

Inventaire et analyse descriptive des maladies auto-immunes et des auto-anticorps associés connus

1^{er} Auteur : Fabien, MAURY, Doctorant, Université Paris-Cité, INSERM U1163, Institut Imagine, Paris, FR

- Maud, DE DIEULEVEULT, CRCN INSERM, Université Paris-Cité, INSERM U1163, Institut Imagine, Paris, FR
- Adrien, COULET, CR INRIA, Inria Paris, Paris, 75012, FR
- Anne-Sophie, JANNOT, MCU-PH, Banque Nationale de Données Maladies Rares, DSI-I&D, AP-HP, Paris, FR
- Julien, NGUYEN VAN, Master2, Université Paris-Cité, INSERM U1016, Institut Cochin, Paris, FR
- Isabelle, DESGUERRE, PU-PH, Département de neurologie pédiatrique, Hôpital Necker Enfants Malades, APHP, Université Paris Cité, Paris, FR
- Nicolas, GARCELON, Département d'informatique, Institut Imagine, Université Paris-Cité, Paris, FR
- Peter, ROBINSON, The Jackson Laboratory for Genomic Medicine, Farmington, USA

1- INTRODUCTION

Les maladies auto-immunes (MAI) sont la conséquence d'un dérèglement du système immunitaire, les auto-anticorps (AAC) étant des marqueurs sérologiques qui dans certains cas ont des effets pathogènes. Il existe plus de 80 types de maladies auto-immunes qui peuvent être systémiques ou affecter un large éventail d'organes spécifiques. Entre 5 et 8 % de la population mondiale est touchée par les MAI, qui sont la quatrième cause de maladie dans les populations occidentales. Les femmes sont plus touchées que les hommes, dans une proportion de 4 pour 1. En revanche, les MAI sont restées moins étudiées. Par exemple, il n'existe donc pas de thésaurus exhaustif des AAC dans la littérature. De plus, la construction d'une représentation complète des relations entre les AAC et les MAI est une étape importante vers une meilleure compréhension et représentation des maladies humaines.

2- MATERIEL & METHODES

- Avec l'aide d'experts médicaux des MAI, des Filières de Santé Maladies Rares françaises, et de l'ERN ERKNet nous avons étendu le thésaurus existant des phénotypes de positivité des auto-anticorps (signes biologiques des MAI) dans la base de données de l'Ontologie des Phénotypes Humains (HPO).
- Nous avons utilisé des expressions régulières pour faire correspondre les AAC de la base de données HPO avec le thésaurus MeSH afin de favoriser l'interopérabilité des données.

Human Phenotype Ontology
L'ontologie des phénotypes humains (HPO) est une ontologie formelle des phénotypes humains avec un vocabulaire normalisé des anomalies phénotypiques rencontrées dans les maladies humaines. Il s'agit d'une base de données collaborative et de nouveaux termes peuvent être ajoutés, édités ou archivés.

Medical Subject Headings
Le thésaurus Medical Subject Headings (MeSH) est un vocabulaire contrôlé et hiérarchisé produit par la National Library of Medicine.

HPO_code	HPO_name	MeSH_code	MeSH_name
HP_0033342	Anti-aquaporin 4 antibody positivity	C000719032	anti-aquaporin 4 autoantibody
HP_0032376	Anti-beta 2 glycoprotein I antibody positivity	C000719296	anti-beta 2 glycoprotein I autoantibody
HP_0033394	Anti-carbonic anhydrase II antibody positivity	C000719297	anti-carbonic anhydrase II autoantibody
HP_0030873	Anti-centromere antibody positivity	C538950	anticentromere antibody
HP_0020151	Anti-dsDNA antibody positivity	C000718988	anti-dsDNA autoantibody

Exemple de 5 lignes de la table de correspondance entre HPO et MeSH

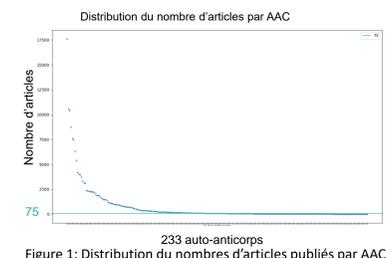
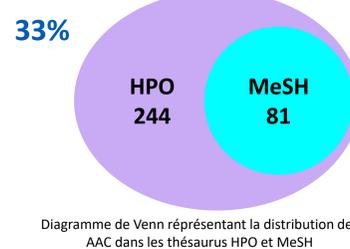
- En utilisant l'API "e-utilities" de PubMed, nous avons collecté les résumés et les métadonnées des articles dans lesquels au moins un AAC était mentionné, en filtrant sur l'espèce « human », « autoantibod- » pour la pertinence. Nous avons étudié les cooccurrences de différents AAC dans les articles (mesuré à l'aide de l'indice de Jaccard modifié) et les avons représentées dans une matrice de corrélation. Enfin, nous avons utilisé le regroupement hiérarchique sur les cooccurrences des AAC pour définir des groupes d'AAC.

Métrique choisie
$$J'(Aac1, Aac2) = \frac{Aac1 \cap Aac2}{\min(Aac1, Aac2)}$$

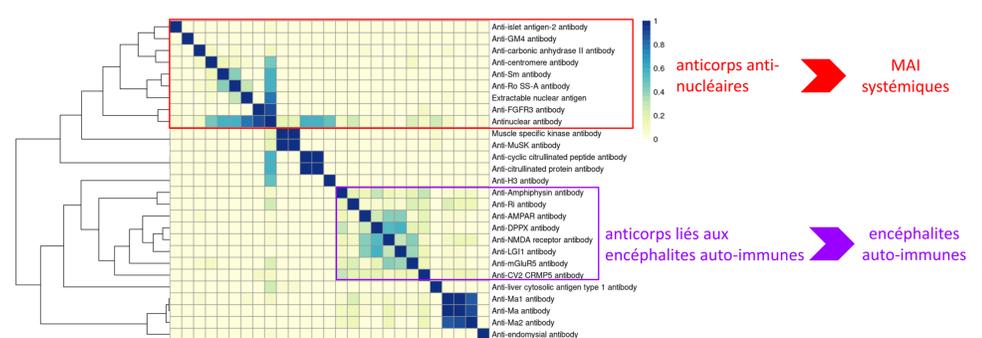
3- RESULTATS PRELIMINAIRES

Le thésaurus comprend actuellement **244 AAC** et la correspondance avec le thésaurus MeSH est de 33%.

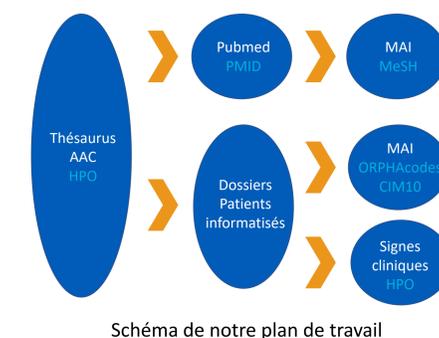
Un premier sous-groupe de 27 AAC a été isolé afin de tester et valider l'approche choisie (Figure 2). Des méthodes de clustering ont permis de former des groupes d'AAC avec des publications communes reflétant partiellement une réalité clinique.



La moitié des AAC sont cités dans plus de 75 articles

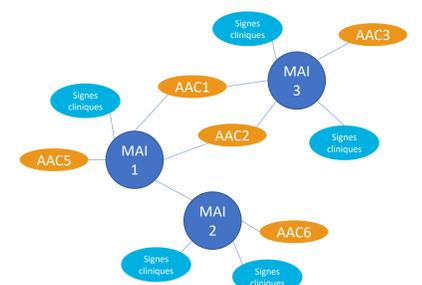


Nous allons ensuite étendre cette méthode à l'ensemble des AAC identifiés dans notre thésaurus. Les MAI associées seront ensuite identifiées. Dans un second temps, nous effectuerons le travail similaire en nous appuyant sur les dossiers de patients informatisés afin d'identifier les liens entre MAI ↔ AAC ↔ signes cliniques.



4- DISCUSSION

Plusieurs limites sont apparues au cours de notre première étude. 1) Une des limites de l'étude est que toute mention d'un AAC dans un article sera considérée comme une occurrence par l'API. Cependant, certains de ces articles concluent que l'AAC n'apparaît jamais dans la maladie, c'est donc un faux positif. 2) Une autre limite est que nous avons fixé un nombre minimal de cooccurrences à 10, ce qui pourrait cacher certains AAC avec un nombre limité d'articles publiés. 3) L'orthographe des AAC reste un challenge puisque aucune nomenclature n'est pour l'instant officiellement reconnue, il en résulte une perte d'articles lors de notre collecte. Nous prévoyons ensuite de construire des graphes de connaissances sur la base de ces informations.



5- CONCLUSION

Nous avons produit le premier thésaurus de AAC (HPO), mis en correspondance avec les ontologies HPO et MeSH, qui englobe les AAC de plusieurs spécialités médicales. Nous avons ensuite utilisé ce thésaurus pour étudier les relations entre les AAC et pour identifier les clusters potentiels trouvés à travers les diverses spécialités médicales. La prochaine étape de notre travail consiste à utiliser les dossiers de patients comme source de connaissance pour les MAI. Ce travail contribuera à une meilleure description des pathologies dans les dossiers médicaux électronique et améliorera la compréhension globale des MAI.