



Canadian Scientific Working Group on DNA Analysis Methods

7 mars 2022

Réponse au rapport de la Commission du droit de l'Ontario intitulé " Intelligence artificielle Étude de cas : Les instruments de génotypage probabiliste dans la justice canadienne " par le SWGDAM canadien

Par la présente, le *Scientific Working Group on DNA Analysis Methods* (SWGDAM) canadien souhaite commenter le rapport de la Commission du droit de l'Ontario (CDO) intitulé Intelligence artificielle Étude de cas : Les instruments de génotypage probabiliste dans la justice canadienne, publié en juin 2021. Le SWGDAM canadien est composé des directeurs/cadres supérieurs et des coordonnateurs scientifiques de quatre entités : les divisions ADN des trois laboratoires publics canadiens, notamment le Centre des sciences judiciaires (CSJ) de l'Ontario, le Laboratoire de sciences judiciaires et de médecine légale (LSJML) du Québec et les Services nationaux de laboratoires judiciaires (SNLJ) de la Gendarmerie Royale du Canada (GRC), desservant les autres provinces et territoires, ainsi que la Banque nationale de données génétiques (BNDG) également de la GRC. Le SWGDAM est une plateforme de discussion et de partage des meilleures pratiques canadiennes en biologie judiciaire.

Les membres du SWGDAM canadien connaissent bien le génotypage probabiliste et l'utilisation qui en est faite au Canada, puisque cette approche est en opération dans deux des trois laboratoires judiciaires depuis plusieurs années (le CSJ a commencé à utiliser le système en 2016 tandis que le LSJML a commencé en 2018) et est également en cours de validation dans le troisième. Nous estimons donc qu'il convient de clarifier et de corriger un certain nombre d'affirmations présentées dans le rapport du LCO.

Génotypage probabiliste

Le génotypage probabiliste (GP) consiste en l'utilisation de méthodes statistiques et d'algorithmes mathématiques pour l'analyse de l'ADN dans un contexte judiciaire. Il s'agit d'une approche qui permet de trouver une réponse mathématique à un résultat ADN. Cette méthode peut être utilisée sur des profils provenant d'un seul individu (souvent référés comme étant des profils purs) ainsi que sur des résultats d'ADN provenant de plusieurs personnes ou individus (souvent nommés comme étant des combinaisons d'ADN). Le logiciel détermine, selon le modèle biologique et les paramètres propres à chaque laboratoire, quels sont les profils génétiques (génotypes) pouvant expliquer le résultat. Ensuite, le logiciel pondère chaque génotype ou ensemble de génotypes selon leur probabilité. Ces probabilités sont par la suite utilisées, lors d'une seconde opération, pour faire le calcul du Rapport de vraisemblance (*Likelihood ratio* ou LR), selon les hypothèses choisies et les études de population pertinentes.

Ainsi, le GP ne permet pas du tout d'évaluer quel « type » de suspect est recherché, tel que formulé à la page 11 du rapport. Il permet plutôt de dresser la liste des génotypes possibles du ou des contributeurs qui pourraient expliquer le résultat ADN obtenu, et ce, sans égard au profil d'un individu ciblé. Ainsi, les outils de GP ne visent pas à « produire de la preuve inculpatoire » tel que décrit à la page 8 du rapport, ils



Canadian Scientific Working Group on DNA Analysis Methods

visent à résoudre le résultat ADN, peu importe la finalité. Ces outils peuvent ainsi permettre l'exclusion autant que l'inclusion d'un profil génétique donné dans un résultat ADN. Aussi, que le GP soit utilisé sur des profils provenant d'un seul individu ou plusieurs individus, les calculs présentent le même niveau de rigueur et de validité.

Divulgence, boîtes noires et transparence

Comme toute technologie, les logiciels de GP doivent faire l'objet d'une validation approfondie avant d'être mise en œuvre dans un laboratoire de science judiciaire. Ce processus prend généralement plusieurs mois et est effectué par des scientifiques compétents dans chacun des laboratoires où le GP est implanté. En tant que laboratoire accrédité aux normes de qualité internationales, ce qui est le cas des trois laboratoires judiciaires publics canadiens, des audits sont effectués par des scientifiques indépendants et impartiaux, externes à l'organisation, au cours desquels les validations sont examinées et évaluées pour s'assurer qu'elles répondent aux normes appropriées. Il n'y a donc aucun obstacle à la transparence des données de validation, comme le laissent entendre les auteurs. En outre, les résumés de validation des laboratoires canadiens sont mis à la disposition des parties juridiques, sur demande et lorsque cela est pertinent.

D'ailleurs, le CSJ a divulgué ses résumés de validation à plusieurs reprises en prévision d'éventuelles contestations. À la page 18 du rapport, il est fait référence à un cas dans lequel un défendeur s'est vu refuser l'accès aux documents de validation du GP à la suite d'une demande de divulgation au Québec. Dans ce cas, le laboratoire concerné a proposé à la défense une consultation de toutes les données, sur place, ainsi que la transmission partielle des données, selon les sections d'intérêt relatives au dossier en question. Il est à noter que cette même validation, réalisée par le LSJML, a également fait l'objet d'une publication avec révision en 2019 [1], qui peut être consultée par tous.

Enfin, l'Institut des sciences et de la recherche environnementales de Nouvelle-Zélande (ESR), fabricant du logiciel de GP utilisé par les trois laboratoires de laboratoires de sciences judiciaires membres du SWGDAM canadien, propose aux procureurs des deux parties, entre autres, un mécanisme d'accès au code et aux détails des algorithmes utilisés, sur demande et gratuitement (<https://www.strmix.com/news/strmix-updates-defense-access-policy/>).

Validité scientifique et partialité

Grâce aux validations effectuées au Canada et dans au moins 90 laboratoires internationaux, il est possible d'affirmer que le GP est une approche objective et valide pour l'usage auquel elle est destinée. Elle permet de considérer tous les aspects d'un résultat d'analyse génétique d'une manière cohérente et mathématique sans considérer le profil du suspect. Le seul intrant non mathématique est l'évaluation du nombre de contributeurs dans le résultat. Il s'agit là d'un élément essentiel de l'interprétation de l'ADN,



Canadian Scientific Working Group on DNA Analysis Methods

pour lequel les experts sont dûment formés et qualifiés, et qui était fait bien avant que le GP ne soit disponible.

Le GP permet une analyse encore plus objective que les méthodes employées précédemment. C'est la raison pour laquelle les auteurs du rapport PCAST de 2016 [2], dont il est question dans le rapport du LCO, encouragent son utilisation. Leur préoccupation ne concernait que les contributeurs présents en faibles proportions (définies par eux comme inférieures à 20% du total) pour lesquels, au moment de la rédaction du rapport, peu de données avaient été formellement publiées. Toutefois, les publications effectuées depuis, ainsi que plusieurs validations non publiées, ont démontré que les LR obtenus pour des contributeurs en aussi faibles proportions reflètent correctement la force d'une association et sont donc fiables [1, 3, 4].

Rapports de vraisemblance

Le Rapport de vraisemblance (LR) est abondamment cité dans le rapport du LCO. Il est toutefois important de noter que l'utilisation du LR n'est pas exclusive au GP. D'ailleurs, des LR sont utilisés pour exprimer la force de preuves génétiques depuis bien avant l'avènement du GP, en relation avec certaines combinaisons d'ADN ainsi que dans les cas de paternité ou autres cas impliquant des comparaisons familiales, tant en matière civile que pénale. La probabilité de concordance aléatoire (Random match probability ou RMP en anglais), qui est considérée comme une norme de référence dans le rapport du LCO et qui est utilisée dans tout le Canada depuis près de 30 ans, est l'équivalent du LR lorsqu'elle est utilisée pour des profils purs. Comme le LR, la RMP est une évaluation des probabilités, pondérées les unes par rapport aux autres. Ce calcul oppose la probabilité du résultat selon deux hypothèses, mutuellement exclusive, et ne nécessite aucune hypothèse de probabilités préalables (c'est-à-dire d'informations non reliées à l'ADN). Dans le cas de la RMP, seules deux hypothèses peuvent être invoquées : 1) l'individu en question est le contributeur du résultat et 2) l'individu en question n'est pas le contributeur du résultat, il s'agit plutôt de quelqu'un d'autre, pris au hasard dans la population qui en est le contributeur.

Lorsque le LR est calculé pour une combinaison d'ADN, le nombre d'hypothèses possibles est tributaire du nombre de contributeurs dans le résultat. Le plus souvent, au moment du calcul du LR, l'hypothèse dite « de défense » n'est pas connue. L'expert doit donc choisir les hypothèses à utiliser pour le calcul du LR. Dans la plupart des cas, les deux hypothèses conduisant au LR le plus conservateur (le plus bas) seront celles retenues. Dans certains cas, un individu sera assumé (ce que l'on appelle conditionner) comme étant un contributeur lorsque, par exemple, il s'agit d'un prélèvement fait sur son corps et qu'il est normal de retrouver son ADN dans un prélèvement fait sur son corps. L'expert choisit donc les hypothèses en étant conservateur, et en tenant compte du contexte de l'affaire. Par ailleurs, chaque laboratoire indique, soit dans le rapport d'analyse, soit dans les documents d'information mis à la disposition des parties, que le LR peut être calculé selon d'autres hypothèses qui peuvent être proposées, sur demande.



Canadian Scientific Working Group on DNA Analysis Methods

Le LR est une évaluation de la probabilité du résultat d'ADN observé selon deux hypothèses et doit finalement être évalué par le juge des faits dans le contexte du reste des preuves dans l'affaire. Par exemple, les preuves génétiques (LR) dans une affaire donnée peuvent étayer l'affirmation selon laquelle le suspect est un contributeur d'une combinaison d'ADN, mais si le suspect n'était pas au pays au moment du crime, n'a jamais été sur le lieu du crime et n'a pas eu de contact avec quelqu'un qui l'a fait, ces facteurs doivent clairement être pris en compte, même si ce n'est pas par l'expert, pour évaluer la véracité de l'affirmation. On peut dire la même chose, que la preuve génétique soit exprimée en tant que LR ou RMP.

Les auteurs du rapport du LCO n'ont pas tort d'affirmer que la présentation de la LR au tribunal peut être un défi. Toutefois, ces difficultés ne sont pas propres au GP ni au LR, mais aux complexités inhérentes aux preuves scientifiques en général. C'est le rôle de l'expert de transmettre les preuves clairement et de ne pas les surestimer ou les sous-estimer. Les principales associations de biologie judiciaire, dont le SWGDAM américain, l'*International Society for Forensic Genetics* (ISFG) et l'*European Network of Forensic Science Institutes* (ENFSI), approuvent toutes l'utilisation de la LR dans un contexte de sciences judiciaires.

En conclusion, nous espérons que les clarifications fournies seront considérées comme utiles et prises en compte lors de la lecture du rapport du LCO.

Cordialement,

Chris Askew
Président du SWGDAM canadien

1. Noël S, Noël J, Granger D, Lefebvre JF, Séguin S. STRmix™ put to the test: 300 000 non-contributor profiles compared to four contributor DNA mixtures and the impact of replicates. *Forensic Sci. Int. Genet.* 2019;41:24-31. doi: 10.1016/j.fsigen.2019.03.017.
2. President's Council of Advisors on Science and Technology, PCAST releases Report on Forensic Science in Criminal Courts. 2016.
3. Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ. Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. *Forensic Sci. Int. Genet.* 2017;29:126–144. doi: 10.1016/j.fsigen.2017.04.004.
4. Bright JA, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix™ – a multi laboratory response to PCAST. *Forensic Sci. Int. Genet.* 2018;34:11–24. doi: 10.1016/j.fsigen.2018.01.003.