

République algérienne démocratique et populaire

Ministère de l'enseignement supérieur et de la recherche scientifique



Université de Blida-1-

Faculté des sciences de la nature et de la vie

Département de biologie et de physiologie cellulaire



**MÉMOIRE**

En vue de l'obtention du diplôme de master en biologie

Option : Bio-informatique

**Thème :**

**Analyse de la fréquence du variant R620W du gène  
PTPN22 dans la population Algérienne**

Présenté par :

**ELOUZERI Safa**

**BENDJELLOUL khadidja**

**Soutenu le : 17/09/2017**

Devant le jury :

-Mr MAHIEEDIN MA(A) Université de Blida-1 Président

-Mme AISSANLR MA(A) Université de Blida-1 Examinatrice

-Mme AMROUN MESBAH.HPROFESSEUR USTHB ALGER Promotrice

**Promotion  
2016-2017**

## Résumé

La polyarthrite rhumatoïde (PR) est une maladie inflammatoire, systémique et auto-immune qui affecte environ 0,3 à 1% de la population générale adulte, avec une prédominance féminine. Elle est caractérisée par une destruction de l'os et du cartilage articulaire avec une hyperplasie de la membrane synoviale. L'existence de facteurs génétiques de la PR est connue depuis longtemps. L'association génétique la plus forte est observée avec les gènes codant pour les molécules HLA-DRB1 qui contribuent à hauteur de 30% dans le risque familial global alors que le gène *PTPN22* est l'un des principaux facteurs génétiques non-HLA codant pour la protéine LYP impliquée dans la régulation des lymphocytes T (LT), acteurs majeurs de l'inflammation synoviale de la PR.

Notre travail comporte deux parties : une méta-analyse et une étude Bio-informatique.

La méta-analyse comporte deux volets. Un volet dont l'objectif est l'étude de la fréquence du polymorphisme C1858T du gène *PTPN22* dans la population normale après cumul des effectifs obtenus à partir de trois études indépendantes effectuées sur la population Algérienne. Le deuxième volet a pour objectif l'analyse de l'association du polymorphisme C1858T du gène *PTPN22* à la PR dans la population Algérienne après cumul des effectifs obtenus à partir de deux études indépendantes. Ces deux études remplissent les critères d'inclusion nécessaires à une méta-analyse. A l'issue de notre méta-analyse, Nos résultats ont montré que l'allèle mineur T du est associé à la PR de façon significative (OR=3.8%CI 95% (1.85-4.74)) ;  $P < 0.0001$ ).

L'étude Bio-informatique est une étude théorique réalisée à partir de la séquence en acides aminés de la protéine LYP, extraite à partir de la banque UniProtKB sur le portail EXPASY. Cette étude est basée sur l'utilisation de logiciels bioinformatiques de prédictions permettant l'annotation des protéines et accessibles gratuitement à partir de portails bio-informatiques publiques. Avec l'outil Clustalomega, nous avons montré que la fonction biologique et la structure biochimique de la protéine LYP sont hautement conservées chez les primates, ce qui nous a permis d'appréhender l'importance du rôle biologique de cette protéine. Nous avons réalisé une caractérisation physicochimique de la protéine LYP avec l'outil PROTPARAM, une prévision de ses structures secondaires et 3D à l'aide des outils PSIPRED et SWISS-MODEL, respectivement. Des motifs fonctionnels en accord avec une activité enzymatique de type PTPase, ont été identifiés à l'aide de l'outil PROSITE. Tous ces résultats démontrent la puissance des outils de la bio-informatique et la nécessité de les intégrer dans toute recherche envisagée en biologie.

## Remerciement

*« Nous tenons tout d'abord à exprimer notre profonde reconnaissance à notre Créateur, à sa force qui a su nous guider et nous apporter réconfort, courage et espoir dans les moments où on en avait le plus besoin.*

*Nous exprimons toute notre gratitude à Mme MESBAH- AMROUN Hamida, Professeur et membre de l'équipe de « Cytokines et NO Synthase » pour la formation de qualité dont on a bénéficié et l'assistance, la collaboration, l'aide et son aimable accueil au sein du laboratoire pour la réalisation de ce travail.*

*Nous remercions notre responsable et chef d'option de bio-informatique du BLIDA notre enseignant DR : « MOHAMED SAID »*

*Nous remercions l'enseignante : « Mme AISSANI .R ».*

*ELOUZERI BELKACEM Safa*

*A mes parents :*

*A mon père Abdelkader, merci pour ton encouragement, ton aide et tes conseils*

*A ma mère Fatiha, son soutien, son confiance, sa tendresse et générosité immense sont le moteur de mon avancé.*

*« A mon mari Mohamed » :*

*Merci de m'avoir soutenu dans les moments difficiles, tes encouragements, ton affection immense, ta patience, m'ont permis d'aller de l'avant dans mon travail*

*A mon chère et adorable fils « Youcef Essedik » que dieu tu protège*

*A mes adorables sœurs « MAROUA & MERIEM », merci pour tes encouragement et l'assistance morale*

*A mes chères frères « AbdelMalek, Mouetez et Ossaid » pour leurs soutiens et leurs affections*

*A ma belle-mère et mon beau-frère « DjamelEddin » pour leurs aides, et leurs patiences.*

*A mes chères amies Sara, Imène, Nessrine, Hadjer et Mouna. Je vous remercie d'avoir été là aussi présents et chaleureuses.*

BENDJELLOUL khadidja

*A mes parents :*

*A mon père, merci pour ton encouragement, ton aide et tes conseils*

*A ma mère, son soutien, son confiance, sa tendresse et générosité immense sont le moteur de mon avancé.*

*A mon mari MOUNIR :*

*Merci de m'avoir soutenu dans les moments difficiles, tes encouragements, ton affection immense, ta patience, m'ont permis d'aller de l'avant dans mon travail*

*A mon adorable fils « MOHAMED ISLAM » :*

*A mes adorables sœurs « SAFAA, CHAYMAA », merci pour tes encouragements et l'assistance morale*

*A mes chères frères « HAMZA, ZAKARIA » pour leurs soutiens et leurs affections*

*A ma belle-famille notamment ma belle-mère et mon beau-père pour leurs aides, et leurs patiences.*

*A ma chère binôme « SAFA » qui je partage ce travail avec elle, qui a été adorable, courageuse et patiente avec moi.*

*A mes chères amies Somia, Imène, Leila, Assia et Selma, meriem, rawaa. Je vous remercie d'avoir été là aussi présents et chaleureuses.*

*Je remercie enfin, tous ceux qui ont participé de près et/ ou de loin pour la réalisation de ce travail.*

## LISTE DES ABRÉVIATIONS:

<b>BCR:</b>	B Cell Receptor
<b>C-cbl:</b>	Casitas B-lineage lymphoma
<b>CD:</b>	Cluster of differentiation
<b>CMH:</b>	Major Histocompatibility Complex
<b>CSK:</b>	C-terminal Src kinase
<b>GWAS:</b>	genome wide association study
<b>Grb2:</b>	Growth factor receptor bound protein 2
<b>HLA:</b>	Human Leucocyte Antigen
<b>OR :</b>	Odd Ratio
<b>IFN-<math>\alpha</math>:</b>	Interferon alpha
<b>IFN-<math>\gamma</math>:</b>	Interferon gamma
<b>IKK<math>\epsilon</math>:</b>	Inhibitor- $\kappa$ B kinase $\epsilon$
<b>IL-17:</b>	Interleukin 17
<b>IRF7:</b>	Interferon Regulatory Factor 7
<b>ITAM:</b>	Immunoreceptor Tyrosine based Activation Motif
<b>LADA:</b>	Latent Autoimmune Diabetes in Adult
<b>LAT:</b>	Linker of Activated T cells
<b>LB:</b>	B Lymphocytes
<b>LCK:</b>	Lymphocyte-specific protein tyrosine kinase
<b>LT:</b>	T Lymphocytes
<b>LYP:</b>	Lymphoïde phosphatase
<b>MAI:</b>	Maladies Auto-immunes
<b>PR :</b>	polyarthrite rhumatoïde
<b>PAMPs:</b>	Pathogen associated molecular pattern
<b>PBMC:</b>	Peripheral Blood Mononuclear Cell
<b>PCR:</b>	Polymerase Chain Reaction
<b>PTPN22:</b>	Protein Tyrosine Phosphatase Non Receptor type 22
<b>R620W:</b>	variant de gène PTPN22
<b>SH2/3:</b>	Src Homology 2/3
<b>SNP:</b>	Single Nucleotide Polymorphism
<b>TBK1:</b>	TANK-Binding Kinase 1

<b>TCR:</b>	T cell Receptor
<b>Th1/17:</b>	T helper 1/17
<b>TLR:</b>	Toll like Receptor
<b>TRAF3:</b>	Tumor necrosis factor receptor-associated factor 3
<b>VAV:</b>	Glutamine nucleotide exchange factor
<b>ZAP-70:</b>	Zeta-chain-associated protein kinase 70
<b>3D:</b>	structure tridimensionnelle

## Liste de figure

Figure 1 :Classification des maladies auto-immunes .....	4
Figure 2 :représentation schématique d'une articulation normale et rhumatoïde.....	5
Figure3 : localisation de gène PTPN22 sur le chromosome humain.....	7
Figure 4 : représentation schématique de la structure de la protéine LYP.....	7
Figure 5 : la fonction de PTPN22 dans la régulation du signal du TCP.....	9
figure 6 : Le rôle de la protéine LYP au niveau des cellules de l'immunité innée.....	10
Figure 7 : Distribution géographique du polymorphisme PTPN22 C1858T.....	13
Figures 8 : résultats obtenus à partir de l'alignement multiples des séquences avec l'outil CLUSTAL OMEGA.....	21
Figures9 : résultats obtenus avec l'Outils PROTPARAM.....	27
Figures10: Résultats obtenus à partir de PSIPRED.....	33
Figures11: résultats obtenus avec l'outil PROSITE.....	35
Figure 12 : résultats obtenus avec l'outil SWISSMODEL.....	37

## Liste des tableaux

Tableau 1:

Caractéristiques démographiques des sujets intégrés dans la méta-analyse.....18

Tableau 2:

Fréquences alléliques et génotypiques du polymorphisme R620W du gène *PTPN22* dans la population Algérienne témoin.....18

Tableau 3 :

Fréquences alléliques et génotypiques du polymorphisme R620W du gène *PTPN22* au cours de la PR et méta-analyse.....19



## Table des matières

I/introduction .....	1
Rappels bibliographiques	
Généralités.....	4
1-les maladies auto immunes .....	4
2-polyarthrite rhumatoïde (PR).....	5
3-facteurs génétiques .....	6
4-intérêt de l'étude des polymorphismes génétiques.....	6
5- Le gène PTPN22 humain (proteine tyrosine phosphatase non-receptor type 22)	
5.1-la localisation du gène PTPN22 .....	7
5.2-la protéine LYP .....	7
5.3-rôle et fonction du produit gène PTPN22 .....	8
5.3.1-au niveau des lymphocytes T : .....	9
5.3.2-au niveau des cellules de l'immunité innée : .....	10
5.4-les polymorphismes du gène PTPN22 .....	11
5.5-distribution géographique du variant R620W .....	12
II/MATERIEL ET METHODES : .....	14
Matériel	
1- Les données bibliographiques relatives au gène PTPN22	
2- Les séquences.....	15
Méthodes.....	15
I-Etude statistique.....	15
- Le recueil de données.....	15
- La méta-analyse.....	15
II- Etude bio-informatique : .....	15
III/RESULTATS & DISCUSSION .....	16
1-méthodes statistiques : résultats obtenus par :.....	17
1.1- outils de bio-statistiques (GRAPH PAD prism7).....	18
1.2- le calcul de fréquences alléliques).....	18
1.3- le calcul de fréquences génotypique ).....	19
1.4- calcul des odd ratio (OR) ).....	19
2-méthodes bio-informatiques :résultats obtenus par :.....	21
2.1-l'outil clustal omega et l' arbre phylogénétique.....	21
2.2-l'outil PROTPARAM :.....	27
2.3-l'outil PSIPRED .....	33
2.4- l'outil PROSITE.....	35
2.5-l'outil SWISS-MODEL.....	37
IV/CONCLUSION .....	39
V/REFERENCES BIBLIOGRAPHIQUES.....	41
VI/ANNEXES .....	48



# **INTRODUCTION**

INTRODUCTION

Les maladies auto-immunes (MAI) sont des maladies multifactorielles qui font intervenir des facteurs environnementaux (agents infectieux, tabac...) et des facteurs génétiques tels que les gènes HLA et non HLA (Well et Batteux, 2003). Parmi les gènes non HLA qui sont impliqués dans le développement des MAI, le gène PTPN22 (protein tyrosine phosphatase non-receptor type 22). Ce gène, localisé en position 1p13, s'étend sur 57949 paires de bases, contient 21 exons et code pour la protéine LYP (Lymphoïde Phosphatase) (Bottini et Peterson, 2014).

La protéine LYP est une phosphatase cytoplasmique exprimée uniquement au niveau des cellules hématopoïétiques (Cloutier et al., 1999). LYP intervient dans différentes voies de signalisations au niveau de différents types cellulaires impliqués dans l'immunité innée et dans l'immunité adaptative (Bottini et Peterson, 2014). De nombreuses études scientifiques ont montré que le SNP (single nucleotide polymorphism) C1858T du gène PTPN22 est associé à de nombreuses MAI dont la polyarthrite rhumatoïde (PR) et ce dans plusieurs populations à travers le monde (Bottini et Peterson, 2014). Le SNP C1858T correspond à la substitution d'une cytosine par une thymine en position 1858 dans l'exon 14 du gène PTPN22. Cette substitution entraîne le remplacement d'une arginine par un tryptophane au niveau de l'acide aminé 620 de la protéine LYP (variant R620W).

Notre travail s'inscrit dans le cadre d'une étude prospective qui a commencé depuis 2013 et qui porte sur la susceptibilité génétique dans les maladies auto-immunes dans la population Algérienne, impliquant le gène PTPN22. La question est de savoir :

- 1- Quelles sont les propriétés physico-chimiques de la protéine LYP?
- 2- Est-ce que le variant R620W du gène PTPN22 existe dans la population Algérienne et quelle serait sa fréquence ?
- 3- Est-ce que le variant R620W du gène PTPN22 est associé à la PR dans la population Algérienne?

Pour réaliser nos objectifs, nous avons adopté la démarche expérimentale suivante :

1-Utilisation des outils de la bioinformatique pour prédire les propriétés physico-chimiques de la protéine LYP humaine et analyse phylogénétique à partir de trois séquences protéiques (homme, chimpanzé, souris) en vue d'établir une relation entre sa structure et sa fonction biologique.

2-Utilisation de la banque bibliographique PubMed pour rechercher des publications autour du variant R620W dans la population Algérienne.

3-Utilisation des outils statistiques pour évaluer la fréquence allélique du variant R620W dans la population Algérienne à travers une méta-analyse regroupant les études ayant porté sur la population Algérienne.

4-Evaluation de l'association du variant R620W à la PR à travers une méta-analyse regroupant les études ayant porté sur la population Algérienne.

## Généralités :

L'inflammation est une réponse physiologique qui a pour but de limiter l'invasion des pathogènes et l'atteinte tissulaire en cas de blessure. La réponse inflammatoire aiguë est un processus strictement contrôlé, avec une phase d'initiation et une phase de résolution. Une maladie inflammatoire chronique est une condition pathologique caractérisée par une inflammation systémique, longue et persistante avec l'absence de phase de résolution.

**1-Les maladies auto-immunes (MAI)** sont la conséquence d'une dérégulation du système immunitaire entraînant une réponse immunitaire inadaptée de l'organisme contre les antigènes du soi à l'origine d'un processus pathologique. Elles sont classées comme **systémiques** ou **spécifiques d'organes**.

Ce sont des affections multifactorielles qui font intervenir des facteurs environnementaux (agents infectieux, tabac...) et des facteurs génétiques tels les gènes HLA et non HLA (Well et Batteux, 2003). Parmi les gènes non HLA, le gène PTPN22 impliqué dans de nombreuses MAI telles que la polyarthrite rhumatoïde (PR).

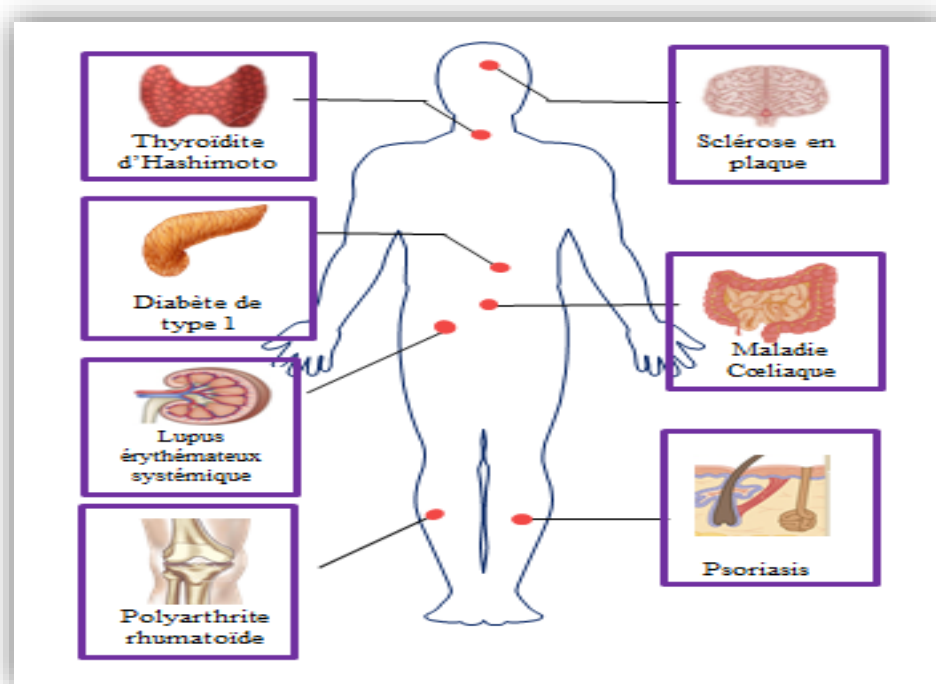
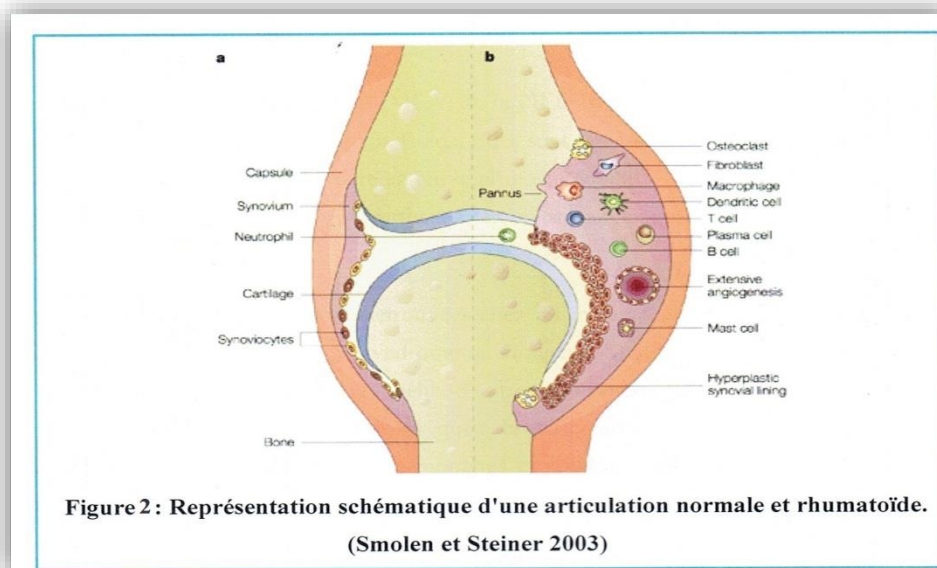


Figure 1 : Classification des maladies auto-immunes associées au gène PTPN22

## 2-La polyarthrite rhumatoïde :

La PR est une maladie inflammatoire auto-immune systémique. Elle est actuellement considérée comme le rhumatisme inflammatoire de l'adulte le plus fréquent avec une prévalence estimée entre 0.3% et 1% de la population générale adulte (Carmona et al. 2010). Elle se caractérise par une atteinte inflammatoire des synoviales articulaires et tendineuses, prédominante aux mains, aux pieds et aux genoux. Elle se définit comme un rhumatisme inflammatoire poly articulaire, évoluant de façon chronique par poussées, pouvant entraîner des déformations et des destructions articulaires (Emery et al.2008).



**Figure 2 : Représentation schématique d'une articulation normale et rhumatoïde.**  
(Smolen et Steiner 2003)

La PR peut également être une maladie systémique entraînant des manifestations extra-articulaires pouvant compromettre le pronostic vital (McInnes et Schett 2011).

Au niveau de l'étymologie, on retrouve le mot (arthrite) signifiant l'inflammation de la membrane synoviale avec sécrétion anormale de liquide synoviale et le préfixe (poly) signifiant que plus de quatre articulations sont généralement touchées au cours de cette maladie.

L'inflammation rhumatoïde, appelée synovite rhumatoïde, se caractérise, comme toutes les inflammations, par une vasodilatation, un œdème et une infiltration cellulaire. Sur le plan histopathologique, de nombreux changements vont s'opérer en fonction de la progression de la maladie (figure) (Choy 2012).

Les capillaires et veinules post-capillaires ainsi que les synoviocytes de la membrane synoviale deviennent hyperplasiques. Une infiltration massive par les monocytes/macrophages, les cellules dendritiques, les lymphocytes T et B ainsi que les plasmocytes à également lieu au sein de la couche

sous-intimale, aboutissant ainsi à la formation d'un tissu bourgeonnant, le pannus synovial qui est à l'origine des téno-synovites, synovites et de l'érosion du cartilage et de l'os sous-jacent (Choy 2012 ; Smolen et Steiner 2003).

### 3-Facteurs génétiques :

Les facteurs génétiques ont été suspectés de jouer un rôle depuis quelques décennies sur la base d'un regroupement de la maladie au sein de familles. Egalement, l'étude du taux de concordance de la PR chez les jumeaux monozygotes a permis de confirmer l'existence d'une prédisposition génétique à la PR (Kurko et al.2013).

Egalement, les études GWAS (voir annexe N°3) ont démontré l'existence de plusieurs gènes candidats qui augmentent la survenue de la PR. Le locus HLA-DR est considéré comme étant le facteur génétique le plus fortement associés à la PR (Stastny 1976). Par la suite, d'autres gènes non-HLA ont été identifiés et associés à la susceptibilité à la PR. Ces gènes candidats sont essentiellement impliqués dans les mécanismes immunopathologiques de la PR (Ruysse-witrand et al.2012).

### 4. Intérêt de l'étude des polymorphismes génétiques :

La caractérisation et la compréhension des variations génétiques constituent un véritable challenge en génétique humaine tant pour les individus indemnes de toute pathologie que pour les personnes malades. Les progrès n'ont été possibles qu'à partir du séquençage de l'ensemble du génome et des avancées technologiques nécessaires à ces travaux. La variabilité génétique se présente sous différentes formes. Le niveau le plus simple est représenté par des variations d'une seule base nucléotidique (A, T, G, C) correspondant aux substitutions nucléotidiques (**Single Nucléotide polymorphism** ou **SNP** en anglais). On distingue également d'autres variations de structure telles que les séquences répétées (mini ou micro-satellites ou **VNTR** pour « Variable Number of Tandem Repeat ») et les insertions/délétions correspondant à la présence/absence d'une ou plusieurs bases et dont la taille est inférieure à 1 kilobase (kb). Ces variations de structures peuvent modifier la quantité d'ADN comme les **CNV (Copy Number Variant)**, qui intéressent des fragments d'ADN de plus de 1 kb (Figure) (**Frazer et al., 2009**).

## 5- Le gène PTPN22 humain (protein tyrosine phosphatase non-receptor type 22).

### 5.1 Localisation du gène PTPN22

Le gène PTPN22 est localisé sur le chromosome 1 en position 1p13. Il s'étend sur 57949 paires de bases et contient 21 exons (Bottini et Peterson, 2014) et s'exprime dans les tissus lymphoïdes.

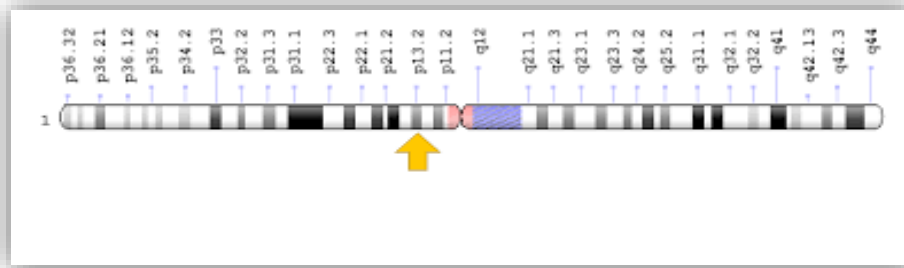


Figure3 : localisation de gène PTPN22 sur le chromosome 1 humain

### 5.2 La protéine LYP:

C'est une phosphatase cytoplasmique exprimée uniquement au niveau des cellules hématopoïétiques (Cloutier et al., 1999). Quatre isoformes de la protéine Lyp, traduites de manière indépendante suite à un épissage alternatif ont été mises en évidence : la protéine LYP1, la protéine LYP2, la protéine LYP3 et la protéine PTPN22.6 (Burn et al., 2011).

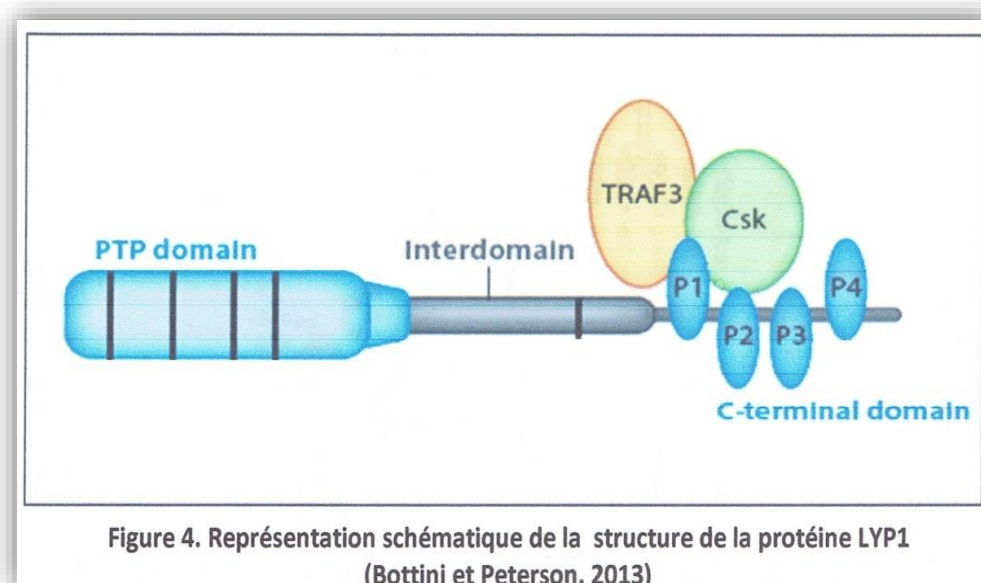


Figure 4. Représentation schématique de la structure de la protéine LYP1 (Bottini et Peterson, 2013)

L'isoforme LYP1 est une protéine composée de 807 acides aminés (aa) et possède un poids moléculaire de 105 KDa. C'est l'isoforme la plus abondante, exprimée par le gène PTPN22. La



protéine LYP est composée de 3 domaines majeurs : un domaine catalytique tyrosine phosphatase dans la partie N-terminal (1-300 aa), un domaine intermédiaire (301-600 aa) et un domaine C-terminal (601-807 aa) contenant 4 motifs riches en prolines (P1, P2, P3 et P4) (Cohen et al., 1999). Le domaine P1 a la capacité de se lier au domaine SH3 (Src homologie 3) de la protéine tyrosine kinase Csk qui est un régulateur négatif de la voie de signalisation du TCR (Ghose et al., 2001). La protéine LYP interagit aussi avec la protéine TRAF3 exprimée au niveau des cellules myéloïdes (Wang et al., 2013) (figure4).

LYP2 est une protéine composée de 691 aa, elle ne présente que le domaine P1 dans la partie C terminale (perte des domaines P2, P3 et P4), elle est essentiellement exprimée par les cellules immunitaires au repos et lors de l'hématopoïèse (Cohen et al., 1999).

LYP3 est une protéine amputée de 28 aa entre le motif P1 et P2. LYP3 paraît être exprimée parallèlement à LYP1 dans les tissus lymphoïdes (Wang et al., 2010).

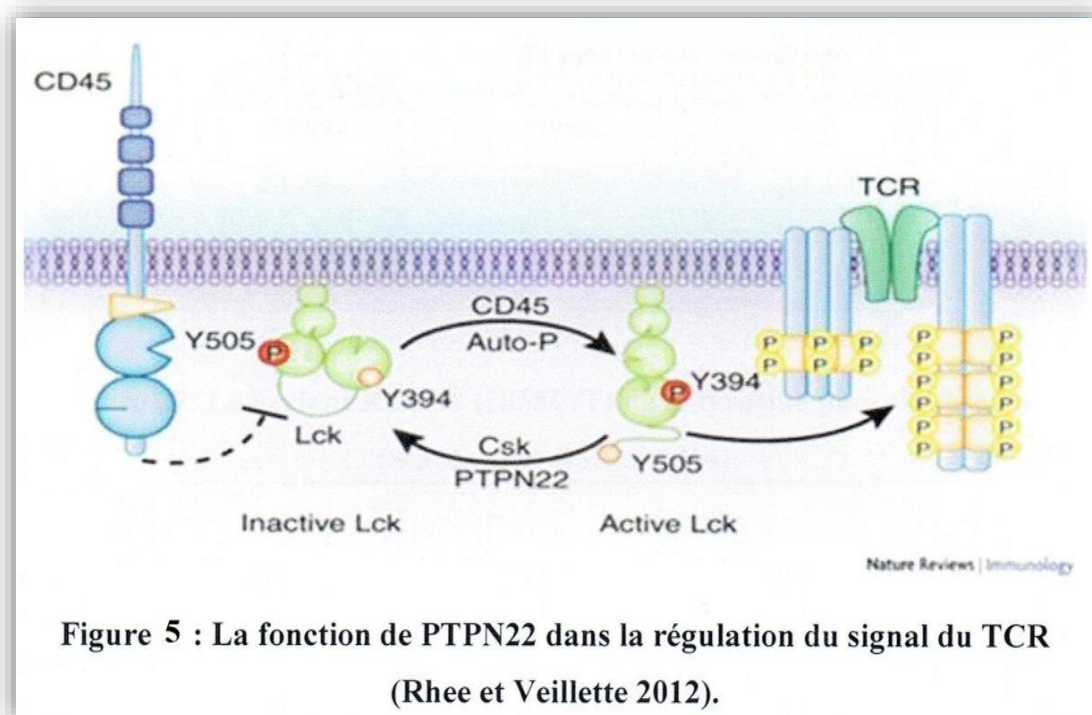
PTPN22.6 est une isoforme de 998 aa amputée de la partie centrale du domaine catalytique. Il apparaît que cette isoforme inhibe l'activité phosphatase de LYP1 (Chang et al., 2012).

### 5.3 Rôles et fonctions du produit du gène PTPN22:

Le gène PTPN22 code pour un membre de la sous-famille non-récepteur de classe 4 de la famille de la protéine-tyrosine phosphatase. La protéine codée est une phosphatase intracellulaire spécifique de lymphocyte qui s'associe à la protéine adaptatrice moléculaire CBL et peut être impliquée dans la régulation de la fonction de CBL dans la voie de signalisation du récepteur de lymphocyte T. Les mutations de ce gène peuvent être associées à une gamme de troubles auto-immuns, y compris le diabète de type 1, la polyarthrite rhumatoïde, le lupus érythémateux systémique et la maladie de Graves, la maladie d'Addison et la myasthénie.

La protéine est surexprimée dans le syndrome coronarien aigu entraînant une baisse des lymphocytes T régulateurs. La protéine LYP intervient dans différentes voies de signalisations au niveau de différents types cellulaires impliqués dans l'immunité innée et dans l'immunité adaptative (Bottini et Peterson, 2014).

### 5.3.1-Au niveau des lymphocytes T :



**Figure 5 : La fonction de PTPN22 dans la régulation du signal du TCR (Rhee et Veillette 2012).**

Le rôle de la protéine LYP est principalement décrit au niveau des lymphocytes T où elle régule négativement la voie de signalisation du TCR. La protéine LYP permet d'inhiber, par déphosphorylation en position tyrosine, l'activité de certaines kinases et molécules adaptatrices se trouvant au début de la cascade de signalisation de TCR. (Wu et al., 2006).

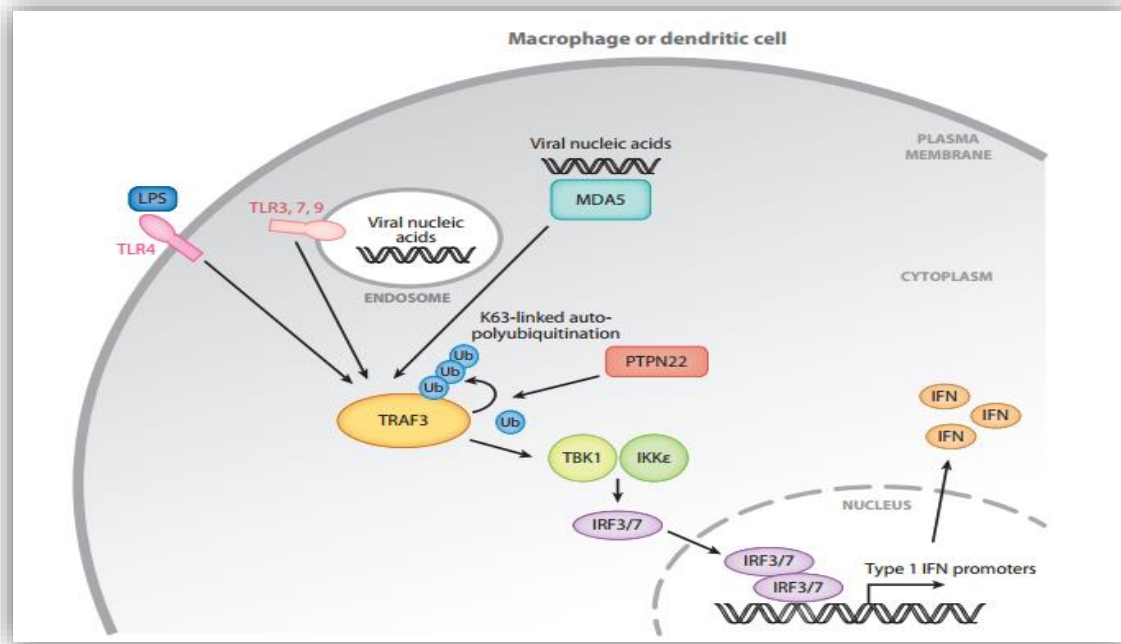
Au repos, l'activité de la protéine LCK, qui est une kinase activatrice de la cascade de signalisation du TCR, est régulée négativement par l'action synergique de la protéine LYP et CSK, en effet, la CSK phosphoryle LCK en position Y505 tandis que LYP déphosphoryle le site Y 394 situé au niveau du domaine catalytique de la protéine LCK. L'action synergique de LYP et de CSK permettra de bloquer le déclenchement de la voie de signalisation du TCR. (figure 5).

Suite à l'engagement du TCR par la reconnaissance du déterminant antigénique présenté par le CMH, l'inactivation de LCK peut être levée par la déphosphorylation du résidu Y 505 par la phosphatase CD 45. Il s'ensuit une autophosphorylation de LCK sur son résidu Y 394. La protéine LCK activée phosphoryle les motifs ITAM des domaines intracytoplasmiques des sous-unités du CD3. Les ITAMS phosphorylés recrutent la kinase Zap-70 via ses deux domaines SH2. Zap-70 est alors phosphorylée par LCK et activée.

Parmi les cibles de Zap70, on trouve l'adaptateur transmembranaire LAT et l'adaptateur cytoplasmique SLP-76. Le recrutement de ces deux molécules adaptatrices permet de constituer la

charpente de base nécessaire à l'initiation des voies de signalisation permettant l'activation et la prolifération des lymphocytes T, en activant d'autres intermédiaires protéiques tels que Grb2 et VAV. La protéine LYP inhibe la signalisation du TCR en déphosphorylant certaines molécules intermédiaires Zap-70, VAV, C-cbl, Grb2 ainsi que les motifs ITAM du corécepteur CD3.

### 5.3.2-Au niveau des cellules de l'immunité innée :



**Figure 6 :** Le rôle de la protéine LYP au niveau des cellules de l'immunité innée (Bottini et Peterson, 2014)

Indépendamment de son activité enzymatique, la protéine LYP favorise la production d'IFN de type 1 par les cellules myéloïdes (macrophages et cellules dendritiques). L'activation de la cascade de signalisation des TLR suite à la reconnaissance des PAMPs provoque l'auto-ubiquitination de TRAF3 et favorise cette ubiquitination. L'ubiquitination de TRAF3 provoque l'activation des sérines-thréonines kinases TBK1 and IKKε. ces deux intermédiaires protéiques phosphorylent les facteurs de transcription IRF3 (dans le cas des infections bactériennes ) et IRF7 (dans le cas des infections virales ). Ces deux facteurs de transcription transloquent vers le noyau et régulent positivement la transcription du gène codant pour les IFN de type 1 (Wang et al., 2013) (figure 6 ).

#### 5.4- Les polymorphismes du gène PTPN22 :

Plusieurs polymorphismes mononucléotidiques du gène PTPN22 sont décrits comme étant associés à de nombreuses maladies auto immunes dont trois polymorphismes majeurs qui sont les plus fréquents, les plus étudiés et les plus liés à l'auto immunité.

Le polymorphisme PTPN22 C1858T correspond à la substitution d'une cytosine par une thymine en position 1858 (C1858) située dans l'exon 14 du gène PTPN22. Ce changement de base étant non synonymique car il entraîne le remplacement d'une arginine par un tryptophane au niveau de l'acide aminé 620 de la protéine LYP. Cette transition en acide aminé se situe au niveau du motif P1 qui permet l'interaction de la protéine LYP avec la protéine CSK. Cette substitution non conservative, appelée polymorphisme R620W, empêche en effet la formation du complexe LYP-CSK (Bottini et al., 2006). De nombreuses études ont montré que le variant C1858T du gène PTPN22 est associée à une augmentation du risque de développer certaines maladies auto-immunes comme le diabète de type 1, la thyroïdite d'Hashimoto, la maladie de Graves, la polyarthrite rhumatoïde, l'arthrite juvénile idiopathique, la myasthénie grave, le vitiligo, le lupus systémique érythémateux, la maladie d'Addison, le purpura thrombopénique idiopathique et la pelade (Stanford et Bottini, 2014). Le variant PTPN22 C1858T prédispose aussi à la cooccurrence de maladies auto-immunes telle que le diabète de type 1 avec la thyroïdite auto-immune (Luo et al., 2012) et la polyarthrite rhumatoïde (Liao et al., 2009). De plus, ce polymorphisme semble être non associé avec la maladie cœliaque, la sclérose en plaque, le psoriasis, l'uvéite auto-immune et la colite ulcéreuse. De plus, ce polymorphisme présente un effet protecteur à l'encontre de la maladie de Crohn et de maladie de Behçet (Stanford et Bottini., 2014). Le polymorphisme C1858T joue aussi un rôle dans la susceptibilité à certaines maladies infectieuses où il augmente le risque du développement de la pneumococcie invasive causée par la bactérie *Streptococcus pneumoniae*. (Chapman et al., 2006) et de la lèpre provoquée par la bactérie *Mycobacterium leprae* (Rani et al., 2009). Cependant ce polymorphisme présente un effet protecteur important contre le développement de la tuberculose causée par la bactérie *Mycobacterium tuberculosis* (Gomez et al., 2005).

Un autre polymorphisme du gène PTPN22 avec une fréquence plus importante en Asie qu'en Europe a été mis en évidence. Ce polymorphisme se situe au niveau de la région promotrice du gène PTPN22 où une cytosine en position -1123 est remplacée par une guanine. L'impact du variant PTPN22-C1123T sur la protéine LYP n'est pas encore caractérisé et ne semble pas affecter le niveau d'expression de cette dernière (Kawasaki et al., 2006). Cependant, ce polymorphisme est souvent co-exprimé avec le variant PTPN22 C1858T chez la population caucasienne, où il contribue faiblement au développement de la polyarthrite rhumatoïde (Dieudé et al., 2008). Chez les populations asiatiques, ce polymorphisme est fortement associé à l'arthrite juvénile idiopathique

(Zhi-Dan Fan et al., 2015), à la polyarthrite rhumatoïde (Feng et al., 2010) à la colite ulcéreuse et au LADA (Liu et al., 2012) chez la population chinoise. Ce polymorphisme a également été associé à la maladie de Grave (Ichimura et al., 2008) et à l'apparition aiguë du diabète de type 1 chez la population japonaise et coréenne (Kawasaki et al., 2006). Le variant PTPN22-C1123T augmente aussi le risque du développement de la spondylarthrite ankylosante chez la population Taïwanaise (Huang et al., 2014).

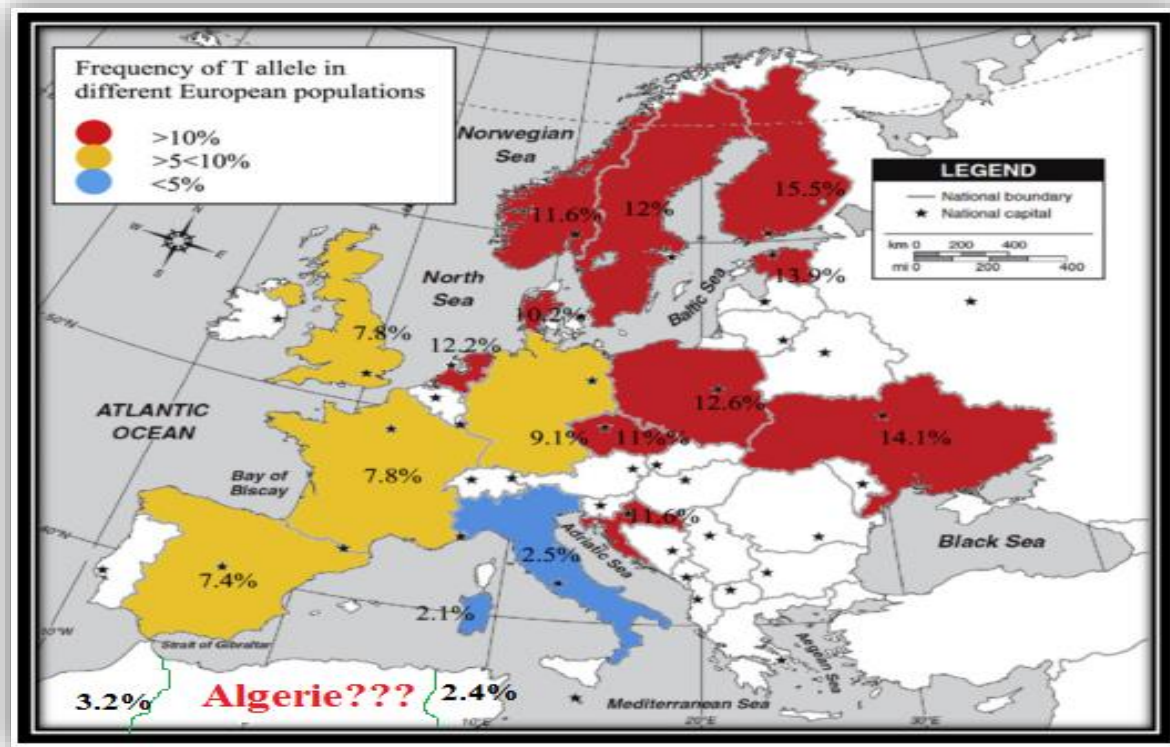
PTPN22 A788G est un troisième polymorphisme mononucéotidique dont la fréquence est inférieure à 7 % chez l'ensemble des populations. Il correspond à la substitution d'une adénine par une guanine en position 788 située au niveau de l'exon 10 du gène PTPN22. Dans la séquence protéique, ce polymorphisme entraîne le remplacement d'une arginine par une glutamine en position 263 (R263Q). Cette transition en acide aminé se situe au niveau du domaine catalytique de la protéine LYP et provoque ainsi une réduction dans son activité phosphatase (Orri et al., 2009). Il a été constaté que ce polymorphisme présente un effet protecteur vis-à-vis du lupus systémique érythémateux (Orri et al., 2009) et de polyarthrite rhumatoïde (Rodriguez et al., 2011) et un effet causal dans la rectocolite ulcéreuse au sein de la population espagnole (Diaz-Gallo et al., 2011). Aucune association n'a été retrouvée entre le variant A788G avec la sclérose en plaque ou avec la maladie de Crohn (Diaz-Gallo et al., 2011). Toutefois, ce polymorphisme peut considérablement augmenter le risque de développement de la tuberculose (Lamsyah et al., 2009).

Plusieurs autres polymorphismes très rares dont la fréquence allélique est inférieure à 1 % ont été mis en évidence tels le variant G2250C qui prédispose, indépendamment de l'allèle C1858T, au diabète de type 1 (Onengut et al., 2006) et le polymorphisme C1108A qui augmente légèrement le risque pour la maladie de Crohn (Rivas et al., 2011).

#### 5.5-Distribution géographique du variant R620W :

La fréquence de l'allèle 1858T du gène PTPN22 varie considérablement d'une population à une autre. En Europe, un gradient de fréquence décroissant nord-sud et est-ouest a été décrit, ce polymorphisme est présent avec une plus grande fréquence en Europe du nord (>10%) avec un pic de fréquence de 15.5% en Finlande et 14.2% en Ukraine. La fréquence du polymorphisme du C1858T est faible en Europe de l'Ouest (7-8%) et encore plus faible en Europe du sud y compris l'Italie et la Sardaigne (<5%), (figure 7). Le variant PTPN22 C1858T est très rare en Afrique Subsaharienne, au Moyen-Orient, chez Amérindiens et chez les populations asiatiques (<1%) (Stanford et Bottini, 2014). Chez les Américains et les Australiens d'origine Européenne le variant C1858T est présent avec une fréquence allélique estimée de 6 à 9 %. En Afrique du nord, la

fréquence de l'allèle PTPN22 1858T se situe à 2.4 % dans la population tunisienne (Zouidi et al., 2014) et à 3.2 % dans la population marocaine (Lamsyah et al., 2009).



**Figure 7 :** Distribution géographique du polymorphisme PTPN22 C1858T (Bottini et Peterson, 2014)

La cause exacte de ces intrigantes différences dans les fréquences alléliques entre les populations demeure encore inconnue. Il a été suggéré que le variant R620W conférerait un avantage de survie, peut être contre les maladies infectieuses comme la tuberculose qui est une maladie fréquente dans les régions froides et qui a entraîné des épidémies graves à travers l'histoire (Gomez et al., 2005).

A decorative graphic of a scroll with a black outline and rounded corners. The scroll is unrolled in the center, with the top and bottom edges curving upwards and downwards respectively. The text is centered within the unrolled portion.

# **Matériels et Méthodes**

Matériels et Méthodes

Ce travail a été réalisé au sein du laboratoire Cytokines et NO-synthase, immunité et pathogénie du Professeur C. Touil-Boukouffa (LBCM-FSB-USTHB), sous la direction du Pr H. Mesbah-Amroun. Notre travail comporte deux parties : une étude statistique et une étude Bio-informatique.

## Matériels

- 1- Les données bibliographiques relatives au gène PTPN22 dans la population Algérienne.
- 2- Les séquences **Q9Y2R2**, **H2PZP7**, **P29352**, correspondant à la séquence de la protéine LYP de l'homme, du chimpanzé et de la souris respectivement, extraites à partir de la banque (Annexe.N°2)

## Méthodes

### Partie I : Etude Statistique

**1- Recueil des données bibliographiques relatives au gène PTPN22** : il est réalisé par interrogation de la banque de données bibliographique PubMed (URL, voir Annexe N°1). La requête utilisée pour extraire les publications a comporté les mots clefs suivants «R620W» AND «PTPN22 gene» AND «Algerian population» AND «rhumatoïd arthritis».

**2-Analyse statistique** : C'est une méta-analyse qui consiste à calculer les fréquences alléliques et génotypiques du variant R620W après cumul des effectifs obtenus à partir des publications. L'analyse de l'association du variant R620W à la PR dans la population Algérienne est également évaluée. Pour cela, ne sont intégrées que les publications ou les études remplissant les 3 critères d'inclusion suivants :

- Etude cas-témoin c.à.d. comportant un groupe témoin et un groupe d'individus atteints de PR dans la population Algérienne
- Etude explorant le même gène PTPN22.
- Etude effectuée sur la même pathologie auto-immune qui est la PR.

La comparaison des fréquences alléliques (allèle sauvage (C) et allèle de risque (T)) et génotypiques du gène PTPN22 entre le groupe de témoins et le groupe de patients a été réalisée par le **test exact de Fisher**. Le calcul des odds Ratio alléliques (OR), avec un intervalle de confiance (IC) de 95% (voir Annexe N°3) a été réalisé en utilisant le logiciel de calculs statistiques **GraphPad Prism7**. La comparaison des moyennes d'âges d'apparition de la PR a été réalisée par le **test de Student**. Les résultats sont considérés comme étant statistiquement significatifs à la valeur seuil de  $p < 0.05$ .

### 2-Partie II : Etude Bio-informatique :

**2.1- Construction de l'arbre phylogénétique** L'analyse phylogénétique de la protéine LYP est réalisée avec l'outil **CLUSTAL OMEGA** (URL, voir Annexe N°1).par construction de l'arbre phylogénétique par la méthode d'alignement multiples des séquences protéiques des trois espèces différentes (homme, chimpanzé, souris).

#### **2.2- Prédiction de propriétés physico-chimiques de la protéine LYP humaine.**

Cette prédiction est réalisée à partir de la séquence primaire en acides aminés de la protéine LYP humaine (numéro d'accèsion **Q9Y2R2**) par utilisation de l'outil **PROTPARAM** (URL, voir Annexe N°1).

#### **2.3- Prédiction des structures 2D et 3D de la protéine LYP humaine**

Les outils **PSIPRED** et **SWISS-MODEL** (URL, Annexe N°1) ont été utilisés pour la prédiction des structures secondaires et 3D respectivement (URL, Annexe N°1).

#### **2.4- Recherche d'une signature peptidique de la protéine LYP humaine**

L'outil **PROSITE** (URL, Annexe N°1) a été utilisé pour rechercher une signature moléculaire ou un motif fonctionnel en accord avec les données connues sur la protéine LYP.





# Résultats et Discussions

## **Partie I : Etude Statistique**

### **1- Recueil des données bibliographiques relatives au gène PTPN22**

La requête utilisée pour extraire de PubMed les publications et comportant les mots clefs suivants «R620W» AND «PTPN22 gene» AND «Algerian population» AND «rhumatoïd arthritis», nous a permis d'extraire une seule publication dont la référence est la suivante:

**- Fodil , M., Benzaoui, A., Zemani - Fodil, F., aberkane, M., Boughrara, W., Saidi-Mehtar, N., Petit-Teixeira E, Boudjema A. (2015). Association of PTPN22 (rs2476601) and STAT4 (rs7574865) polymorphisms with Rheumatoid Arthritis in the Western Algerian population. Acta Rheumatol Port; 40(1):56-62.**

C'est une étude de type Cas-Témoin, réalisée sur un échantillon de 307 sujets représentatifs de la population Oranaise (110 témoins et 197 sujets PR). La fréquence du variant R620W est de 2% dans le groupe témoin et son association à la PR est hautement significative.

Etant donné le faible nombre de références extraites à partir de la base de données PubMed, nous avons décidé d'intégrer à notre étude les résultats issus de deux travaux de thèses soutenus à l'USTHB (Université des Sciences et de la technologie Houari Bomedienne). Il s'agit de :

**- Boukercha, A. "Contribution des polymorphismes des gènes de cytokines, du gene *NOD2* et du gene *PTPN22* dans la susceptibilité et la sévérité au cours des maladies inflammatoires chroniques de l'intestin dans la population Algérienne" (2015). Thèse de doctorat Biotechnologie et pathologies moléculaires, 3<sup>ème</sup> cycle LMD, FSB, USTHB.**

Dans cette étude Cas-Témoin, réalisée sur échantillon de 100 sujets témoins représentatifs de la population Algéroise, l'auteur Boukercha A. a rapporté une fréquence du variant R620W égale à 2% dans la population normale.

**- Louahchi, S. «Etude immunogénétique de la polyarthrite rhumatoïde en Algérie»(2016), Thèse de doctorat Biotechnologie et pathologies moléculaires, 3<sup>ème</sup> cycle LMD:FSB, USTHB.**

Dans cette étude, Cas-Témoin, réalisée sur échantillon de 688 sujets représentatifs de la population Algérienne, l'auteur Louahchi S. a rapporté une fréquence du variant R620W égale à 5.55 % dans le groupe témoin, ainsi que son association non significative à la PR.

Les caractéristiques démographiques des sujets intégrés à l'étude sont représentées dans le tableau1.

**Tableau 1:** Caractéristiques démographiques des sujets intégrés dans la méta-analyse.

Les études	Louahchi, 2016		Fodilet <i>et al.</i> , 2015		Boukercha, 2015
	Témoins	Patients PR	Témoins	Patients PR	Témoins
<b>Effectifs</b>	345	343	197	110	100
<b>Moyenne d'âge</b>	37±11.6	48.4±13	40±22	48±30.5	23.99±9.9
<b>Sex-ratio (H/F)</b>	1/5	1/6	89/108	1/9	68/32

## 2- Analyse statistique

### - Calcul de la fréquence allélique et génotypique du polymorphisme R620W obtenues dans le cadre des trois études sur la population Algérienne :

L'étude de la fréquence du polymorphisme C1858T du gène *PTPN22* dans la population normale est réalisée en calculant les fréquences alléliques et génotypiques après cumul des effectifs obtenus pour trois groupes témoins obtenus à partir des trois études indépendantes : l'étude de (Boukercha A, 2015), l'étude de (Fodil *et al.*, 2015) ainsi que l'étude de (Louahch S, 2016). Les résultats obtenus sont décrits dans le tableau 2.

**Tableau 2:** Fréquences alléliques et génotypiques du polymorphisme R620W du gène *PTPN22* dans la population Algérienne témoin.

Gène (SNP)	Groupes analysés			Fréquences cumulées des 3 études (méta-analyse)
PTPN22 (R620W) Rs2476601 C1858T (C→T)	Témoins (N=100) (Boukercha A, 2015)	Témoins (N=197) (Fodil <i>et al.</i> , 2015)	Témoins (N=306) (Louahchi S, 2016)	Groupe témoin cumulé (N= 603)
Allèles	200	394	612	1206
C	196 (98%)	386 (98%)	578 (94.44%)	1160 (96,2%)
T	4 (2%)	8 (2%)	34 (5,55%)	46 ( <b>3,8%</b> )
Génotypes	100	197	306	603
CC	96 (96%)	189 (96%)	275 (90%)	560(92,9%)
CT	4 (4%)	8 (4%)	28 (9%)	40(6,6%)
TT	0 (0%)	0 (0%)	3 (1%)	3(0,5%)

Les résultats ont montré que l'allèle muté (**T**) est présent dans la population normale avec une fréquence de **3,8%**. Le génotype homozygote sauvage (**CC**) est majoritaire (96%) alors que le génotype homozygote muté (**TT**) est rare (0.5%). En effet, ce génotype n'a été détecté ni dans le groupe témoin de Fodil *et al.*, ni dans celui de Boukercha A. Le génotype hétérozygote muté (**CT**) est retrouvé à une fréquence de 6.6%

**- Evaluation de l'association du variant R620W à la PR :**

L'étude de l'association du polymorphisme C1858T du gène *PTPN22* à la PR dans la population Algérienne est réalisée en calculant les fréquences alléliques et génotypiques après cumul des effectifs obtenus pour seulement les deux études indépendantes : (Fodil *et al.*, 2015) et (Louahch S, 2016), l'étude de (Boukercha , 2015) n'ayant pas porté sur la PR. Les résultats obtenus sont décrits dans le tableau3

**Tableau 3 :** Fréquences alléliques et génotypiques du polymorphisme R620W du gène *PTPN22* au cours de la PR et méta-analyse.

Gène (SNP)	Louahchi et al, 2016				Fodil et al, 2015				Meta-analyse				
	Témoins N=306	Sujets PR N=300	P	OR 95% CI	Sujets Témoins N=197	Sujets PR N=110	P	OR 95% CI	Témoins (306+197=503)	Sujets PR (300+110=410)	P	OR 95% CI	
PTPN22 (R620W) Rs2476601													
C	594 (97%)	576 (96%)	0.314	1.37 (0.76-2.49)	386 (98%)	185 (84%)	<0.0001	9.83 (4.28-22.56)	980 (97.41%)	761 (92.8%)	<0.0001****	2.92 (1.85-4.74)	
T	18 (3%)	24 (4%)			8(2%)	35 (16%)			26 (2.58%)	59 (7.2%)			
CC	275 (90%)	279(93%)	0.169		189(96%)	76(69%)			464 (92.2%)	355 (86.6%)			
CT	28 (9%)	18(6%)			0.143	8(4%)			33(30%)	36 (7.15%)			51 (12.4%)
TT	3 (1%)	3 (1%)			0.981	0 (0%)			1(1%)	3 (0.6%)			4 (1%)
TT+CT	31	21	0.192	0.66 (0.38-1.21)	8(4%)	34(31%)	<0.0001	10.57 (4.74-22.46)	39 (7.75%)	55 (13.4 %)	0.006**	1.84 (1.20-2.84)	
CC	275(90%)	279(93%)			189(96%)	76(69%)							

Les résultats de la méta-analyse ont révélé que l'allèle R620W est significativement associées à la PR (OR= 2.92 ; IC =1.85-4.74 ;  $p^{****}<0.0001$ ). En outre, la méta-analyse démontre que la présence d'une copie ou plus de l'allèle T est significativement différente chez les patients PR et les contrôles (OR=1.84 ; IC=1.20-2.84 ; $p^{**}=0.006$ )

Nous nous sommes intéressés à l'analyse d'un polymorphisme du gène PTPN22. Ce dernier est considéré comme étant le deuxième facteur impliqué dans la susceptibilité à la PR après le locus HLA. Il code pour une tyrosine phosphatase Lyp (Lymphoïde phosphatase), impliquée dans la régulation négative des lymphocytes T via le TCR.

Le gène PTPN22 est situé sur le chromosome 1p13.2. le polymorphisme recherché dans notre étude est un SNP exonique (exon14) ; il correspond au remplacement d'une cystéine (C) par une tyrosine (T) en position 1858 (C1858T), au niveau de la région codante du gène PTPN22 (figure). Cette variation au niveau nucléotidique se traduit au niveau protéique par un changement d'acide aminé, d'une arginine à un tryptophane sur le codon 620 (R620W) au niveau du premier domaine riche en proline de la protéine.

L'association de ce polymorphisme à la PR a été confirmée par des études menées chez plusieurs populations européennes incluant les anglais (Harrison et al.2006), les finlandais (Seldin et al.2005), les suédois (Plenge et al.2005) et les espagnoles (Orozco et al.2005). De même, cette association a été retrouvée chez la population de l'Amérique du Nord incluant les mexicains (Torres-Carrillo et al.2012), les canadiens (Van Oene et al.2005) et les américains (Begovich et al.2004). Egalement, une absence d'association a été notée chez la population Grecque (Plnt et al.2010) et Turque (Sahin et al.2009).

En Tunisie, deux études concernant ce polymorphisme montrent des résultats contradictoires. En effet, une première étude a démontré l'absence d'association de ce SNP à la PR (OR=0.85 ; CI 95% (0.59-1.20) ;P= 0 .33) (Ghanzi Chabchoub et al. 2009). Tandis que la deuxième a rapporté sa forte association à la survenue de la maladie (OR=1.83 ; CI 95% (1.56-2.15) ;P=10<sup>-3</sup>) (Sfar et al.2009).

Le résultat obtenu dans l'étude de Louahchi *et al.*, 2016 a montré l'absence d'une association de ce SNP à la survenue de la PR (OR=1.37 ; CI 95% (.76-2.49) ; P=0.314). Ceci suggère que le polymorphisme rs2476601 PTPN22 pourrait avoir une implication dans le développement de la PR dans notre population (OR=1.37 >1) mais que les résultats n'ont pas atteint un seuil de significativité statistique (P=0.314 >0.05), en raison d'un effectif pas assez élevé. En effet, cette analyse avait une puissance inférieure à 50% pour détecter l'association allélique de ce SNP à la PR (à un niveau de signification < à 0.05). Compte tenu de la fréquence de l'allèle mineur T du variant PTPN22 chez les témoins (3%), à peu près 4200 sujets seraient nécessaires pour détecter un large effet au niveau de ce locus, avec une puissance de 80%, à un p inférieur à 0.05. De ce fait, la réplication de cette étude sur un échantillon d'effectif plus important était nécessaire pour confirmer ce résultat. D'où l'intérêt d'une méta-analyse. L'étude faite par Fodil et al.2015 sur 110 malades PR et 197 sujets sains de la population Algérienne de l'ouest a démontré une très forte association de ce SNP à la survenue de la maladie (OR=9.83 ;CI 95% (4.28-22.56) ; P= 3.32\*10<sup>-11</sup>). A l'issue de notre méta-analyse, nous obtenons les résultats suivants (OR=; 2.92CI 95% (1.85-4.74)) ; P<0.0001). Ceci signifie que pour un effectif cumulé de (306+197=503) témoins et (306+197=503) sujets PR, le variant R620W du gène PTPN22, est retrouvé associé à la PR dans la population Algérienne, avec un seuil de significativité P<0.0001.

Ainsi, la méta-analyse nous a permis de gagner en significativité statistique par rapport aux résultats de Louahchi *et al.*, 2016, puisqu'on est passé de (OR=1.37 ; CI 95% (.76-2.49) ; P=0.314) à (OR=; 2.92CI 95% (1.85-4.74)) ; P<0.0001).

## 2- Partie II : Etude Bio-informatique :

### 2-1. Les résultats obtenus à partir de l'alignement multiples des séquences avec l'outil Clustal Omega

The screenshot shows the Clustal Omega web interface. The browser address bar displays the URL: [www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobid=clustalo-I20170412-135714-0302-26195653-pg&tool=clustalo&showColors=true](http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobid=clustalo-I20170412-135714-0302-26195653-pg&tool=clustalo&showColors=true). The page title is "Results for job clustalo-I20170412-135714-0302-26195653-pg". Below the title, there are tabs for "Alignments", "Result Summary", "Phylogenetic Tree", and "Submission Details". The "Alignments" tab is selected. The main content area shows a multiple sequence alignment of protein tyrosine phosphatase (PTPN22) from various species. The alignment is displayed in a color-coded format, with conserved residues highlighted in red and yellow. The species listed include: AB04841ERS | AB04841ERS\_DAIRES, F61631 | F61631\_MONDO, F77716 | F77716\_MACIU, E99110 | E99110\_HUMAN, F54258 | F54258\_HUMAN, G30874 | G30874\_HUMAN, AB04841TES | AB04841TES\_HUMAN, H22P27 | H22P27\_PANTR, Q9Y2R2 | P1N22\_HUMAN, AB04841S7 | AB04841S7\_HUMAN, D442D5 | D442D5\_RAT, P293S2 | P1N22\_MOUSE, and E9Q4S3 | E9Q4S3\_MOUSE. The alignment shows conserved residues across all species, with gaps indicated by dashes.

The screenshot shows the Clustal Omega web interface, displaying a different region of the multiple sequence alignment for protein tyrosine phosphatase (PTPN22). The browser address bar displays the URL: [www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobid=clustalo-I20170412-135714-0302-26195653-pg&tool=clustalo&showColors=true](http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobid=clustalo-I20170412-135714-0302-26195653-pg&tool=clustalo&showColors=true). The page title is "Results for job clustalo-I20170412-135714-0302-26195653-pg". Below the title, there are tabs for "Alignments", "Result Summary", "Phylogenetic Tree", and "Submission Details". The "Alignments" tab is selected. The main content area shows a multiple sequence alignment of protein tyrosine phosphatase (PTPN22) from various species. The alignment is displayed in a color-coded format, with conserved residues highlighted in red and yellow. The species listed include: AB04841ERS | AB04841ERS\_DAIRES, F61631 | F61631\_MONDO, F77716 | F77716\_MACIU, E99110 | E99110\_HUMAN, F54258 | F54258\_HUMAN, G30874 | G30874\_HUMAN, AB04841TES | AB04841TES\_HUMAN, H22P27 | H22P27\_PANTR, Q9Y2R2 | P1N22\_HUMAN, AB04841S7 | AB04841S7\_HUMAN, D442D5 | D442D5\_RAT, P293S2 | P1N22\_MOUSE, and E9Q4S3 | E9Q4S3\_MOUSE. The alignment shows conserved residues across all species, with gaps indicated by dashes.

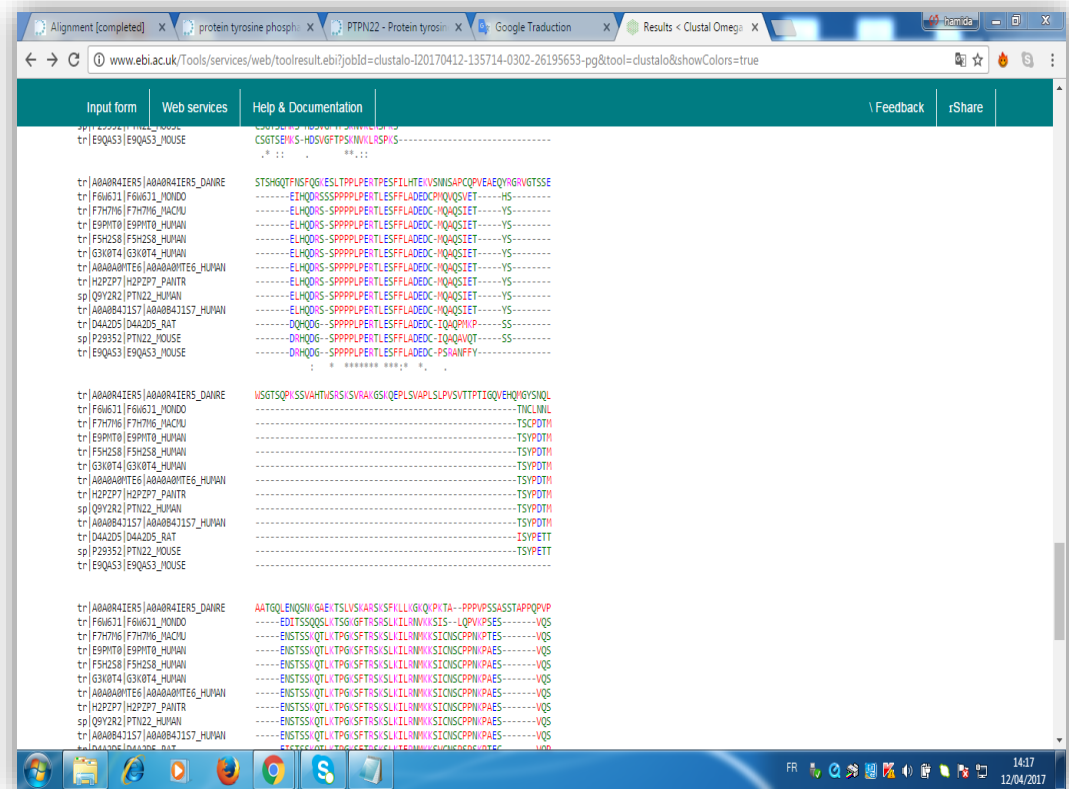
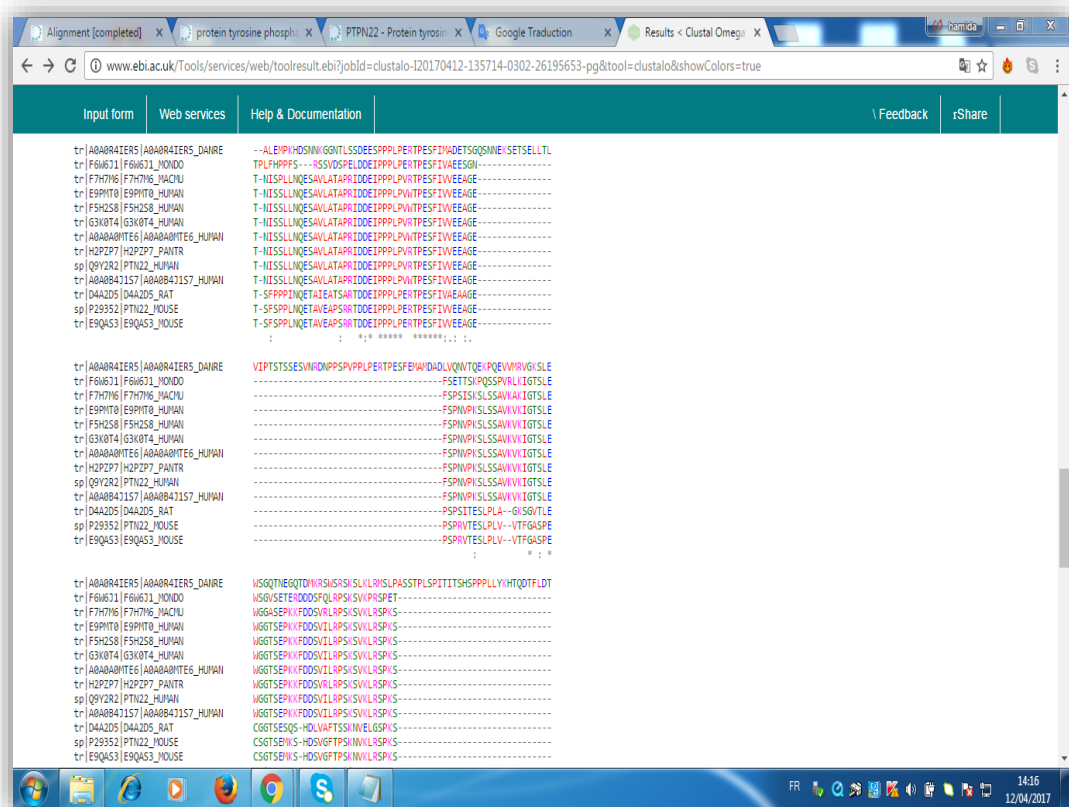
Captures d'écran de résultats obtenus par l'outil clustal omega (l'alignement multiples).

# Les résultats obtenus à partir de l'alignement multiples des séquences avec l'outil Clustal Omega



Captures d'écran de résultats obtenus par l'outil clustal omega (l'alignement multiples).

# Les résultats obtenus à partir de l'alignement multiples des séquences avec l'outil Clustal Omega



Captures d'écran de résultats obtenus par l'outil clustal omega (l'alignement multiples).

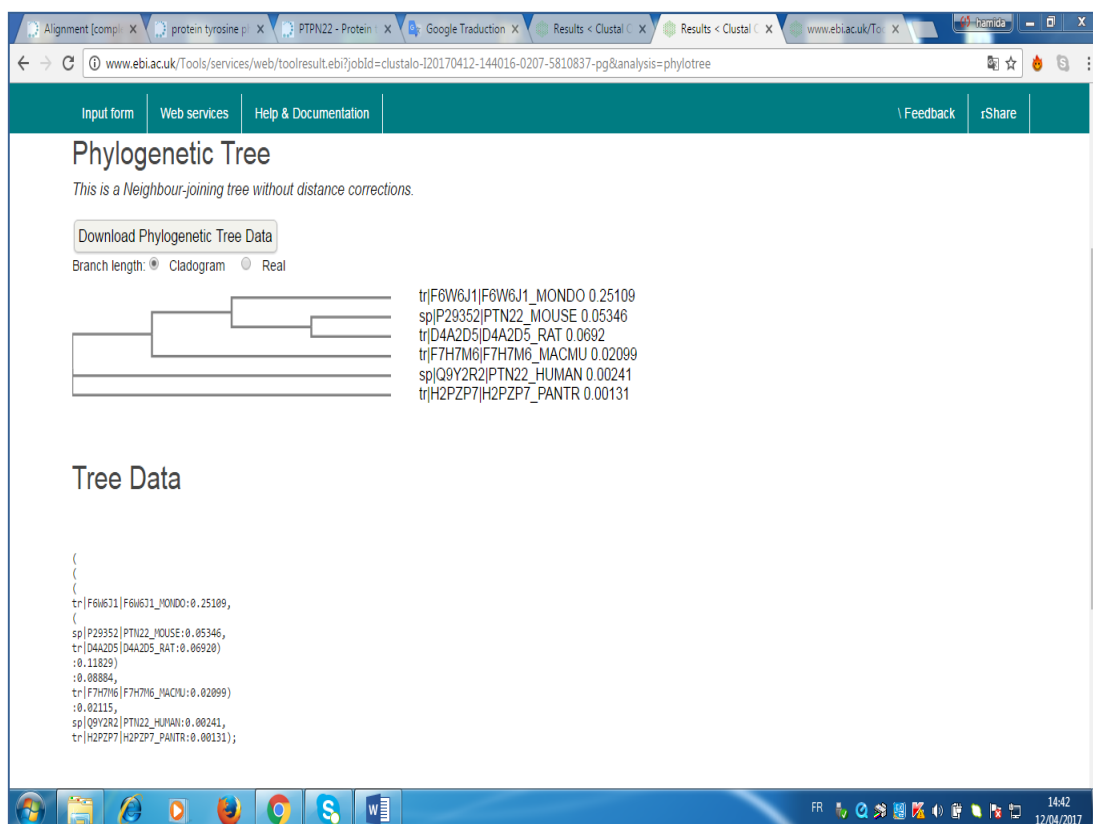
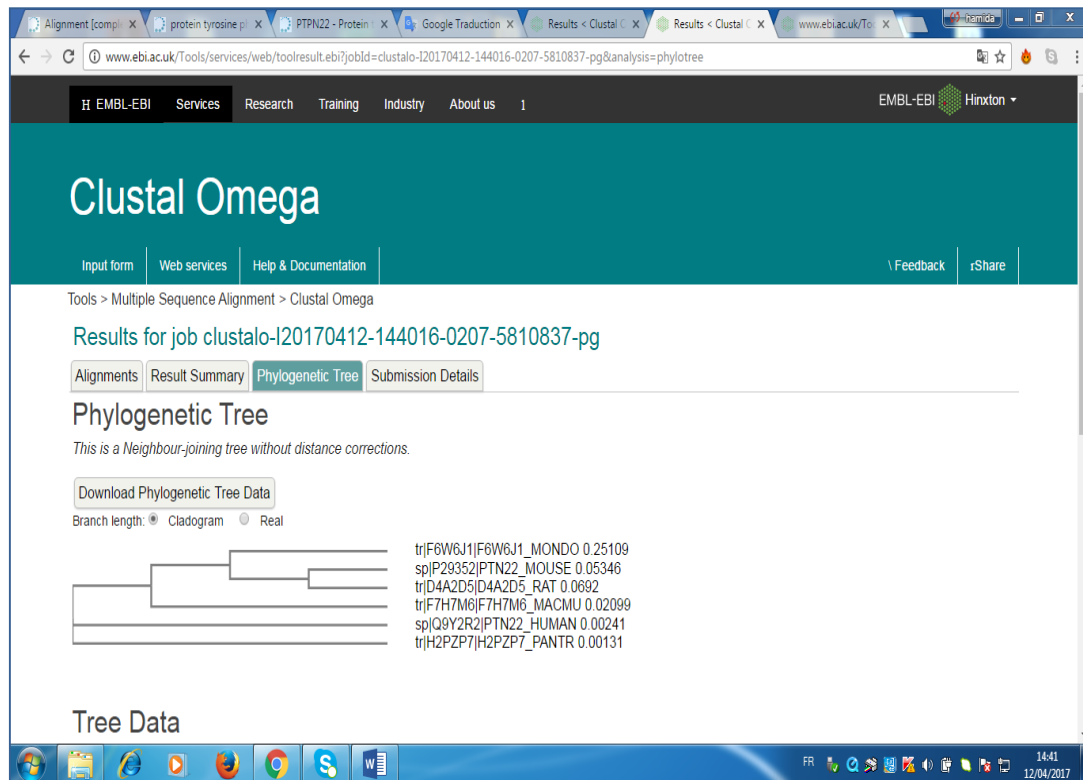


## Les résultats obtenus à partir de l'alignement multiples des séquences avec l'outil Clustal Omega

The screenshot displays the Clustal Omega web interface. The browser address bar shows the URL: [www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-120170412-135714-0302-26195653-pg&tool=clustalo&showColors=true](http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-120170412-135714-0302-26195653-pg&tool=clustalo&showColors=true). The page features a navigation bar with 'Input form', 'Web services', 'Help & Documentation', 'Feedback', and 'rShare'. The main content area shows two blocks of multiple sequence alignment. The first block is a full-length alignment of the protein tyrosine phosphatase (PTN22) across 15 species, with the first sequence (tr|A0A0R4IERS|A0A0R4IERS\_DAIRES) highlighted in red. The second block is a zoomed-in view of a specific region, showing conserved motifs like 'TYGPGVLLFSGFTNFGIP(G)PNI(PET)N-' and 'NHSNSFLNFGAINFSP(G)P(N)PPTINII' across the same species. A note below the alignments states: 'PLEASE NOTE: Showing colors on large alignments is slow.' The footer includes the EMBL-EBI logo and navigation links for 'Services', 'Research', 'Training', 'Industry', and 'About EMBL-EBI'. The system tray at the bottom shows the date and time as 14:17 on 12/04/2017.

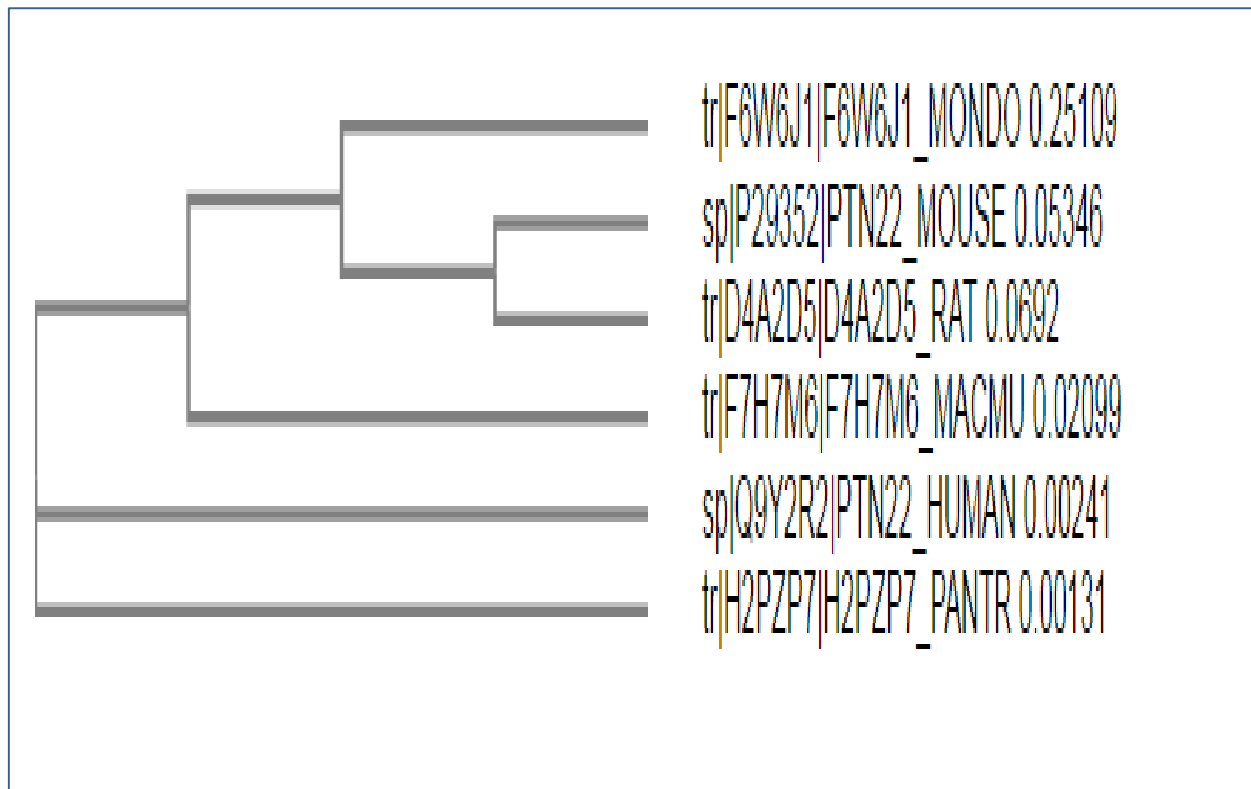
Captures d'écran de résultats obtenus par l'outil clustal omega (l'alignement multiples ).

## - Construction de l'arbre phylogénétique avec l'outil Clustal Omega



Captures d'écran de résultats obtenus par l'outil clustal omega (l'arbre phylogénétique).

Arbre phylogénétique obtenu avec l'outil Clustal Omega (aggrandissement)



**Captures d'écran de résultats obtenus par l'outil clustal omega (l'arbre phylogénétique).**

Cette analyse bio-informatique avec l'outil Clustal omega nous permet d'appréhender l'importance du rôle biologique de la protéine LYP. En effet, les résultats obtenus indiquent que les séquences de la protéine LYP de l'homme et du chimpanzé (Q9Y2R2-HUMAN et H2PZP7-PANTR, respectivement) sont identiques à 100%. Ceci signifie que la structure biochimique de la protéine LYP et donc sa fonction biologique sont hautement conservées chez les primates.

## 2.2- Les résultats obtenus avec l'Outils protparam

**Outil ProtParam**

ProtParam ( [Références / Documentation](#) ) est un outil qui permet de calculer divers paramètres physiques et chimiques pour une protéine donnée stockée dans [Swiss-Prot](#) ou [TrEMBL](#) ou pour une séquence protéique entrée par l'utilisateur . Les paramètres calculés comprennent le poids moléculaire, la pI théorique, la composition d'acides aminés, la composition atomique, le coefficient d'extinction, la demi-vie estimée, l'indice d'instabilité, l'indice aliphatique et la grande moyenne d'hydropathicité (GRAVY) ( [Disclaimer](#) ).

Veillez noter que vous ne pouvez remplir que l' **un** des champs suivants à la fois.

Entrez un numéro d'accèsion (AC) de Swiss-Prot / TrEMBL (par exemple **P05130** ) ou un identificateur de séquence (ID) (par exemple **KPC1\_DROME** ) :

Ou vous pouvez coller votre propre séquence d'acides aminés (dans le code d'une lettre) dans la case ci-dessous:

```

TCAACTCTAATTTTATGTAGTAATAAATGGCAGGTAATTGTTTTACAAGAATCCAC
CTGACTTCCC
CTAATGCATTAAAAATATTTTATTTAAATAACTTTATTATAACTTTTGAAGAACATGTA
GTATTGTTTA
AACATCATTGTTCTTCAGTATTTTTCATTTGGAAGTCCAATAGGGCAAATGAATGAAG
TATTATTATC
TGTCCTTGAGTACAATGTATCCAACAGCACTCAATAAATTTTGGTTGTTAAACTG
    
```

**Genomic context**

Location: 1p13.2 [See PTPN22 in Genome Data Viewer](#) [Map Viewer](#)

Exon count: 24

Annotation release	Status	Assembly	Chr	Location
<a href="#">108</a>	current	GRCh38.p7 ( <a href="#">GCF_000001405.33</a> )	1	NC_000001.11 (113813811..113871761, complement)
<a href="#">105</a>	previous assembly	GRCh37.p13 ( <a href="#">GCF_000001405.25</a> )	1	NC_000001.10 (114356433..114414375, complement)

**Chromosome 1 - NC\_000001.11**

Genomic map showing genes: RPS2P14, PHTF1, RSN1, RP4B1-RS1, PTPN22, BCL2L15, RP4B1, and DCLRE1B.

**Genomic regions, transcripts, and products**

Go to [reference sequence details](#)

Genomic Sequence: [NC\\_000001.11 Chromosome 1 Reference GRCh38.p7 Primary Assembly](#)

Captures d'écran de résultats obtenus par l'outil PROTPARAM

## 2.2- Les résultats obtenus avec l'Outils protparam (suite)

https://www.ncbi.nlm.nih.gov/gene/26191

Cette page en **anglais** a été traduite en **français** Afficher l'original

**Genomic regions, transcripts, and products**

Go to [reference sequence details](#)

Genomic Sequence: **NC\_000001.11** Chromosome 1 Reference GRCh38.p7 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

**Expression**

HPA RNA-seq normal tissues

- Project title: HPA RNA-seq normal tissues
- Description: RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes
- BioProject: [PRJEB4337](#)
- Publication: [PMID 24309898](#)
- Analysis date: Wed Jun 15 11:32:44 2016

**Bibliography**

Related articles in PubMed

- [Suppression of protein tyrosine phosphatase PTPN22 gene induces apoptosis in T-cell leukemia cell line \(Jurkat\)](#)

BioAssay, by Gene target  
BioAssays, RNAi Target, Active  
BioAssays, RNAi Target, Tested  
BioProjects  
BioSystems  
CCDS  
ClinVar  
Conserved Domains  
dbVar  
EST  
Full text in PMC  
Full text in PMC\_nucleotide  
Functional Class  
Gene neighbors  
Genes with a similar H3K4me3 profile  
Genome

https://www.ncbi.nlm.nih.gov/nucleotide/AF001846

REFERENCE 1 (bases 1 to 3058)  
AUTHORS Cohen,S., Dadi,H., Shaoul,E., Sharfe,N. and Roifman,C.M.  
TITLE Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp  
JOURNAL Blood 93 (6), 2013-2024 (1999)  
PUBMED [10068674](#)

REFERENCE 2 (bases 1 to 3058)  
AUTHORS Roifman,C.M.  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-1997) Immunology and Allergy, The Hospital For Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

FEATURES

source Location/Qualifiers  
1..3058  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_type="thymocytes"

CDS  
42..2468  
/codon\_start=1  
/product="lymphoid phosphatase LyP1"  
/protein\_id="AAD00904.1"  
/translation="MDQREILQKFLDEAQS...  
FTTVAENAKNIKKNRYKIDILPYDYSRVLSLITSDSSYINANFIRKGVYGPAYIAT  
QGPLSTLLDFWRMIWEYSVLIIVMACMEYEMGKKKERYWAEFGEMQLEFGPFSVSC  
EAEKRSYDIIRTLKVKFNSETRTIYQFHYKNWPDHDPVSSIDPILELINDVRCYQED  
DSVPICIHCSAGCGRTGIVCAIVDYTWMLLDGIIIPENFSVFSLIREMRTQRPFLVQT  
QEQVELVYNVAVLELFRQMDVIRDKHSGETESQAKHCIPEKNHTLQADSYSNLPKSTT  
KAARKMNQQRKMEIKESSEDFRTSEISAKEELVLHPAKSSTISDFLELNYSFDKNS  
DITMKWQTKAFPIVGEPLQKHQSLDLGSLLEFGCSNSKFPVNAAGRYFNSKVPITRTKS  
TFELIOORETKEVDSKENFSYLESOEHDSCFVEMOAKVMHVSSAELNYSLPYDSKH

Articles about the PTPN22 gene  
Suppression of protein tyrosine phosphatase [Biomed Pharmacother. 2017]  
Associations of single nucleotide polymorphisms of P [Hum Immunol. 2017]  
The role of human leukocyte antigen DRB1-DQB1 haplotypes [Hum Immunol. 2017]  
See all...

Pathways for the PTPN22 gene  
Translocation of ZAP-70 to Immunological synapse  
Phosphorylation of CD3 and TCR zeta chains  
TCR signaling  
See all...

Reference sequence information  
RefSeq alternative splicing

Captures d'écran de résultats obtenus par l'outil PROTPARAM

## 2.2- Les résultats obtenus avec l'Outils protparam (suite)

Reference sequence information  
RefSeq alternative splicing  
See 11 reference mRNA sequence splice variants for the PTPN22 gene.

More about the PTPN22 gene  
This gene encodes of member of the non-receptor class 4 subfamily of the protein-tyrosine phosphatase family. The encoded protein is a lymph...  
Also Known As: LYP, LYP1, LYP2, PEP, P...

Homologs of the PTPN22 gene  
The PTPN22 gene is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, rat, chicken, and frog.

Related information  
BioSystems

ORIGIN

1	tcocctcaacc	tactttataga	ctatTTTTtct	tgctotgog	catggacca	agagaaattc
61	tgcgaagatt	cctggatgag	gcccaaaaga	agaaaaattc	taaaaggagg	tttgccaattg
121	aattctgaa	gctgaaaagg	caatctacca	agtacaaggc	agacaaaacc	tatcctacaa
181	ctgtgctga	gaatgccaag	aatatcaaga	aaaacagata	taaggatatt	ttgcoctatg
241	attatagccg	ggtagaacta	tcocctgata	cctctatgta	ggattccacc	tacatcaattg
301	ccaacttoat	taaggagatt	tatggacca	aggottatat	tgccaccacc	ggtcctttat
361	ctacaacctc	cctggacttc	tgaggatga	tttgggaata	tagtctcctt	atcattgtta
421	tggcactgat	ggagtatgaa	atgggaaaga	aaaagtgtga	gogctactgg	gctgagccag
481	gagagatgca	gctggaattt	ggcctctct	ctgtatcctg	tgaagctgaa	aaaaggaaat
541	ctgattatat	aatcaggact	ctaaaagtta	agttcaatag	tgaactctga	actatctacc
601	agtttccatta	caagaattgg	ccagaccatg	atgtacottc	atctatagac	cctattcttg
661	agctcatctg	ggatgtaagt	tgttaccaag	aggatgacag	tgttcccata	tgcattccat
721	cgactgctgg	ctgtggaagg	actggtgta	tttgtctat	tggtgattat	acatggatgt
781	tgctaaaaga	tgggataatt	cctgagaact	tcagtgtttt	cagtttgatc	cgggaaatgc
841	ggacacagag	gocctcatta	gttcaaacgc	aggaaacaata	tgaactgtgc	tacaatgtg
901	tattagaact	attaagaga	cagatggatg	ttatcagaga	taaacattct	ggaacagaga
961	gtcaagcaaa	gcattgtatt	cctgagaaaa	atcaactctc	ccaagcagac	tctattcttc
1021	ctaatttacc	aaaaatgacc	acaaaagcag	caaaaatgat	gaaccaacaa	agcacaacaa

Nombre d'acides aminés: 808

Poids moléculaire : 91763.83

PI théorique: 7.20

Composition d'acides aminés:  Ala (A) 37 4,6%

Arg (R)	32	4,0%
Asn (N)	38	4,7%
Asp (D)	40	5,0%
Cys (C)	14	1,7%
Gln (Q)	33	4,1%
Glu (E)	59	7,3%
Gly (G)	25	3,1%
His (H)	15	1,9%
Ile (I)	43	5,3%
Leu (L)	62	7,7%
Lys (K)	67	8,3%
Met (M)	19	2,4%
Phe (F)	36	4,5%
Pro (P)	57	7,1%
Ser (S)	103	12,7%
Thr (T)	48	5,9%
Trp (W)	10	1,2%
Tyr (Y)	31	3,8%
Val (V)	39	4,8%
Pyle (O)	0	0,0%
Sec (U)	0	0,0%

Captures d'écran de résultats obtenus par l'outil PROTPARAM

## 2.2- Les résultats obtenus avec l'Outils protparam (suite)

(B) 0 0,0%  
(Z) 0 0,0%  
(X) 0 0,0%

Nombre total de résidus chargés négativement (Asp + Glu): 99  
Nombre total de résidus chargés positivement (Arg + Lys): 99

Composition atomique:

Carbon C 4078  
Hydrogène H 6360  
Nitrogène N 1082  
Oxygène O 1260  
Soufre S 33

Formule: C 4078 H 6360 N 1082 O 1260 S 33  
Nombre total d'atomes: 12813

Coefficients d'extinction:

Les coefficients d'extinction sont en unités de  $M^{-1} \text{ cm}^{-1}$ , à 280 nm mesurée dans l'eau.

Ext. Coefficient 102065  
Abs 0,1% (= 1 g / l) 1,112, en supposant que toutes les paires de résidus Cys forment des cystines

Ext. Coefficient 101190  
Abs 0,1% (= 1 g / l) 1,103, en supposant que tous les résidus Cys sont réduits

Les coefficients d'extinction sont en unités de  $M^{-1} \text{ cm}^{-1}$ , à 280 nm mesurée dans l'eau.

Ext. Coefficient 102065  
Abs 0,1% (= 1 g / l) 1,112, en supposant que toutes les paires de résidus Cys forment des cystines

Ext. Coefficient 101190  
Abs 0,1% (= 1 g / l) 1,103, en supposant que tous les résidus Cys sont réduits

La demi-vie estimée:

Le N-terminal de la séquence considérée est M (Met).

La demi-vie estimée est: 30 heures (réticulocytes mammifères, in vitro).  
> 20 heures (levure, in vivo).  
> 10 heures (Escherichia coli, in vivo).

Indice d'instabilité:

L'indice d'instabilité (II) est calculé comme étant 47,77  
Cela classe la protéine comme instable.

Indice aliphatique: 69,26

Grande moyenne d'hydropathicité (GRAVY): -0,598

Captures d'écran de résultats obtenus par l'outil PROTPARAM

Selon l'outil **PROTPARAM**, nous avons pu avoir les prédictions des propriétés physico-chimiques suivantes:

-Nombre d'acide aminés : 808 aa

-Poids moléculaire : 91763,83 daltons

-Nombres de résidus chargés négativement (asp+glu) :99

-Nombre total de résidus chargés positivement (arg+Lys) :99

-Nombre d'atome : 12813

-Coefficients d'extinction : abs = 0 ,1%

- Estimation de la demi-vie estimée : La demi-vie est une prédiction du temps qu'il faut pour que la moitié de la quantité de protéine dans une cellule disparaisse après sa synthèse dans la cellule.

PROTPARAM s'appuie sur « la règle du N-terminal », qui relie la demi-vie d'une protéine à l'identité de son résidu N-terminal. La prédiction est donnée pour 3 organismes modèles (humain, levure et E.coli). La règle du N-terminal provient des observations que l'identité du résidu N-terminal d'une protéine joue un rôle important dans la détermination de sa stabilité in vivo. Le N-terminal de la séquence considérée est la Méthionine (Met).

La demi-vie estimée est : 30 heures (réticulocytes mammifères, in vitro).

>20 heures (levure, in vivo).

>10 heures (Escherichia coli, in vivo)

-Indice d'instabilité : Une protéine dont l'indice d'instabilité (II) est inférieur à 40 est prédite comme stable, une valeur supérieure à 40 prédit que la protéine peut être instable.

L'indice d'instabilité est calculé comme étant égal à 47,77, ce qui classe la protéine LYP comme étant instable.

- L'Indice aliphatique :69,26

L'indice aliphatique d'une protéine est défini comme le volume relatif occupé par les chaînes latérales aliphatiques (alanine, valine, isoleucine et leucine). Il peut être considéré comme un facteur positif pour l'augmentation de la thermostabilité des protéines globulaires. L'indice aliphatique d'une protéine est calculé selon la formule suivante

Indice aliphatique = X (Ala) + a \* X (Val) + b \* (X (Ile) + X (Leu))

Où X (Ala), X (Val), X (Ile) et X (Leu) sont en pourcentage en moles (fraction molaire de 100 X) d'alanine, de valine, d'isoleucine et de leucine. Les coefficients a et b sont le volume relatif de la chaîne latérale valine (a = 2,9) et Leu/Ile (b = 3.9).

-Indice d'hydropathie (GRAVY) : -0,598.

L'indice GRAVY (Grand average of hydropathicity) ou indice d'hydropathie, représente le profil d'hydropathie, autrement dit l'affinité de la protéine pour l'eau (hydrophilie ou hydrophobicité).



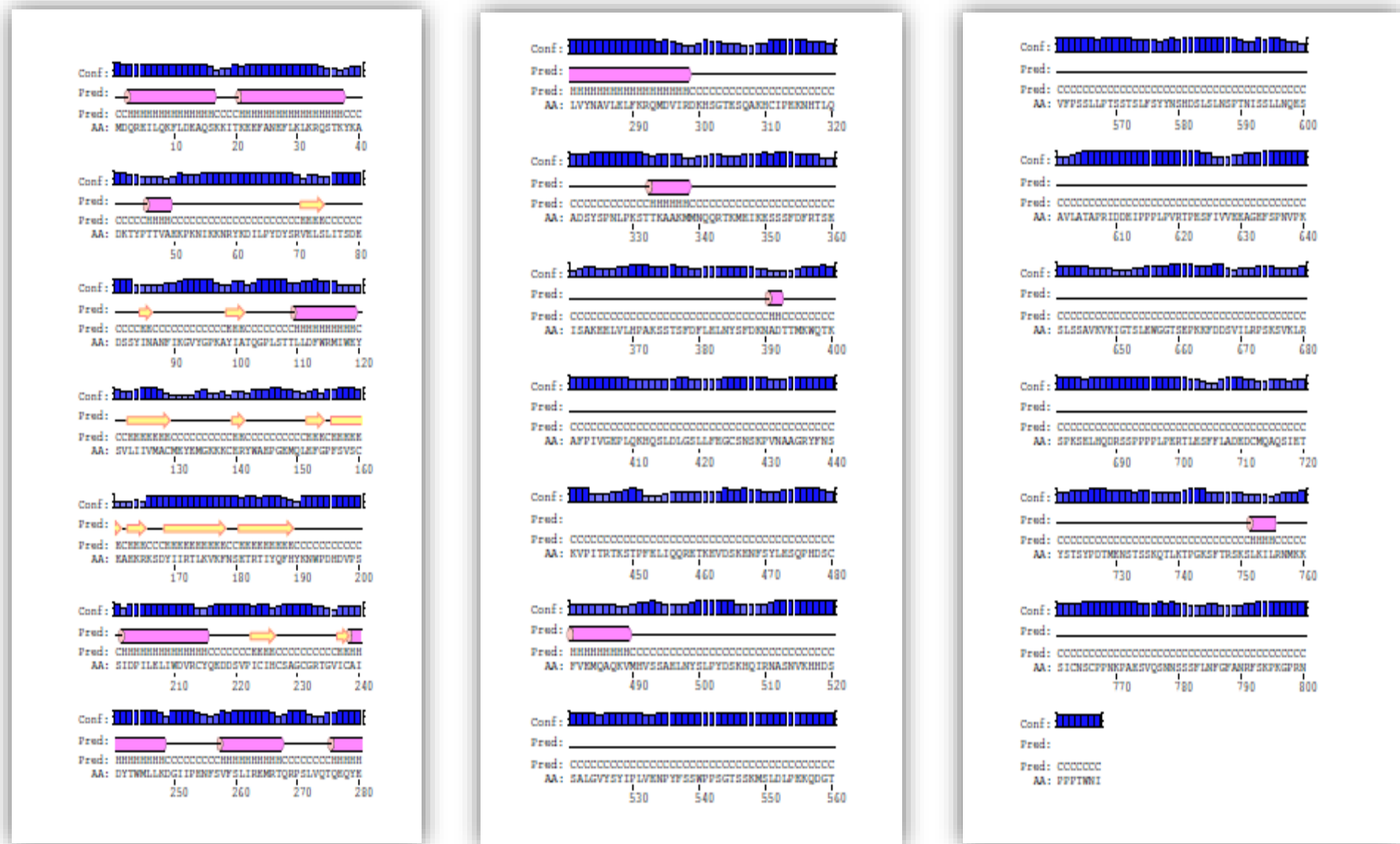
L'indice GRAVY est calculé selon l'échelle de Kyte et Doolittle (Kyte and Doolittle, 1982). Si la valeur est positive, la protéine est majoritairement hydrophobe. Si la valeur est négative la protéine est majoritairement hydrophile. La moyenne de l'index Gravy pour les protéines solubles doit être autour de -0,3.

La séquence de la protéine LYP donne un index Gravy de -0,598, ce qui est plutôt élevé pour une protéine soluble. Ce résultat classe la protéine LYP parmi les plus hydrophiles que la moyenne, ce qui est en accord avec sa localisation cytosolique.

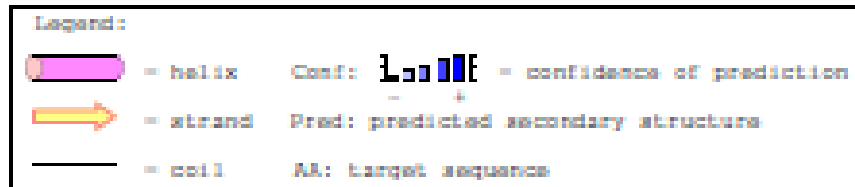
Par comparaison, la gp160 du HIV (la protéine d'enveloppe du HIV) a un index de -0,24 ce qui la classerait comme "moins hydrophile" que la protéine LYP, ce qui est normal puisque c'est une protéine membranaire qui est insoluble dans l'eau du fait de son domaine transmembranaire.

Les repliements locaux ou structures secondaires que prennent les acides aminés dans une protéine sont de trois grandes catégories : les hélices  $\alpha$ , qui établissent des liaisons hydrogène avec l'eau, les feuillets  $\beta$  à caractère hydrophobe ainsi que les boucles ou coudes (coil) qui provoquent des changements de direction de la chaîne peptidique.

### 2.3-Les Résultats obtenus à partir de PSIPRED



Captures d'écran de résultats obtenus par l'outil PSIPRED



### Légende et clés de lecture des résultats obtenus par l'outil PSIPRED

Nous remarquons que l'analyse de la structure 2D par l'outil **PSIPRED** a donné une prédiction de structure hélices  $\alpha$  (helix) majoritaires par rapport aux feuillets  $\beta$  (strands). La protéine LYP a un caractère hydrophile à la surface qui est en contact avec le cytosol et un cœur à l'intérieur de la protéine, qui aurait un caractère hydrophobe composé de feuillets  $\beta$ . De plus, nous pouvons dire que les résultats obtenus avec l'outil PSIPRED vont dans le même sens que ceux obtenus avec PROTPARAM, et sont en accord avec la localisation cytosolique de la protéine LYP.

## 2.4- Les résultats obtenus avec l'outil PROSITE

prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=98143312835.scan.gz

Cette page en anglais a été traduite en français. Afficher l'original

ExPASy Bioinformatics Resource Portal PROSITE Accueil | Contact

Accueil | ScanProsites | ProRule | Documents | Téléchargements | Des liens | Financement

### prosite ScanProsites Results Viewer

Format Output: vue graphique - cette vue affiche les résultats de ScanProsites avec les fonctionnalités intra-domaine prédites selon ProRule [ aide ].

**Hits pour tous les motifs PROSITE (version 2017\_05) sur la séquence sp-Q9Y2R2-PTN22\_HUMAN:**

Trouvé: 3 hits en 1 séquence

Sp-Q9Y2R2-PTN22\_HUMAN (807 aa)

```

MDQREILQKFLDEAQS KKIITKEEFANEFLKLRQSTKYKADKTYPTTVAEKPKNIKKNRYKDILPY
DYSRVELSLITSDSDSSYINANFIKGVYGPKAYIATQGPLSTILLDFWRMIWEYSVLIIVMACMEY
EMGKKKCCERYWAEPEGEMQLFEGPFSVSCAEKRKSDYIIRTLKVKFNSETRTIYQPHYKNWPDHDV
PSSIDPILELILWVRCYQEDDSVPICIHCSAGCGRTGVICAIIDYTWMLLKDGIIPENFSVFSLEIE
MRTIQRPSSLVQIQEQYE L VYNVAVLELFKRQMDVIRDKHSGTESQAKHCIPKNNHTLQADSYSNLPK
STTKAAKMMNQRTKMEIKESSSDFRITSEISAKEELVLHPAKSSTSFDFLELNYSFDKNADITMK
WQTKAFPIVGEPLQKHQSLDLDLGLLFEFGCSNSKPVNAAGRYFNSKVPITRTKSTPFELIQQREITKE
VDSKENEFSYLFESORHDSCEVEMOQKVMHVSSEAFINYSIRYDSZKHOTRNASNVKHHDSALIGVYSY
    
```

prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=98143312835.scan.gz

```

LLNQESAVLATAPRIDDEIPPLPVRTPEFSFIVVEEAGEFSPNVPKLSLSAVKVKIGTSLWGGTIS
EPKKFDDSVILRPSKSVKLRSPKSELHQDRSSPPPLPERTLESFFLADEDCMQAQSIETYSTSY
DTMENSTSSKQTLKTPGKSFTRSKSLKILRNMKKSICNSCPPNPKPAESVQSNSSSFLNFGFANRF
SKPKGPRNPPPTWNI
    
```

**Legend:**

- disulfide bridge
- active site
- other 'ranges'
- other sites

Veuillez noter que les représentations graphiques des domaines présentés ci-après sont uniquement à titre d'illustration et que leurs couleurs et leurs formes ne sont pas destinées à indiquer l'homologie ou la fonction partagée. Pour plus d'informations sur la façon dont ces représentations graphiques sont construites, allez à <http://prosite.expasy.org/mydomains/>.

**Hits par profils: [2 hits (par 2 profils distincts) sur 1 séquence]**

La majuscule représente les positions de correspondance, les positions d'insertion des minuscules et le symbole '-' représente les suppressions relatives au profil correspondant.

règle: 1 100 200 300 400 500 600 700 800 900 1000

Sp-Q9Y2R2-PTN22\_HUMAN (807 aa)

TYR\_PHOSPHATASE\_PTP

TYR\_PAT

Captures d'écran de résultats obtenus par l'outil PROSITE

**SCANPROSITE** est un outil utilisé pour identifier ou prédire la fonction biologique de toute nouvelle séquence protéique inconnue, traduite à partir de séquences potentiellement codante enregistrées dans les banques de séquences (séquences génomiques, EST ou ADNc). L'outil PROSITE se compose d'une base de données PROSITE dédiée aux motifs fonctionnels et signatures peptidiques associées à des fonctions biologiques connues et significativement formulées, de telle sorte qu'avec des outils informatiques appropriés, tels que la méthode de comparaison par alignement de séquences, il est possible d'identifier rapidement et de manière fiable à quelle famille connue de protéines (le cas échéant) appartient la séquence requête.

L'analyse par SCANPROSITE conduit à trois résultats possibles obtenus par alignement:

- TYR-PHOSPHATASE-PTP profil de la famille des protéines du type PTP :  
Domaine a.a. [24-289] avec un score d'alignement de : 52.789  
Caractéristiques prédites : Domaine [24-289] tyrosine-protéine phosphatase  
Site Actif en position [227] correspondant à une activité phosphocystéine intermédiaire
- TYR-PHOSPHATASE-2 profil de la famille des protéines phosphatase spécifique de la tyrosine :  
Domaine a.a. [203-280] avec un score d'alignement de : 18.278
- TYR-PHOSPHATASE-1 site actif des protéines phosphatase spécifique de la tyrosine :  
Domaine a.a [225-235]  
Caractéristique prédites : site Actif en position [227] correspondant à une activité phosphocystéine intermédiaire

Pour rappel, les protéines phosphatases spécifiques de la tyrosine (EC 3.1.3.48) (PTPase) (1,2,3,4,5) sont des enzymes qui catalysent l'élimination d'un groupe phosphate attaché à un résidu tyrosine. Ces enzymes sont très importantes dans le contrôle de la croissance cellulaire de la prolifération, de la différenciation et de la transformation. Les formes multiples de PTPase ont été caractérisées et peuvent être classées en deux catégories. Les PTPase solubles et les protéines réceptrices transmembranaires qui contiennent des domaines PTPase.

D'après les résultats obtenus par PROSITE, nous retrouvons que LYP est une protéine soluble ayant un site actif en position [227] avec une activité enzymatique de type protéines phosphatase spécifique de la tyrosine (PTPase) et une activité phosphocystéine intermédiaire.

## 2.5- Les résultats obtenus avec l'outil SWISS-MODEL

**PV - JavaScript Protein Viewer**

**Sur**

PV est un visualiseur JavaScript pour visualiser les structures protéiques directement dans les navigateurs. Il est super-rapide et facile à intégrer à votre site Web et ne nécessite aucun plug-ins à installer.

PV est la protéine-visualiseur par défaut disponible dans **SWISS-MODEL**.

**Voir en action**

Intéressé de voir ce que PV peut faire pour vous? Ensuite, regardez cette page de démonstration simple disponible [ici](#).

**C'est Open Source**

<https://swissmodel.expasy.org/interactive/rBXpW/templates/>

Accession	Protein Name	Progress	Resolution (Å)	Monomer	Count
4j51.1.A	Tyrosine-protein phosphatase non-receptor type 22	100.00	2.3Å	monomer	1 x N75 <sup>cf</sup>
3h2x.1.A	Tyrosine-protein phosphatase non-receptor type 22	100.00	2.2Å	monomer	None
2p6x.1.A	Tyrosine-protein phosphatase non-receptor type 22	100.00	1.9Å	monomer	None
2p6x.2.A	Tyrosine-protein phosphatase non-receptor type 22	100.00	1.9Å	monomer	None
1ygu.1.A	Leukocyte common antigen	31.15	2.9Å	monomer	1 x PRO-THR-PTR-SER <sup>cf</sup>
1ygu.2.A	Leukocyte common antigen	31.15	2.9Å	monomer	1 x PRO-THR-PTR-

3brh.2.A

3brh.2.A

3brh.1.A

Captures d'écran de résultats obtenus par l'outil SWISS MODEL

SWISS-MODEL est une application dédiée à la modélisation 3D des protéines. Il guide l'utilisateur dans la construction de modèles 3D par homologie protéique et ceci avec différents niveaux de complexité.

La construction d'un modèle d'homologie comprend quatre étapes principales: (i) l'identification du (des) modèle (s) structurel (s), (ii) l'alignement de la séquence cible et de la structure du modèle, (iii) la construction du modèle et (iv) l'évaluation de la qualité du modèle.

Ces étapes nécessitent des logiciels spécialisés et intègrent des bases de données séquentielles privées et des bases de données publiques. Chacune des étapes ci-dessus peut être répétée de manière interactive jusqu'à ce qu'un résultat de modélisation satisfaisant soit atteint.

Si le modèle souhaité pour la modélisation est connu et disponible dans la bibliothèque de modèles SWISS-MODEL (SMTL), un alignement de modèle cible au format FASTA peut être utilisé pour démarrer le processus de modélisation, ignorant ainsi la recherche de modèle.

Le mode Alignement permet à l'utilisateur avancé d'invoquer l'étape de modélisation à partir d'alignements alternatifs et d'évaluer la qualité de ces modèles alternatifs.

La bibliothèque de modèles SWISS-MODEL est une base de données structurale importante de structures de protéines déterminées expérimentalement dérivées de la PDB (Protein Data Bank) et des Séquence protéiques de la banque UniProtKB. La séquence d'acides aminés de la protéine cible peut être soumise soit en texte clair, soit en format FASTA.

Dans notre cas, les résultats de modélisation 3D montrent la présence de structures en hélices  $\alpha$  à la surface de la protéine LYP. Les résultats obtenus par l'outil SWISS-MODEL confirment donc ceux obtenus par les autres outils notamment PROTPARAM et PSIPRED, lesquels sont en accord avec le caractère hydrophile de la protéine en raison de sa localisation cytosolique.



# Conclusion



## Conclusion

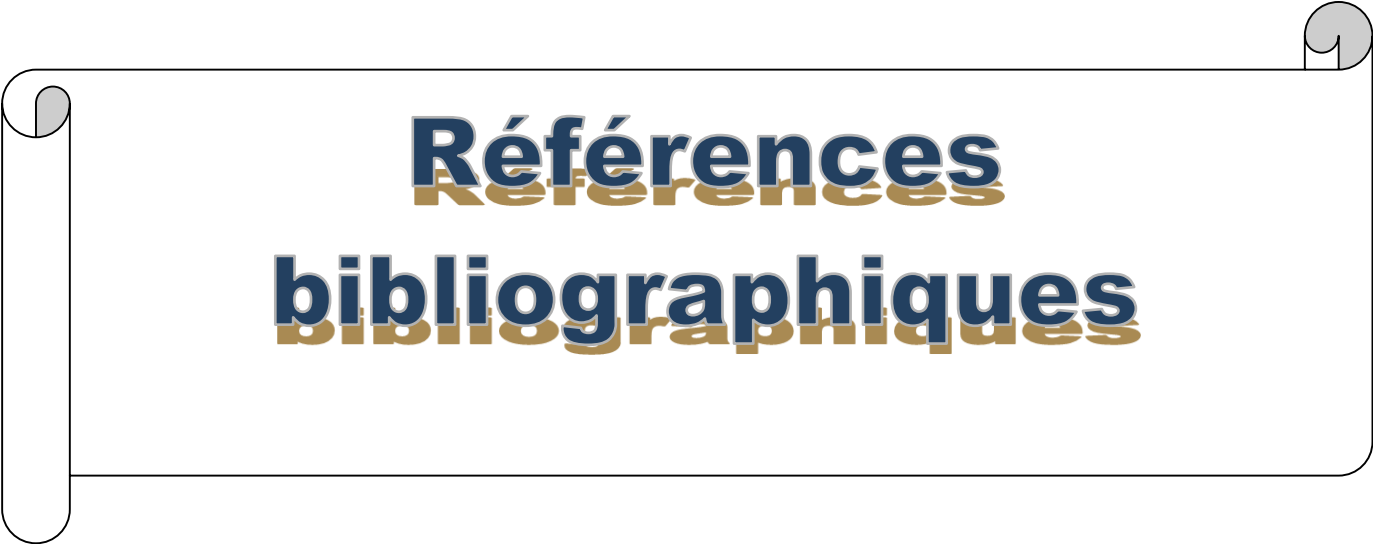
En conclusion, nous avons montré à l'issue de notre étude que l'allèle muté C1858T du gène PTPN22 est présent dans la population algérienne à une fréquence de 3,8%. Cette fréquence est proche des fréquences décrites dans les populations caucasienne européenne et confirme ainsi l'existence d'un gradient Nord-Sud décroissant de l'allèle muté, étendant ce gradient vers l'Afrique du nord.

L'Association de l'allèle muté C1858T du gène PTPN22 à la PR dans la population algérienne à l'ouest et au centre du moins, est ici démontrée grâce à la méta-analyse. Les résultats ont montré que le génotype homozygote sauvage CC est majoritaire (96%) alors que le génotype homozygote muté (TT) est rare (0.5%). En effet, ce génotype n'a été détecté ni dans le groupe témoin de Fodil *et al.*, ni au cours de l'étude de Boukercha, dans le groupe témoin. Le génotype hétérozygote muté (CT) est retrouvé à une fréquence de 6.6%. Ces résultats bien que préliminaires nous ont permis de confirmer l'association du gène PTPN22 à la PR dans la population Algérienne et montrent l'importance du rôle biologique de la protéine LYP dans les mécanismes pathophysiologies dans les maladies auto-immunes de façon générale.

L'importance du rôle biologique de la protéine LYP a pu également être appréhendée à travers notre analyse bio-informatique. En effet, les résultats obtenus avec Clustal omega indiquaient que la fonction biologique et la structure biochimique de la protéine LYP sont hautement conservées chez les primates, ce qui démontre l'importance de cette protéine.

Egalement, après avoir récupéré la séquence en acides aminés de la protéine LYP à partir de la base de données de séquence UniProtKB sur le portail EXPASY, nous avons pu réaliser une caractérisation physicochimique de la protéine LYP (PRTOPARAM), une prévision de ses structures secondaires (PSIPRED), une prévision de sa structure 3D (SWISS-MODEL) ainsi que l'identification de motifs fonctionnels (PROSITE) en accord avec une activité enzymatique de type PTPase. Ces résultats obtenus uniquement sur la base de la structure primaire en acides aminés de la protéine LYP démontrent toute la puissance de la bio-informatique et la nécessité d'intégrer ses l'outil dans toutes recherche envisagée en biologie.

En perspective, il serait intéressant d'augmenter l'effectif en élargissant l'étude du polymorphisme R620W à d'autres régions de l'Algérie.



**Références**  
références

**bibliographiques**  
bibliographiques

## **B**

Begovich, B., Victoria E, H., Carlton, LA., Honigberg, S. J., Schrodi, A P., Chokkalingam, H C., Kristin, G., Ardlie, A. Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis. *American Journal of Human Genetics*, (2004), 75 (2): 330-37.

Boukercha, A. "Contribution des polymorphismes des gènes de cytokines, du gène *NOD2* et du gène *PTPN22* dans la susceptibilité et la sévérité au cours des maladies inflammatoires chroniques de l'intestin dans la population Algérienne" (2015). Thèse de doctorat, 3<sup>ème</sup> cycle LMD, USTHB.

Bottini, N., Peterson, E. J. (2014). Tyrosine phosphatase PTPN22 : multifunctional regulator of immune signaling, development, and disease. *Annu Rev Immunol* ; 32 : 83-119.

Bottini, N., Vang, T., Cucca, F., Mustellin, T. (2006). Role of PTPN22 in type 1 diabetes and other autoimmune disease. *Semin Immunol*; 18(4): 207-213.

Bottini, N., Gary, S., Firestein. Duality of Fibroblast-like Synoviocytes in RA : Passive Responders and Imprinted Aggressors. *Nature Reviews Rheumatology*, (2013), 9(1) : 24-33.

Burn, G. L., Svensson, L., Sanchez-Blanco, C., Saini, M., Cope, A. P. (2011). Why is PTPN22 a good candidate susceptibility gene for autoimmune disease ? *FEBS Lett* ; 585(23) : 3689-3698.

## **C**

Chang, H. H., Tai, T. S., Lu, B., Iannaccone, C., Cernadas, M., Weinblatt, M., Ho, I.C. (2012). PTPN22.6, a dominant negative isoform of PTPN22 and potential biomarker of rheumatoid arthritis. *PLoS One* ; 7(3) : e 33067.

Chapman, S. J., Khor, C. C., Vannberg, F. O., Maskell, N.A., Davies, C. W., Hedley, E.L., Hill, A.V. (2006). PTPN22 and invasive bacterial disease. *Nat Genet*; 38(5):499-500.

Choy, E. Understanding the Dynamics: Pathways Involved in the Pathogenesis of Rheumatoid Arthritis. *Rheumatology*, (2012), 51 (5): 3-11.

Cloutier, J. F., Veillette, A. (1999). Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* ; 189(1) : 111-121.

Cohen, S., Dadi, H., Shaoul, E., Sharfe, N., Roifman, C. M. (1999). Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. *Blood* ; 93(6) : 2013-2024.

## *D*

Diaz-Gallo, L. M., Espino-Paisan, L., Fransen, K., Gomez-Garcia, M., Van Sommeren, S., Cardena, C., Martin, J. (2011). Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis* ; 17(11) : 2287-2294.

Dieudé, P., Guedj, M., Wipff, J., Avouac, J., Hachulla, E., Diot, E., Allanore, Y. (2008). The PTPN22 620W allele confers susceptibility to systemic sclerosis : findings of a large case-control study of European Caucasians and a meta-analysis. *Arthritis Rheum* ; 58 (7) : 2183-2188.

## *E*

Emery, P., I.B. McInnes, R. van Vollenhoven, et M. C. Kraan. Clinical Identification and Treatment of a Rapidly Progressing Disease State in Patients with Rheumatoid Arthritis. *Rheumatology*, (2008), 47 (4) : 392-98.

## *F*

Feng, X., Li, Y. Z., Zhang, Y., Bao, S. M., Tong, D. W., Zhang, S. L., & Hu, C.J. (2010). Association of the PTPN22 gene (-1123G>C) polymorphism with rheumatoid arthritis in Chinese patients. *Tissue Antigens* ; 76(4) :297-300.

Fodil, M., Benzaoui, A., Zemani - Fodil, F., aberkane, M., Boughrara, W., Saidi-Mehtar, N., Peti t-Teixeira E, Boudjema A. (2015). Association of PTPN22 (rs2476601) and STAT4 (rs7574865) polymorphisms with Rheumatoid Arthritis in the Western Algerian population. *Acta Rheumatol Port* ; 40(1) :56-62.

Frazer, Kelly A., Sarah S. Murray, Nicholas J. Schork, et Eric J. Topol. Human Genetic Variation and Its Contribution to Complex Traits. *Nature Reviews. Genetics*, (2009), 10 (4) : 241-51.

## *G*

Gomez, L. M., & Martin, J. (2005). Genetic influence of PTPN22 R620W polymorphism in tuberculosis. *Hum Immunol*; 66(12): 1242-1247.

Ghazi, Chabchoub, Elisabeth P.T., Maalej, A., Mariem Ben Hamad, Zouheir Bahloul, Cornelis, F., Ayadi, H. The R620W Polymorphism of the Protein Tyrosine Phosphatase 22 Gene in Autoimmune Thyroid Diseases and Rheumatoid Arthritis in the Tunisian Population. *Annals of Human Biology*, (2009), (3) : 342-49.

## H

Harrison, P., Pointon, C., Farrar, M., Brown, A., Wordsworth, P. J. Effects of PTPN22 C1858T Polymorphism on Susceptibility and Clinical Characteristics of British Caucasian Rheumatoid Arthritis Patients. *Rheumatology*, (2006), 45 (8): 1009-11.

Hung, C. H., Wei, J. C., Chen, C. C., Chuang, C. S., Chou, C. H., Lin, Y. J., Wong, R. H. (2014). Associations of PTPN22 and CTLA-4 genetic polymorphisms with Taiwanese ankylosing spondylitis. *Reumatologia*;34(5):683-691.

## I

Ichimura, M., Kaku, H., Fukutani, T., Koga, H., Mukai, T., Miyake, I., Hiromatsu, Y. (2008). Associations of protein tyrosine phosphatase non receptor 22 (PTPN22) gene polymorphisms with susceptibility to Graves' disease in a Japanese population. *Thyroid*; 18(6):625-630.

## K

Kurko, J., Tímecá B., Judít Laki, T. T., Mikecz, K., Szekanecz, Z. Genetics of Rheumatoid Arthritis- A Comprehensive Review. *Clinical Reviews in Allergy & Immunology*, (2013), 45 (2): 170-79.

Kawasaki, E. A., Ikegami, H., Kobayashi, T., Maruyama, T., Nakanishi, K., Eguchi, K. (2006). Systematic Search For Single Nucleotide Polymorphisms In A Lymphoid Tyrosine Phosphatase Gene (PTPN22): Association Between A Promoter Polymorphism and Type 1 Diabetes in Asian Populations. *Am J Med Genet A*; 140(6):586-593.

## L

Lamsyah, H., Rueda, B., Baassi, L., Elaouad, R., Bottini, N., Sadki, K., & Martin, J. (2009). Association of PTPN22 gene functional variants with development of pulmonary tuberculosis in Moroccan population. *Tissue Antigens*; 74(3): 228-232.

Liao, K. P., Gunnarsson, M., Kallberg, H., Ding, B., Pleng, R. M., Padyukov, L., Karlson, E. W., Klareskog, L., Askling, J., Alfredsson, L. (2009). Specific association of type 1 diabetes mellitus with anti-cyclic citrullinated peptide-positive rheumatoid arthritis. *Arthritis Rheum*; 60(3):653-600.

Liu, F., Liu, J., Zheng, T. S., Li, Q., Wang, C., Pan, X. P., Zhao, Y. W. (2012). The -1123G>C variant of PTPN22 gene promoter is associated with latent autoimmune diabetes in adult Chinese Hans. *Cell Biochem Biophys*; 62(2):273-279.

Louahchi, S., (2016). Thèse de doctorat, 3<sup>ème</sup> cycle LMD: Etude immunogénétique de la polyarthrite rhumatoïde en Algérie, 3(1): 63-65.

Louifi, H. Brahim, W. Rôle du gène PTPN22 dans le Diabète de Type 1. Etude sur une cohorte de la population Algérienne. Master 2 de Biologie et Pathologie moléculaire, USTHB, promotion (2014-2015).

Luo, L., Cai, B., Liu, F., Hu, X., & Wang, L. (2012). Association of Protein Tyrosine Phosphatase Nonreceptor 22 (PTPN22) C1858T gene polymorphism with susceptibility to autoimmune thyroid disease: a meta-analysis. *Endocr J*;59(5):439-445.

### *M*

McInnes, I B. (2011). Georg Schett. The pathogenesis of rheumatoid arthritis. *New England Journal of Medicine*, 365 (23) : 2205-19.

### *O*

Onengut-Gumuscu, S., Buchner, J. H., Cancannon, P. (2006). A haplotype-based analysis of the PTPN22 locus in type 1 diabetes. *Diabetes*;55(10): 2883-2889.

Orru, V., Tsai, S. J., Rueda, B., Fiorillo, E., Stanford, S. ., Dasgupta, J., Bottini, N. (2009). A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. *Hum Mol Genet*; 18(3): 569-579.

Orozco, Gisela, Elena Sanchez ,Miguel A .Gonzalez-gay ,Miguel A .Lopez-Nevot, Belén Torres, Rafael Cáliz Norberto Ortego-Centeno, et al . Association of a Functional Single-nucleotide Polymorphism of PTPN22, Encoding Lymphoid Protein Phosphatase , with Rheumatoid Arthritis and Systemic Lupus Erythematosus . *Arthritis and Rheumatism*(2005),52(1):219-24.

### *R*

Rani, R., Singh, A., Israni, N., Singh, A., Sharma, P., Kar, H. K. (2009). The role of polymorphic protein tyrosine phosphatase non-receptor type 22 in leprosy. *J Invest Dermatol*; 129(11): 2726-2728.

Rhee, Inmoo, et André Veillette. Protein Tyrosine Phosphatases in Lymphocyte Activation and Autoimmunity. *Nature Immunology*, (2012), 13 (5) : 439-47.

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., Daly, M. J. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet*; 43(11): 1066-1073.

Rodriguez- Rodriguez, L., Taib, W. R., Topless, R., Steer, S., Gonzalez-Escribano, M. F., Balsa, A., Martin, J. (2011). The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in caucasian case-control samples. *Arthritis Rheum*; 63(2): 365-372.

Ruysen-Witrand, A., A. Constantin, A. Cambon-Thomsen, et M. Thomsen. New Insights into the Genetics of Immune Responses in Rheumatoid Arthritis. *Tissue Antigens*, (2012), 80 (2): 105-18.

*S*

Sahin N, Gunduz F, Inanc N, Direskeneli H, Saruhan- Direskeneli G. No association of PTPN22 gene polymorphism with rheumatoid arthritis in Turkey. *RheumatolInt*, (2009), (30):81-83.

Seldin, M. F., R. Shigeta, K. Laiho, H. Li, H. Saila, A. Savolainen, M. Leirisalo-Repo, et al. Finnish Case-Control and Family Studies Support PTPN22 R620W Polymorphism as a Risk Factor in Rheumatoid Arthritis, but Suggest Only Minimal or No Effect in Junvenile Idiopathic arthritis. *Genes and Immunity*, (2005), 6 (8): 720-22.

Smolen, JS., Gunter, S. (2003). Therapeutic Strategies for Rheumatoid Arthritis. *Nature Reviews. Drug Discovery*, 2 (6): 473-88.

Stahl, E A., Raychaudhuri ,S., Elaine, F ., Emmers,R., Xie,G., Eyre,S., Thomson, B P. , Li, Y., et al. Genome-Wide Association Study Meta-Analysis Identifies Seven New Rheumatoid Arthritis Risk Loci. *Nature Genetics*, (2010) 42 (6): 508-14.

Stanford, S. M., Bottini, N. (2014). PTPN22: the archetypal non-HLA autoimmunity gene. *Nat Rev Rheumatol*, 10(10):602-611.

Stastny, P. Mixed Lymphocyte Cultures in Rheumatoid Arthritis. *The Journal of Clinical Investigation*, (1976), 57 (5): 1148-57.

Stastny, P. Association of the B-Cell Alloantigen DRw4 with Rheumatoid Arthritis. *The New England Journal of Medicine*, (1978) 298 (16): 869-71.

*T*

Torres,C., Megdalena,N., Ruiz-Noa, Y., Esther Martinez-Bonilla, G., Leyva-Torres ,D S., et al. The +1858C/T PTPN22 Gene Polymorphism Confers Genetic Susceptibility to Rhumatoid Arthritis in Mexican Population from the Westren Mexico. *Immunology Letters*, (2012), 147 (1-2): 41-46.

## V

Van Oene, M., Richard .,FW, Xiangdong ,L., Yazdanpanah, M., Xiangjun G., Newman,B., Kwan,A., et al. Association of the Lymphoid Tyrosine Phosphatase R620W Variant with Rheumatoid Arthritis, but Not Crohn's Disease, in Canadian populations. *Arthritis & Rheumatism*, (2005), 52 (7): 1993-98.

Vang, T., Congia, M., Macis, MD., Musumeci, L., Orru, V., Zavattari, P., Nika, K., Tautz,L., Tasken, K., Cucca, F., Mustelin, T., Bottini, N. (2005). Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet*; 37(12): 1317-1319.

## W

Wang, S., Dong, H., Han, J., Ho, WT., Fu, X., Zhao, ZJ.(2010). Identification of a variant form of tyrosine phosphatase LYP. *BMC MolBiol* ; 2 ;11 :78.

Wang, Y., Shaked, I., Stanford, S. M., Zhou, W., Curtsinger, J. M., Mikulski, Z., Peterson, E. J.(2013). The autoimmunity – associated gene PTPN22 potentiate toll-like receptor-driven, type 1 interferon – dependent immunity. *Immunity* ; 39(1) :111-112.

## Z

Zhi-Dan Fan, Z. D., Wang, F.F., Huang, H., Huang, N., Ma, H. H., Guo, Y. H., Yu, H. G. (2015). STAT4, rs7574865 G/T and PTPN22 rs2488457 G/C polymorphisms influence the risk of developing juvenile idiopathic arthritis in Han Chinese patients. *PLoSOne*, 10(3); e 0117389.

Zikherman,J., Hermiston, M., steiner, D. , Hasegawa ,K., Chan, A., Weiss, A .PTPN22 Dificiency Cooperates With the CD45 E613R Alléle to Break Tolerance on A Non- Autoimmune Background .*J Immunol*,(2009),182:4093-4106.

Zouidi, F., Stayoussef, M., Fourati, H., Abida, O., Ayed, M. B., Masmoudi, H. (2014). Contribution of PTPN22, CD28,CTLA-4 and ZAP-70 variants to the risk of type 1 diabetes in Tunisians. *Gene*. 5 3 3(1): 420-426.





# **Annexes**

**Annexes**

## ANNEXE N°1

### URL des Portails, banques et outils bio-informatiques (logiciels)

- Banque bibliographique PubMed:

<http://www.pubmed.gov>

- Portail NCBI (National Center for Biotechnology Information) :

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

- Banque de séquence de gènes

[www.ncbi.nlm.nih.gov/gene](http://www.ncbi.nlm.nih.gov/gene)

- Portail EBI (European Bioinformatics Institute) ;

<http://www.ebi.ac.uk/>

- Portail Expasy (Expert Protein Analysis System)

<http://www.expasy.org>

- Logiciel PROTPARAM

<http://www.expasy.org/protparam>

- Logiciel SWISSMODEL

<https://swissmodel.expasy.org>

- Logiciel PSIPRED

[www.bioinf.cs.ucl.ac.uk/psipred/result/7b](http://www.bioinf.cs.ucl.ac.uk/psipred/result/7b)

- Logiciel PROSITE

[www.prosite.expasy.org/cgi-bin/prosite](http://www.prosite.expasy.org/cgi-bin/prosite)

- Logiciel CLUSTAL OMEGA

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

## ANNEXE N°2

### Séquences protéiques au format FASTA.

#### Séquence protéique de la protéine PTPN22 humaine

>sp|Q9Y2R2|PTN22\_HUMAN Tyrosine-protein phosphatase non-receptor type 22 OS=Homo sapiens GN=PTPN22 PE=1 SV=2

MDQREILQKFLDEAQSKKITKEEFANEFLKLRQSTKYKADKTYPTTVAEKPKNIKKNRY  
KDILPYDYSRVELSLITSEDEDSSYINANFIKGVYGPKAYIATQGPLSTLLDFWRMIWEY  
SVLIIVMACMEYEMGKKKCERYWAEPGEMQLEFGPFSVSCEAEKRKSDYIIRTLKVKFNS  
ETRTIYQFHYNWPDHDPSSIDPILELIWDVRCYQEDDSPICIHCSAGCGRTGVICAI  
DYTWMLLKDGIIPENFSVFSLIREMRTQRPSLVQTQEYELVYNAVLELFKRQMDVIRDK  
HSGTESQAKHCIPEKNHTLQADSYSPNLPKSTTKAAKMMNQQRKMEIKESSSDFDRTSE  
ISAKEELVLHPAKSSTSFDFLELNYSFDKNADTTMKWQTKAFPIVGEPLQKHQSLDLGSL  
LFEGCSNSKPVNAAGRYFNSKVPITRTKSTPFELIQQRETKEVDSKENFSYLESQPHDSC  
FVEMQAQKVMHVSSAELNYSLPYDSKHQIRNASNVKHHDSALGVYSYIPLVENPYFSSW  
PPSGTSSKMSLDLPEKQDGTVPSSLLPTSSTSLFSYYNSHDSLSLNSPTNISSLLNQES  
AVLATAPRIDDEIPPPLPVRTPESFIVVEEAGEFSPNVPKSLSSAVKVKIGTSLEWGGTS  
EPKKFDDSVILRPSKSVKLRSPKSELHQDRSSPPPPLPERTLESFFLADEDCMQAQSIET  
YSTSYPDTMENSTSSKQTLKTPGKSFTRSKSLKILRNMKKSICNSCPPNKPAESVQSNNS  
SSFLNFGFANRFSKPKGPRNPPPTWNI

**Séquence protéique de la protéine PTPN22 du chimpanzé**

**>tr|H2PZP7|H2PZP7\_PANTR Protein tyrosine phosphatase, non-receptor type 22  
(Lymphoid) OS=Pan troglodytes GN=PTPN22 PE=2 SV=1**

MDQREILQKFLDEAQSKKITKEEFANEFLKLRQSTKYKADKTYPTTVAEKPKNIKKNRY  
KDILPYDYSRVELSLITSEDEDSSYINANFIKGVYGPKAYIATQGPLSTLLDFWRMIWEY  
SVLIIVMACMEYEMGKKKCERYWAEPGEMQLEFGPFVSCEAEKRKSDYIIRTLKVKFNS  
ETRTIYQFHYKNWPDHDVPSSIDPILELIWDVRCYQEDDSPICIHCSAGCGRTGVICAI  
DYTWMLLKDGIIPENFSVFSLIQEMRTQRPSLVQTQEYELVYNAVLELFKRQMDVIRDK  
HSGTESQAKHCIPEKNHTLQADSYSPNLPKSTTKAAKMMNQRTKMEIKESSSDFDRTSE  
ISAKEELVLHPAKSSTSDFLELNYSFDKNADTTMKWQTKAFPIVGEPLQKHQSLDLGSL  
LFEGCSNSKPVNAAGRYFNSKVPITRTKSTPFELIQQRETKEVDSKENFSYLESQPHDSC  
FVEMQAQKVMHVSSAELNYSLPYDSKHQIRNASNVKHHDSALGVYSYIPLVENPYFSSW  
PPSGTSSKMSLDLPEKQDGTVPSSLLPTSSTSLFSYYNSNDSLSLNSPTNISSLLNQES  
AVLATAPRIDDEIPPPLPVRTPESFIVVEEAGEFSPNVPKSLSSAVKVKIGTSLEWGGTS  
EPKKFDDSVRLRPSKSVKLRSPKSELHQDRSSPPPPLPERTLESFFLADEDCMQAQSIET  
YSTSYPDTMENSTSSKQTLKTPGKSFTRSLSLKILRNMKKSICNSCPPNKPAESVQSNNS  
SSFLNFGFANRFSKPKGPRNPPPTWNI

**Séquence protéique de la protéine PTPN22 de la souris.**

**>sp|P29352|PTN22\_MOUSE Tyrosine-protein phosphatase non-receptor type 22 OS=Mus musculus GN=PTPN22 PE=1 SV=1**

MDQREILQQLLKEAQKKKLNSEEFASEFLKLRQSTKYKADKIYPTTVAQRPKNIKKNRY  
KDILPYDHSLVELSLLTSDDEDSSYINASFYKGVYGPYKAYIATQGPLSTLLDFWRMIWEY  
RILVIVMACMEFEMGKKKCYRYWAEPGETQLQFGPFSISCEAEKKKSDYKIRTLKAKFNN  
ETRIYQFHYNWPDHDPSSIDPILQLIWDMRCYQEDDCVPICIHCSAGCGRTGVICAV  
DYTWMLLKDGIIPKNFSVFNLIQEMRTQRPSLVQTQEYELVYSVLELFKRHMDVISDN  
HLGREIQAQCSIPEQSLTVEADSCPLDLPKNAMRDVKTNNQHSKQGAEAESTGGSSLGLR  
TSTMNAEEELVLHSAKSSPSFNCLELNCGCNNKAVITRNGQARASPVVGEPLQKYQSLDF  
GSMLFGSCPSALPINTADRYHNSKGPVKRTKSTPFELIQQRKTNDLAVGDGFSCLESQFH  
EHYSLRELQVQRVAHVSSEELNYSLPGACDASCVPRHSPGALRVHLYTSLAEDPYFSSSP  
PNSADSKMSFDLPEKQDGATSPGALLPASSTTSFFYSNPHDSLVMNTLTSFSPPLNQETA  
VEAPSRRTDDEIPPPLPERTPESFIVVEEAGEPSRVTESLPLVVTFGASPECSGTSEMK  
SHDSVGFTPSKNVKLRSFKSDRHQDGSPPPPLPERTLESFFLADEDCIQAQAVQTSSTSY  
PETTENSTSSKQTLRTPGKSFTRSKSLKIFRNMKKSVCNSSSPSKPTERVQPKNSSSFLN  
FGFGNRFSKPKGPRNPPSAWNM

## ANNEXE N°3

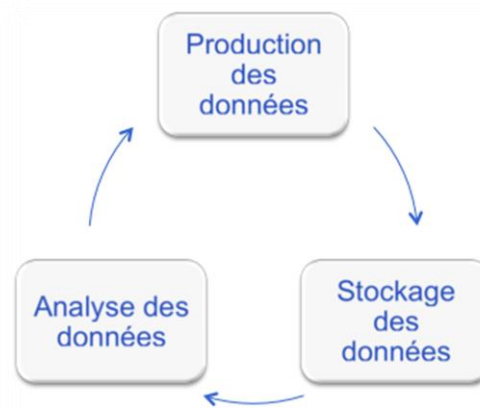
### Glossaire

**Le système immunitaire** est un ensemble de cellules et voies métaboliques conduisant à l'élimination d'une grande variété de pathogènes. Ce système repose sur la notion centrale du soi opposé au *non-soi* ainsi qu'au *soi modifié*.

**L'auto-immunité** est une situation au cours de laquelle il y a une rupture des mécanismes de tolérance conduisant à l'action pathogène du système immunitaire vis-à-vis de constituants naturels de l'organisme.

### Définition de la bioinformatique :

La bioinformatique est un domaine de recherche qui analyse et interprète des données biologiques, au moyen de méthodes informatiques, afin de créer de nouvelles connaissances et résoudre des problèmes en biologie.



**Définition de la génomique :** La génomique est l'étude des génomes, de leur organisation et de leur évolution. C'est aussi l'étude de l'expression et de la fonction de tous les gènes du génome.



la bio-informatique est utilisée à différentes étapes du séquençage des génomes pour assurer la lecture des séquences à la sortie des séquenceurs, assemblage des génomes à partir des fragments séquencés, recherche des répétitions pour corriger les mauvais assemblages, même le regroupement des séquences appartenant à un même gène, elle est utilisée pour comparer les séquences obtenues 2 à 2, ou multiple, ou encore une séquence contre une banque.

**SNP (Single Nucleotide Polymorphism):** les SNPs constituent la forme la plus abondante de variations génétiques dans le génome humain. Ils représentent plus de 90% de toutes les différences entre individus. C'est un type de polymorphisme de l'ADN où un seul nucléotide, le composant de base de l'ADN, est modifié.

**Polymorphisme synonyme** : appelé aussi synonyme silencieux ou non fonctionnel, ce sont des changements nucléotidiques qui induisent l'utilisation d'un codon différent lors de la traduction mais l'acide aminé reste le même. La séquence protéique est donc inchangée.

**Polymorphisme non synonyme** : le changement d'un nucléotide induit un changement de l'acide aminé utilisé lors de la traduction et donc de la séquence protéique.

**Allèle** : Une des différentes formes que peut prendre un même gène. Les allèles occupent la même position (locus) sur les chromosomes homologues.

**Génotype** : c'est l'ensemble ou une partie donnée de l'information génétique (composition génétique d'un individu. Il existe deux formes du gène X : l'allèle sauvage Xa et l'allèle muté Xb, alors le génotype d'un individu pour le gène X peut être soit homozygote sauvage Xa/Xa, homozygote muté Xb/Xb ou hétérozygote Xa/Xb.

**Phénotype** : Ensemble des caractères observables d'un individu. Le phénotype correspond à la réalisation du génotype (expression des gènes) mais aussi des effets du milieu, de l'environnement.

**GWAS (Genome Wide Association Study)** : une étude GWAS est une étude d'association à très grande échelle. La méthode consiste à génotyper un maximum de marqueurs génétiques (de 317 000 à 1 000 000 selon le type de puce à ADN) chez un grand nombre de sujets atteints de la maladie génétique "complexe" et un grand nombre de sujets témoins. Les variations d'un nucléotide ou SNP malgré leur nature bi-allélique qui contraste avec le caractère multi allélique des microsatellites, ont été retenues compte tenu de leur très grand nombre sur le génome humain offrant ainsi une très bonne couverture du génome.

**Risque relatif** : Le risque relatif ou RR est une mesure de probabilité, souvent utilisé en épidémiologie, qui permet de mesurer le risque de développer une maladie par rapport à un facteur de risque. Le RR se calcule en faisant le rapport entre le fait de développer la maladie si l'on est exposé au facteur de risque et le fait de développer la maladie si l'on n'est pas exposé au facteur de risque. Cette mesure permet de se rendre compte de la corrélation entre certains facteurs de risque (manger gras, fumer...) et la survenue de certaines maladies.

**Odd Ratio (OR)** : L'odd ratio (dont une traduction littérale en français peut être « rapport des cotes ») est le rapport de l'odd de l'événement dans le groupe traité divisé par l'odd de l'événement dans le groupe contrôle. Dans le cadre d'une étude d'association génétique l'Odd Ratio se calcule comme suit :

$$OR = \frac{\text{Nombre de patients porteurs d'allèle sauvage}}{\text{Nombre de patients porteurs d'allèle muté}}$$

L'odds ratio s'interprète de façon similaire au risque relatif. Un odds ratio de 1 correspond à l'absence d'effet. En cas d'effet bénéfique, l'odds ratio est inférieur à 1 et il est supérieur à 1 en cas d'effet délétère. Plus l'odds ratio est éloigné de 1, plus l'effet est important.

**Risque attribuable** : Le **risque attribuable** indique le nombre de malades parmi les personnes qui ont été exposées à un facteur qui peut être attribué à cette exposition.

Incidence (chez les personnes exposées) – Incidence (chez les personnes non-exposées). Ceci nous indique l'impact causal dû à cette exposition.

**Intervalle de confiance** : L'intervalle de confiance (IC) à 95% est un intervalle de valeurs qui a 95% de chance de contenir la vraie valeur du paramètre estimé. Avec moins de rigueur, il est possible de dire que l'IC représente la fourchette de valeurs à l'intérieur de laquelle nous sommes certains à 95% de trouver la vraie valeur recherchée. L'intervalle de confiance est donc l'ensemble des valeurs raisonnablement compatibles avec le résultat observé (l'estimation ponctuelle). Il donne une visualisation de l'incertitude de l'estimation.\*\*