

**CONFÉRENCE  
CIFAR-GETX LE  
MEILLEUR DES  
MONDES: UTILISATION  
DES RÉSEAUX  
GÉNÉTIQUES POUR  
DÉCODER LA MALADIE**

**4 ET 5 JUIN 2018**

EN PARTENARIAT AVEC

**CIFAR**



**McLAUGHLIN**  
CENTRE



# JOUR 1

## ALLOCUTION PRINCIPALE

### GEORGE CHURCH, UNIVERSITÉ HARVARD – PERSONNALISATION DU GÉNOME

- Le séquençage du génome humain nous a permis de mieux comprendre la nature des variantes génétiques qui influencent le risque de maladie et comment les gènes influencent la réaction d'une personne aux médicaments.
- Depuis 2005, le Projet du génome personnel (PGP) offre le seul répertoire en libre accès sur la génomique humaine, les données environnementales et reliées aux caractères, ouvrant la voie à la médecine de précision.
- Dans les vingt dernières années, le coût du séquençage d'un génome humain est passé de 3 milliards de dollars à environ 1000 dollars, et pourrait tomber à 100 dollars par génome d'ici cinq ans. Cette réduction des coûts est essentielle pour accroître la participation à des projets tels que le PGP et, par le fait même, accroître les données disponibles pour renforcer le pouvoir prédictif de l'interprétation du génome humain.
- Les gens sont encore réticents à faire séquencer leur génome. Un des mécanismes pour permettre aux scientifiques d'avoir accès à davantage de données serait d'offrir aux gens un incitatif financier pour faire séquencer leur génome.

## SÉANCE 1 : FACTEURS COMPLEXES QUI INFLUENCENT L'EXPRESSION DES GÈNES

### ANIMATEUR :

Hannah Carter, Chercheuse mondiale CIFAR-Azrieli, Université de San Diego

### PRÉSENTATEURS :

David Kelley, Calico – Interprétation de la fonction à partir de la séquence à l'aide de l'apprentissage profond

Trey Ideker, Université de la Californie à San Diego – Utilisation de réseaux pour traduire le génotype en phénotype

Jasper Rine, Université de la Californie à Berkeley – Espace : La dernière frontière du silençage génique

- De nombreuses variantes génétiques qui contribuent à des maladies complexes se trouvent dans des régions régulatrices non codantes du génome. L'avènement de l'intelligence artificielle (IA) et de l'apprentissage automatique a fourni un outil puissant pour analyser de vastes ensembles de données et prédire la fonction d'éléments régulateurs auparavant inconnus.
- Grâce à l'analyse locale de séquences régulatrices, de meilleurs modèles et des volumes de données plus importants favorisent les avancées et rehaussent notre compréhension de la façon dont ces séquences se comporteront dans les dix prochaines années. Dans dix ans, nous devrions pouvoir cartographier tous les réseaux biologiques pour construire automatiquement un modèle de la cellule.
- La traduction du génotype d'un patient en phénotype constitue encore un défi. Au lieu d'essayer de lier les nombreuses variations du génome directement au phénotype du patient, le recours aux réseaux génétiques en tant

qu'architecture qui nous éclaire sur un réseau neuronal peut rehausser notre compréhension de la connexion entre le génotype et le phénotype. Grâce à cette méthode, il est possible de fonder la modélisation prédictive sur la biologie cellulaire réelle et d'accroître ainsi la puissance prédictive pour les soins de santé et d'autres domaines.

- On ne sait pas vraiment si les algorithmes d'apprentissage automatique sont assez intelligents pour faire le relevé de toute la biologie cellulaire, même avec suffisamment de données. On se demande de quoi dépendent les modèles en cours d'élaboration, comme la nature des

données produites par les expérimentateurs. Conséquemment, il est encore nécessaire de bien comprendre la biologie réelle dans le développement et l'analyse des résultats de ces modèles.

- En plus de la séquence d'ADN, nous devons également tenir compte de l'environnement local dans la régulation de l'expression génétique. Les changements épigénétiques peuvent produire une autre dimension de complexité dans l'expression et la régulation des gènes, et doivent être pris en compte lors de la conception de traitements qui tirent parti du silençage génique.

## SÉANCE 2 : INTERACTIONS GÉNÉTIQUES, PÉNÉTRANCE ET MALADIE HUMAINE

### ANIMATEUR :

Stephen Scherer, Boursier principal du CIFAR, Centre McLaughlin de médecine moléculaire, Université de Toronto

### PRÉSENTATEURS :

Adam Shlien, Hôpital pour enfants malades – Repérer des motifs de mutation dans le cancer pédiatrique

Jason Moffat, Boursier principal du CIFAR, Université de Toronto – Identification de gènes essentiels par la cartographie des interactions génétiques

Shamil Sunyaev, École de médecine de Harvard – Variation génétique et pénétrance

- Les mutations somatiques peuvent entraîner une rechute du cancer. Des méthodes et des outils novateurs, comme la modélisation prédictive, sont en cours de développement pour suivre l'évolution et la trajectoire des mutations somatiques tout au long de la vie d'une tumeur, permettant ainsi aux médecins de prendre une longueur d'avance sur la progression tumorale.
- L'étude des interactions génétiques pourrait rehausser notre compréhension de gènes dont la fonction est actuellement inconnue. L'examen de l'effet de combinaisons de mutations dans de multiples gènes permettra de comprendre comment les gènes se combinent et entraînent des maladies communes et complexes. CRISPR permettra de cartographier systématiquement les interactions génétiques à un rythme sans précédent et donnera lieu à de nouvelles annotations de la fonction des gènes et à l'ordonnancement des gènes dans les voies cellulaires pour la première fois.
- Les études utilisent souvent la variation génétique pour repérer des SNP qui sont associés à des symptômes pathologiques et qui sont statistiquement plus susceptibles d'être associés à une maladie. Il faut faire attention quand on utilise des termes comme « cooccurrence » ou « association » relativement à ces gènes de maladies éventuels, en grande partie à cause du facteur de pénétrance. Il faut tenir compte de la pénétrance dans la conception de méthodes pour prédire l'effet fonctionnel des variantes génétiques, ainsi que dans la détermination des variantes qui sont importantes dans la maladie.
- Des études moléculaires d'une mutation considérée comme bénigne pour l'ensemble du phénotype

peuvent démontrer que ses interactions mènent à une condition qui la rend délétère. Une mutation est considérée comme pathogène lorsqu'elle présente un lien de causalité avec le phénotype de la maladie en question à une pénétrance relativement élevée.

- Chaque nouvelle découverte d'un gène associé à

la maladie est une ancienne découverte d'un gène associé à la maladie dont l'expression phénotypique est différente. Si une variante est jugée d'importance inconnue, il se peut que le bon phénotype ne soit pas à l'étude ou que le phénotype ne porte pas encore de nom clinique. Il s'agit là d'un défi quotidien.

## SÉANCE 3 : TECHNOLOGIES TRANSFORMATRICES

### ANIMATEUR :

Hu Li, Clinique Mayo

### PRÉSENTATEURS :

Chad Myers, Boursier du CIFAR, Université du Minnesota – Méthodes computationnelles pour mieux comprendre la maladie complexe

Jennifer Listgarten, Université de la Californie à Berkeley – L'apprentissage automatique rencontre l'édition génomique

Prashant Mali, Université de la Californie à San Diego – Repousser les frontières des applications de CRISPR

- Toute maladie complexe est associée à plusieurs variantes et il semblerait que chaque variante contribue à la maladie dans une certaine mesure. Pourtant, la modélisation de toutes les variantes n'expliquerait pas complètement la maladie. La maladie complexe met probablement en jeu des interactions génétiques, car les mutations peuvent se combiner et devenir pathogènes. Ce domaine est trop vaste pour être mesuré par voie expérimentale; il nous faut des méthodes computationnelles pour tirer des inférences sur les réseaux d'interactions génétiques et concevoir des modèles prédictifs afin d'approfondir notre compréhension des interactions entre les fonctions des gènes.
- La quantité sans précédent de données produites par les études à haut débit va révolutionner notre compréhension de la biologie. Cependant, l'espace combinatoire est trop grand pour la mesure expérimentale; par exemple, les mutations par paires créent 441 millions de mesures expérimentales. Des méthodes informatiques sophistiquées sont nécessaires pour interpréter les données et cela permettrait la création de modèles cellulaires holistiques. De nouvelles méthodes d'apprentissage automatique pourraient inférer des modèles prédictifs à partir de réseaux d'interactions génétiques et ainsi rehausser notre compréhension de la fonction et de l'interaction des gènes.
- De nouveaux outils d'apprentissage automatique pourraient accroître l'efficacité de CRISPR-Cas9 pour qu'il maintienne sa cible et qu'il cible la séquence génétique optimale afin d'obtenir l'effet d'inactivation désiré. Pour entraîner les algorithmes, les informaticiens doivent collaborer avec les biologistes pour mesurer systématiquement l'efficacité de CRISPR pour plusieurs gènes. Par la suite, les algorithmes peuvent tirer des leçons de données issues d'exemples du monde réel et prédire les sites cibles pour des gènes complètement différents qui n'ont pas fait l'objet de mesures
- Le génome lui-même est maintenant une cible « droguable ». Dans le contexte de l'oncologie de précision, on utilise CRISPR pour découvrir de nouvelles interactions létales synthétiques. Cela peut favoriser la découverte de nouvelles polythérapies où il est possible d'exploiter la mutation oncogène et de la cibler en association avec un médicament qui cible une voie parallèle. L'interaction létale synthétique produite peut donc induire la mort spécifique des cellules cancéreuses tout en épargnant les cellules normales.

- En ce qui concerne la découverte de médicaments, une grande partie de ce qui s'applique aux interactions génétiques s'applique aussi aux interactions chimiques-génétiques. De nombreuses méthodes d'interaction génétique mises au point dans des organismes modèles comme la levure sont maintenant utilisées comme modèle pour les cellules humaines. Les scientifiques peuvent utiliser les interactions chimiques-génétiques pour établir des profils des effets des composés et, en retour, concevoir des cartes de plus en plus précises de la cible des composés dans la cellule.
- L'intégration des données est complexe : les données peuvent sembler robustes, mais les artefacts deviennent apparents. De multiples méthodes sont nécessaires pour résoudre ce problème, comme l'analyse de nombreuses lectures différentes, l'utilisation de la validation croisée et la reproduction des résultats dans plusieurs groupes de recherche.
- Pour ce qui est de l'avenir, il sera important de trouver le bon cadre pour effectuer un dépistage de la maladie. Par exemple, il faudra comprendre si un dépistage tient compte de l'hétérogénéité de fond et est représentatif de la tumeur. La mise au point d'un cadre véritablement représentatif du système pathologique pourrait permettre l'obtention de résultats plus significatifs.

## SYNTHÈSE ET PERSPECTIVES : JOUR 1

- Le « meilleur des mondes » est déjà là. Nous pouvons lire la séquence génomique, écrire la séquence génomique en utilisant la biologie synthétique et modifier la séquence génomique en utilisant CRISPR.
- Maintenant que les technologies sont meilleures et moins coûteuses, il nous faut des données pangénomiques plus exhaustives.
- L'accessibilité des données est un élément important – notamment, un accès plus large aux ensembles de données publiés provenant d'organismes modèles.
- Pour tirer profit des réseaux génétiques, il faut mettre davantage l'accent sur la façon d'intégrer les leçons apprises des différents organismes modèles – de la levure aux organismes multicellulaires.
- Il faut créer des équipes interdisciplinaires fonctionnelles pour contribuer à cerner les questions qui se trouvent à la frontière de ce meilleur des mondes. La formation interdisciplinaire représente un énorme défi; les biologistes et les experts en informatique doivent communiquer entre eux, et les systèmes doivent permettre à ces interactions de se multiplier

## JOUR 2

### SÉANCE 4 : VARIATION NATURELLE ET GÉNÉTIQUE POPULATIONNELLE

#### ANIMATEUR :

Philip Awadalla, directeur, biologie computationnelle, Institut ontarien de recherche sur le cancer

#### PRÉSENTATEURS :

Nancy Cox, Université Vanderbilt – Examen du paysage phénotypique de la population

Frederick Roth, Boursier principal et codirecteur d'un programme du CIFAR; Université de Toronto – Interprétation de variantes génétiques rares

- Les biobanques sont des moteurs de découverte des relations gènes-phénotypes. La force d'une biobanque réside dans sa capacité d'examiner l'ensemble d'un phénomène médical à la fois. En recueillant des données médicales et des échantillons auprès de vastes populations de millions de participants, elles permettent de réaliser de puissantes études pour améliorer notre compréhension de la maladie complexe.
- L'interprétation du risque de maladie associé à des variantes dans des génomes humains individuels constitue un défi majeur en médecine. Comme nous connaissons seulement les effets d'une petite portion des variantes cataloguées, il est encore difficile de cerner les variantes génétiques importantes. Vu la difficulté que représente la réalisation d'études humaines du type et de la taille nécessaires à l'interprétation de variantes rares, un système de « levure humanisée » permet de cerner le dysfonctionnement moléculaire de toutes les variantes cliniques chez les patients atteints d'une pathologie donnée. Cela permettra de mieux comprendre la relation entre les variantes faux-sens rares et la fonction protéique.
- Au fil de l'amélioration des technologies de séquençage, il faudra passer du séquençage ciblé au séquençage pangénomique. Vu les problèmes d'extensibilité, l'analyse et l'interprétation du séquençage pangénomique posent un défi qui s'ajoute à ceux du coût et du temps.
- De nouveaux outils informatiques pour interpréter le séquençage du génome peuvent fournir aux cliniciens des données fonctionnelles et des prédicteurs computationnels. Toutefois, quand les cliniciens utilisent les résultats de recherche en clinique, ils veulent que des résultats probants viennent étayer une décision finale qui intègre leur expertise, leur connaissance des antécédents familiaux et le phénotype. Ils ne veulent pas de termes, comme « probablement pathogène » ou « bénin ». Les rapports de vraisemblance, qui indiquent l'augmentation de la probabilité découlant de résultats probants sur la pathogénicité, seront plus utiles

## SÉANCE 5 : SUPPRIMER LE CLOISONNEMENT ENTRE LE MILIEU UNIVERSITAIRE, L'INDUSTRIE ET LA CLINIQUE – DES SYSTÈMES MODÈLES AUX POPULATIONS HUMAINES

### ANIMATEUR :

Andrew Hessel, Humane Genomics Inc

### PRÉSENTATEURS :

Calum MacRae, Université Harvard – Génétique humaine et l'étape limitante : Le phénotype

Marc Fiume, DNA Stack – L'Internet est le meilleur outil pour le partage d'information

Amit Deshwar, DeepGenomics – Création d'un nouvel univers de médicaments génétiques

- L'apparence de leur génome importe peu aux patients; ce qui leur importe c'est de se sentir mieux et de vivre plus longtemps. Il faut intégrer les phénotypes et la génomique aux interventions qui appuieront ces résultats.
- Nous mesurons les mauvaises choses. Nous devons réfléchir systématiquement à l'endroit où pourrait se trouver de l'information utile (p. ex., les mécanismes cellulaires) et à la façon dont nous pouvons créer des plateformes pour recueillir ces lectures et ces données à l'échelle nécessaire pour tirer des conclusions significatives.
- Nous devons exploiter les progrès technologiques récents dans des domaines aussi divers que la reconnaissance faciale, les produits technovestimentaires et la réaction à court terme aux médicaments pour recueillir des données et avoir une vision systémique d'une maladie.

Les chercheurs doivent formuler des règles généralisables pour interpréter ces données afin de contribuer aux modèles de prédiction de la maladie. De la sorte, la biologie et la physiologie cellulaires pourraient se retrouver au chevet du patient.

- Le développement pharmaceutique se penchera d'abord sur les maladies rares, car la cartographie génotype-phénotype est très claire et le besoin médical est évident, mais ces produits pharmaceutiques pourraient un jour traiter toute sorte de maladies complexes. Des projets de séquençage à grande échelle nous permettront d'identifier davantage de valeurs aberrantes dans des phénotypes très extrêmes et, une fois la mutation identifiée, nous pourrions créer une cible thérapeutique.
- Le système actuel de la R. et D. dans le domaine pharmaceutique ne fonctionne pas. Le processus de découverte est lent et coûteux pour faire passer un médicament de la découverte à l'expérimentation et à l'essai. Des milliers de maladies n'ont pas de traitement connu et beaucoup d'autres sont associées à d'importants besoins médicaux non satisfaits. Malgré d'énormes percées en génomique, en séquençage et en biologie des systèmes, le rendement du capital investi dans la R. et D. pharmaceutique n'a cessé de diminuer au cours des 25 dernières années pour se situer sous le seuil de l'investissement en capital.
- Nous avons grand besoin de l'automatisation complète du processus de découverte de médicaments; la capacité prédictive des outils d'apprentissage automatique peut nous aider en la matière.
- La génomique n'a pas accès à des ensembles de données de haute qualité et de grand volume, et il n'y a pas de volonté ni d'engagement en ce qui concerne le partage des données. La production de données est le premier facteur limitant : en raison de

la fragmentation de l'infrastructure dans le paysage de la génomique, il est difficile d'utiliser les données, ainsi que d'y avoir accès et de s'y fier, ce qui entraîne un très faible RCI pour la production de données.

- Le deuxième facteur limitant est la capacité de partage des données. Le partage des données est essentiel, car aucune organisation ne peut à elle seule produire suffisamment de données pour guérir toutes les maladies complexes et rares. L'Internet a démontré que nous pouvons partager des données et, dans presque tous les secteurs, nous voyons que cela a un effet transformateur.
- L'Internet serait à la base d'un monde meilleur pour la génomique afin de permettre à l'apprentissage automatique de favoriser des découvertes plus rapides et plus complexes qui augmentent le rendement de l'investissement associé à la production de données. Un Internet sécurisé pour la génomique ouvrirait la voie à des intégrations techniques qui permettraient aux données de circuler de façon sûre et efficace entre les systèmes, de fournir un accès sur demande, de briser les silos de données traditionnels et de contribuer à optimiser la valeur des données génomiques et de santé.

---

Jing Hou, Ph.D., a contribué à la rédaction de ce rapport.