

## Projet LORIA Capsid-Pixel : HLA-3D-DIFF Sujet CDD ingénieur

Structuration, stockage et fouille de données biomédicales dans un système d'information. Application à l'analyse de la compatibilité HLA donneur/receveur d'un service de transplantation.

### Détails du CDD

- Durée : 6 mois, prolongation possible
- Démarrage souhaité : avant le 31 décembre 2020
- Lieu de travail habituel : LORIA, équipe Capsid, Vandoeuvre-les-Nancy
- Rémunération : selon le niveau d'études et la grille de salaires ingénieur de l'Université de Lorraine
- Contact : Marie-Dominique Devignes, équipe Capsid

### Contexte

Le complexe majeur d'histocompatibilité (CMH) humain est principalement composé du système de groupes sanguins ABO et des protéines du "human leukocyte antigen" (HLA). Les différences entre les protéines HLA du donneur et du receveur constituent l'une des principales limites de la transplantation d'organes : en cas d'incompatibilité, il peut se déclencher chez le patient une réponse immunitaire pouvant aboutir au rejet de l'organe greffé. Le système HLA est en outre indispensable au bon fonctionnement de la réponse immunitaire physiologique aux agents pathogènes et aux cellules tumorales. En effet, les protéines HLA servent à présenter de petits fragments peptidiques de « protéines étrangères » aux lymphocytes T pour l'activation d'une réponse immunitaire salutaire pour l'organisme.

De nos jours, la majorité des pertes de greffons est due à des anticorps (Ac) anti-HLA produits par le receveur contre une ou plusieurs protéines HLA du donneur appelées historiquement les « antigènes HLA » parce qu'ils réagissent avec des anticorps dans la réaction anticorps-antigène. Cette réaction est une interaction entre deux protéines : l'anticorps et l'antigène, favorisée par la complémentarité structurale entre l'anticorps et l'antigène.

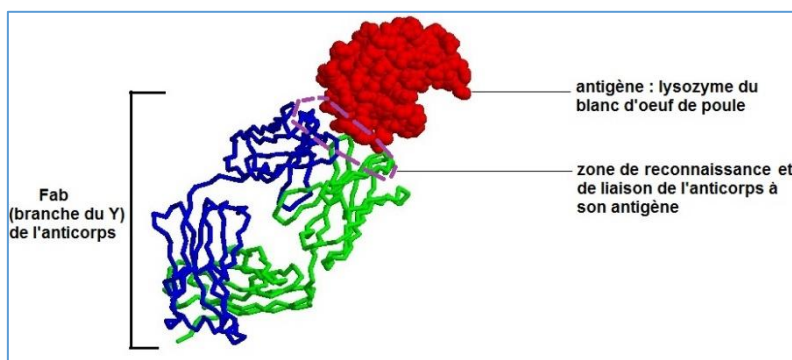


Fig 1 : Exemple de reconnaissance anticorps-antigène : l'épitope est la partie de l'antigène reconnue par le fragment Fab de l'anticorps (zone de reconnaissance et de liaison sur le schéma ci-contre).

<http://acces.ens-lyon.fr/acces/thematiques/immunité-et-vaccination/ressources-logicielles/sequences/anticorps-circulants/Cote-eleves-specificite-anticorps-demarche-dinvestigation>

L'analyse de la compatibilité donneur-receveur consiste en particulier à comparer le répertoire HLA du donneur et du receveur, côté antigène et côté anticorps. Puisqu'un individu ne produit pas d'anticorps contre ses propres antigènes HLA, on peut supposer que si le donneur et le receveur ont exactement le même répertoire HLA il n'y aura pas de rejet. Cependant cette situation est rarement observée du fait de la multiplicité des antigènes HLA. Chaque individu possède jusqu'à 22 protéines HLA distinctes par individu, issues des 11 gènes connus à ce jour : A, B, C pour la classe I et DRB1, DRB3-5, DRA1,

DQB1 et DQA1, DPB1 et DPA1 pour la classe II, dont l'individu a hérité une version de chacun de ses parents. De plus, il existe des milliers de formes variantes (allèles) pour chacun de ces gènes, ce qui crée une combinatoire quasi infinie des répertoires HLA. En pratique certains allèles sont rares, et il est estimé qu'il n'existe pas plus de 1 000 épitopes diversement répartis dans les protéines HLA. En pratique, les transplantations sont presque toujours effectuées avec un niveau d'incompatibilité significatif (mais pas quantifié), c'est-à-dire dire l'existence de différences entre les répertoires HLA des donneurs et des receveurs.

**Le but du travail** est d'organiser de façon structurée les données issues des tests d'anticorps anti-HLA dans les sérums consécutifs des receveurs en attente de greffe ou greffés, afin d'étudier par fouille de données (1) les relations entre les profils de réactivités et le ou les épitopes reconnus par ces sérums, que l'on associera à la structure 3D des antigènes porteurs de ces épitopes, et (2) les conditions immunologiques conduisant au rejet ou à l'acceptation d'un greffon en fonction des différences observées dans les allèles HLA du donneur et du receveur.

### Données disponibles

Concernant le point (1), les données à notre disposition sont des profils d'anticorps anti-HLA obtenus par analyse Luminex des sérums de milliers de patients (receveurs avant et après la greffe), testés contre une batterie de 200 allèles HLA dont on connaît la séquence en aa. Pour chaque allèle, une mesure de fluorescence (MFI) donner l'intensité de la reconnaissance par le ou les anticorps présents dans le sérum.

Concernant le point (2), les données sont, pour chaque receveur ou paire receveur-donneur, les allèles HLA obtenus par séquençage des gènes HLA du receveur et du donneur, et l'évolution au cours du temps du profil anticorps anti-HLA du receveur avant et après la greffe.

Des bases de données publiques comme IMGT/HLA<sup>1</sup> ou HLA Epitope Registry<sup>2</sup> seront aussi utiles pour avoir accès aux connaissances du domaine.

### Etapas du travail

#### 1. Définition des use-case : cas d'étude du système d'information

Ces use-cases seront formulés par les participants du projet HLA-3D-Diff. A titre d'exemple on pourra avoir les use-case suivants

- Calculer les statistiques des valeurs de MFI pour l'allèle DQA1\*03:03 (moyenne, écart-type, min, max, quantiles au choix)
- Extraire tous les sérums présentant un profil donné : par exemple une MFI > 5000 pour un ou plusieurs allèles d'une liste d'intérêt
- Lister tous les allèles qui ont le même épitope
- Repérer des combinaisons d'allèles qui n'apparaissent jamais ou très rarement
- Visualiser les épitopes à partir d'un profil anticorps anti-HLA donné
- Représenter l'évolution dans le temps des profils anticorps anti-HLA d'un receveur donné

#### 2. Conception et construction du système d'information (base de données et fonctionnalités d'interrogations)

Après une analyse des outils existants et de leurs limites, il faudra concevoir et développer le système d'information permettant de traiter les use-cases listés et de mettre en œuvre la fouille de données

---

<sup>1</sup> <http://www.imgt.org/IMGTindex/HLA.php>

<sup>2</sup> <https://www.epregistry.com.br/>

3. **Développement d'une interface utilisateur et de démonstrateurs** pour l'utilisation du système.
- Mise en place d'un tableau de bord du SI
  - Calculs de statistiques et visualisation des données
  - Production de jeux de données pour la fouille de données
  - Expériences de fouilles de données

#### Profil souhaité

- Formation en informatique niveau M2
- Langages SQL, Python, R,
- Notions de programmation web
- Intérêt pour la fouille de données (R, WEKA)
- Intérêt pour les applications biomédicales

#### Candidature

Envoyer CV et lettre de motivation à Marie-Dominique Devignes :

[marie-dominique.devignes@loria.fr](mailto:marie-dominique.devignes@loria.fr)

avant le 23 novembre 2020.