAT TGTAAGTCTT TCACTTCTTT GGTAAATTT ATTTT
1q21.3	rs12123821	152206676	1	0,0133438	C>T	TGGAGTTAGA CAGATTTAGG TATAAAACCT GACTTTACTA ATATAATATA CTTGACCTTT TAAAGTCTCA GTTTGCTCTT CAGTAAATA ACATTTATAA CCACAGGGTT GCTAT
1q24.2	rs7523907	167458010	1	0,297096	C>A/ C>T	CCATAAAATG AAGGAATTTT GCCCCTCACA TTTTCAAGAA CCTATGAGTT CGAAGAGACC TGAAAGAACA TTTTATGTTG AGCCTACACC AGTGCTATCA CCTAAAGTGC GTGAC
1q21.3	rs71625139	151829917	1	0,025314	A>G	CCATGTCACC AACTGTCACC TAACTGCTGT AATTGTGTCA TTGTCTATGA AATGGGTCTT TCCTGCCAAA TTGTAAAGCC CTGTGAAGGC AGGGACTGTG TTGGTCTTAT TTACT
1q21.3	rs55875222	151790858	1	0,00588697	C>G/ C>T	ACCGAGGAGT TTTCTATCTT CTAACCCCTG GAGGAAAATC TGCGCTCCAG CAGCTGGGAC CGAGGCATGA GAGCGTTGAA TCATGAGTGG CACCGTTTCG AAAAATTGGT GTCTT
1q21.3	rs12144049	152468434	1	0,264325	C>A/ C>G/ C>T	TGGGGCTTAC TTTTGTGCT TTTCTTTTCT CAAGGATTAT AATCCTCTGT CGGTTTCTGT TCAATGCTTG AAAACATTTG CTCTGTATCT TTTATTCAGT TTCTTAGTTG TTTAT
1q21.3	rs138726443	152307547	1	NA	G>A/ G>T/ G>C	ATGAATGGTG TCCTGACCCT CTTGGGACGT TGAGTGCCTG GAGCTGTCTC GTGCCTGCTT GTGGTGGGAT CCTTGTCTTC CTCCAGTGCT GGTCCCGGTC CGTCCATGGG CGGAC
1q21.3	rs1552994	152198985	1	0,0117739	T>C	ATGAATGGTG TCCTGACCCT CTTGGGACGT TGAGTGCCTG GAGCTGTCTC GTGCCTGCTT GTGGTGGGAT CCTTGTCTTC CTCCAGTGCT GGTCCCGGTC CGTCCATGGG CGGAC

2. ANALYSE DES SNPs

Nous remarquons que, vu le nombre élevé des SNPs retenus pour cette étude (199 variants), nous présentons un cas d'analyse typique pour un seul variant choisi aléatoirement.

L'annotation des variants a été effectuée avec deux outils pour mettre en exergue les mutations et leurs conséquences au sein des variants génétiques associés à l'asthme :

- i- **MutationT@ster** : Outil bioinformatique conçu pour prédire les effets fonctionnels des mutations génétiques, principalement les variants nucléotidiques simples (SNPs) et les indels (Insertions/Délétions). Le MutationT@ster analyse un variant génétique et prédit s'il est bénin (non pathologique/polymorphisme) ou pathologique (causeur de maladie), ou inconnu dans la base de données ou alors probablement bénin ou pathogène.
- ii- **gnomAD** : Base de données de séquençage d'exomes et de génomes humains issus de projets internationaux, afin de fournir les fréquences alléliques dans différentes populations. Son utilisation permet de filtrer les variants génétiques probablement bénins (car trop fréquents), ou pour contextualiser un SNP trouvé chez un patient (est-il courant dans la population ?) ou encore pour comparer les fréquences entre différentes populations (ethnies) surtout dans les maladies multifactorielles.

2-1/ Cas du variant rs2070901 : Situé sur le chromosome 1 au sein du gène *FCER1G* (Fc fragment of IgE receptor Ig). Ce gène code pour la chaîne gamma du récepteur de haute affinité pour l'immunoglobuline E (IgE), un composant essentiel dans la médiation des réponses inflammatoires allergiques, notamment via l'activation des mastocytes et des basophiles. Le SNP rs2070901, caractérisé par une substitution nucléotidique G/T, est localisé dans une région régulatrice du gène, suggérant un potentiel impact sur l'expression de *FCER1G*.

A- Analyse sur NCBI/SNP : L'annotation de rs2070901 a été effectuée via le portail NCBI⁵ en interrogeant la base de données SNP, pour désigner la position chromosomique du SNP et pour préciser la position de la mutation causatrice de son dysfonctionnement. Le paramètre recherché dans cette interrogation est le MAF (Minor Allele Frequency), car c'est un indicateur de variabilité génétique dans la population étudiée (individus asthmatiques, dans notre cas). Ce paramètre aide à distinguer, au sein de nos variants génétiques, ceux qui sont rares (MAF<1%) et donc potentiellement plus délétères ou spécifiques à certaines pathologies (l'asthme en l'occurrence).

B- Analyse avec MutationT@ster2025 : La localisation de la mutation a été corroborée, dans certains SNPs avec l'outil bioinformatique MutationT@ster version 2025. Certains de nos SNPs ne sont pas encore insérés dans sa base de données.

Le programme MutationT@ster permet d'estimer des paramètres statistiques et moléculaires pour prioriser les mutations potentiellement pathologiques dans les gènes associés à l'asthme et de soutenir nos interprétations biologiques par une analyse *in silico*. Le programme est utilisé pour :

- i- Prédire l'effet du variant génétique : pathologique ou bénin (polymorphisme),
- ii- Désigner les conséquences moléculaires : mutation synonyme, non-sens, gain ou perte de codon STOP.
- iii- Déterminer l'effet sur la protéine : changement d'acides aminés, gain ou perte de domaine fonctionnel.
- iv- Mesurer la conservation évolutive : par calculs de scores de conservation tels que phyloP car un variant dans une région ADN hautement conservée est plus susceptible d'être pathologique.

C- Analyse avec gnomAD : L'utilisation du programme bioinformatique gnomAD va permettre d'estimer les fréquences des variants génétiques au sein d'une population, car sachant que l'asthme est une maladie multigénique, il est alors essentiel pour notre analyse de savoir si un variant est rare ou fréquent. De plus, les paramètres tels que CADD (Combined Annotation Dependent Depletion) et REVEL pour prédire l'impact biologique des mutations. Ainsi la prédiction du score de CADD (ou CADD score phred, le plus

⁵ <https://www.ncbi.nlm.nih.gov/snp/?term=rs2070901>

utilisé en bioinformatique) va nous permettre de comparer la nocivité potentielle d'un variant génétique donné :

- i- Un score phred CADD = 10 : le variant est parmi les 10% les plus délétères.
- ii- Un score de 20 : parmi les 1% les plus délétères.
- iii- Un score de 30 : parmi les 0,1% les plus délétères.

3. IDENTIFICATION DES VARIANTS ASSOCIÉS À L'ASTHME

L'identification des SNPs pertinents repose sur une sélection rigoureuse des variants situés dans des gènes déjà associés à l'asthme ou dans des régions chromosomiques connues pour leur implication dans cette maladie.

Afin d'identifier les variants les plus fortement associés à l'asthme parmi les 80 rsID de notre liste, une analyse bioinformatique a été réalisée en interrogeant plusieurs bases de données génétiques et études d'association.

Les rsID ont été comparés aux informations disponibles dans les bases dbSNP, ClinVar, GWAS Catalog, NCBI Variation Viewer ainsi que dans les portails spécialisés tels que DisGeNET⁶ et Open Targets Genetics⁷ (figure 4). Ces plateformes permettent d'examiner les annotations fonctionnelles des variants (type de mutation, gène concerné, région génomique), ainsi que leur association statistique avec des phénotypes spécifiques comme l'asthme (p-value, odds ratio, études cliniques ou populationnelles). Les variants ont été filtrés sur la base de leur association validée ou significative avec l'asthme, notamment ceux retrouvés dans des études GWAS publiées. Seuls les variants présentant une implication directe ou une forte suspicion dans la physiopathologie de l'asthme ont été retenus pour discussion (Tableau 6), les autres SNP n'ont pas été retenus (tableau 7).

⁶ <https://disgenet.com/>

⁷ <https://genetics.opentargets.org/>

Analyse des variants SNPs (rsID) associés à l'asthme

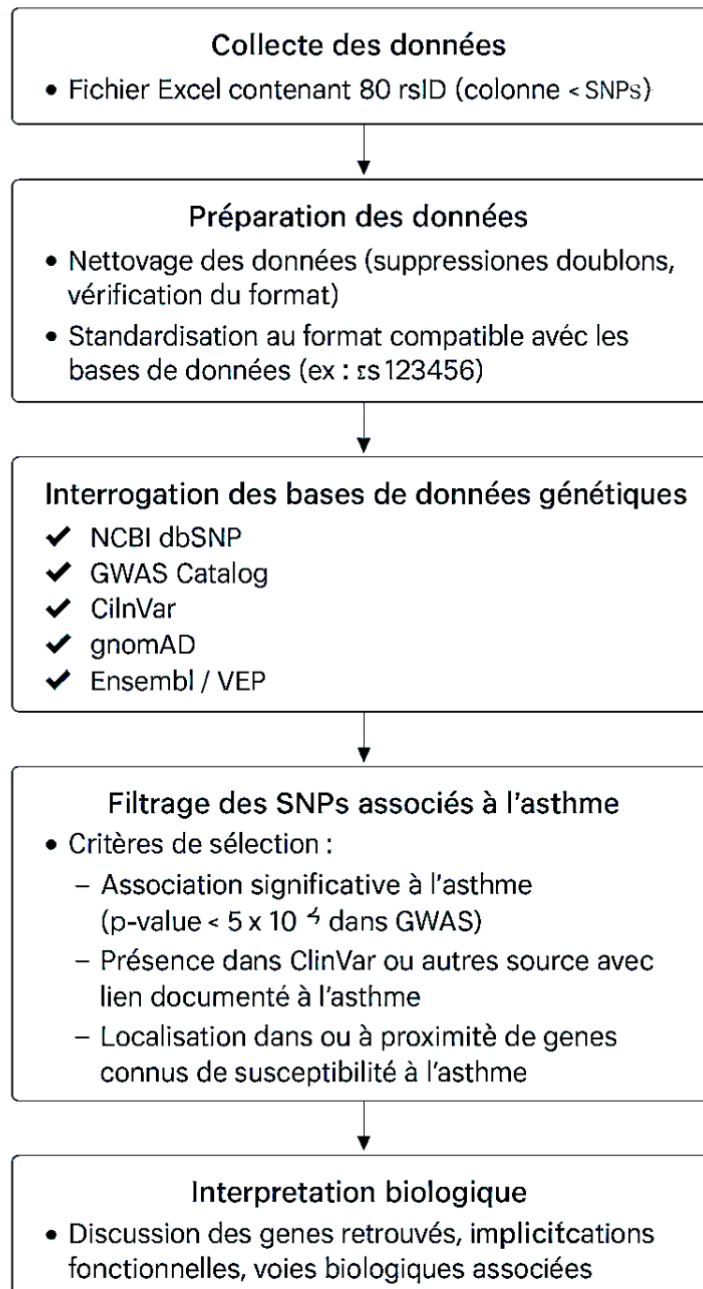


Figure 4 : Étapes d'identification des variants associés à l'asthme à partir d'une liste de SNPs

Cet organigramme illustre les différentes étapes suivies pour identifier les variants génétiques potentiellement associés à l'asthme à partir d'un ensemble de 80 SNPs (rsID). L'analyse a consisté en une interrogation croisée de bases de données spécialisées (GWAS Catalog, ClinVar, dbSNP, gnomAD), suivie d'un filtrage basé sur les niveaux d'évidence et les fréquences alléliques.

4. ANALYSE PHYLOGÉNÉTIQUE DES VARIANTS GÉNÉTIQUES

Pour analyser les relations phylogénétiques entre les différents SNPs, un alignement multiple des séquences a été réalisé à l'aide du logiciel MAFFT⁸ (Multiple Alignment using Fast Fourier Transform), accessible via la plateforme en ligne.

Les séquences au format FASTA, correspondant aux régions flanquantes (environ 200 pb en amont et en aval) de chaque SNP, ont été préalablement extraites de la base de données dbSNP (NCBI). L'alignement a été effectué en utilisant le mode L-INS-i en fonction du nombre et de la diversité des séquences. Les paramètres standards ont été conservés, notamment l'utilisation du système pénalité de gap fixée à 1,53 ; ce qui permet un compromis fiable entre rapidité et précision.

Une fois l'alignement multiple obtenu, les données ont été exportées au format ClustalW pour servir à la construction de l'arbre phylogénétique. Celui-ci a été généré à l'aide de la méthode Neighbor-Joining (NJ), une approche distance-based largement utilisée pour inférer des relations évolutives à partir de matrices de distances corrigées. Cette méthode a été choisie pour sa rapidité de calcul et sa robustesse avec des séquences génétiques relativement courtes et peu divergentes, comme c'est souvent le cas pour les régions SNP. L'arbre résultant, au format Newick, a ensuite été visualisé et annoté sur la plateforme du logiciel MAFFT. Ce processus va permettre de regrouper les SNPs en clades distincts, révélant des structures phylogénétiques interprétables en fonction de leur proximité génétique, ce qui peut suggérer une origine commune ou des fonctions biologiques convergentes.

⁸ <https://mafft.cbrc.jp/alignment/server/index.html>.

PARTIE 3 : RÉSULTATS & DISCUSSION

1. ANALYSE DES SNPs

Il est à noter que tous nos SNPs ont tous été annotés comparativement à la dernière version du génome référence humain GRCh38 (Genome Reference Consortium Human Build 38), ou hg38 en bioinformatique. Ce génome est la 38e version officielle du génome humain de référence, publiée en décembre 2013 par le Genome Reference Consortium (GRC) et versionnée le 3 février 2022. Elle remplace l'ancienne version GRCh37/hg19, largement utilisée jusque-là. Le hg38 est disponible pour le téléchargement sur l'URL : https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/.

1-1/ Annotation NCBI/SNP : L'annotation du SNP rs2070901, sur NCBI, a révélé deux allèles C et T en position chr1: 161215268G/C/T. Cette mutation ne concerne pas la CDS (coding sequence : 5'-Met ...STOP-3') du gène, mais dans la région régulatrice ; ce qui suggère qu'il pourrait influencer l'expression de ce gène. Des variations dans l'expression de *FCER1G* peuvent affecter la signalisation des récepteurs IgE, modulant ainsi la réponse immunitaire aux allergènes. Cela pourrait expliquer son association avec des maladies allergiques telles que l'asthme, comme cela a été noté dans les travaux de Kim et al. (2023).

Des études récentes ont établi une association significative entre rs2070901 et des maladies allergiques telles que l'asthme, le rhume des foins et l'eczéma. Par exemple, une analyse multinomiale a révélé que rs2070901, en combinaison avec d'autres SNPs, est significativement associé à des réactions d'hypersensibilité médicamenteuse, renforçant l'idée de son rôle dans la modulation de la réponse immunitaire allergique. De plus, une étude a identifié rs2070901 comme l'un des SNPs associés à l'asthme dans une cohorte de patients, soulignant son importance dans la susceptibilité génétique à cette maladie.

1-2/ Analyse du SNP (rs2070901) sur MutationT@ster n'a pu être effectuée car ce variant n'est pas enregistré sur cette base de données.

1-3/ Analyse sur gnomAD : le traitement du variant rs 2070901 (Tableau 3) a révélé les résultats suivants :

Tableau 3 : Résumé des paramètres clés de l'analyse du rs2070901 sur gnomAD

Allèle	Mineur	MAF	CADD G>C	phyloP G>C	CADD G>T	phyloP G>T
G/C/T	T	0,389914	13,5	1,14	13,4	1,14

1-4/ Autres variants : L'annotation du SNP rs4286272 sur NCBI/SNP a révélé qu'il s'agit d'un variant de type SNV (Single Nucleotide Variation), localisé sur le GRCh38 à la position 2:197670954 (Figure 5). La représentation graphique du polymorphisme rs4286272 localisé sur le chromosome 2 indique un variant, de type SNP (T/C). Sur cette figure, il est visualisé dans son contexte génomique à l'aide du Genome Data Viewer. Il est situé dans une région intronique entourée d'autres variants, illustrant sa position relative aux séquences voisines et aux éléments génétiques annotés.

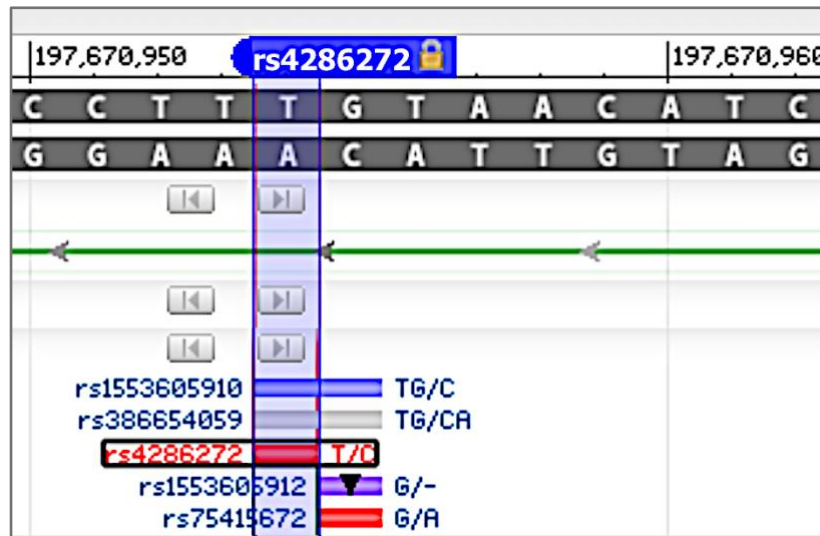


Figure 5 : Cartographie du variant rs4286272 sur le chromosome 2 via le Genome Data Viewer (NCBI).

Le SNP rs4286272 a été annoté avec l'outil MutationT@ster version 2025 et a révélé les résultats suivants :

- Variant type : Single base exchange
- Gene region : Intron
- DNA changes : c.139+4366 A>G ; g.5092A>G
- Pathogenic variant (ClinVar) : Variant not listed in ClinVar as (likely) pathogenic.
- Variant DBs : dbSNP ID rs4286272
- Phylogenetic conservation : PhyloP = 0,49
- Strand : -1
- Original gDNA sequence snippet :
GTAGATATGACAGATGTTACAAAGGAGGAACTTAGCCTGTG
- Altered gDNA sequence snippet :
GTAGATATGACAGATGTTACG*AAGGAGGAACTTAGCCTGTG

Le paramètre DNA changes : c.139+4366A>G révèle les informations suivantes :
c. : Position sur l'ADNc (transcrit codant)

139+4366 : 4366 nucléotides dans l'intron après l'exon 139

A>G : Substitution d'une adénine par une guanine (transition) : mutation intronique silencieuse.

Phylogenetic conservation : PhyloP = 0,49 signifie un score d'une analyse comparative entre espèces pour évaluer la conservation évolutive de cette position (A>G) dans le génome.

Ce Phylogenetic p-value est un indice calculé à partir d'arbres phylogénétiques qui mesure la conservation ou l'accélération de l'évolution d'un site nucléotidique :

- Si PhyloP > 0 : le site est conservé (évolution plus lente que prévu) ; indication de fonction biologique probable.
- Si PhyloP < 0 : Site accéléré (évolution plus rapide que prévu) ; suggère l'absence de pression sélective ou adaptation positive.
- Si PhyloP ≈ 0 : Évolution neutre (pas de conservation particulière).

L'indicateur phyloP mesure séparément la conservation des bases individuelles, ignorant les effets de leurs voisines. Les sites dont la conservation est prévue reçoivent des scores positifs, tandis que les sites dont l'évolution est rapide reçoivent des scores négatifs. Les scores affichés dans MutationTaster2025 sont basés sur l'alignement de 100 espèces de vertébrés.

Pour le cas du SNP rs42866272, la valeur de PhyloP égale à 0,49 indique une légère conservation du nucléotide. Ce site évolue un peu plus lentement que prévu ; ce qui suggère une possible fonction régulatrice ou structurelle, mais ce n'est pas très significatif. Ce marqueur SNP n'est probablement pas soumis à une forte pression sélective, en raison de l'effet régulateur modéré attribué à ce marqueur.

Le résultat de cette analyse pour les marqueurs rs 4286272, rs 61816761 et rs 3130160 est résumé dans le tableau 4 pour des variants situés sur différents chromosomes.

Tableau 4 : Caractéristiques fonctionnelles et évolutives des variants génétiques.

Variant	Chr	Région	Mutation HGVS	PhyloP
rs4286272	2	Intron	c.139+4366A>G ; g.5092A>G	0,49
rs61816761	1	CDS	c.1501 C>T	2,21
rs3130160	6	Intron	c.481+962 A>T	0,019

HGVS⁹ : Human Genome Variation Society.

Le score PhyloP du variant rs61816761(ou Arg501Ter ou R501X) indique que la position où a eu lieu la mutation de C par T est conservée à l'intérieur des espèces ; il s'agit d'une forte pression sélective ; ce qui suggère que le nucléotide C est biologiquement important pour le code de la protéine résultante. En effet, la mutation subie est de type non sens du gène FLG qui code pour la protéine fillagrine qui a introduit un codon STOP prématuré, entraînant une perte de fonction de la protéine essentielle à la barrière cutanée et contribuant à l'aggrégation des filaments de kératine dans la cellule

⁹ <https://hgvs-nomenclature.org/stable/>

épidermique assurant ainsi la cohésion et l'intégrité de la peau. La perte de la fonction de cette protéine compromet cette barrière, rendant la peau plus perméable aux allergènes et aux irritants.

Le variant rs 3130160 a exprimé un PhyloP égal à 0,019. Cela signifie que cette position génomique est faiblement conservée au cours de l'évolution et que les mutations y sont tolérées. Par conséquent, le variant serait neutre sur le plan fonctionnel surtout que la mutation est située dans un intron.

Bien que le variant rs3130160 présente un score PhyloP faible (0,019), suggérant une faible conservation évolutive, il a néanmoins été associé à l'asthme dans plusieurs études d'association. Ce paradoxe peut s'expliquer par un rôle spécifique à l'humain dans la régulation de gènes immunitaires, ou encore par une liaison génétique avec un autre polymorphisme fonctionnel situé à proximité. Le variant **rs3130160** a été identifié comme étant associé à l'asthme dans plusieurs études de génétique des populations. Sa faible valeur de PhyloP n'exclut pas une implication fonctionnelle dans des processus biologiques spécifiques à l'humain. Par exemple, une étude publiée dans *BMC Pulmonary Medicine* a intégré des données de GWAS et de eQTL pour identifier des gènes et SNPs associés à l'asthme modéré à sévère. Cette analyse a mis en évidence plusieurs loci, dont certains pourraient inclure ou être en liaison avec des variants tels que rs3130160 (Dong, 2020).

L'annotation du marqueur rs4286272 sur la base de données gnomADv4.1.0 révèle que 22 441 individus de cette base de données sont homozygotes (C/C) pour ce variant et plus de 32 000 sont hétérozygotes. Ces chiffres indiquent donc que cette mutation est très courante dans la population mondiale; ce qui suggère que ce variant serait rarement causal de pathologie. Il présente des fréquences alléliques variables selon les populations géographiques et le sexe (tableau 5) ; ce qui suggère des éléments intéressants dans le contexte de la génétique des maladies complexes comme l'asthme.

L'annotation du marqueur rs61816761 sur la base de données gnomADv4.1.0 révèle que 315 individus sont homozygotes (A/A) pour ce variant et portent donc deux copies du variants et 26731 sont hétérozygotes porteurs d'une copie de l'allèle de référence et une copie de l'allèle variant.

L'allèle A (variant) est présent à haute fréquence dans la population globale, mais la présence de plusieurs homozygotes (n=315) suggère que le variant n'est probablement pas létal à l'état homozygote, mais peu être associé à un risque accru pour certaines pathologies dans certains contextes génétiques ou environnementaux. En effet, ce variant est associé à des maladies atopiques (L'atopie est une prédisposition génétique à développer des allergies, notamment des allergies cutanées, respiratoires ou alimentaires) comme l'eczéma et potentiellement l'asthme dans certaines populations. Sa distribution suggère une large tolérance au variant, bien que des études antérieures aient démontré son association avec des maladies allergiques telles que la dermatite atopique. Ce profil de fréquence a également été observé dans une méta-analyse (Rodriguez, 2009) qui a révélé que la haploinsuffisance de FLG augmente fortement le risque d'eczéma (odds ratio (OR), 3,12 ; IC à 95 %, 2,57–3,79) et est également significativement associée à l'asthme (OR), 1,48 ; IC à 95 %, 1,32–1,66). OR est l'Odds Ratio (rapport des cotes). La valeur 1,48 indique que les individus porteurs du variant ont 1,49 fois plus de chances de développer un asthme par rapport aux non-porteurs. Donc ce variant augmente de 48% le risque relatif de développer l'asthme.

Tableau 5 : Fréquences alléliques du variant rs4286272 en fonction de la géographie et du sexe¹⁰.

Groupe génétique ancêtre	Occurrences de l'allèle mineur	Nombre total d'allèles	Nombre d'homozygotes	Fréquence allélique
Moyen orient	126	294	33	0,4286
Afrique	20727	41396	5116	0,5007
Europe	35843	67900	9497	0,5279
Asie majeure	4003	5162	1561	0,7755
XX	41572	77670	11214	0,5352
XY	40404	74160	11214	0,5448

¹⁰ https://gnomad.broadinstitute.org/variant/1-161215268-G-C?dataset=gnomad_r4

Le calcul des fréquences est illustré par l'exemple l'Asie majeure où 4003 allèles mineurs ont été observés sur 5162 allèles au total dans l'ensemble de la cohorte gnomAD.

Ainsi l'expression de p va donner : $p = \frac{\text{Occurences allèle mineur}}{\text{nombre d'allèles total}} = \frac{4003}{5162} \approx 0,7755$

Ce variant est majoritaire en Asie (>77%). Cela traduirait une sélection naturelle locale ou une dérive génétique historique. Cependant sa fréquence en Europe qui vaut 63% pourrait indiquer une dilution au fil des générations, mais au Moyen-Orient, la fréquence diminue pour atteindre 42%. Cette baisse correspondrait à une diversité génétique plus importante. En fin, les différences entre les sexes sont faibles (moins de 1%) et non significatives statistiquement.

2- IDENTIFICATION DES VARIANTS ASSOCIÉS À L'ASTHME

Les variants les plus pertinents (tableau 6) ont été retenus pour une interprétation fonctionnelle et une annotation biologique.

Tableau 6 : Identification des SNPs en lien avec l'asthme à partir d'une analyse croisée de bases de données génétiques.

rsID	Gène	Type de variant	Lien avec l'asthme
rs61816761	<i>FLG</i>	Mutation non-sens (p.Arg501Ter)	Association établie avec l'asthme atopique (surtout en Europe).
rs20541	<i>IL13</i>	Mutation faux-sens (Arg130Gln)	Associé à l'asthme et à des taux élevés d'IgE.
rs1800925	<i>IL13</i>	Région promotrice	Corrélié à l'asthme allergique.
rs3819024	<i>IL17A</i>	Région promotrice	Lien établi avec des formes allergiques de l'asthme.
rs3024656	<i>IL4R</i>	Intronique	Associé à l'asthme et à l'hyperréactivité bronchique.
rs1029489	<i>IL4R</i>	Région 3' proche du gène	Influence la susceptibilité à l'asthme.
rs6498012	<i>IL4R</i>	Intronique	Variant impliqué dans des études GWAS sur l'asthme.
rs2099903	<i>MBL2</i>	3' UTR	Associé à une réponse inflammatoire chez les asthmatiques.
rs8178880	<i>PLAT</i>	Intronique	Impliqué dans la perméabilité vasculaire et l'asthme.
rs4252053	<i>PLG</i>	Région 5'	Études suggèrent une contribution à l'asthme allergique.
rs1879026	<i>TLR3</i>	Intronique	Corrélié à l'asthme via les réponses antivirales.
rs10025405	<i>TLR3</i>	Région intergénique	Présent dans des loci associés à l'asthme.
rs5743303	<i>TLR3</i>	Région 5' proche du gène	Influence potentielle sur la régulation de l'expression de TLR3.
rs5743305	<i>TLR3</i>	Région 5' proche du gène	Implication dans l'immunité innée et l'asthme.
rs7668666	<i>TLR3</i>	Intronique	Présent dans plusieurs cohortes asthmatiques.
rs2236624	<i>ADORA2A</i>	Intronique	Gène cible de médicaments contre l'asthme.
rs11053624	<i>DECTIN1</i>	Région 5'	Impliqué dans la réponse immune et l'asthme allergique.
rs7959451	<i>DECTIN1</i>	3' UTR	Rôle dans la régulation post-transcriptionnelle.
rs72677651	<i>JAK1 / AK4</i>	Région 1p31.3	Identifié dans une étude GWAS sur l'asthme.
rs76335186	<i>NDUFA4</i>	Région 7p21.3	Signal génétique dans des cas d'asthme infantile.
rs67049014	<i>RNASE2 / METTL17</i>	Région 14q11.2	Lien avec les niveaux d'IgE chez les asthmatiques.
rs56675562	<i>CDK20 / SPATA31C2</i>	Région 9q22.1	Variant enrichi dans les populations asthmatiques.

Tableau 7 : Variants sans association significative avec l’asthme.

rsID	Gène	variant	association avec l’asthme
rs4286272	RFTN2	Intronique	Aucune
rs7240102	LOC100130003/C18orf26	Inconnu	Aucune
rs6581895	YEATS4/FRS2	Inconnu	Aucune
rs12405429	FAM89A/TRIM67	Inconnu	Aucune
rs1420101	IL18R1/IL1RL1	Inconnu	Aucune
rs117639512	KLK5	Inconnu	Aucune
rs10404342	ZNF71	Inconnu	Aucune
rs4879926	TLN1	Inconnu	Aucune

3- ANALYSE PHYLOGÉNÉTIQUE

Le fichier PIR obtenu à partir de l’alignement multiple présente une structure typique, où chaque séquence est identifiée par un nom de variant SNP (ex. : rs2070901, rs3132954) suivi de sa séquence alignée. Il s’agit ici de SNPs associés à l’asthme, extrait de la base de données dbSNP.

Un fichier PIR est un format textuel utilisé principalement pour représenter des alignements de séquences biologiques. Son nom vient de la base de données Protein Information Ressource (PIR). Il est utilisé pour décrire les séquences alignées, chacune précédée d’un en-tête qui suit une syntaxe particulière.

L’alignement (figure 6 et figure 7) met en évidence plusieurs caractéristiques importantes. On observe d’abord des zones fortement gappées en début ou en fin de séquence, comme chez rs3132954, traduisant des insertions ou suppressions par rapport à d’autres SNPs. À l’inverse, certaines régions centrales sont bien conservées, avec des motifs répétés tels que **ggtcca** partagés entre rs34492353 et rs443863, indiquant une possible fonction commune.

L’analyse biologique de ces observations suggère que les zones conservées pourraient correspondre à des régions fonctionnellement importantes, tandis que les régions variables, comportant des substitutions ou des indels, sont le reflet de la diversité génétique propre aux SNPs.

Cette structure permet donc non seulement de visualiser la variabilité entre les variants génétiques, mais aussi de cibler les positions les plus susceptibles d'avoir un impact biologique, en particulier si elles affectent des domaines conservés.

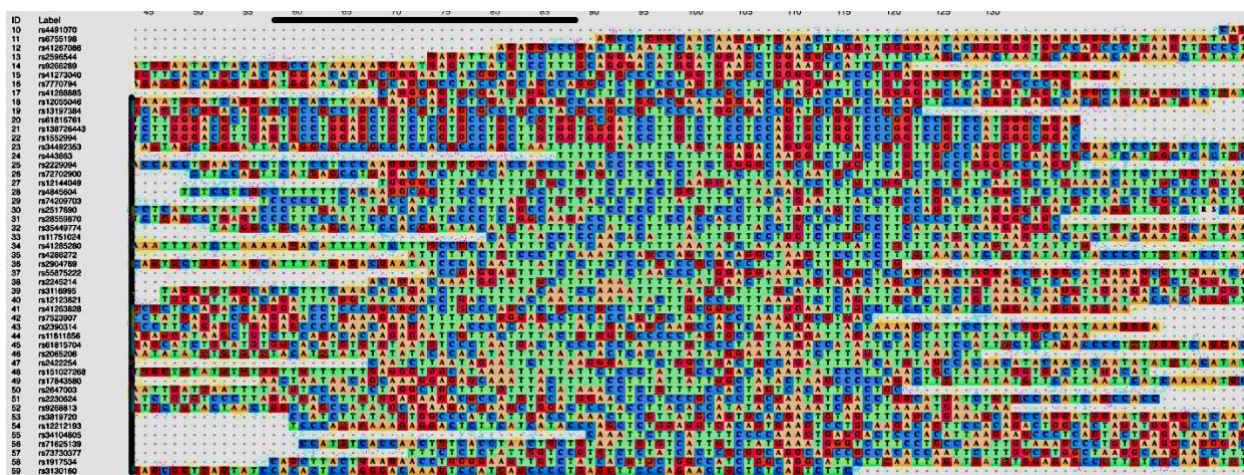


Figure 6 : Vue globale du résultat de l'alignement multiple des variants génétique.

- Adénine (A) : orange ou rouge
- Cytosine (C) : bleu
- Guanine (G) : vert
- Thymine (T) : beige

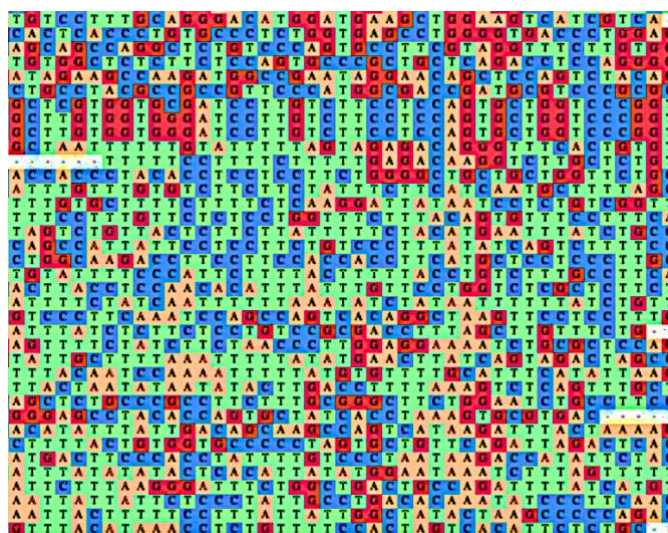


Figure 7 : Vue agrandie d'une partie centrale de l'alignement multiple.

L'analyse phylogénétique des variants SNPs associés à l'asthme constitue un outil essentiel pour mieux comprendre l'histoire évolutive et la distribution de ces polymorphismes dans les populations humaines. En identifiant les relations de parenté entre les différents variants, cette approche permet de retracer leur origine, de détecter des événements de sélection naturelle ou de dérive génétique, et d'évaluer leur conservation entre espèces proches.

Dans le contexte de l'asthme, maladie multifactorielle impliquant des interactions complexes entre facteurs génétiques et environnementaux, la phylogénie permet d'identifier les SNPs potentiellement fonctionnels ou ancestraux, ceux susceptibles d'avoir un rôle clé dans la susceptibilité à la maladie. Cette démarche peut également orienter les recherches vers des cibles génétiques pertinentes pour le diagnostic, la prévention ou le développement de traitements personnalisés.

L'arbre phylogénétique construit à partir des 60 SNPs (figure 8) révèle une structuration marquée en deux clades (groupes) principaux, suggérant des proximités évolutives différenciées entre les variants génétiques.

Le premier clade est composé majoritairement de variants tels que *rs2070901*, *rs7770794*, *rs2904789*, *rs2647003*, *rs3130160*, *rs13197384*, *rs2229094*, *rs41263828*, *rs72702900* et *rs11811856*, qui forment des clades compacts à faibles distances génétiques. Par exemple, le clade formé par *rs2070901* et *rs7770794* présente une longueur de branche très courte (0.91713), traduisant une forte proximité évolutive. De même, les SNPs *rs2904789* et *rs2647003* s'associent avec *rs3130160* dans un sous-clade interne, mettant en évidence leur proximité relative. Un autre clade distinct dans ce groupe comprend *rs13197384* associé aux SNPs *rs2229094* et *rs41263828*, ce qui pourrait indiquer une conservation fonctionnelle ou une co-évolution dans des loci génomiques proches. Ce groupe semble regrouper des SNPs possiblement conservés et impliqués dans des mécanismes similaires, notamment dans l'immunité ou l'inflammation.

Le deuxième clade principal comprend, quant à lui, des variants tels que *rs3828058*, *rs4845604*, *rs74209703*, *rs4286272*, *rs2517690*, *rs28559870*, *rs35449774*, *rs61815704*, *rs71625139*, ainsi que les SNPs *rs34104805*, *rs55875222*, *rs73730377*, *rs3819720*, *rs11751024*, *rs3116995*, *rs12123821*, *rs6755198*, *rs12055046*, *rs7523907* et *rs4491070*.

Ces SNPs forment des clades profondément ancrés avec des branches plus longues, traduisant une diversité évolutive plus grande. Par exemple, *rs4845604* est étroitement lié à *rs74209703* et à *rs4286272*, un variant précédemment identifié comme fortement associé à l'asthme. Un autre clade notable dans ce groupe inclut *rs61815704* et *rs71625139*, suggérant une possible co-sélection. Enfin, les SNPs *rs34104805* et *rs41267086* sont groupés dans un sous-clade bien distinct, signalant une relation évolutive cohérente.

La distinction entre ces deux groupes souligne la diversité des signatures phylogénétiques des SNPs analysés, pouvant refléter des origines évolutives différentes ou des pressions de sélection spécifiques.

Une étude de Kim et al. (2022) a identifié des SNPs dans le gène *TNFSF15* associés à l'asthme persistant chez les enfants coréens. Les variants *rs7856856*, *rs7848647*, *rs7848648* et *rs7848649* ont montré une association significative avec l'asthme, suggérant une implication de *TNFSF15* dans la pathogenèse de la maladie. Le travail de Dong et al. (2020) a intégré des données de GWAS et d'eQTL pour identifier des gènes et des SNPs associés à l'asthme modéré à sévère. Parmi les gènes identifiés figurent *GNGT2*, *TLR6* et *TTC19*, soulignant l'importance de ces loci dans la susceptibilité à l'asthme.

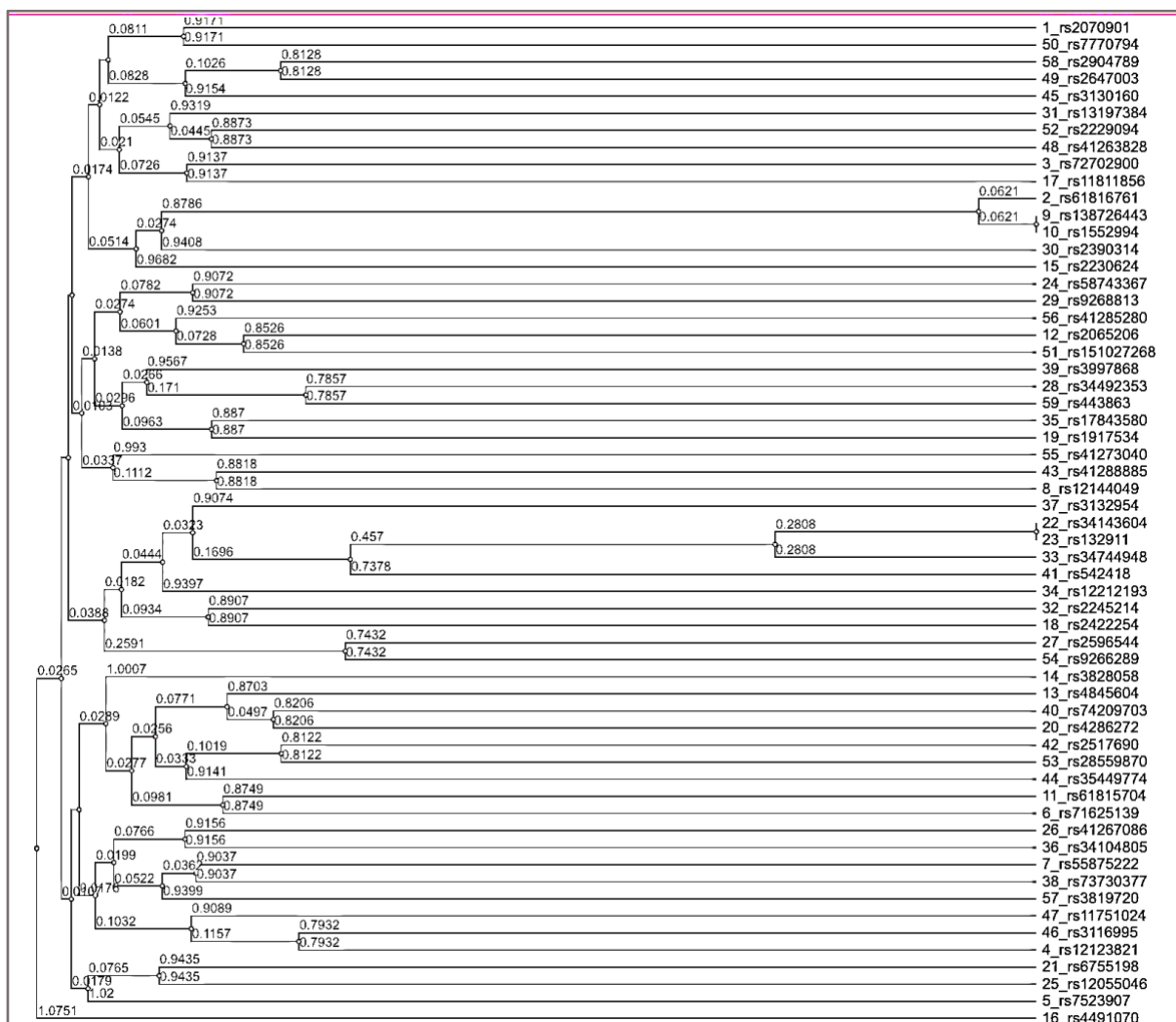


Figure 8 : Organisation phylogénétique des variants génétiques liés à l’asthme.

CONCLUSION

Dans le cadre de ce mémoire, nous avons entrepris une étude approfondie visant à identifier et à caractériser les variants génétiques associés à l'asthme à l'aide d'approches bioinformatiques et génétiques complémentaires. L'analyse de données issues de plusieurs bases de données de référence (NCBI/dbSNP, gnomAD, MutationT@ster) a permis de mettre en évidence un ensemble de SNPs d'intérêt potentiel, tant sur le plan fonctionnel qu'évolutif.

Parmi ces variants, rs2070901 a retenu l'attention en raison de sa structure tri-allélique (G, C et T), de sa fréquence allélique mineure élevée (MAF = 0,39) et de sa valeur PhyloP = 1,14, suggérant une conservation évolutive non négligeable. À l'inverse, le variant rs4286272 présente une valeur PhyloP plus modérée (0,49), indiquant une pression de sélection plus faible.

L'exploration des données fonctionnelles a également permis d'identifier des variants étroitement associés à la pathologie asthmatique : à titre de citation, le variant rs61816761, de type mutation non-sens, susceptible de provoquer une perte de fonction protéique, et le variant rs2099903, localisé dans la région 3'-UTR, avec un impact potentiel sur la régulation post-transcriptionnelle de l'expression génique.

Enfin, l'analyse phylogénétique des séquences nucléiques correspondant à ces SNPs a révélé l'existence de deux clades (groupes) distincts. Le premier groupe, comprenant notamment rs2070901, se distingue du second groupe, dans lequel figure, par exemple, rs3828058, ce qui témoigne d'une divergence évolutive plausible entre ces clusters de variants.

Ces résultats mettent en lumière l'apport considérable des outils bioinformatiques dans l'analyse intégrée des variants associés à des maladies complexes comme l'asthme. Ils offrent également une base solide pour de futures études fonctionnelles ou cliniques visant à valider l'implication de ces SNPs dans la physiopathologie de la maladie.

RÉFÉRENCES BIBLIOGRAPHIQUES

Braman S. S. 2006. The global Burden of Asthma. *CHEST Journal*. 130 (1): 4S-12S. DOI : 10.1378/chest.130.1_suppl.4S.

Bouzigon E., Demenais F., Kauffmann F. 2005. Génétique de l'asthme et de l'atopie : combien de gènes identifiés ? *Bull. Acad. Nat. Méd.* **7** : 1435-1448.

Bouzigon E. 2010. Asthme : du phénotype aux génotypes. *Revue française de l'allergie*. **50** : 193-196. DOI: 10.1016/j.reval.2010.01.033.

Dahmani D. I., Chila N., Abdelouahab F., Bouyoucef H., Bougrida M., Rouabah L. et al. 2020. L'association entre les gènes HLA II et la susceptibilité à l'asthme chez une population constantinoise sélectionnée. *Pan. Afr. Med. J.* **18** : 35-48. DOI: 10.11604/pamj.2020.35.48.13340.

Demenais F., Margaritte-Jeanin P., Barnes K. C., Cookson W. O. C., Moffatt M. F., Ober C., Nicolae D. L. 2018. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nature Genetics*. **50**: 42-53. DOI: 10.1038/s41588-017-0014-7.

Dong Z., Ma Y., Zhou H., Shi L., Ye G., Yang L. et al. 2020. Integrated genomics analysis highlights important SNPs and genes implicated in moderate-to-severe asthma based on GWAS and eQTL datasets. *BMC Pulmonary Medicine*. **20**: 270-286. DOI: 10.1186/s12890-020-01303.

Dos Santos B. A., Gbary A. R., Kpozehouen A., Kassa F. 2015. Facteurs associés à l'asthme sévère chez les patients asthmatiques suivis au centre national hospitalier de pneumo-phtisiologie de Cotonou (Bénin) en 2014. *Pan. Afr. Med. J.* **7**: 11-22. DOI: 10.11604/pamj.2015.22.11.5625.

Han Y., Jia Q., Jahani P. S., Hurrell B. P., Pan C., Gukasyan J. et al. 2020. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat. Commun.* **11** (1): 1-13. DOI: 10.1038/s41467-020-15649-3.

Holloway J. W., Yang I. A., Holgate S. T. 2010. Genetics of allergic disease. *J. Allergy Clin. Immunol.* **125** (2) : 581-594. DOI:10.1016/j.jaci.2009.10.071.

Khatri P., Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. 2005. *Bioinformatics.* **21**:3587–95. DOI: 10.1093/bioinformatics/bti565.

Kim S. H., Cho B. Y., Park C. S., Shin E. S., Cho E. Y., Yang E. M. et al. 2009. Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. *Clin. Exp. Allergy.* **39** (2):203-212. DOI : 10.1111/j.1365-2222.2008.03117.x.

Kim S. H., Forno E., Yan Q., Jiang Y., Zhang R., Boutaoui N. et al. 2020. SNPs identified by GWAS affect asthma risk through DNA methylation and expression of cis-genes in airway epithelium. *Eur. Respir. J.* **55** (4): 1902079. DOI: 10.1183/13993003.02079-2019.

Kim K. W., Kim D. Y., Yoon D., Kim K. K., Jang H., Schoettler N. 2022. Genome-wide association study identifies TNFSF15 associated with childhood Asthma. *Allergy.* **77** (1): 218-229. DOI:10.1111/all.14952.

Kim D. J., Lim J. E., Jung H., Chung J. Y., Baek E. J., Jung H. et al. 2023. Identification of asthma-related genes using asthmatic blood eQTLs of Korean patients. *BMC Med. Genomics.* **16**: 259. DOI: 10.1186/s12920-023-01677-7.

Knudsen S. 2004. *Guide to analysis of DNA microarray data*. John Wiley & Sons, Inc.; Hoboken, NJ.

Moffatt M. F., Kabesch M., Liang L., Dixon A. L., Strachan D., Heath S. et al. 2007. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. **44**:470-473. DOI: 10.1038/nature06014.

Palikhe N. S., Haji Q., Mack E., Sinnatamby T., Sandford A. J., Cameron L. et al. 2024. Association of single nucleotide polymorphisms in the F2RL1 gene with clinical and inflammatory characteristics of patients with asthma. *Allergy Asthma Clin. Immunol.* **20**: 8. DOI: 10.1186/s13223-024-00873-z.

Rodriguez E., Bourecht H., Herberich E., Wagenpfeil S., Brown S. J., Cordell H. J. et al. 2009. Meta-analysis of flaggrin polymorphism in eczema and asthma: Robust risk factors in atopic disease. *Atopic dermatitis and skin disease*. **123** (6):1361-1370. DOI: 10.1016/j.jaci.2009.03.036

Stikker B. S., Hendriks R. W., Stadhouders R. 2023. Decoding the genetic and epigenetic basis of asthma. *Allergy*. **78**: 940-956. DOI: 10.1111/all.15666.

Tarca A. L., Romero B., Draghici S. 2006. Analysis of microarray experiments of gene expression profiling. *Am. J. Obstet. Gynecol.* **195** (2): 373-388. DOI:10.1016/j.ajog.2006.07.001.

Tung N., Lin N. U., Kidd, J., et al. 2022. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing. *JAMA Oncology*, **8** (1) : 1–8. DOI:10.1001/jamaoncol.2021.4530.

Schena M. 2000. Microarray biochip technology. Eaton Publishing; Sunnyvale, CA.
Vercelli D. Discovering susceptibility genes for asthma and allergy. *Nat. Rev. Immunol.* **8**: 169-82. DOI: 10.1038/nri2257

Villeneuve T., Guillemainault L. 2020. Asthma and obesity in adults. *Revue de maladies respiratoires*. **37** : 60-74. DOI : 10.1016/j.rmr.2019.03.016.

Wachinou A. P., Agodokpessi G., Kintossou E., Imorou N., Sagbo G. 2020. Prévalence de l'asthme et facteurs associés chez les adolescents âgés de 13 à 14 ans en milieu scolaire dans deux villes du Bénin. *Revue des maladies respiratoires actualités*. **12** (1) : 178. DOI : 10.1016/j.rmra.2019.11.399.

Yang Y. H., Buckley M. J., Dudoit S., Speed T. P. 2002. Comparison of methods for image analysis on cDNA microarray data. *J. Comput Graph Stat*. **11**:108-36.

Yuan L., Toa J., Wang J., She W., Zou Y., Li R. et al. 2021. Global, regional, national, burden of asthma from 1990 to 2021, with projections of incidence to 2050: a systematic analysis of the global burden of disease study 2021. *EClinicalMedicine*. 80:103051.

Zhang X, Jafari N, Barnes RB, Confino E, Milad M, Kazer RR. 2005. Studies of gene expression in human cumulus cells indicate pentraxin 3 as a possible marker for oocyte quality. *Fertil Steril*. **83**:1169–79. DOI: 10.1016/j.fertnstert.11.030.