



JMP® Genomics

Découvrez les caractéristiques biologiques des données génomiques grâce à deux références en matière de logiciels d'analyse : SAS et JMP.

Qu'est-ce que JMP® Genomics ?

JMP Genomics est un logiciel de découverte statistique issu de deux références en matière de logiciels d'analyse : SAS et JMP. Les laboratoires de recherche l'utilisent pour identifier des modèles pertinents dans de volumineuses bases de données de génétique à haut débit, d'expression, de nombre de copies et de protéomique. Des graphiques interactifs facilitent l'exploration des relations entre les données grâce à une série complète d'algorithmes statistiques classiques ou avancés.

Pourquoi ce logiciel est-il essentiel ?

Les laboratoires de recherche souhaitent rentabiliser au mieux les investissements en temps, capitaux et ressources nécessaires pour générer des données génomiques d'excellente qualité. Seules des analyses statistiques spécialisées permettent de découvrir les « trésors » dissimulés dans les longues listes de biomarqueurs ou gènes candidats. Qu'il soit utilisé pour identifier de possibles cibles médicamenteuses, explorer la biologie d'un organisme modèle ou développer un modèle de maladie prédictif, JMP Genomics aide les chercheurs à acquérir un avantage concurrentiel en identifiant rapidement des protéines ou des gènes essentiels.

À qui s'adresse-t-il ?

JMP Genomics s'adresse aux biologistes, biostatisticiens, spécialistes de la génétique statistique et étudiants qui analysent les volumineux jeux de données très courants dans la recherche génomique. Il réunit de nombreuses méthodes d'analyse dans un même progiciel pour postes de travail. Il n'est pas toujours facile d'adopter un nouveau logiciel dans un établissement de grande taille. C'est la raison pour laquelle JMP Genomics regroupe des outils capables d'analyser de nombreux types de données, ce qui facilite l'exploration de nouveaux domaines génomiques à mesure que les recherches évoluent.

jmp.com/lifesciences-resources

JMP® Genomics de SAS est un progiciel complet d'exploration et d'analyse des données génomiques, qui réunit la toute puissance de la plate-forme de découverte statistique JMP et des outils SAS® Analytics adaptés aux jeux de données les plus volumineux.

Les biologistes, biostatisticiens et spécialistes de la génétique statistique du monde entier l'utilisent pour analyser les intensités, counts ou génotypes continus générés à partir de plates-formes de puces à ADN classiques et résumés au moyen de technologies next gen sequencing.

L'approche visuelle de JMP Genomics permet de lier dynamiquement les statistiques aux graphiques afin d'offrir une vue détaillée des résultats d'analyse. Même les étudiants peu expérimentés en matière d'analyse génomique peuvent rapidement découvrir les tendances importantes et les valeurs aberrantes que recèlent leurs données grâce à des boîtes

« À mon sens, nous manquons cruellement de logiciels offrant un tel niveau de souplesse et de puissance statistique. »

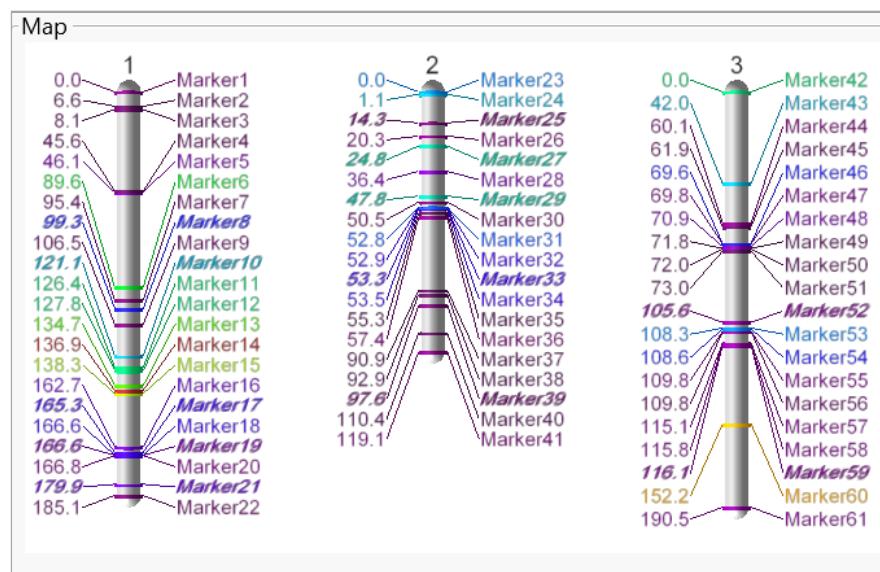
Erik Sulman

Chercheur en oncologie, Houston

de dialogue simplifiées et des workflows personnalisés ne nécessitant aucune compétence en programmation ni formation statistique particulière. De plus en plus d'enseignants intègrent d'ailleurs JMP Genomics à leurs programmes.

Un simple pointer-cliquer suffit pour être opérationnel.

JMP Genomics met la puissance de SAS au service de l'étude des données génomiques, quelle qu'en soit la forme – criblage d'associations significatives, recherche de modèles pertinents dans des études d'expression, évaluation des



Visualisez les cartes de liaisons créées dans JMP Genomics ou importées depuis un autre logiciel.

différences de nombre de copies, etc. À mesure que vous élargissez vos domaines de recherche, vous pouvez explorer de nouvelles données dans un environnement familier, sans perdre de temps à vous former à plusieurs logiciels et à passer d'un jeu de données à l'autre.

Outre une vaste bibliothèque de graphiques prédéfinis, JMP Genomics offre un accès complet aux nombreuses fonctionnalités analytiques et graphiques de la plate-forme JMP. Vous pouvez ainsi concevoir des expériences aussi étendues qu'efficaces et créer différents graphiques interactifs. Grâce à des outils tels que le Constructeur de graphiques de type « glisser-déposer » et le filtre de données interactif, chaque utilisateur est en mesure de créer des vues personnalisées de ses données.

La fenêtre d'accueil personnalisable de JMP Genomics permet aux utilisateurs, qu'ils soient débutants ou chevronnés, d'accéder rapidement aux outils adaptés à leurs besoins d'analyse. L'Assistant JMP Genomics vous aide à importer des échantillons d'informations et de données à partir des plates-formes de données génomiques et formats texte les plus courants. Les options graphiques et celles d'analyse de suivi sont organisées en rapports tabulés, les tables sous-jacentes étant masquées afin de simplifier la présentation des résultats d'analyse complexes. Vous pouvez toutefois rappeler facilement ces tables afin de vérifier un détail, ou fermer d'un clic la totalité des tables et graphiques.

Expression

Des workflows de base et intermédiaires conviviaux simplifient l'analyse des nombres de lectures et intensités d'expression à partir d'études RNA-seq. Grâce aux méthodes de filtrage, de normalisation, de modélisation et de découv-

verte de configurations pilotées par SAS Analytics, vous pouvez manipuler des jeux de données volumineux impossibles à traiter avec la simple mémoire disponible. JMP Genomics 6 propose différentes options de normalisation et de modélisation utilisables pour l'expression par puces à ADN et les données de counts brutes ou transformées.

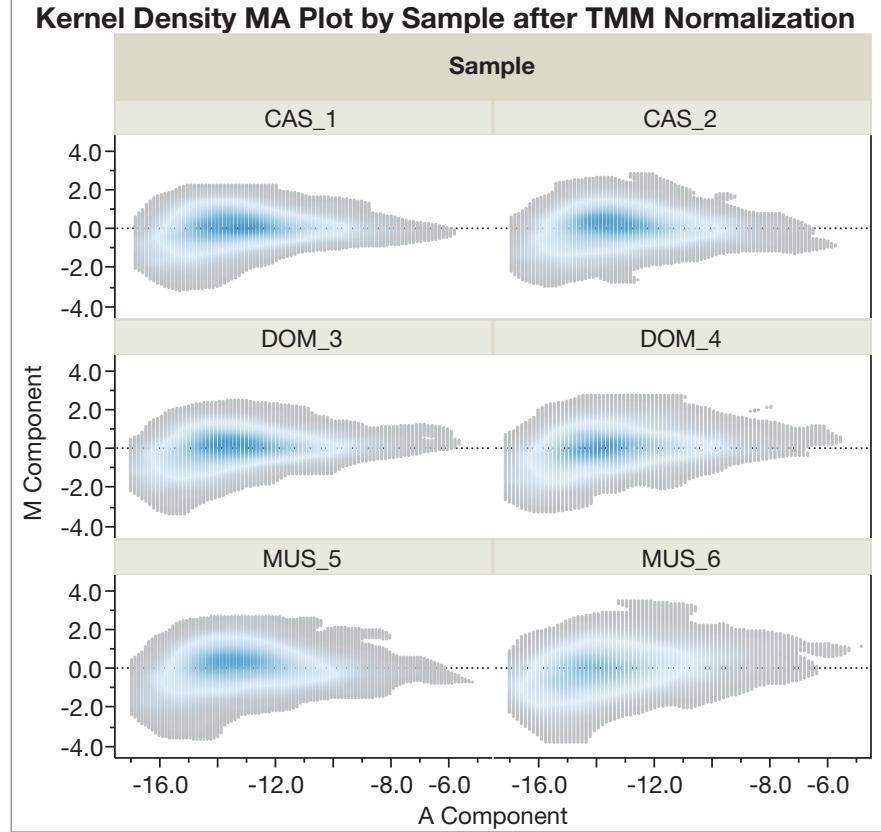
Génétique

JMP Genomics offre des outils d'exception aux spécialistes de la génétique statistique, des simples analyses d'association de cas-témoins et de déséquilibre de liaison aux tests d'association les plus complexes impliquant différents types

de traits. Vous pouvez ainsi examiner les associations entre des SNP et plusieurs caractères continus, corriger la structure de la population et explorer des interactions SNP-SNP. Les workflows de type « pointer-cliquer » simplifient l'analyse du modèle mixte Q-K et celle d'association des variantes rares et courantes.

Cartes de liaisons et analyse QTL

JMP Genomics intègre une série de processus de création, d'optimisation et de visualisation des cartes de liaisons de marqueurs utilisées pour améliorer les cultures ou orienter les travaux de sélection animale. L'identification des groupes

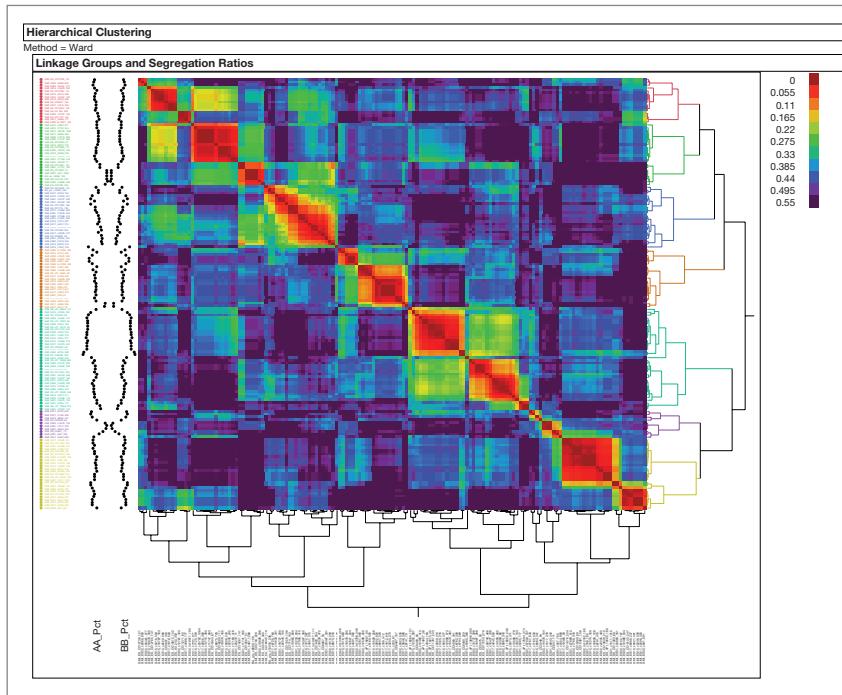


Mettez à l'échelle les données de count de l'ensemble des échantillons par normalisation TMM, comparez les facteurs TMM de différents échantillons et visualisez des graphiques de densité Kernel de données normalisées.

« On a tendance à considérer les fonctions graphiques comme normales, mais il est vraiment formidable de pouvoir visualiser la séparation [des données de haute dimension]. Les écarts importants sautent tout simplement aux yeux. »

Faye Schilkey

Centre national américain de ressources génomiques



JMP Genomics offre des options automatisées et interactives d'identification des groupes de liaison et de classement des marqueurs.

de liaison peut être automatique ou interactive et vous pouvez désigner des groupes de consensus et des marqueurs dont la position est connue tout en optimisant l'ordre des marqueurs. Vous pouvez en outre visualiser et filtrer les cartes de marqueurs sous forme de graphiques interactifs simples ou créer des vues multichromosomes sophistiquées. Pour finir, il est possible de résumer les informations phénotypiques, d'explorer les interactions entre génotype et environnement dans le cadre d'essais multi-environnements et de réaliser des analyses de marqueur unique QTL ou IM/CIM avec option de permutation.

Nombre de copies

Repérez aisément les différences de nombre de copies entre plusieurs groupes ou parmi des individus. Évaluez la qualité des données au niveau des

échantillons afin d'identifier les valeurs aberrantes potentielles et de filtrer les caractéristiques selon des critères statistiques. Vous pouvez en outre ajuster les données LOH et de type nombre de copies à l'aide d'échantillons de référence groupés ou appariés. Utilisez la segmentation binaire circulaire (CBS) pour identifier les points de rupture ou des méthodes ANOVA pour déceler les différences majeures entre plusieurs groupes ou par rapport à un groupe de référence. Les graphiques interactifs facilitent quant à eux l'identification des zones de variation communes.

Modélisation prédictive

JMP Genomics propose huit méthodes de modélisation prédictive intégrant des options de filtre/verrouillage de régresseur et de validation croisée. Le logiciel permet d'identifier les biomarqueurs les

plus significatifs au sein de jeux de données de plusieurs types – SNP, expression ou nombre de copies. Les stratégies d'itération et de duplication intégrées au logiciel visent à réduire les biais, tandis que les approches de validation croisée permettent d'évaluer simultanément les performances relatives de centaines de modèles.

Next Gen Sequencing

JMP Genomics offre des fonctionnalités d'analyse statistique sophistiquées en aval des données de résumé issues de pipelines d'analyse de séquence de pointe. Importez des dénominations de séquences au format texte ou résumez ceux issus de fichiers SAM, BAM et Eland pour tirer parti des méthodes de normalisation et de modélisation linéaire généralisée. Les workflows de base simplifient le processus d'analyse.

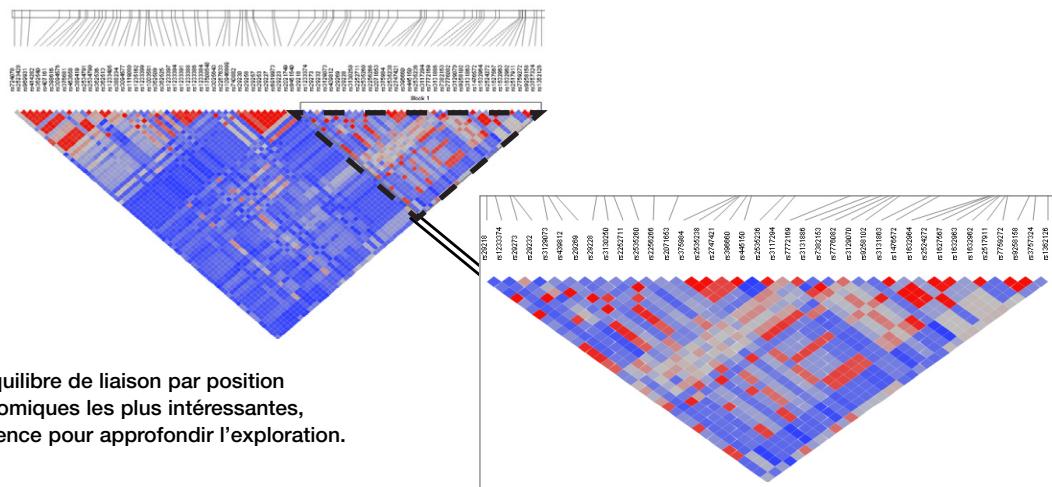
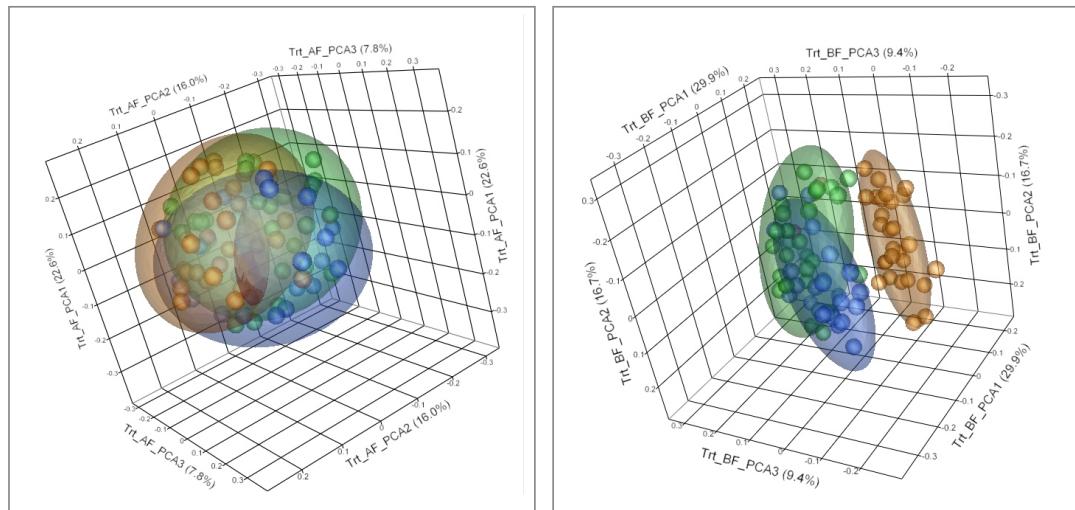
Vous pouvez également importer des génotypes directement depuis différents formats texte ou fichiers VCF, ou faire appel à des variantes issues de fichiers BAM à l'aide d'un génome de référence. Effectuez une analyse d'association des variantes SNP rares et courantes ou identifiez les régions d'état identique (IBS) communes à des individus liés ou indépendants.

Tirez parti de la richesse des informations fournies par les expériences de séquençage pour découvrir les corrélations entre différents types de données par criblage. Les résultats des analyses statistiques ou dénominations de séquences peuvent être visualisés dans le navigateur de JMP Genomics et les histogrammes ou graphiques de chaleur, peuvent être superposés à des résumés de niveau groupe ou individuel afin de compléter les données connues des gènes et SNP.

Découvrez les effets de lots au niveau de vos données et supprimez-les avant de procéder à une analyse statistique.

JMP Genomics offre plusieurs options de normalisation de lots. À gauche, les échantillons recueillis dans différents lots se regroupent étroitement, compensant ainsi les effets du traitement.

À droite, nous pouvons voir ces mêmes échantillons après suppression de l'effet de lot.

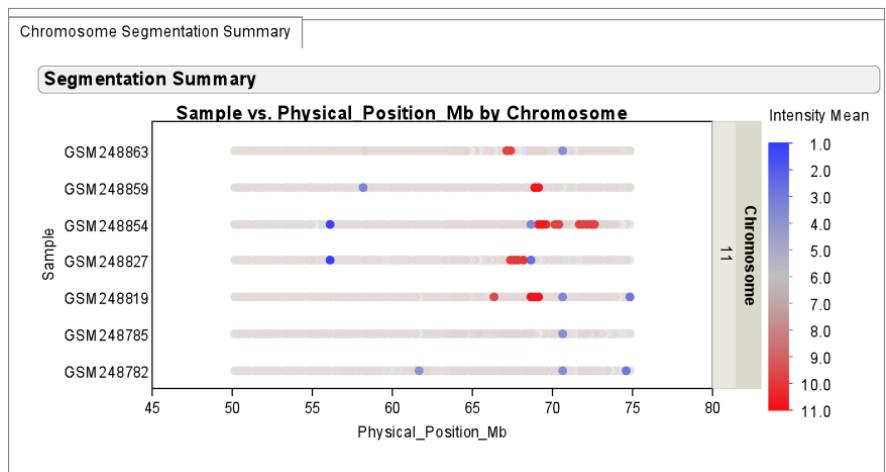


Examinez les motifs de déséquilibre de liaison par position afin d'identifier les zones génomiques les plus intéressantes, puis mettez les blocs en évidence pour approfondir l'exploration.

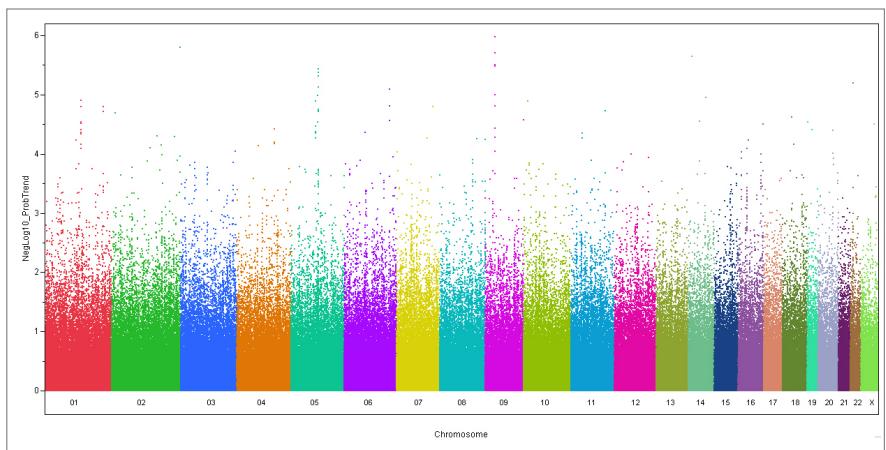
« Lorsque l'on passe d'une dizaine de gènes observés à plusieurs milliers, il est difficile d'interpréter les résultats. C'est même impossible sans outils permettant de visualiser et d'explorer facilement l'annotation des résultats. JMP est excellent dans ce domaine. »

Tom Juenger, University of Texas, Austin

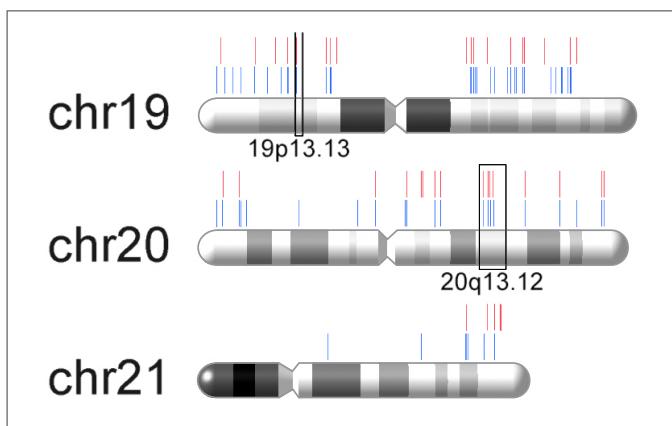
Les graphiques résumant la segmentation peuvent être filtrés de manière interactive afin d'identifier les zones communes de perte ou gain de nombre de copies.



Affichez les p-values des tests statistiques par chromosome, ou créez des vues multichromosomes personnalisées.



Créez des thèmes de couleur de génome personnalisés et superposez les résultats statistiques, puis effectuez un zoom avant pour approfondir l'exploration et visualiser les données de suivi des gènes et SNP.



Principales fonctionnalités de JMP® Genomics 6.0

Importation de données dans JMP Genomics à partir de nombreux formats :

- Lecture de dénominations de séquences, d'intensités ou de génotypes contenus dans un seul ou plusieurs fichiers texte
- Lectures alignées au format SAM ou BAM et variantes de fichiers VCF
- Pipeline génomique complet et fichiers de sortie *testvariant*
- Fichiers de résumé indel et SNP CLC bio
- Données de type expression Illumina BeadStudio ou GenomeStudio, SNP, nombre de copies, etc.
- Formats de fichiers Affymetrix CEL et CHP, BAR, LOHCHP, CNCHP, CNAT et Cytogenetics CEL et CHP
- Fichiers Agilent à une ou deux couleurs, GenePix et QuantArray
- Fichiers Excel et délimités par des virgules, y compris des formats de données issus de plusieurs plates-formes NimbleGen

Workflows souples adaptés à l'ensemble des utilisateurs :

- Assistant JMP Genomics pour l'importation de nouveaux jeux de données
- Workflows de base pour les données de type expression, exon, RNA-seq et nombre de copies et la création des cartes de liaisons
- Options intermédiaires pour l'analyse et le contrôle de la qualité des expressions
- Workflows d'association de variantes rares et Q-K
- Constructeur de workflows permettant de créer des workflows personnalisés

Intégration de statistiques aux workflows d'analyse Next Gen Sequencing pour :

- Normaliser et analyser les dénominations de séquence générées au niveau de l'exon, de la transcription ou des gènes par différents pipelines
- Évaluer les données RNA-seq au moyen de workflows de type « pointer-cliquer »
- Tester les possibilités d'association entre les traits et les variantes rares ou courantes
- Effectuer une analyse de corrélation croisée pour associer les dénominations de séquence à d'autres mesures génomiques numériques

Évaluation de jeux de données de variantes à l'échelle d'un génome pour :

- Examiner les configurations de données manquantes pour les individus et les marqueurs
- Résumer les propriétés de marqueur – fréquences des allèles et génotypes, HWE, hétérozygosité et diversité
- Filtrer les jeux de données par attributs de marqueur ou d'échantillon
- Explorer les associations avec un ou plusieurs traits binaires ou quantitatifs
- Calculer des interactions et effectuer des ajustements pour les covariables
- Tester les associations à l'aide des données SNP imputées
- Calculer et visualiser le déséquilibre de liaison (LD) et les blocs LD
- Visualiser et ajuster la structure de la population via une analyse PCA ou une mise à l'échelle MDS
- Réaliser une métanalyse GWAS à l'aide de p-values ou d'effets

Extension des options d'analyse des données de marqueur pour incorporer les éléments suivants :

- Tests de variantes SNP groupées dans un locus ou une voie métabolique
- Tests d'association de variantes rares avec options de permutation
- Matrices de distance génétique pour des individus ou populations
- Calcul des matrices des relations de partage d'allèles, d'IBD et d'IBS
- Compression de matrices K pour réduire les temps de traitement
- Corrections des tests d'association relatifs à la structure de la population et aux liens de parenté
- Identification des zones génomiques partagées à l'identique
- Estimation des haplotypes et associations haplotype-caractère
- Sélection de tagSNP pour les haplotypes ou le déséquilibre de liaison
- Réconciliation des différences de brins entre études

Amélioration des stratégies de sélection des cultures et du bétail par :

- Visualisation des distributions de phénotypes continus et catégoriels entre individus, génotypes ou lignées
- Évaluation des génotypes dans le cadre d'essais multi-environnements
- Identification des groupes de liaison à partir de données croisées expérimentales
- Classement des marqueurs au sein de groupes de liaison via des algorithmes d'optimisation évolués
- Consultation des cartes de liaisons créées dans JMP Genomics et d'autres progiciels cartographiques
- Réalisation de cartographies QTL d'intervalle composite (Composite Interval Mapping, CIM), d'intervalle simple (Interval Mapping, IM) et de marqueur unique avec option de permutation

Évaluation de la qualité de jeux de données d'expression volumineux pour :

- Identifier les problèmes de qualité des données et supprimer les valeurs aberrantes
- Repérer les facteurs expliquant la variance de vos données
- Visualiser les distributions, les diagrammes PCA et les clusters d'échantillons
- Normaliser l'ensemble des échantillons
- Supprimer les effets de lots dus à des variables techniques
- Ajuster les distributions de dénominations de séquences par normalisation TMM et KDMM
- Utiliser les méthodes de normalisation loess, des quantiles, RMA, GCRMA et ANOVA ou diverses statistiques pour la standardisation

Application de méthodes statistiques éprouvées pour :

- Découvrir des différences majeures grâce aux modèles linéaires généralisés et ANOVA
- Appliquer différentes sortes d'ajustements pour les tests multiples
- Effectuer des ajustements pour les covariables et effets aléatoires
- Cibler l'expression spécifique d'allèle
- Analyser les données de survie censurées
- Générer des profils d'expression par groupe ou échantillon avec sélection et filtrage dynamiques
- Réaliser une classification hiérarchique et un partitionnement par la méthode des K-means

Utilisation d'outils d'analyse avancés pour la modélisation prédictive afin d'autoriser :

- L'identification de biomarqueurs à partir de jeux de données volumineux
- La sélection de régresseurs parmi plusieurs types de données
- Le filtrage de régresseur personnalisé lors de l'élaboration du modèle
- Le verrouillage des principaux régresseurs catégoriques ou continus
- La comparaison des performances au travers de huit méthodes différentes
- La validation croisée à l'aide de nombreuses options d'itération et de « hold-out »
- L'analyse de la courbe d'apprentissage afin d'évaluer l'impact de la taille de l'échantillon

Évaluation des jeux de données de type nombre de copies pour :

- Examiner la qualité des données avec l'analyse PCA et des distributions
- Ajuster les mesures de nombre de copies à l'aide de contrôles groupés ou appariés
- Déceler les segments analysés par segmentation binaire circulaire (CBS)
- Visualiser les configurations partagées de perte ou gain de nombre de copies
- Identifier les zones génomiques montrant des différences statistiques significatives entre groupes, ou entre individus et un groupe de contrôle

(à suivre)

Principales fonctionnalités de JMP® Genomics 6.0 (suite)

Exploitation des outils d'annotation de JMP Genomics pour :

- Combiner les informations fonctionnelles avec les résultats statistiques
- Transférer les résultats vers le logiciel IPA (Ingenuity Pathways Analysis) pour rechercher les points d'interaction entre les listes de SNP, de gènes et de protéines et les réseaux métaboliques en couleur
- Effectuer des analyses d'enrichissement en utilisant les informations fonctionnelles provenant du logiciel IPA
- Apparier les informations sur les voies métaboliques issues des bases de données MsigDB, KEGG ou autres afin de procéder à des analyses d'enrichissement ou au scoring des groupes de gènes
- Visualiser des groupes de gènes co-régulés dans des voies métaboliques KEGG
- Télécharger des fichiers de bibliothèque et d'annotation depuis Affymetrix NetAffx
- Comparer l'appartenance à des listes pour un maximum de cinq groupes et afficher les superpositions avec des diagrammes de Venn à l'aide d'identifiants de gènes communs

Création de vues de génome pour :

- Visualiser les chromosomes avec des thèmes de couleur personnalisables

- Comparer plusieurs expériences pour identifier des régions de significativité commune
- Superposer les résultats au niveau des gènes, SNP, diagrammes en barres et palettes de couleurs
- Utiliser diverses mesures continues pour résumer les p-values
- Obtenir des informations détaillées sur des zones intéressantes pour représenter graphiquement les p-values et afficher les données de suivi des gènes, SNP, diagrammes en barres et palettes de couleurs

Caractéristiques de la plate-forme logicielle JMP :

- Constructeur de graphiques autorisant l'exploration visuelle des motifs de données
- Création de graphiques personnalisés par simple pointer-cliquer
- Copier-coller de données dans Word et PowerPoint
- Intégration de scripts de capture et partage des découvertes
- Modules d'extension dédiés aux analyses externes (R, SAS, etc.)

Graphiques interactifs générés automatiquement lors de l'analyse :

- Organisation sous forme de rapports dynamiques liés à des données sous-jacentes
- Sélection par pointer-cliquer et création aisée de sous-ensembles
- Interrogation dynamique à l'aide du filtre de données de JMP
- Convertibilité en rapports statiques



JMP® Academic Suite

SAS est fier de collaborer avec des établissements d'enseignement supérieur pour faire avancer la réflexion statistique et préparer les étudiants à une économie mondiale de plus en plus concurrentielle et fondée sur le savoir. Les tarifs préférentiels des licences JMP® accordés aux établissements et aux départements d'enseignement supérieur

permettent d'utiliser cette solution dans les cours théoriques, laboratoires et autres structures de formation et de recherche. Si votre établissement est déjà titulaire de licences JMP Academic Suite, vous pouvez y ajouter JMP Genomics ou son alter ego, JMP Clinical, en bénéficiant d'une remise supplémentaire.

En savoir plus : jmp.com/fr/academic

Licences

Les licences JMP Genomics pour les entreprises, l'administration et les établissements d'enseignement sont disponibles sous forme d'abonnement annuel.

Webinaires gratuits

Pour en savoir plus sur notre logiciel et les **webinaires de prise en main** gratuits, consultez le site Web : jmp.com/lifesciences-resources

Systèmes d'exploitation

JMP Genomics est compatible avec la plupart des versions professionnelles 32 et 64 bits des systèmes d'exploitation Windows XP, Vista et Windows 7 pour serveurs et postes de travail.



JMP® Genomics s'inscrit progressivement dans le quotidien des étudiants

JMP Genomics est utilisé depuis des années dans les laboratoires de recherche universitaires. Aujourd'hui, ce logiciel dynamique et interactif de SAS est aussi de plus en plus présent dans les salles de classe des établissements d'enseignement supérieur.

Susan Singer, professeur de sciences naturelles titulaire de la chaire Laurence McKinley Gould au Carleton College de Northfield (Minnesota), attribue cette évolution à la « profonde mutation » de l'étude des données génomiques. Les étudiants sont aujourd'hui mis au défi de « tirer parti de l'ampleur des données disponibles », explique-t-elle.

Les études portant sur le génome impliquent d'obtenir une vision globale, puis détaillée. C'est justement cette vision que JMP Genomics offre aux étudiants, grâce à une combinaison unique de graphiques JMP interactifs et de puissantes fonctionnalités SAS® Analytics sur poste de travail.

En savoir plus : jmp.com/jmpgoncampus

