



DOCUMENT D'INFORMATION

Résultats du Concours 2017 en bio-informatique et en génématique (BI/G) (projets du volet multisectoriel)

Génome Canada est heureux d'annoncer les 25 projets retenus en vue d'un financement à la suite de son Concours 2017 en bio-informatique et en génématique. Au total, 23,1 millions de dollars sont investis dans ces projets qui produiront la nouvelle génération d'outils et de méthodologies nécessaires au traitement des gros volumes de données produits par les technologies modernes de la génomique et fourniront à la communauté des chercheurs un large accès à ces outils.

Les bailleurs de fonds de la recherche au Canada reconnaissent expressément que la bio-informatique et la génématique sont une priorité partagée, car ces domaines seront la clé grâce à laquelle on réussira à extraire le sens des ensembles de données de plus en plus complexes.

ALBERTA

Trousse complète d'outils d'analyse et navigateur génomique à haut rendement pour une caractérisation rapide, fiable et approfondie des génomes bactériens

Directeurs de projet : Paul Stothard, Université de l'Alberta; Gary Van Domselaar, Agence de la santé publique du Canada – Laboratoire national de microbiologie

Centre de génomique : Genome Alberta

Volet : multisectoriel

Financement total : 940 000 \$

Le séquençage de l'ADN est devenu plus facile et moins coûteux, ce qui mène à l'expansion rapide des bases de données génomiques et à une meilleure compréhension de la vie en général. Pourtant, même si la séquence d'un génome donne des indices importants sur l'histoire et les caractéristiques d'un organisme, les données produites par les technologies actuelles de séquençage ne viennent pas dans une forme facile à interpréter. Les résultats sont souvent produits en petits fragments, un peu comme les mots ou les expressions qui forment une recette ou description plus significative, mais confuse. De plus, les définitions d'un grand nombre de mots et d'expressions que l'on trouve dans l'information sur les séquences d'ADN souvent ne sont pas claires. Les génomes, comme les recettes, sont plus faciles à comprendre

s'ils sont écrits dans un langage familier, présentés dans un ordre logique et accompagnés d'images utiles.

De nombreux chercheurs ont aujourd'hui de la difficulté à traiter leurs données de séquençage en une forme qui convient à leurs applications de recherche et se tournent vers les bio-informaticiens pour trouver de l'aide. Ce recours à d'autres est toutefois inefficace, en particulier dans le domaine de la génomique microbienne pour lequel il existe peu de bio-informaticiens chevronnés.

Paul Stothard, Ph. D., de l'Université de l'Alberta et Gary Van Domselaar, Ph. D., de l'Agence de la santé publique du Canada supervisent la mise au point de *Proksee*, un système logiciel grâce auquel les chercheurs pourront convertir les données brutes de séquençage de bactéries en assemblages pangénomiques de haute qualité, hautement décrits et faciles à interpréter (autrement dit, des recettes de la vie microbienne). Ces ressources d'analyse puissantes seront gratuitement mises à la disposition de tous les chercheurs, y compris ceux qui travaillent dans des petites entreprises canadiennes sans l'expertise ou les ressources pour transformer en un format utilisable les données brutes de séquençage.

En aidant la communauté scientifique canadienne à transformer l'information des séquences d'ADN en nouvelles connaissances sur la biologie d'importantes espèces bactériennes, la plateforme perfectionnée et conviviale *Proksee* appuiera des découvertes scientifiques et des innovations dans divers domaines, dont la création d'aliments plus sûrs et plus sains, de nouvelles techniques de lutte contre la pollution et de nouvelles méthodes de fabrication de produits importants.

ATLANTIQUE

Résistance antimicrobienne : émergence, transmission et écologie (ARETE)

Directeurs de projet : Robert Beiko, Université Dalhousie; Fiona Brinkman, Université Simon Fraser

Centres de génomique : Genome Atlantic, Genome British Columbia

Volet : multisectoriel

Financement total : 1 400 000 \$

Les bactéries résistent de plus en plus aux agents antimicrobiens, ce qui est inquiétant pour l'industrie agroalimentaire où la surutilisation et la mauvaise gestion des antibiotiques ont joué un rôle important. Une seule solution n'inversera pas cette tendance et ne redynamisera pas le pipeline de découverte d'agents antimicrobiens. Il est donc devenu prioritaire, partout dans le monde, de mieux comprendre les gènes qui rendent les bactéries résistantes et leur mode de propagation. Le partage des gènes de résistance entre les bactéries pathogènes est un élément clé de la transmission de la résistance. Les bactéries résistantes se déplacent également entre les habitats, par exemple le sol agricole et les animaux de ferme, mais il faut mieux comprendre les principaux points de transmission pour définir les priorités dans les activités de surveillance et la réglementation.

Robert Beiko, Ph. D., de l'Université Dalhousie et Fiona Brinkman, Ph. D., de l'Université Simon Fraser dirigent une grande équipe composée de partenaires universitaires et gouvernementaux qui s'efforcent de déterminer quels gènes sont partagés, quelles bactéries partagent les gènes et le mode de déplacement des bactéries entre les habitats. L'équipe met au point des algorithmes et des logiciels de bio-informatique qui transformeront la manière d'examiner la résistance antimicrobienne (RAM) qui ne sera plus un « instantané statique », mais une vue dynamique de la transmission de la RAM. Elle validera ses outils en utilisant des milliers de génomes de la salmonelle, de l'*E. coli* et d'autres bactéries pathogènes recueillis par les partenaires de l'Agence de la santé publique du Canada et d'Agriculture et Agroalimentaire Canada. Le pipeline de logiciels intégrés qui résultera de ce projet sera ouvert et gratuit.

Grâce au cadre bio-informatique rigoureux créé par le projet, des approches plus éclairées pourront être adoptées pour réduire au minimum les risques posés par la résistance et favoriser la mise en œuvre d'un cadre national d'application de la génomique et de la bio-informatique au continuum « de la ferme à la table » en ce qui concerne la RAM.

COLOMBIE-BRITANNIQUE

AnnoVis : Annotation et visualisation des assemblages *de novo* de génomes et de transcriptomes

Directeur de projet : Inanc Birol, British Columbia Cancer Agency

Centre de génomique : Genome BC

Volet : 2 (multisectoriel)

Financement total : 1 000 000 \$

Il est possible, de nos jours, d'en savoir plus que jamais auparavant sur une espèce en décryptant son génome (ADN) et son transcriptome (ARN). Il est cependant plus difficile de traiter et d'interpréter ces données sous une forme cohérente que de les recueillir. La caractérisation des gènes d'une espèce donnée joue un rôle indispensable dans l'interprétation. Par exemple, la présence de certains gènes dans un génome donné peut renseigner sur la qualité du bois des produits forestiers locaux ou donner une idée des modèles infectieux d'un agent pathogène dans la chaîne alimentaire. L'identification et l'annotation exhaustives et exactes des gènes dépendent toutefois de l'intégralité et de l'exactitude de l'information sous-jacente.

L'utilité des données génomiques et transcriptomiques dépend de deux facteurs : la conception de l'expérimentation et les méthodes d'analyse utilisées. La conception de l'expérimentation est bien définie; Inanc Birol, Ph. D., du BC Cancer Agency se concentre sur les méthodes d'analyse, un aspect encore difficile. Son équipe et lui veulent créer des méthodes qui amélioreront la qualité des génomes et des transcriptomes assemblés en détectant les séquences mal assemblées. Ils mettront également au point des outils et des ressources qui faciliteront l'annotation des gènes et en prédiront la fonction. Ils créeront de plus des outils de visualisation pour évaluer la qualité des assemblages et leurs annotations connexes.

Les outils seront mis à la disposition des chercheurs par le truchement du portail de logiciels de M. Birol et ils aideront à mieux comprendre le monde qui nous entoure à l'échelle génomique et transcriptomique.

Pipeline de quantification amélioré et orienté par la protéogénomique (PIGQpipe) : protéomique ciblée et normes internes des peptides protéogénomiques pour quantifier les variantes révélées dans des expériences protéogénomiques

Directeurs de projet : Christoph Borchers, Yassene Mohammed, Université de Victoria

Centre de génomique : Genome BC

Volet : 1 (santé)

Financement total : 556 472 \$

La mesure des protéines dans des fluides biologiques comme l'urine ou le sang est indispensable au diagnostic, au traitement et à la surveillance des maladies chez les humains. Actuellement, pour cette mesure, les chercheurs et les cliniciens partout dans le monde utilisent les réseaux de spectrographie de masse pour surveiller les réactions multiples et parallèles. Les approches de cette surveillance permettent de quantifier avec précision et exactitude les protéines dans les échantillons, mais elles demeurent dépendantes des bases de données de référence sur les séquences de protéines humaines qui peuvent ne pas contenir des variantes génomiques spécifiques à certaines maladies. Pour utiliser davantage ces méthodes, il faut donc accélérer le développement et la validation des biotests, tout comme la cadence des analyses des échantillons. De plus, l'analyse des données doit être normalisée et automatisée pour permettre une mise en œuvre solide de grande envergure.

Il n'existe actuellement aucun cadre logiciel perfectionné et intuitif pour aider les chercheurs à concevoir des essais à haut rendement en spectrométrie de masse, à traiter les données et à interpréter les résultats, ce qui rend difficile, laborieuse et chronophage l'utilisation de la spectrométrie de masse.

Christoph Borchers, Ph. D., et Yassene Mohammed, Ph. D., de l'Université de Victoria développent un nouveau pipeline de logiciels, PIGQpipe, qui intégrera les données issues des propres expériences des chercheurs aux données des bases des données publiques en ligne afin de créer un « guichet unique » pour tous les aspects de la conception des essais en spectrométrie de masse, d'analyse et d'interprétation des données. À l'aide d'une seule interface Web, PIGQpipe reliera, intégrera et automatisera le flux des essais (y compris le choix des cibles et les conditions optimales d'expérimentation), à l'aide de données pour calculer les valeurs de concentration des protéines, évaluer sur le plan statistique les différences entre les groupes de traitement ou les états pathologiques et présenter les données.

En améliorant la productivité, l'ampleur, la qualité et la portée de la recherche dans ce domaine, PIGQpipe devrait permettre de nouvelles découvertes, dont des outils pour la médecine de précision; faciliter et accélérer l'adoption des technologies de spectrométrie de masse en milieu clinique et bâtir l'expertise dans la communauté scientifique.

Méthodes d'apprentissage machine pour la prédiction de la résistance aux médicaments de bactéries pathogènes

Directeurs de projet : Leonid Chindelevitch, Maxwell Libbrecht, Université Simon Fraser et Jesse Shapiro, Université de Montréal

Centre de génomique : Genome BC

Volet : 1 (santé)

Financement total : 1 000 000 \$

Les options de traitement pour les personnes infectées de bactéries résistantes aux antibiotiques sont limitées de nos jours et pourraient le devenir encore plus au fil du temps. Le risque d'une ère post-antibiotique, pendant laquelle les infections même mineures ou les interventions médicales courantes pourraient être fatales, est réel. Au-delà des coûts personnels, le fardeau économique de la résistance aux médicaments est élevé et selon les estimations, il coûte plus de 1 milliard de dollars par année en Amérique du Nord seulement.

Le dépistage génomique des agents pathogènes pour en déterminer l'identité et la sensibilité aux antibiotiques pourrait être à la fois plus précis et efficace que d'autres méthodes de diagnostic des maladies infectieuses et de choix des traitements appropriés. La difficulté réside toutefois dans deux domaines : l'analyse du grand volume de données produites par ce séquençage pour comprendre les facteurs génomiques qui sous-tendent la résistance aux médicaments et la création d'un modèle précis de prédiction de la résistance aux médicaments à partir des données de séquençage.

Leonid Chindelevitch, Ph. D., et Maxwell Libbrecht, Ph. D., de l'Université Simon Fraser et Jesse Shapiro, Ph. D., de l'Université de Montréal mettent au point des outils informatiques basés sur l'apprentissage machine qui permettront de révéler les relations complexes entre les séquences des génomes bactériens et la résistance aux antibiotiques, d'après les données des génomes bactériens qui existent déjà dans les banques de données publiques et celles que possèdent leurs collaborateurs.

Le projet, qui fait suite à deux projets financés dans le cadre du Concours 2015 en bio-informatique et en géomatique, aboutira à deux outils : un modèle prédictif exhaustif de la résistance à divers médicaments de bactéries précises, qui se mettra à jour par lui-même à mesure que de nouvelles données seront connues; et une interface Web conviviale qui permettra aux chercheurs, y compris les chercheurs en milieu clinique et en santé publique, de télécharger en toute sécurité et d'analyser des données génomiques de bactéries pathogènes et ainsi de prédire avec exactitude la résistance aux médicaments. Cet outil vise à stimuler la recherche sur la résistance aux médicaments et la mise au point de meilleurs outils cliniques, au profit finalement des patients atteints de maladies infectieuses.

Éclairer la matière sombre du métabolome à l'aide de réseaux neuronaux convolutifs

Directeurs de projet : Leonard Foster, Université de la Colombie-Britannique; David Wishart, Université de l'Alberta

Centres de génomique : Genome BC; Genome Alberta

Volet : 1 (santé)

Financement total : 500 000 \$

Le corps humain contient des centaines de milliers de produits chimiques naturels et il est exposé à des centaines de milliers d'autres au cours de la vie quotidienne. Ces produits chimiques, appelés métabolites, participent à des processus biologiques clés chez les humains et d'autres grands organismes. La métabolomique est la discipline scientifique qui permet aux chercheurs de mesurer ces petites molécules à grande échelle. Elle peut révéler des molécules qui interviennent dans des maladies, produire des tests diagnostiques et pronostiques et prédire comment les patients réagiront à des médicaments d'ordonnance précis. On a beaucoup fait avec les outils informatiques actuellement utilisés qui ne peuvent identifier qu'une petite fraction des métabolites dans un échantillon. De meilleurs outils informatiques qui pourraient identifier les autres métabolites accéléreraient de façon spectaculaire le rythme de la recherche en métabolomique.

Leonard Foster, Ph. D., de l'Université de la Colombie-Britannique et David Wishart, Ph. D., de l'Université de l'Alberta mettent au point des outils informatiques basés sur une technique d'intelligence artificielle appelée « apprentissage profond » pour traiter l'immense volume de données produit par les expériences en métabolomique. Le premier outil, DeepMet, augmentera le nombre de molécules qui peuvent être identifiées dans des expériences en métabolomique. Le deuxième, MetUnknown, aidera à attribuer des structures chimiques à des molécules encore inconnues. Ensemble, ces outils aideront à mieux comprendre la majeure partie du métabolome négligée par les outils actuels.

DeepMet et MetUnknown seront offerts dans trois formats différents pour que les chercheurs de différents domaines puissent les utiliser sans formation spécialisée. Au bout du compte, ces outils aideront les chercheurs à identifier des cibles thérapeutiques pour des maladies complexes telles que le cancer et à mettre au point de nouveaux tests qui aideront les médecins à personnaliser les traitements médicaux.

Reconstruction, à l'échelle mondiale, des voies métaboliques à partir des génomes de l'environnement

Directeur de projet : Steven Hallam, Université de la Colombie-Britannique

Centre de génomique : Genome BC

Volet : 2 (multisectoriel)

Financement total : 1 200 000 \$

On peut penser qu'il ne s'agit que de nous, les humains, mais depuis plus de 3,5 milliards d'années, des microorganismes sont la forme de vie dominante sur Terre. Au cours de la dernière décennie, le séquençage à haut débit et la spectrométrie de masse ont produit des données sur l'ADN, l'ARN, les protéines et les métabolites de ces microbes. Toutes ces données nous renseignent sur les fonctions et l'identité des microbes et les relient à un large éventail de fonctions écosystémiques à l'échelle des individus, des populations et des communautés. Il

existe cependant peu d'outils logiciels évolutifs pour extraire et surveiller ces ensembles de données environnementales et s'en servir. Cette lacune est particulièrement contrariante à une époque où nous, en tant que société, devons résoudre les problèmes des changements climatiques mondiaux, alors que les communautés microbiennes offrent un plan virtuel détaillé pour reconstruire notre avenir mondial de manière plus durable.

Steven Hallam, Ph. D., de l'Université de la Colombie-Britannique dirige la création de l'encyclopédie des génomes de l'environnement (EngCyc), un recueil de plans métaboliques détaillés de communautés microbiennes à l'aide d'outils logiciels à haut rendement. Un portail Web donnera accès à l'EngCyc aux fins de consultation et l'utilisateur pourra, il y aura un portail Web et l'utilisateur pourra construire des plans détaillés qu'il aura définis de manière automatisée et évolutive, ce qui permettra de découvrir des gènes et des voies. Ces options d'exploration des données favoriseront l'acquisition et l'application de connaissances.

La combinaison du portail Web et des outils logiciels permettra aux chercheurs, à l'échelle nationale et internationale, d'explorer avec plus d'efficacité les données et de mobiliser les pouvoirs cachés des communautés microbiennes au profit des secteurs du bioraffinage, de l'exploration minière et de l'énergie, pour ne nommer que ceux-là.

Outils bio-informatiques pour le partage et l'analyse fédérés de données d'épidémiologie génomique en temps réel dans le cadre de l'initiative *One Health*

Directeurs de projet : William Hsiao, Université de la Colombie-Britannique; Gary Van Domselaar, Agence de la santé publique du Canada

Centres de génomique : Genome BC; Genome Prairie

Volet : 2 (multisectoriel)

Financement total : 1 164 488 \$

Les maladies infectieuses comme l'influenza, la maladie à virus Ebola, la listériose et la salmonellose, de même que d'autres agents pathogènes peuvent être dévastateurs pour les animaux et les humains, nuire aux économies et paralyser le commerce. L'initiative *One Health* (une seule santé) reconnaît que la santé des humains, des animaux et de l'environnement est étroitement liée et qu'il faut une approche collective pour détecter les éclosions, y réagir et les prévenir efficacement.

La génomique a transformé la détection et la caractérisation des agents pathogènes, accéléré la mise au point de tests diagnostiques et de vaccins et accru nos connaissances sur ces agents pathogènes. L'utilisation des données génomiques pour la surveillance en temps réel chez nos partenaires est cependant difficile en raison de l'information contextuelle peu uniforme associée aux échantillons génomiques, du manque de plateforme de confiance et sécurisée pour le partage des données et de l'insuffisance des outils pour les analyses génomiques localisées et collaboratives.

William Hsiao, Ph. D., de l'Université de la Colombie-Britannique et du BC Centre for Disease Control Public Health Laboratory et Gary Van Domselaar, Ph. D. de l'Agence de la santé

publique du Canada créent de nouvelles plateformes de partage des données et des outils de traitement des données pour permettre le partage en temps réel des données entre plusieurs instances, ce qui permettra aux chercheurs de différents domaines – et aux logiciels qu’ils utilisent – de communiquer l’information plus rapidement, avec plus d’exactitude et en toute sécurité. Trois centres canadiens de collaboration et des partenaires internationaux mettront à l’essai les outils pour s’assurer de leur pertinence et de leur utilité.

Les outils logiciels mis au point dans le cadre de cette collaboration visent à transformer les modes de partage et d’analyse des données sur les maladies infectieuses, ce qui mènera à une meilleure surveillance de l’émergence et de la propagation des agents pathogènes dans la faune, les aliments et les animaux de boucherie, ainsi que dans l’environnement, ce qui réduira le fardeau des maladies, préviendra les embargos sur le commerce agroalimentaire et réduira au minimum les rappels coûteux de produits alimentaires. Ils amélioreront la communication entre les agences de la santé publique et les établissements agricoles et rehausseront la collaboration à l’échelle provinciale, nationale et internationale. En fin de compte, le projet améliorera la santé et le bien-être de la population canadienne.

Élaboration et mise en œuvre d’outils bio-informatiques pour les plateformes de surveillance phylogénétique du VIH et du VHC

Directeurs de projet : Jeffrey B. Joy, Julio S.G. Montaner, Université de la Colombie-Britannique

Centre de génomique : Genome BC

Volet : 1 (santé)

Financement total : 1 200 000 \$

L’Organisation mondiale de la santé (OMS) a fixé des cibles pour l’éradication du VIH/sida en tant que pandémie d’ici 2020 et du virus de l’hépatite C (VHC) d’ici 2030. Dans le monde industrialisé, où les épidémies sont bien gérées, il devient de plus en plus difficile d’identifier les poches restantes de transmission continue du VIH et du VHC, ce qui rend difficiles l’atteinte et le dépassement de ces cibles. Pour faciliter leur atteinte, nous avons besoin de nouveaux outils bio-informatiques qui peuvent identifier les communautés à haut risque d’infection continue par le VIH et le VHC et être utilisés pour appuyer la mise en œuvre des interventions de santé publique et des services de santé.

La surveillance phylogénétique – soit la mobilisation de grandes bases de données de séquences à croissance rapide – peut être une solution pour identifier les grappes de VIH ou de VHC diagnostiquées en temps quasi réel. L’élaboration et la mise en œuvre d’un tel système exigent également de nouveaux outils bio-informatiques qui faciliteront le partage des résultats des analyses avec les représentants de la santé publique, tout en maintenant la confidentialité et la protection des renseignements personnels des patients. Il faudrait aussi que ces outils puissent profiter des données de séquençage de la prochaine génération.

Jeffrey B. Joy, Ph. D., et Julio S.G. Montaner, Ph. D., de l’Université de la Colombie-Britannique élaborent et mettent en œuvre un système semblable. Il servira de plateforme nationale de surveillance phylogénétique en temps quasi réel, capable d’intégrer les données de séquençage

de la prochaine génération et sera associé à des outils bio-informatiques pour présenter et distribuer en toute sécurité les résultats aux agences de la santé publique. L'équipe met au point de nouvelles méthodes de surveillance des épidémies et des infections. Ces outils seront mis gratuitement à la disposition des chercheurs et permettront de cibler avec plus d'efficacité les interventions en santé publique, ce qui nous fera progresser plus rapidement vers l'atteinte des cibles fixées par l'OMS pour limiter et éliminer le VIH et le VHC.

ONTARIO

BridGE-SGA : une nouvelle plateforme informatique pour la découverte d'interactions génétiques à la base de maladies humaines

Directeurs de projet : Charles Boone, Université de Toronto; Chad L. Myers, Université du Minnesota

Centre de génomique : Ontario Genomics

Financement total : 990 910 \$

La capacité de séquencer tout un génome humain à un coût de moins en moins élevé a mené à un changement fondamental en recherche biomédicale. Il existe cependant un écart entre la quantité de données disponibles et notre capacité de comprendre et d'interpréter ces données. Il est essentiel de combler cette lacune pour concrétiser les promesses de la médecine de précision.

Charles Boone, Ph. D., Brenda Andrews, Ph. D., du Donnelly Centre for Cellular and Biomolecular Research de l'Université de Toronto et Chad Myers, Ph. D., de l'Université du Minnesota ont découvert, en collaborant, qu'une part importante de notre incapacité à interpréter les données génomiques découle probablement du fait avéré que les maladies découlent généralement d'interactions génétiques complexes. Même si tous les humains ont essentiellement le même ensemble de gènes, la plupart ont environ cinq millions de variantes génétiques uniques. L'effet d'une variante, quelle qu'elle soit, dépend de ses interactions avec les autres. Il nous faut donc comprendre non seulement les millions de différences génétiques qui influencent la fonction d'un gène, mais aussi les interactions de tous ces gènes les uns avec les autres. Les méthodes et les technologies informatiques actuelles n'ont pas la puissance statistique de le faire.

Les trois chercheurs ont établi la première carte complète des interactions génétiques de n'importe quel organisme et bâti une méthode informatique, BridGE, pour découvrir des interactions génétiques. Ils travaillent maintenant à l'élaboration d'une plateforme informatique novatrice pour les données de séquençage génomique, BridGE-SGA, qui permettra de découvrir des interactions génétiques associées à des maladies à partir de données génotypiques humaines à grande échelle. Les chercheurs veulent ainsi découvrir les interactions génétiques de diverses maladies. L'identification et la compréhension de ces interactions génétiques clés amélioreront notre capacité d'interpréter les données du séquençage pangénomique et d'identifier de nouvelles cibles génétiques pour la découverte et la mise au point de médicaments.

Outils informatiques pour l'acquisition indépendante de données (DIA) en protéomique et en métabolomique quantitatives

Directeurs de projet : Anne-Claude Gingras, Lunenfeld-Tanenbaum Research Institute; Hannes Röst, Donnelly Centre for Cellular & Biomolecular Research, Université de Toronto

Centre de génomique : Ontario Genomics

Volet : 1 (santé)

Financement total : 1 000 000 \$

Lorsque les cellules cessent de se comporter comme elles le devraient ou de communiquer correctement avec les autres cellules, il peut s'en suivre des maladies comme le diabète ou le cancer. Les protéines et les métabolites des petites molécules déterminent le comportement des cellules, de sorte qu'il est indispensable d'identifier et de quantifier ces molécules pour comprendre l'apparition de maladies et trouver les moyens de les prévenir.

La spectrométrie de masse est devenue l'outil de travail de prédilection en protéomique et en métabolomique. Anne-Claude Gingras, Ph. D., du Lunenfeld-Tanenbaum Research Institute et Hannes Röst, Ph. D., du Donnelly Centre for Cellular & Biomolecular Research de l'Université de Toronto utilisent une technologie appelée acquisition indépendante de données indépendantes (DIA – Data-Independent Acquisition) au moyen de laquelle le spectromètre de masse identifie et quantifie systématiquement les protéines et les métabolites présents dans un échantillon. La technologie a montré qu'elle pouvait améliorer l'exactitude quantitative, la reproductibilité et le rendement comparativement à d'autres méthodes. Depuis son adoption, toutefois, elle n'a été appliquée qu'à des études de petite envergure et dans un nombre relativement petit de laboratoires. Les limites de cette méthode s'expliquent par l'absence de logiciels conviviaux qui permettraient une analyse évolutive des données complexes produites par la recherche biomédicale et médicale à grande échelle.

Le projet repose sur la solidité avérée de l'équipe en analyse des données de cette technologie de l'acquisition indépendante des données et de la création de logiciels et l'équipe produira un ensemble intégré d'outils offerts aux termes d'une licence d'accès libre. Pour encourager l'adoption de ces outils, de la documentation, des webinaires et des ateliers seront mis à la disposition des utilisateurs potentiels. Les résultats du projet pourraient avoir des répercussions durables sur le secteur canadien de la santé en facilitant la recherche sur les causes fondamentales des maladies et en aidant à résoudre des questions cliniques telles que la stratification des patients.

SYNERGx : cadre informatique pour la prédiction de la synergie créée par l'association des médicaments

Directeur de projet : Benjamin Haibe-Kains, Princess Margaret Cancer Centre

Centre de génomique : Ontario Genomics

Volet : 1 (santé)

Financement total : 972 700 \$

Lorsqu'un seul médicament est utilisé pour traiter le cancer, le patient peut ne pas y réagir ou développer une résistance. La polythérapie, dans laquelle le traitement prévoit deux médicaments ou plus, a plus de chances de réussir. Il est pourtant impossible de tester toutes les associations de médicaments dans les essais cliniques en raison du coût élevé des ressources nécessaires et de certaines considérations éthiques. Des techniques informatiques sont donc nécessaires pour modéliser la grande quantité de données disponibles pour améliorer les stratégies actuelles de traitement du cancer et proposer des associations plus efficaces des médicaments.

Benjamin Haibe-Kains, Ph. D., du Princess Margaret Cancer Centre met au point SYNERGx, une nouvelle plateforme informatique qui intégrera des ensembles de données pharmacogénomiques multiples. Ces ensembles de données serviront à prédire les associations possibles de médicaments connus qui peuvent agir en synergie, ce qui signifie que leur efficacité thérapeutique combinée est supérieure à la somme de leurs effets individuels.

La plateforme utilisera des outils d'analyse pour améliorer la modélisation des effets synergiques des médicaments. Les utilisateurs auront accès à des données pharmacogénétiques de haute qualité sur l'association de médicaments et à un pipeline d'apprentissage machine en accès libre pour la prédiction des synergies médicamenteuses. SYNERGx mettra également en œuvre une nouvelle façon d'optimiser les études de sélection préliminaire des médicaments pour identifier les nouvelles associations synergiques qui pourront ensuite être validées dans des études précliniques et cliniques.

SYNERGx offrira un moyen efficace de mettre à profit les investissements considérables faits dans les études pharmacogénomiques en permettant l'intégration d'ensembles de données autrement disparates. Il s'agit d'une importante évolution dans la conception de nouvelles stratégies thérapeutiques de lutte contre le cancer.

Logiciel visant à identifier et à quantifier les peptides dans les mégadonnées de la spectrométrie de masse obtenues à l'aide de la méthode d'acquisition indépendante des données

Directeurs de projet : Bin Ma, Université de Waterloo; Michael Moran, Hospital for Sick Children

Centre de génomique : Ontario Genomics

Volet : 2 (santé)

Financement total : 925 987 \$

La médecine de précision donne la possibilité d'adapter les décisions médicales et les décisions de traitement à l'échelle de chaque patient pour maximiser les résultats et réduire au minimum les effets indésirables. Elle peut être utilisée pour traiter un large éventail de maladies, dont le cancer. Les décisions sont souvent fondées sur la présence et la quantité de biomarqueurs tels que des protéines dans des échantillons de sang ou de tissu.

Les progrès des instruments de spectrométrie de masse ont permis de découvrir et de mesurer des biomarqueurs protéiques, mais les chercheurs n'ont pas les logiciels bio-informatiques nécessaires pour analyser les données. Bin Ma, Ph. D., de l'Université de Waterloo et Michael Moran, Ph. D., de l'Hospital for Sick Children mettent au point un logiciel qui permettra d'identifier et de quantifier avec plus de précision et d'exactitude les données de la spectrométrie de masse produites au moyen d'une méthode appelée acquisition indépendante des données (DIA – Data-independent Acquisition). Les chercheurs espèrent que leur logiciel accroîtra de beaucoup le nombre total de protéines identifiées et quantifiées comparativement au logiciel d'analyse existant de l'acquisition indépendante des données. Leur logiciel sera tout particulièrement efficace pour les modifications post-translotionnelles qui sont des biomarqueurs indispensables dans la fonction et la dégradation d'une protéine.

Le libre accès au logiciel pour les laboratoires universitaires, tout comme son rendement supérieur, peut aider les chercheurs en santé à découvrir et à suivre les biomarqueurs de la maladie. Au cours de la prochaine décennie, le logiciel pourrait devenir un outil indispensable dans de nombreux laboratoires de protéomique qui effectuent des analyses à l'aide de l'acquisition indépendante des données partout dans le monde. Le nouveau logiciel pourra aider les partenaires commerciaux à créer de nouveaux produits, services et emplois à valeur ajoutée.

Au bout du compte, les résultats du projet mèneront à des améliorations de la santé humaine et à la réduction des coûts de santé, ce qui permettra de dépister hâtivement et de diagnostiquer la maladie et facilitera le choix du meilleur traitement pour chacun des patients.

CReSCENT : Trousse d'outils d'expression des cellules du cancer à l'échelle de la cellule unique

Directeurs de projet : Trevor Pugh, Princess Margaret Cancer Centre; Michael Brudno, The Hospital for Sick Children

Centre de génomique : Ontario Genomics

Financement total : 1 000 000 \$

Les tumeurs sont des mélanges complexes de cellules cancéreuses, immunes et normales qui interagissent et changent durant le traitement. L'interaction des trois types de cellules peut dicter l'évolution du cancer au fil du temps, de même que la réaction ou la résistance aux traitements. Des progrès récents des technologies de séquençage microfluidique et de séquençage de l'ADN ont permis aux chercheurs d'analyser simultanément des dizaines de milliers de cellules individuelles provenant de tissus complexes, y compris des tumeurs. L'interprétation de ces données est difficile en raison du manque d'ensembles de références de haute qualité pour chaque type de cellule dans le corps et l'absence de méthodes pour relier ces données à la biologie des tumeurs.

Trevor Pugh, Ph. D., du Princess Margaret Cancer Centre et Michael Brudno, Ph. D., de l'Hospital for Sick Children mettent au point la trousse d'outils d'expression des cellules du cancer à l'échelle de la cellule unique (CReSCENT), un ensemble évolutif et normalisé de méthodes algorithmiques, d'outils et de portail de données nouveaux mis au point dans une

infrastructure nuagique. Pour comparer les cellules dans des tissus cancéreux et sains, le système regroupera les données génomiques d'une cellule unique, produites par des chercheurs spécialistes du cancer, et les reliera à des données de référence internationales produites par des experts de partout dans le monde dans le cadre de l'atlas des cellules humaines. Ce système de partage et de regroupement des données est un facteur clé de différenciation de CReSCENT qui augmentera la productivité des chercheurs en accélérant l'exécution et la comparaison des méthodes informatiques, tout en fournissant des données contextuelles qui faciliteront la compréhension du comportement des cellules dans des tissus tumoraux.

Cette plateforme, que tout chercheur sur n'importe quelle plateforme informatique pourra utiliser, constituera une ressource de données cruciale pour la prochaine vague de recherches en génomique du cancer à l'échelle de la cellule unique. CReSCENT regroupera des chercheurs d'un large éventail de domaines scientifiques et types de maladies et accroîtra l'impact des données produites d'un programme de recherche à l'autre. À long terme, ce système ouvrira la voie à de nouveaux diagnostics à l'échelle d'une cellule unique et la découverte de nouvelles stratégies pharmaceutique pour améliorer les soins de santé.

Dockstore 2.0 : améliorer une plateforme communautaire de partage d'outils de recherche infonuagiques agnostiques

Directeurs de projet : Lincoln Stein, Institut ontarien de recherche sur le cancer; Mark Fiume, DNASTack

Centre de génomique : Ontario Genomics

Volet : 1 (santé)

Financement total : 875 269 \$

Avec l'aide de Génome Canada, Lincoln Stein, Ph. D., de l'Institut ontarien de recherche sur le cancer a mis au point avec succès Dockstore, un système qui permet d'exécuter avec fiabilité des algorithmes informatiques complexes en biologie, reproductibles sur diverses plateformes. L'Alliance mondiale pour la génomique et la santé en a fait sa principale plateforme de progiciels et de nombreux groupes tiers en bio-informatique l'utilisent aussi. DNASTack, une entreprise canadienne, a fait l'acquisition de Dockstore et M. Stein travaille maintenant avec Mark Fiume de cette entreprise, pour maximiser l'utilité de Dockstore.

Ces améliorations visent à favoriser une collaboration et un partage accrus entre les concepteurs de logiciels en génématique. Plus précisément, elles faciliteront l'utilisation de Dockstore, rendront ses progiciels plus puissants et expressifs, amélioreront leur interopérabilité et faciliteront leur fonctionnement sur un large éventail de systèmes et d'architectures informatiques.

Ces logiciels seront utiles aux bio-informaticiens et aux génématiciens et les résultats de recherche obtenus seront reproductibles, portables et réutilisables.

D'ePlant à eEcosystem : de nouveaux cadres et outils pour partager, consulter, explorer et intégrer les données « omiques » des végétaux

Directeurs de projet : Nicholas Provart, Université de Toronto; Jörg Bohlmann, Université de la Colombie-Britannique

Centre de génomique : Ontario Genomics, Genome BC

Volet : 2 (multisectoriel)

Financement total : 1 000 000 \$

D'importants progrès ont été réalisés en biologie végétale au cours de la dernière décennie, en grande partie grâce aux nouvelles technologies de séquençage de l'ADN et de phénotypage (c.-à-d. la cartographie de l'expression physique des traits génétiques). Les ensembles de données obtenus permettent aux chercheurs de déterminer comment différents végétaux se développent et réagissent aux changements dans leur environnement. Toutefois, pour profiter de l'énorme quantité de nouvelles données, il faut des outils novateurs pour intégrer et visualiser le nombre de points de données individuels dans différents ensembles de données. Le système original ePlant, mis au point dans le cadre d'un projet précédent de Genome Canada, intègre de nombreux types de données, mais il n'était pas configuré pour les données phénotypiques. Parmi leurs multiples applications, les données phénotypiques fournissent des renseignements importants sur les caractéristiques qui intéressent surtout les phytogénéticiens.

Nicholas Provart, Ph. D., de l'Université de Toronto et Jörg Bohlmann, Ph. D. de l'Université de la Colombie-Britannique mettent au point un nouveau module pour intégrer dans ePlant la grande diversité des données disponibles, y compris les données écosystémiques, les phénotypes et les génotypes. Leurs travaux s'appliqueront aux espèces déjà existantes dans ePlant et à toute nouvelle espèce qui y sera ajoutée dans le cadre du présent projet. Les chercheurs ouvriront aussi le système ePlant à la communauté des chercheurs pour constituer un écosystème plus large de données. Ce système en ligne servira de ressource aux phytobiologistes qui pourront ainsi partager leurs ensembles de données.

Au bout du compte, ces outils peuvent accélérer la tâche d'identification de gènes utiles pour nourrir, abriter et alimenter un monde de neuf milliards de personnes d'ici 2050.

Extraire le signal du bruit : grande analyse de la biodiversité à partir de données de séquençage à haut débit

Directeurs de projet : Sarah Adamowicz, Paul Hebert, Université de Guelph

Centre de génomique : Ontario Genomics

Volet : 2 (multisectoriel)

Financement total : 507 231 \$

La surveillance de la biodiversité est indispensable à la santé de l'environnement et à la gestion des ressources naturelles. Elle aide à évaluer les répercussions de la mise en valeur des ressources, mais également à identifier rapidement et à peu de frais les insectes ravageurs, les espèces envahissantes et les agents pathogènes. Elle est essentielle à la croissance économique

du Canada dans les secteurs de la foresterie, de l'agriculture et des pêches, de même qu'à la prise de décisions en santé publique. Les méthodes génétiques de surveillance de la biodiversité, par exemple le séquençage à haut rendement, sont maintenant largement adoptées, mais la bio-informatique n'a pas suivi le rythme des données produites. De plus, les méthodes actuelles sont axées sur les bactéries et les organismes semblables et non sur les plantes multicellulaires et les animaux qu'il faut également surveiller.

Sarah Adamowicz, Ph. D., et Paul Hebert, Ph. D., de même que d'autres collègues de l'Université de Guelph, créent de nouveaux outils bio-informatiques qui faciliteront le traitement rapide et exact des données génétiques découlant du séquençage à haut débit. Ces outils permettront d'analyser en même temps des échantillons globaux constitués de nombreuses espèces différentes. Ils comprendront un outil qui éliminera le bruit pour détecter les erreurs; une méthode pour regrouper les séquences génétiques en unités semblables aux espèces afin de permettre l'analyse de la biodiversité; et une méthode d'attribution des données de séquençage à des catégories taxonomiques supérieures pour mettre au jour l'information biologique fonctionnelle. L'équipe combinera ces différents outils en un pipeline informatique de la biodiversité qui peut être intégré aux plateformes existantes sur le Web afin que divers utilisateurs puissent s'en servir.

Les nouveaux outils informatiques en biodiversité appuieront la recherche universitaire à grande échelle sur la biodiversité; des évaluations efficaces, exactes et économiques de l'environnement pour les secteurs de l'exploitation minière et des pâtes et papiers; l'amélioration de la capacité et de l'exactitude de la réglementation; et des données sur la biodiversité plus exactes et obtenues plus rapidement pour les décideurs gouvernementaux et privés.

QUÉBEC

Outils bio-informatiques pour une épigénomique 3D intégrative

Directeurs de projet : Mathieu Blanchette, Jacek Majewski et Jérôme Waldispühl,
Université McGill

Centre de génomique : Génome Québec

Volet : 1 (santé)

Financement total : 1 122 405 \$

Depuis le séquençage du génome humain au début du millénaire, les scientifiques ont accompli d'énormes progrès, en ce qui a trait à la compréhension du rôle que les gènes jouent dans notre identité, les différences entre les personnes, et notre sensibilité à des maladies comme le cancer. Les outils bio-informatiques et informatiques sont essentiels à la visualisation et à l'analyse des données génétiques, mais ils ne permettent pas d'effectuer des analyses tridimensionnelles (3D) de la façon dont l'ADN se replie sur lui-même pour se nicher dans le noyau, ce qui constitue son architecture de la chromatine. Cette structure 3D définit le fonctionnement normal du génome durant les premières étapes de la différenciation cellulaire

et du développement, et elle est reconnue pour provoquer de nombreuses anomalies du développement, ainsi que le cancer.

Les chercheurs ont un besoin pressant d'avoir accès à des outils informatiques de nouvelle génération qui intégreraient nos connaissances de ces états tridimensionnels et leur permettraient d'utiliser de façon optimale les quantités en croissance rapide de données produites par le séquençage moderne de l'ADN. Mathieu Blanchette, Ph. D., et son équipe de l'Université McGill créeront et mettront sur le marché des outils informatiques et statistiques améliorés pour l'analyse des données 3D dans leur contexte 3D originaire. Leurs nouveaux outils seront intégrés à la plateforme de visualisation 3D de l'équipe et aideront les scientifiques à explorer les données, à avancer de nouvelles hypothèses et à mettre ces dernières à l'épreuve dans des cadres de travail statistique rigoureux.

Les résultats de ce projet seront rendus largement accessibles par le biais d'un logiciel ouvert qui permettra d'effectuer des analyses statistiquement fiables de données génomiques 3D individuelles et regroupées, qui fournira un visualisateur génomique 3D fondé sur la réalité permettant une visualisation intégrée des données génomiques, et qui comprendra une trousse d'outils pour l'exploration intégrée des données génomiques et épigénomiques dans leur contexte 3D au sein du génome. Une version grand public de la plateforme de visualisation servira également pour la communication avec le public à travers des expositions dans les écoles et les musées.

Plateforme sécurisée de partage de données épigénomiques pour analyses intégratives (EpiShare)

Directeurs de projet : Guillaume Bourque et Yann Joly, Université McGill

Centre de génomique : Génome Québec

Volet : 1 (santé)

Financement total : 1 000 000 \$

L'épigénétique est l'étude des modifications réversibles du matériel génétique des cellules qui perturbent les mécanismes d'expression génétique. Ces modifications sont partiellement héréditaires et partiellement attribuables à l'environnement et aux habitudes de vie. Grâce aux progrès du séquençage de l'ADN de nouvelle génération et aux nombreux travaux menés à l'échelle internationale, les chercheurs ont désormais accès à une vaste quantité de données épigénétiques humaines. Ces données peuvent fournir des explications et des renseignements sur l'interprétation d'études d'association pangénomiques (GWAS), soit l'étude des variantes génétiques dans un seul génome. Toutefois, l'accès à ces données peut représenter un défi important, en raison, entre autres, des fichiers de très grande taille conservés dans de nombreux endroits et du fait que les données doivent être stockées derrière des mécanismes de contrôle de sécurité afin de garantir la protection des renseignements personnels des participants.

L'Alliance mondiale pour la génomique et la santé (AMGS) a déjà élaboré des outils et des normes pour le partage sûr et éthique des données génomiques. À partir de ce cadre de travail,

les professeurs Guillaume Bourque et Yann Joly de l'Université McGill, mettront au point des mécanismes pour rendre plus flexible le processus d'accès et d'analyse des données épigénomiques, tout en intégrant les exigences touchant l'éthique, la sécurité et la protection de la vie privée en ce qui a trait au partage des données. Pour ce faire, ils créent la plateforme EpiShare afin que les chercheurs n'aient pas à télécharger et à nettoyer des données et des métadonnées de différents projets épigénomiques, un travail chronophage qui nuit à l'efficacité des ressources informatiques. Ainsi, un portail Web facilitera la détection des données et permettra d'effectuer des analyses multi-omiques de ces ensembles de données à accès contrôlé, sur leur lieu de stockage. La plateforme EpiShare permettra également de garantir la protection des renseignements personnels des participants au cours du processus.

***TooT Suite* : Prédiction et classification des protéines de transport membranaire**

Directeur de projet : Gregory Butler, Université Concordia

Centre de génomique : Génome Québec

Volet : 2 (multisectoriel)

Financement total : 600 000 \$

L'association de l'accroissement de la population et de l'amélioration du niveau de vie menace la suffisance de notre approvisionnement en nourriture. En effet, en 2050, lorsqu'on s'attend à ce que la population mondiale atteigne neuf milliards de personnes, la production agricole devra augmenter de 60 à 70 % pour nourrir tous ces humains. Comme la plupart des terres arables de la planète sont déjà en exploitation, la solution doit provenir de la hausse du rendement, tant des cultures que des animaux.

Les plantes et les animaux font partie d'un écosystème complexe. Ils cohabitent en symbiose avec des communautés microbiennes, les « microbiomes », qui vivent dans leurs organismes ou près d'eux et influent sur leur santé et leur croissance et, par conséquent, leur productivité en tant que source alimentaire. À l'heure actuelle, les scientifiques utilisent la génomique et la métagénomique pour étudier ces microbiomes afin de mieux comprendre comment les interactions entre les microbiomes et leurs hôtes influent sur la santé et la croissance de ces derniers. Ces interactions se produisent par des échanges de composés chimiques, favorisés par des protéines de transport qui permettent aux composés de traverser les membranes d'une cellule.

Gregory Butler, Ph. D., de l'Université Concordia, met au point *TooT Suite*, une façon d'annoter les protéines de transport membranaire dans un organisme, que ce soit une plante ou un animal, et dans un microbiome, fournissant ainsi des renseignements au sujet d'interactions potentielles entre eux. *TooT Suite* sera un ensemble ouvert d'outils bio-informatiques faciles à utiliser qui aidera les chercheurs en génomique agricole à mieux comprendre ces interactions. Ils fonctionneront à partir de différents systèmes de classification existants, en prédisant le terme le plus approprié pour une protéine dans chacun de ces systèmes, palliant ainsi le manque d'uniformité. *TooT Suite* ouvrira une ère de recherche agricole fondée sur la pensée systémique et la métagénomique qui seront requises pour relever les défis futurs de l'approvisionnement et la sécurité alimentaires.

Mettre à profit des modèles de prévision de la production laitière grâce à la bio-informatique et à l'intelligence artificielle

Directeurs de projet : Abdoulaye Baniré Diallo, Université du Québec à Montréal; Marc-André Sirard, Université Laval

Centre de génomique : Génome Québec

Volets : 2 (multisectoriel)

Financement total : 1 004 258 \$

Au Canada, la production laitière représente un important marché, avec un cheptel de près de 1,4 million de vaches (dont la majeure partie au Québec et en Ontario) qui génère des revenus nets de 6,17 milliards de dollars chaque année. Ce marché est également unique, puisque la production laitière canadienne est régie par un système de quotas et soutenue par plus de 13 000 fermes, familiales pour la plupart. L'optimisation de la productivité, la maximisation des ressources et la limitation des dépenses sont directement associées à la rentabilité des fermes. Diverses innovations, comme la sélection génétique, des efforts de gestion accrus et un contrôle des variables touchant la production, ont contribué à maximiser la rentabilité au cours des dernières années. Depuis 40 ans, on recueille des données de production, comme la quantité de lait produite, les composants du lait, les caractéristiques des animaux et les conditions d'élevage, mais ces données n'ont jamais été intégrées aux données génomiques.

Abdoulaye Baniré Diallo de l'Université du Québec à Montréal et Marc-André Sirard de l'Université Laval s'affairent à élaborer des outils qui permettront une telle intégration, stimulant la mise au point de nouvelles pratiques d'élevage et fournissant une estimation individuelle précise de la productivité à vie des vaches. Leurs trousseaux d'outils d'exploration de données et d'apprentissage machine produiront des modèles de prévision de la production laitière qui influenceront sur les pratiques d'élevage et optimiseront la rentabilité des fermes laitières. Le projet profite des avancées en génomique qui permettent d'établir un profil génétique détaillé de chaque vache laitière, à un coût relativement bas, optimisant les décisions en matière d'alimentation, de reproduction, de traitement et de soins. Les résultats seront applicables à l'ensemble des fermes du Canada. Le projet sera aussi un tremplin vers l'inclusion de nouveaux caractères de sélection aux programmes de reproduction, comme la réponse aux diverses conditions d'hébergement, la croissance des génisses et l'adaptation aux systèmes de traite robotisés. Cette recherche est menée par une équipe multisectorielle et transdisciplinaire formée de scientifiques de l'Université du Québec à Montréal, de l'Université Laval et de l'Université McGill, en partenariat avec Valacta.

Médecine de précision en épigénomique cellulaire

Directeurs de projet : Celia Greenwood, Institut Lady Davis de recherches médicales;
Karim Oualkacha, Université du Québec à Montréal

Centre de génomique : Génome Québec

Volet : 1 (santé)

Financement total : 660 512 \$

Une grande part du risque d'être atteint d'une maladie réside dans nos gènes, mais d'autres facteurs y contribuent. Il arrive que l'environnement ou l'exposition modifie le fonctionnement des gènes. La méthylation de l'ADN est une des causes de modification de l'activité génétique. Il s'agit d'un processus lors duquel une molécule de méthyle se lie à un nucléotide cytosine dans l'ADN. Des mécanismes précis de méthylation sont nécessaires au développement normal; chez l'humain, la perturbation de la méthylation joue un rôle dans certaines maladies, comme le cancer et les maladies auto-immunes. Depuis peu, il est techniquement et financièrement possible de mesurer la méthylation de l'ADN avec une résolution de la taille d'un nucléotide simple, à une vaste échelle dans tout le génome, au moyen d'une méthode qu'on appelle « séquençage au bisulfite ». Toutefois, les données obtenues par ce séquençage présentent de nombreuses mesures imprécises ou manquantes, ce qui les rend difficiles à interpréter et limite le potentiel des études pour décrire le rôle de l'épigénétique dans une maladie.

Celia Greenwood, Ph. D., de l'Institut Lady Davis de recherches médicales, et Karim Oualkacha, Ph. D., de l'Université du Québec à Montréal, ont réuni une équipe d'experts en vue de créer un ensemble comprenant un algorithme et un logiciel pour analyser des données sur la méthylation de l'ADN à grande échelle et de grande dimension, afin que nous puissions profiter du potentiel caché de ces données. Les méthodes et la trousse se pencheront d'abord sur la sclérodémie, une maladie auto-immune invalidante causant la cicatrisation de nombreux tissus et dont la réponse aux traitements est imprévisible.

L'équipe a le soutien d'un organisme de défense des patients, Sclérodémie Québec, et de deux entreprises de Montréal, une ayant mis au point des méthodes d'apprentissage machine et l'autre fournissant une plateforme informatique à haut rendement. Ces méthodes et cet ensemble logiciel révolutionneront la compréhension de la méthylation de l'ADN et de ses contributions aux maladies.

Arrimage moléculaire de nouvelle génération tirant avantage de technologies d'intelligence artificielle pour comprendre de vastes ensembles de données sur la liaison aux ligands

Directeur de projet : Rafaël Najmanovich, Université de Montréal

Centre de génomique : Génome Québec

Volet : 1 (santé)

Financement total : 500 000 \$

Dans la nature, rien n'existe de façon isolée, de la plus petite molécule aux plus gros arbres ou animaux. Tout est en interaction. Rafaël Najmanovich, Ph. D., de l'Université de Montréal, se préoccupe des plus petites molécules, particulièrement des interactions entre les petites

molécules et les protéines qui gouvernent l'essentiel du métabolisme et de la signalisation cellulaire. Il veut comprendre la « reconnaissance moléculaire » qui se produit entre ces petites molécules et les protéines, afin de mieux saisir les processus biologiques visant à mettre au point de nouveaux médicaments.

Les ligands (généralement, de petites molécules) se lient aux protéines, dans des cavités à leur surface. La prédiction de ces interactions se fait au moyen de simulations d'arrimage. De nos jours, les méthodes d'arrimage sont largement utilisées et acceptées, mais n'ont pas encore atteint le niveau de réalisme, de rapidité et de précision requis pour réaliser leur plein potentiel en vue de concevoir des médicaments. Le laboratoire de M. Najmanovich a déjà mis au point FlexAID (pour Flexible Artificial Intelligence Docking), une technique d'arrimage flexible fondée sur l'intelligence artificielle qui surpasse les méthodes d'arrimage largement utilisées. Il met actuellement au point un logiciel d'arrimage de nouvelle génération qui augmentera la précision biologique, grâce à laquelle les chercheurs pourront modéliser des processus d'arrimage, accélérer les simulations pour le classement des petites molécules et se concentrer sur des familles de protéines particulières qui sont importantes pour la conception de médicaments.

Les méthodes d'arrimage en cours de création seront à la base d'un logiciel ouvert qui permettra de comprendre, sur le plan structural, la découverte de médicaments, et d'améliorer la compréhension de la reconnaissance moléculaire. Ce projet contribuera aussi à faire progresser la recherche pharmaceutique de façon à ce que les Canadiens en tirent des bienfaits sociaux et économiques.

Plateforme intégrative pour la métabolomique et la biologie des systèmes

Directeurs de projet : Jianguo Xia et Guillaume Bourque, Université McGill; Pierre-Étienne Jacques, Université de Sherbrooke

Centre de génomique : Génome Québec

Volet : 1 (santé)

Financement total : 1 094 607 \$

Bien que relativement récente, la métabolomique – l'étude approfondie des métabolites (de petites molécules) dans les liquides biologiques, comme le sang et l'urine – est de plus en plus utilisée en association avec d'autres technologies « omiques » pour comprendre les maladies et les processus biologiques complexes. Afin de gérer les mégadonnées de la métabolomique, les chercheurs ont un besoin urgent d'outils bio-informatiques à haut rendement et faciles à utiliser pour le traitement, l'interprétation et l'intégration aux autres données « omiques » des données brutes.

À la lumière de leur réussite antérieure avec des plateformes en ligne pour l'analyse de données « omiques » sur la santé humaine, Jianguo Xia, Ph. D., et Guillaume Bourque, Ph. D., de l'Université McGill, ainsi que Pierre-Étienne Jacques, Ph. D., de l'Université de Sherbrooke, élaborent SystemsAnalyst, une plateforme informatique de nouvelle génération intégrant des ressources de Calcul Canada et du nuage public aux plus récentes technologies d'analytique

visuelle. Leur objectif est de créer un guichet unique permettant d'effectuer des analyses efficaces, transparentes et reproductibles de grandes quantités de données en vue de soutenir la métabolomique des systèmes et l'intégration des données multi-omiques.

SystemsAnalyst profitera directement aux cliniciens et aux biologistes de laboratoire en fournissant une puissante plateforme conviviale qui viendra pallier les lacunes actuelles en bio-informatique pour les études métabolomiques et multi-omiques. Des études de cas sur la malaria et la maladie intestinale inflammatoire faciliteront l'identification d'importants biomarqueurs, de voies clés et de processus biologiques qui jouent des rôles de premier plan dans ces maladies, en vue de contribuer à l'élaboration de nouvelles stratégies thérapeutiques. Le projet permettra de former au moins 10 employés hautement qualifiés, dont le milieu a besoin de toute urgence pour l'analyse et la modélisation de données multi-omiques. SystemsAnalyst peut aussi facilement être appliqué à d'autres maladies complexes.

Création et validation d'une plateforme en ligne pour les données « omiques » et la toxicologie environnementale

Directeurs du projet : Jianguo Xia et Niladri Basu, Université McGill

Centre de génomique : Génome Québec

Volet : 2 (multisectoriel)

Financement total : 1 047 507 \$

L'évaluation des risques environnementaux est en phase de passer rapidement aux outils « omiques » et aux techniques de biologie des systèmes afin d'évaluer les répercussions des facteurs de stress sur la croissance, la reproduction et la survie des animaux. Malgré un vif intérêt parmi les groupes d'intervenants, il est évident que les scientifiques de la vie et de l'environnement ne peuvent pas gérer convenablement le type et les quantités de données en rapide émergence. Il n'existe pas d'outils bio-informatiques acceptés et normalisés pour organiser, analyser, visualiser et interpréter les données « omiques » des espèces des principales études.

Au cours de la dernière décennie, Jianguo Xia, Ph. D., de l'Université McGill, a mis au point un ensemble d'outils nuagiques pour l'analyse, la visualisation et l'interprétation des données « omiques ». Ces puissants outils conviviaux se sont avérés hautement populaires dans la communauté de chercheurs en santé humaine, ralliant des milliers d'utilisateurs à travers le monde. Un besoin urgent se fait sentir pour une trousse semblable d'outils « omiques » afin d'appuyer le nombre croissant d'études « omiques » dans les sciences de la vie et de l'environnement, notamment dans des domaines tels que l'exploitation minière, l'agriculture, la foresterie, l'aquaculture et la gestion de la qualité de l'eau.

Jianguo Xia et son collègue, Niladri Basu, Ph. D., mettent au point eco. OmicsAnalyst, un outil nuagique intuitif permettant l'analyse et la visualisation de telles données, en commençant par 12 espèces bio-indicatrices clés, dont des poissons, des oiseaux, des mammifères et des invertébrés. Afin de perfectionner et de valider les outils, ils entreprendront une série d'études

de cas sur une classe particulièrement importante de facteurs de stress, soit la pollution chimique. Leur travail tiendra compte des utilisateurs finaux au sein des organismes gouvernementaux et des universités. Un manuel d'utilisation technique sera produit afin de favoriser l'adoption par les utilisateurs finaux.

En raison des étroites relations qu'entretient l'équipe de recherche avec les intervenants, il est très probable que le nouvel outil soit adopté et qu'il permette de surmonter les obstacles au traitement des données « omiques » auxquels se butent actuellement les scientifiques de la vie et de l'environnement. Cet outil permettra, entre autres, d'améliorer la qualité des données, l'efficacité du processus décisionnel et l'utilisation des ressources.