



HAL
open science

Quand le Deep-learning prédit l'impact d'une mutation sur l'expression d'un gène

Julie Demars, Raphaël Mourad, Justine Bonnery

► **To cite this version:**

Julie Demars, Raphaël Mourad, Justine Bonnery. Quand le Deep-learning prédit l'impact d'une mutation sur l'expression d'un gène. 2024. hal-04803954

HAL Id: hal-04803954

<https://hal.inrae.fr/hal-04803954v1>

Submitted on 26 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Génétique

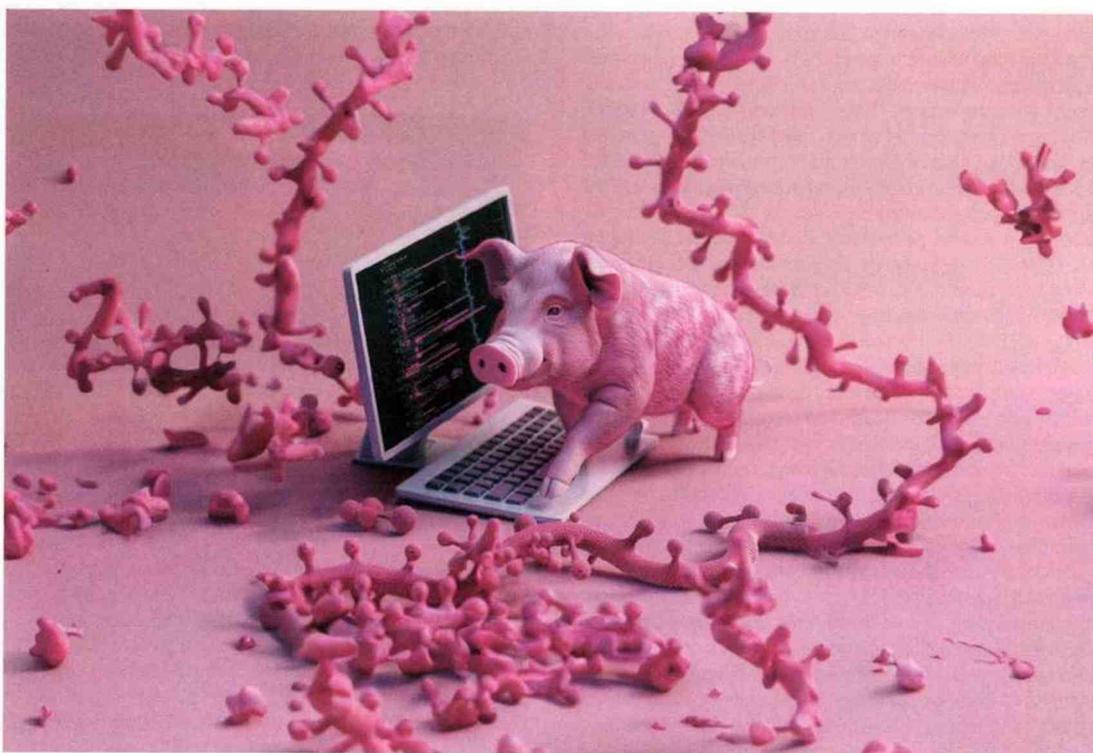
Intelligence artificielle Quand le Deep-learning prédit l'impact d'une mutation sur l'expression d'un gène

Il y a quelques années, la sélection génétique animale a connu une véritable révolution grâce à l'avènement de la génomique, qui a permis de séquencer les génomes et donc de sélectionner beaucoup plus facilement certains traits de caractère essentiels aux programmes de sélection : les phénotypes. Cependant, relier des variants génétiques détectés à des phénotypes d'intérêt reste toujours compliqué. Le programme de recherche Obama - porté par deux laboratoires du centre [Inrae](#) Occitanie-Toulouse, Génétique, physiologie et systèmes d'élevage (GenPhySE), et Mathématiques et informatique appliquées de Toulouse (Miat) - combine l'intelligence artificielle et la génomique animale pour améliorer la compréhension de l'influence des variants génétiques sur les phénotypes. « Comment peut-on, dans un protocole de génétique, revisiter des résultats que nous avons acquis il y

a longtemps, avec de nouvelles approches, et être beaucoup plus précis et résolutifs que ce que nous étions capables de faire jusqu'à présent ? », questionnent la biologiste Julie Demars ([GenPhySE/Inrae](#)) et l'enseignant-chercheur, maître de conférences à l'université Toulouse III Paul Sabatier Raphaël Mourad ([Miat/Inrae](#)) qui mènent ce projet.

Protocole de génétique PorcQTL

« Nous repérons les régions du génome associées à des caractères d'intérêt et pour cela, il nous faut disposer de variants génétiques et de phénotypes », explique Julie Demars. Les phénotypes proviennent de l'unité expérimentale GenESI (Élevages porcins innovants) localisée sur le site du Magneraud en Charente-Maritime où un premier programme français de génétique, appelé PorcQTL, avait croisé des animaux de races Large White et Meishan. « La race Meishan présente de bonnes aptitudes pour des caractères de reproduction et de qualités maternelles et les truies sont capables de sevrer beaucoup de porcelets. La race Large White est reconnue pour ses caractères de production. L'idée est de tirer parti de ce croisement pour prendre le meilleur des deux pour différents caractères et de voir, en termes de génétique, quelles sont les régions du génome qui peuvent être impliquées dans des caractères d'intérêt pour la filière porcine. » Si dans les années 1990, les caractères de production reposaient plutôt sur le taux de muscles, de carcasse, de viande



C'est quoi cette image bizarre ? C'est la vision qu'a l'intelligence artificielle de la génétique porcine.



maigre, de lipides intramusculaires, sur les qualités organoleptiques et de reproduction, aujourd'hui la sélection intègre également des caractères de comportement. « Par rapport à la demande sociétale en bien-être animal, certains caractères comme la docilité et le comportement maternel attentif envers ses porcelets sont de plus en plus importants dans les élevages, tout en ne détériorant pas les caractères de production. Et l'intérêt de ce protocole PorcQTL est que les mesures de comportement ont été faites. »

Association statistique : études d'associations pangénomiques

« La question est de savoir, parmi tous les variants génétiques, quels sont ceux qui vont avoir un lien et un impact vis-à-vis du phénotype considéré ? Si nous cherchons à augmenter la masse grasseuse par exemple, la majorité des variants génétiques n'ont aucun lien, mais peut-être qu'une petite fraction en aura un... Donc l'objectif va être d'établir ce lien de manière statistique : nous allons aller chercher l'association statistique entre certains variants génétiques identifiés et le phénotype. » Ainsi, parce que le ou les variants qui ont un rôle causal sur le phénotype directement ne peuvent pas être identifiés, Julie Demars et Raphaël Mourad vont identifier une région et l'associer au phénotype d'intérêt. « Dans ce protocole PorcQTL, nous avions une région sur le chromosome 1 que nous savons être impliqué dans la croissance et l'engraissement du porc. Seulement, cette région fait encore 300 000 paires de bases, nous avons 6 000 variants et parmi eux, il n'y en a peut-être qu'un seul voire deux ou trois qui ont un rôle causal. C'est là que Julie Demars arrive à une étape où elle est bloquée », explique Raphaël Mourad. Julie Demars a réussi à aller à la résolution du QTL mais n'a pas pu faire le criblage des mutations causales.

C'est à présent qu'intervient l'intelligence artificielle, et plus précisément l'apprentissage profond ou Deep-learning. « Si avec l'intelligence artificielle, nous arrivons à en mettre une dizaine en avant, et qu'avec quelques manipulations fonctionnelles, nous parvenons à dire que deux ou trois ont un effet sur un mécanisme moléculaire, ce sera déjà très bien ! »

L'apprentissage profond ou Deep-learning

« Le génome de l'homme et celui du porc sont si similaires, que tout ce que nous avons développé en génétique humaine - les connaissances, les outils et les méthodes - de surcroît avec souvent 10 ans d'avance, peut être utilisé chez le porc. L'apprentissage profond ne va pas prédire directement l'impact sur le phénotype, mais va prédire l'impact de mutations sur des phénomènes moléculaires : la transcription des gènes, l'activité de la chromatine, la fixation de facteurs de transcription sur l'ADN... Nous disposons de très grandes quantités de données fonctionnelles au niveau moléculaire chez l'homme, mais beaucoup moins chez le porc. Donc, en partant du principe que les propriétés du génome de l'homme au porc changent peu, nos prédictions devraient être assez bonnes. L'intérêt est de considérablement réduire la quantité d'expériences fonctionnelles à faire », témoigne Raphaël Mourad. Ainsi, parmi les 6 000 variants,

ils vont pouvoir prédire l'impact de toutes ces mutations au niveau moléculaire. Dérégulent-elles l'expression d'un gène ? « Avec l'apprentissage profond et les données de l'homme et de la souris, nous sommes capables de dire si un variant génétique empêche la fixation d'un facteur qui lui, va être impliqué dans l'expression du gène d'à côté. Cela permet de réduire la liste des variants qui ont un effet », explique Julie Demars qui est donc passée de 6 000 mutations sur ses premières données à une centaine. « Là, en effet cela débloque vraiment la situation, mais il faut garder en tête que cela ne va pas tout résoudre. Nous sommes censés maintenant aller beaucoup plus vite et faire un bond en termes de connaissances, cependant je reste vigilante en termes d'applications pour la filière sur l'identification de mutations car tant que nous n'avons pas d'explication du mécanisme, nous ne pouvons pas répondre à cette question. Mais je n'ai pas de doute qu'on puisse tomber sur un gène très connu dans la littérature. »

Avec ce projet, Julie Demars voit un intérêt pour les schémas génétiques car « ils génèrent de la donnée de génotypage et de séquençage qui peut nous permettre d'améliorer la finesse des modèles à développer. Si nous identifions quelques variants intéressants pour des caractères souhaités par la filière et en sélection, ils seront potentiellement ajoutés sur les puces de sélection génomique afin d'améliorer les valeurs génomiques des reproducteurs et les index de sélection ». ● Justine Bonnery



« Nous allons chercher l'association statistique entre certains variants génétiques identifiés et le phénotype. »

Raphaël Mourad, enseignant-chercheur, maître de conférences à l'université Toulouse III Paul Sabatier



« Les modèles d'apprentissage profond peuvent prédire l'impact d'une mutation sur des phénomènes moléculaires. »

Julie Demars, biologiste à l'unité GenPhySE/Inrae



Le métaprogramme Digit-bio : quand la biologie numérique explore le vivant