

Cahier des charges

Mirna MARIE-JOSEPH, Théo GAUVRIT, Kévin MERCHADOU

Février 2019

Chapitre 1

Contexte

Le laboratoire de génétique moléculaire du Centre Hospitalier d'Urgence (CHU) de Bordeaux effectue quotidiennement des diagnostics prénataux. Le kit PP16 est utilisé dans ce type d'étude. Il permet, grâce à l'utilisation de seize microsatellites (séquences formées d'un motif nucléotidique d'ADN répété), de déterminer la présence ou non d'ADN maternel dans un échantillon foetal. En cas de contamination, deux formules permettent de calculer le taux de contamination de l'échantillon. Ces formules sont discriminées selon le profil allélique du foetus. L'étude de cette contamination repose sur le principe mendélien que tout individu hérite d'un allèle de chacun de ses deux parents.

L'analyse des profils alléliques s'effectuent à l'aide du logiciel GeneMapper. Les profils de la mère et du foetus sont comparés entre eux. Dans certains cas, le profil du père peut être également ajouté dans l'analyse. Les allèles présents sur le profil sont sélectionnés à l'aide du logiciel et exportés sous forme de fichier texte (extension .txt). Ce fichier est alors copié/collé dans une feuille Excel permettant d'obtenir une conclusion quant à la contamination ou non de l'échantillon. On considère cinq conclusions possibles lors de l'étude du profil allélique de chaque microsatellite :

Homozygote (HMZ) non contaminé : le foetus est homozygote (le même allèle présent en deux exemplaires) et on ne décèle pas de contamination.

Pour ce cas, on constate un seul pic allélique sur le profil du foetus.

Hétérozygote (HTZ) non contaminé : le foetus est hétérozygote (deux allèles distincts) et on ne décèle pas de contamination. On constate deux pics alléliques sur le profil du foetus.

Homozygote (HMZ) contaminé : le foetus est homozygote et on décèle une contamination de l'échantillon. Ce cas se présente sous la forme d'un pic allélique provenant de la mère et d'un second moins prononcé provenant également de la mère.

Hétérozygote (HTZ) contaminé : le foetus est homozygote et on décèle une contamination de l'échantillon. Elle se représente sous forme de trois

pics alléliques au sein du profil.

Non-informatif : on ne peut conclure quant à la contamination de l'échantillon. Ce type de résultat est obtenu dans les cas où la mère est homozygote pour le microsatellite concerné, quand le fœtus possède les deux mêmes allèles que la mère, quand l'allèle non hérité de la mère tombe dans l'écho. C'est à dire, lorsque l'allèle non hérité se situe à une distance de quatre nucléotides d'un des allèles que possède le fœtus.

Néanmoins, l'action humaine est encore nécessaire pour conclure à partir des résultats obtenus avec l'aide de la feuille Excel.

Dans le but d'améliorer l'automatisation du procédé, le développement d'une application est demandée.

L'objectif de cette application est d'assurer une automatisation complète de l'analyse des profils alléliques. L'application doit permettre d'amener à la conclusion d'un résultat sans que l'action humaine ne soit nécessaire.

Chapitre 2

Besoins

L'application devra être capable à partir du fichier texte exporté depuis le logiciel GeneMapper d'effectuer l'analyse et d'afficher les conclusions à l'écran sans intervention humaine.

L'application devra donc pouvoir discriminer un fichier texte contenant des profils alléliques mère/foetus d'un fichier texte contenant des profils alléliques mère/père/foetus.

De plus l'application devra proposer une interface simple d'utilisation. Elle devra donc être conçue dans l'objectif d'être le plus ergonomique possible. Les attentes suivantes ont été formulées concernant l'application :

- Sauvegarde des résultats obtenus.
- Comparaison de résultats.
- Exportation des résultats dans un format prédéfini.
- Impression des résultats obtenus.

L'ensemble des ordinateurs du laboratoire étant sous Windows 7, cette application devra être compatible avec le système d'exploitation. Une mise à jour prochaine des ordinateurs de Windows 7 vers Windows 10 implique également que l'application puisse fonctionner une fois la mise à jour effectuée.

Enfin, bien que ne nécessitant pas l'intervention de la main humaine, l'application doit permettre aux biologistes d'accéder au cheminement effectué pour arriver à la conclusion proposée. Cette fonctionnalité assure un contrôle humain sur le résultat prédit par l'application.

Chapitre 3

Conception

L'application sera implémentée à l'aide du langage de programmation Python. Notre choix s'est porté sur ce langage car il est aujourd'hui extrêmement populaire et très utilisé à travers le monde. Ainsi, il existe une documentation très fournie et un nombre très impressionnant de bibliothèque. Parmi ces bibliothèques, certaines permettent de réaliser des applications graphiques. Nous avons donc décidé de conserver un langage homogène entre l'implémentation de l'interface de graphique et l'implémentation de l'algorithme permettant de conclure sur le degré de contamination de l'échantillon proposé. De plus, python est fonctionnel indépendamment du système d'exploitation utilisé.

Lors du choix de la bibliothèque ... nous avons dû garder en tête la contrainte de rétro-compatibilité Windows 7/Windows 10. Il a donc fallu choisir une bibliothèque encore mise à jour pour Windows 7. Cependant, la mise à jour de Windows 7 à Windows 10 étant prévu sur les ordinateurs du laboratoire, nous avons décidé la mise en place d'un patch permettant d'assurer le bon fonctionnement de l'application une fois sous Windows 10.

De plus, nous avons fait en sorte de choisir une bibliothèque gratuite permettant d'en utiliser toutes les fonctionnalités sans avoir à déboursier le moindre crédit. Il était également nécessaire d'utiliser une bibliothèque populaire. En effet, cette popularité nous assure que la bibliothèque est constamment mise à jour et qu'il existe beaucoup de documentations et d'aides sur cette dernière.

Notre choix s'est également porté sur cette bibliothèque car elle est *cross-platform*, ce qui signifie qu'elle fonctionne tout aussi bien sous différents systèmes d'exploitation (Mac OS, Windows, Linux). Ainsi, nous assurons la portabilité de l'application sous différents systèmes d'exploitation.

L'interface graphique sera construite à partir de la bibliothèque... Le résultat suivant est attendu :

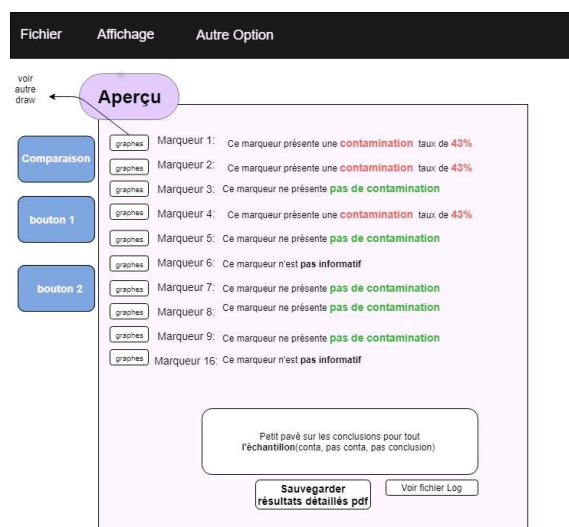


FIGURE 3.1 – Représentation schématique de l'interface graphique de l'application.

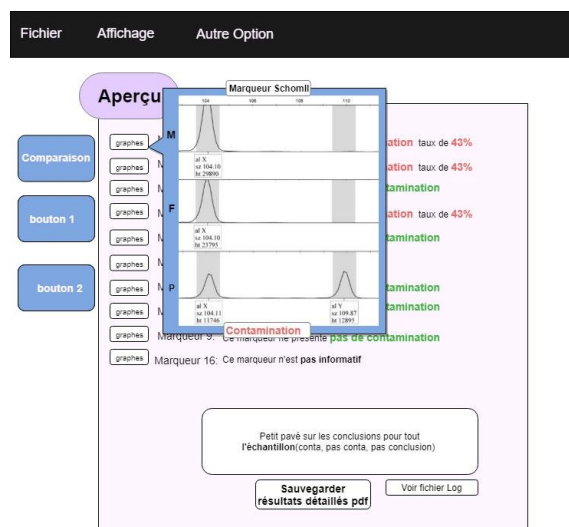


FIGURE 3.2 – Représentation schématique de l'interface graphique de l'application suite.

Un bouton permettra de sauvegarder les résultats aux formats pdf ce qui les rendra plus facilement partageable entre laboratoires. Il sera également possible d'exporter ces résultats sous un tout autre format (texte par exemple). Cette fonctionnalité permettra l'utilisation de ces résultats au sein d'une autre application. Enfin, comme présenté en figure 3.2, un passage de la souris sur le bouton "graphes" permettra l'affichage du pop-up montrant le profil allélique correspondant au résultat obtenu pour chaque microsatellite. Cette fonction permettra aux utilisateurs de l'application de pouvoir constater la bonne conclusion de l'analyse en la comparant avec leur ressenti visuel sur le profil allélique. Le bouton comparaison permettra de diviser la fenêtre en deux afin d'aligner deux résultats pour comparaison. Cet outil permettra également de lancer plusieurs diagnostics en même temps.

Le second code correspondra au fichier log. Ce fichier permettra de retracer le cheminement effectué pour obtenir le résultat présenté par l'application. Un bouton dans l'interface permettra d'accéder à ce fichier pour garantir une totale transparence des résultats. Il pourra également être modifié dans le but d'adapter la recherche à de nouveaux objectifs.

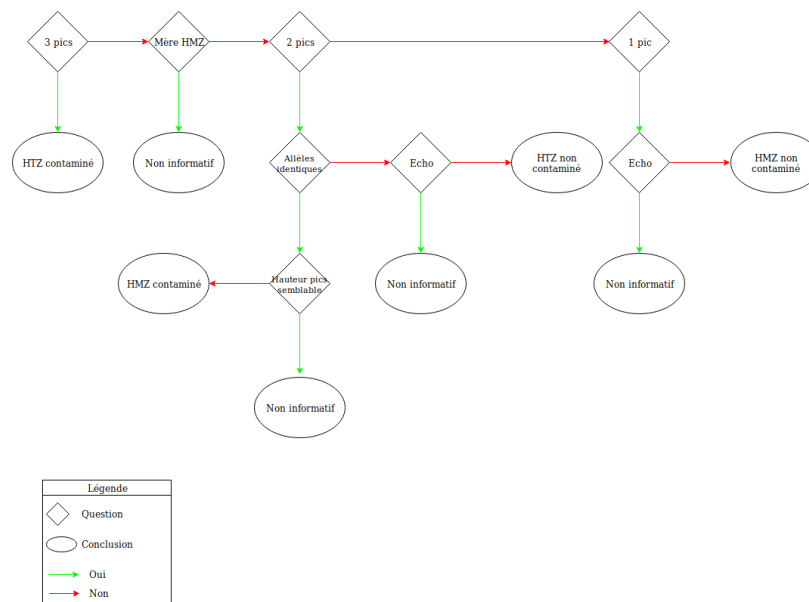


FIGURE 3.3 – Représentation schématique de l'arbre de décision du fichier log

En se basant sur les informations obtenues à partir du contexte d'étude, l'arbre ci-dessous peut être obtenu. Il permet d'aboutir à des conclusions sur l'échantillon étudié. A l'aide de ce cheminement, il est possible de discriminer toutes les conclusions possibles. Notamment, celles liées au caractère non informatif du résultat. Ainsi, à partir de l'arbre on peut déterminer si la notion non informative du résultat est due à un écho, à un profil allélique semblable ou à une mère homozygote.