

Резюмета на рецензираните публикации на Румяна Кирова Йорданова на английски и български език

Публикации за конкурса

1. Zhelyazkova, M., Yordanova, R., Mihaylov, I., Kirov, S., Tsonev, S., Danko, D., Mason, C., Vassilev, D. Origin Sample Prediction and Spatial Modeling of Antimicrobial Resistance in Metagenomic Sequencing Data, (2021) Frontiers in Genetics, 12, art. no. 642991, DOI: 10.3389/fgene.2021.642991

Abstract

The steady elaboration of the Metagenomic and Metadesign of Subways and Urban Biomes (MetaSUB) international consortium project raises important new questions about the origin, variation, and antimicrobial resistance of the collected samples. CAMDA (Critical Assessment of Massive Data Analysis, <http://camda.info/>) forum organizes annual challenges where different bioinformatics and statistical approaches are tested on samples collected around the world for bacterial classification and prediction of geographical origin. This work proposes a method which not only predicts the locations of unknown samples, but also estimates the relative risk of antimicrobial resistance through spatial modeling. We introduce a new component in the standard analysis as we apply a Bayesian spatial convolution model which accounts for spatial structure of the data as defined by the longitude and latitude of the samples and assess the relative risk of antimicrobial resistance taxa across regions which is relevant to public health. We can then use the estimated relative risk as a new measure for antimicrobial resistance. We also compare the performance of several machine learning methods, such as Gradient Boosting Machine, Random Forest, and Neural Network to predict the geographical origin of the mystery samples. All three methods show consistent results with some superiority of Random Forest classifier. In our future work we can consider a broader class of spatial models and incorporate covariates related to the environment and climate profiles of the samples to achieve more reliable estimation of the relative risk related to antimicrobial resistance.

Абстракт

Непрекъснатото развитие на международния проект на консорциума Metagenomic и Metadesign of Subways and Urban Biomes (MetaSUB) повдига нови важни въпроси за произхода, вариациите и антимикробната резистентност на събраните проби. Форумът CAMDA (Critical Assessment of Massive Data Analysis, <http://camda.info/>) организира ежегодни съревнования, където различни биоинформатични и статистически подходи се тестват върху проби, събрани по целия свят, за класифициране на бактериите и прогнозиране на географския им произход. В тази работа се предлага метод, който не само предсказва местоположението на неизвестни проби, но и оценява относителния риск от антимикробна резистентност чрез пространствено моделиране. Ние въвеждаме нов компонент в стандартния анализ, тъй като прилагаме Бейсов пространствен конволюционен модел, който отчита пространствената

структура на данните, определена от географската дължина и ширина на пробите, и оценява относителния риск от таксони с антимикробна резистентност в различните региони, което е от значение за общественото здраве. След това можем да използваме оценения относителен риск като нова мярка за антимикробна резистентност. Ние също така сравняваме ефективността на няколко метода за машинно обучение, като Gradient Boosting Machine, Random Forest и Neural Network, за да предскажем географския произход на мистериозните проби. И трите метода показват съвместими резултати с известно превъзходство на класификатора Random Forest. В бъдещата ни работа ние можем да разгледаме по-широк клас пространствени модели и да включим ковариати, свързани с околната среда и климатичните профили на пробите, за да постигнем по-надеждна оценка на относителния риск, свързан с антимикробната резистентност.

2. Putman, A.H., Wolen, A.R., Harenza, J.L., Yordanova, R.K., Webb, B.T., Chesler, E.J., Miles, M.F., Identification of quantitative trait loci and candidate genes for an anxiolytic-like response to ethanol in BXD recombinant inbred strains, (2016) Genes, Brain and Behavior, 15 (4), pp.367-381.

Abstract

Genetic differences in acute behavioral responses to ethanol contribute to the susceptibility to alcohol use disorder and the reduction of anxiety is a commonly reported motive underlying ethanol consumption among alcoholics. Therefore, we studied the genetic variance in anxiolytic-like responses to ethanol across the BXD recombinant inbred (RI) mouse panel using the light-dark transition model of anxiety. Strain-mean genetic mapping and a mixed-model quantitative trait loci (QTL) analysis replicated several previously published QTL for locomotor activity and identified several novel anxiety-related loci. Significant loci included a chromosome 11 saline anxiety-like QTL (Salanq1) and a chromosome 12 locus (Etanq1) influencing the anxiolytic-like response to ethanol. Etanq1 was successfully validated by studies with BXD advanced intercross strains and fine-mapped to a region comprising less than 3.5 Mb. Through integration of genome-wide mRNA expression profiles of the mesocorticolimbic reward circuit (prefrontal cortex, nucleus accumbens and ventral midbrain) across the BXD RI panel, we identified high priority candidate genes within Etanq1, the strongest of which was *Ninein* (*Nin*), a *Gsk3 β* -interacting protein that is highly expressed in the brain.

Абстракт

Генетичните различия в острите поведенчески реакции към етанол допринасят за податливостта към алкохолизъм и намаляването на тревожността е често съобщаван мотив, който е в основата на консумацията на етанол сред алкохолиците. Поради това, проучихме генетичните различия в анксиолитично-подобните реакции към етанол в панела на рекомбинантните инбредни (RI) мишки BXD, като използвахме модела на тревожност при преход светлина-тъмнина. Генетичното мапиране на средните и анализът на количествени локуси (QTL) със смесен модел възпроизведе няколко публикувани по-рано QTL за локомоторна активност и идентифицира няколко нови локуса, свързани с тревожността. Значимите локуси включват QTL за тревожност, подобна на индуцирана от физиологичен разтвор, на хромозома 11 (Salanq1) и локус на хромозома 12 (Etanq1), влияещ върху анксиолитично подобна реакция към етанол. Etanq1 беше успешно валидиран чрез проучвания с

напреднали междуполови щамове BXD и е точно картографиран в регион, обхващащ по-малко от 3,5 Mb. Чрез интегриране на профилите на експресия на мРНК в целия геном на мезокортиколимбичната верига за възнаграждение (префронтална кора, nucleus accumbens и вентрален среден мозък) в панела BXD RI, ние идентифицирахме високоеприоритетни кандидат-гени в рамките на Etanq1, най-силният от които беше Ninein (Nin), *Gsk3 β* -взаимодействащ протеин, който е силно експресиран в мозъка.

3. Ha, T., Swanson, D., Larouche, M., Glenn, R., Weeden, D., Zhang, P., Hamre, K., Langston, M., Phillips, C., Song, M., Ouyang, Z., Chesler, E., Duvvurru, S., Yordanova, R., Cui, Y., Campbell, K., Ricker, G., Phillips, C., Homayouni, R., Goldowitz, D., Cerebellar gene regulation in time and space, (2015) Developmental Biology, 397 (1), pp. 18-30., DOI: 10.1016/j.ydbio.2014.09.032
Abstract

The mammalian CNS is one of the most complex biological systems to understand at the molecular level. The temporal information from time series transcriptome analysis can serve as a potent source of associative information between developmental processes and regulatory genes. Here, we introduce a new transcriptome database called, Cerebellar Gene Regulation in Time and Space (CbGRiTS). This dataset is populated with transcriptome data across embryonic and postnatal development from two standard mouse strains, C57BL/6J and DBA/2J, several recombinant inbred lines and cerebellar mutant strains. Users can evaluate expression profiles across cerebellar development in a deep time series with graphical interfaces for data exploration and link-out to anatomical expression databases. We present three analytical approaches that take advantage of specific aspects of the time series for transcriptome analysis. We demonstrate the use of CbGRiTS dataset as a community resource to explore patterns of gene expression and develop hypotheses concerning gene regulatory networks in brain development.

Абстракт

ЦНС на бозайниците е една от най-сложните биологични системи за разбиране на молекулярно ниво. Времевата информация от анализа на транскриптомни времеви серии може да послужи като мощен източник на асоциативна информация между процесите на развитие и регулаторните гени. Тук представяме нова транскриптомна база данни, наречена Cerebellar Gene Regulation in Time and Space (CbGRiTS). Тази база данни се състои от транскриптомни данни от ембрионалното и постнаталното развитие на два стандартни щамове мишки, C57BL/6J и DBA/2J, няколко рекомбинантни инбредни линии и церебеларни мутантни щамове. Потребителите могат да оценяват експресионните профили на мозъчното развитие в дълбоки времеви серии с помощта на графични интерфейси за изследване на данни и да свързват това с анатомични експресионни бази данни. Ние представяме три аналитични подхода, които се възползват от специфични аспекти на времевите серии за анализ на транскриптомите. Ние демонстрираме използването на CbGRiTS като ресурс на общността за изследване на моделите на генна експресия и разработване на хипотези относно генните регулаторни мрежи в развитието на мозъка.

4. Orozco, L.D., Bennett, B.J., Farber, C.R., Ghazalpour, A., Pan, C., Che, N., Wen, P., Qi, H.X., Mutukulu, A., Siemers, N., Neuhaus, I., Yordanova, R.,

Gargalovic, P., Pellegrini, M., Kirchgessner, T., Lusi, A.J., Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages, (2012) Cell, 151 (3), pp. 658-670., DOI: 10.1016/j.cell.2012.08.043

Abstract

Many common diseases have an important inflammatory component mediated in part by macrophages. Here we used a systems genetics strategy to examine the role of common genetic variation in macrophage responses to inflammatory stimuli. We examined genome-wide transcript levels in macrophages from 92 strains of the Hybrid Mouse Diversity Panel. We exposed macrophages to control media, bacterial lipopolysaccharide (LPS), or oxidized phospholipids. We performed association mapping under each condition and identified several thousand expression quantitative trait loci (eQTL), gene-by-environment interactions, and eQTL "hot spots" that specifically control LPS responses. We used siRNA knockdown of candidate genes to validate an eQTL hot spot in chromosome 8 and identified the gene 2310061C15Rik as a regulator of inflammatory responses in macrophages. We have created a public database where the data presented here can be used as a resource for understanding many common inflammatory traits that are modeled in the mouse and for the dissection of regulatory relationships between genes.

Абстракт

Много често срещани заболявания имат важен възпалителен компонент, който отчасти се модулира от макрофагите. Тук използвахме стратегията на системната генетика, за да изучим ролята на общите генетични вариации в отговора на макрофагите към възпалителни стимули. Изследвахме нивата на транскриптите в целия геном на макрофаги от 92 щамове от панела на хибридни мишки (Hybrid Mouse Diversity Panel). Ние подложихме макрофагите на контролна среда, бактериален липополизахарид (LPS) или окислени фосфолипиди. Извършихме асоциирано мапиране при всяко условие и идентифицирахме няколко хиляди локуса с количествени характеристики на експресията (eQTL), взаимодействия между гени и среда и eQTL "горещи точки които специфично контролират отговорите при LPS. Използвахме siRNA нокдаун на кандидат-гени, за да валидираме една eQTL гореща точка на хромозома 8 и идентифицирахме гена 2310061C15Rik като регулатор на възпалителните реакции в макрофагите. Създадохме публична база данни, в която представените тук данни могат да се използват като ресурс за разбиране на много общи възпалителни признаци, които се моделират при мишките и за разчленяване на регулаторните връзки между гените.

5. Ha, T.J., Swanson, D.J., Kirova, R., Yeung, J., Choi, K., Tong, Y., Chesler, E.J., Goldowitz, D., Genome-wide microarray comparison reveals downstream genes of Pax6 in the developing mouse cerebellum, (2012) European Journal of Neuroscience, 36 (7), pp. 2888-2898., DOI: 10.1111/j.1460-9568.2012.08221.x

Abstract

The Pax6 transcription factor is expressed in cerebellar granule cells and when mutated, as in the Sey/Sey mouse, produces granule cells with disturbed survival and migration and with defects in neurite extension. The impact of Pax6 on other genes in the context of cerebellar development has not been identified. In this study, we performed

transcriptome comparisons between wildtype and Pax6-null whole cerebellar tissue at embryonic day (E) 13.5, 15.5 and 18.5 using Affymetrix arrays (U74Av2). Statistical analyses identified 136 differentially regulated transcripts (FDR 0.05, 1.2-fold change cutoff) over time in Pax6-null cerebellar tissue. In parallel we examined the Math1-null granulo-prival cerebellum and identified 228 down-regulated transcripts (FDR 0.05, 1.2-fold change cutoff). The intersection of these two microarray datasets produced a total of 21 differentially regulated transcripts. For a subset of the identified transcripts, we used qRT-PCR to validate the microarray data and demonstrated the expression in the rhombic lip lineage and differential expression in Pax6-null cerebellum with in situ hybridisation analysis. The candidate genes identified in this way represent direct or indirect Pax6-downstream genes involved in cerebellar development.

Абстракт

Транскрипционният фактор Pax6 се експресираща в клетките на церебеларните гранули и когато е мутирал, както при мишката *Seu/Seu*, се получават гранулни клетки с нарушена преживяемост и миграция и с дефекти в неуритното удължаване. Влиянието на Pax6 върху други гени в контекста на развитието на церебеларния мозък не е установено. В това проучване ние извършихме транскриптомни сравнения между див тип и Pax6-null на цяла церебеларна тъкан в ембрионален ден (E) 13,5, 15,5 и 18,5, като използвахме масиви на Affymetrix (U74Av2). Статистическите анализи идентифицираха 136 диференциално регулирани транскрипта (FDR 0,05, 1,2-fold change) с течение на времето в Pax6-null церебеларната тъкан. Успоредно с това, ние изследвахме Math1-нулевия гранулопривичен церебелум и идентифицирахме 228 транскрипта с понижена експресия (FDR 0,05, 1,2-fold change граница). Сечението на тези два набора от микрочипови данни продуцират общо 21 диференциално регулирани транскрипта. За подгрупа от идентифицираните транскрипти, ние използвахме qRT-PCR, за да валидираме данните от микрочиповете и демонстрирахме експресия в "rhombic lip lineage" и диференциална експресия в Pax6-нулевия церебелум с хибридизационен анализ *in situ*. Идентифицираните по този начин кандидат-гени представляват директни или индиректни Pax6-downstream гени, участващи в развитието на малкия мозък.

6. Ji, R.-R., Ott, K.-H., Yordanova, R., Bruccoleri, R.E., FDR-FET: An optimizing gene set enrichment analysis method, (2011) *Advances and Appl. in Bioinformatics and Chemistry*, 4 (1), 37-42, DOI: 10.2147/AABC.S15840

Abstract

Gene set enrichment analysis for analyzing large profiling and screening experiments can reveal unifying biological schemes based on previously accumulated knowledge represented as "gene sets". Most of the existing implementations use a fixed fold-change or P value cutoff to generate regulated gene lists. However, the threshold selection in most cases is arbitrary, and has a significant effect on the test outcome and interpretation of the experiment. We developed a new gene set enrichment analysis method, ie, FDR-FET, which dynamically optimizes the threshold choice and improves the sensitivity and selectivity of gene set enrichment analysis. The procedure translates experimental results into a series of regulated gene lists at multiple false discovery rate (FDR) cutoffs, and computes the P value of the overrepresentation of a gene set using a Fisher's exact

test (FET) in each of these gene lists. The lowest P value is retained to represent the significance of the gene set. We also implemented improved methods to define a more relevant global reference set for the FET. We demonstrate the validity of the method using a published microarray study of three protease inhibitors of the human immunodeficiency virus and compare the results with those from other popular gene set enrichment analysis algorithms. Our results show that combining FDR with multiple cutoffs allows us to control the error while retaining genes that increase information content. We conclude that FDR-FET can selectively identify significant affected biological processes. Our method can be used for any user-generated gene list in the area of transcriptome, proteome, and other biological and scientific applications.

Абстракт

"Анализът на обогатяването на генните множества" (GSEA) за анализиране на големи профилиращи и скринингови експерименти може да разкрие обединяващи биологични схеми, основани на предварително натрупани знания, представени като "генни набори" (генни множества). Повечето от съществуващите реализации използват фиксирана граница на промяна (fold change) или стойност на P (p-value), за да генерират списъци с регулирани гени. Изборът на праг обаче в повечето случаи е произволен и оказва значително влияние върху резултата от теста и интерпретацията на експеримента. Ние разработихме нов метод за анализ на обогатяването на генни множества, т.е. FDR-FET, който динамично оптимизира избора на праг и подобрява чувствителността и селективността на анализа на обогатяването на генни множества. Процедурата превръща експерименталните резултати в поредица от списъци с регулирани гени при различни гранични стойности на коефициента на фалшиво откриване (FDR) и изчислява P-стойността на свръхпредставянето на даден набор от гени с помощта на точния тест на Фишер (FET) във всеки от тези списъци от гени. Най-ниската стойност на P се запазва, за да представи значимостта на набора от гени. Също така приложихме подобрени методи за определяне на по-подходящ глобален референтен набор за FET. Ние демонстрираме валидността на метода, като използваме публикувано микрочипово изследване на три протеазни инхибитора на човешкия имунодефицитен вирус и сравняваме резултатите с тези от други популярни алгоритми за анализ на обогатяването на генни множества. Нашите резултати показват, че комбинирането на FDR с множество гранични стойности ни позволява да контролираме грешката, като същевременно запазваме гени, които увеличават информационното съдържание. Ние стигаме до заключението, че FDR-FET може селективно да идентифицира значими засегнати биологични процеси. Нашият метод може да се използва за всеки генериран от потребителя списък с гени в областта на транскриптома, протеома и други биологични и научни приложения.

7. Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.-Z., Brewer, H., Weitz, K., Camp II, D.G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D., Lusk, A.J., Comparative analysis of proteome and transcriptome variation in mouse, (2011) PLoS Genetics, 7 (6), art. no. e1001393, DOI: 10.1371/journal.pgen.1001393

Abstract

The relationships between the levels of transcripts and the levels of the proteins they encode have not been examined comprehensively in mammals, although previous work in plants and yeast suggest a surprisingly modest correlation. We have examined this issue using a genetic approach in which natural variations were used to perturb both transcript levels and protein levels among inbred strains of mice. We quantified over 5,000 peptides and over 22,000 transcripts in livers of 97 inbred and recombinant inbred strains and focused on the 7,185 most heritable transcripts and 486 most reliable proteins. The transcript levels were quantified by microarray analysis in three replicates and the proteins were quantified by Liquid Chromatography-Mass Spectrometry using O(18)-reference-based isotope labeling approach. We show that the levels of transcripts and proteins correlate significantly for only about half of the genes tested, with an average correlation of 0.27, and the correlations of transcripts and proteins varied depending on the cellular location and biological function of the gene. We examined technical and biological factors that could contribute to the modest correlation. For example, differential splicing clearly affects the analyses for certain genes; but, based on deep sequencing, this does not substantially contribute to the overall estimate of the correlation. We also employed genome-wide association analyses to map loci controlling both transcript and protein levels. Surprisingly, little overlap was observed between the protein- and transcript-mapped loci. We have typed numerous clinically relevant traits among the strains, including adiposity, lipoprotein levels, and tissue parameters. Using correlation analysis, we found that a low number of clinical trait relationships are preserved between the protein and mRNA gene products and that the majority of such relationships are specific to either the protein levels or transcript levels. Surprisingly, transcript levels were more strongly correlated with clinical traits than protein levels. In light of the widespread use of high-throughput technologies in both clinical and basic research, the results presented have practical as well as basic implications.

Абстракт

Връзките между нивата на транскриптите и нивата на протеините, които те кодират, не са изследвани обстойно при бозайниците, въпреки че предишни работи при растенията и дрождите показват изненадващо слаба корелация. Ние разгледахме този въпрос, като използвахме генетичен подход, при който естествените вариации бяха използвани за да променят нивата както на транскриптите така и на протеините сред инбредни щамове мишки. Ние определихме количествено над 5000 пептида и над 22 000 транскрипта в черния дроб на 97 инбредни и рекомбинантни инбредни щамове и се съсредоточихме върху 7185-те най-наследствени транскрипта и 486-те най-надеждни белтъка. Нивата на транскриптите бяха определени количествено чрез микрочипов анализ с три повторения, а протеините бяха определени количествено чрез течна хроматография-масспектрометрия, използвайки подход за изотопно маркиране, базиран на O(18)-референция. Показваме, че нивата на транскриптите и протеините се корелират значително само за около половината от изследваните гени, със средна корелация от 0,27, а корелациите на транскриптите и протеините варират в зависимост от клетъчното местоположение и биологичната функция на гена. Разгледахме техническите и биологичните фактори, които биха могли да допринесат за незначителната корелация. Например,

диференциалният сплайсинг явно влияе на анализа за някои гени, но въз основа на дълбокото секвениране това не допринася съществено за общата оценка на корелацията. Използвахме също така геномни асоциативни анализи, за да открием локуси, контролиращи нивата както на транскриптите така и на протеините. Изненадващо, наблюдава се малко припокриване между протеин-свързани локуси и транскрипти-свързани локуси. Типизирахме многобройни клинично значими фенотипи сред щамовете, включително мастна тъкан, нива на липопротеини и тъканни параметри. Използвайки корелационен анализ, установихме, че малък брой клинични фенотипни връзки се запазват между белтъчните и мРНК генните продукти и че по-голямата част от тези връзки са специфични или за белтъчните нива, или за нивата на транскрипта. Нивата на транскриптите са по-силно свързани с клиничните признаци отколкото нивата на протеините. В светлината на широкото използване на високопроизводителни технологии както в клиничните, така и във фундаменталните изследвания, представените резултати имат както практическо, така и фундаментално значение.

8.Romanoski, C.E., Lee, S., Kim, M.J., Ingram-Drake, L., Plaisier, C.L., Yordanova, R., Tilford, C., Guan, B., He, A., Gargalovic, P.S., Kirchgessner, T.G., Berliner, J.A., Lusi, A.J., Systems Genetics Analysis of Gene-by-Environment Interactions in Human Cells, (2010) Amer. Journal of Human Genetics, 86 (3), pp. 399-410., DOI: 10.1016/j.ajhg.2010.02.002

Abstract

Gene by environment (GxE) interactions are clearly important in many human diseases, but they have proven to be difficult to study on a molecular level. We report genetic analysis of thousands of transcript abundance traits in human primary endothelial cell (EC) lines in response to proinflammatory oxidized phospholipids implicated in cardiovascular disease. Of the 59 most regulated transcripts, approximately one-third showed evidence of GxE interactions. The interactions resulted primarily from effects of distal-, trans-acting loci, but a striking example of a local-GxE interaction was also observed for FGD6. Some of the distal interactions were validated by siRNA knockdown experiments, including a locus involved in the regulation of multiple transcripts involved in the ER stress pathway. Our findings add to the understanding of the overall architecture of complex human traits and are consistent with the possibility that GxE interactions are responsible, in part, for the failure of association studies to more fully explain common disease variation

Абстракт Взаимодействията между гени и околна среда (GxE) са очевидно важни за много човешки заболявания, но се оказва, че е трудно да бъдат изследвани на молекулярно ниво. Ние докладваме за генетичен анализ на хиляди характеристики на експресирани транскрипти в човешки първични ендотелни клетъчни линии (ЕК) в отговор на провъзпалителни окислени фосфолипиди, участващи в сърдечносъдови заболявания. От 59-те най-регулирани транскрипта приблизително една трета показаха доказателства за GxE взаимодействия. Взаимодействията са резултат предимно от ефектите на дистално, транс-действащи локуси, но поразителен пример за локално GxE взаимодействие е наблюдаван и за FGD6. Някои от дисталните взаимодействия бяха потвърдени чрез експерименти с нокдаун на siRNA,

включително локус, участващ в регулирането на множество транскрипти, включени в ER стресовия механизъм. Нашите резултати допринасят за разбирането на цялостната архитектура на сложните човешки фенотипове и са в съответствие с възможността, че GxE взаимодействията са отчасти отговорни за неуспеха на асоциативните проучвания да обяснят по-пълно вариациите при честите заболявания.

9. Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., Truong, A., Yang, W.-P., He, A., Kayne, P., Gargalovic, P., Kirchgessner, T., Pan, C., Castellani, L.W., Kostem, E., Furlotte, N., Drake, T.A., Eskin, E., Lusis, A.J., A high-resolution association mapping panel for the dissection of complex traits in mice, (2010) *Genome Research*, 20 (2), pp. 281-290., DOI: 10.1101/gr.099234.109

Abstract

Systems genetics relies on common genetic variants to elucidate biologic networks contributing to complex disease-related phenotypes. Mice are ideal model organisms for such approaches, but linkage analysis has been only modestly successful due to low mapping resolution. Association analysis in mice has the potential of much better resolution, but it is confounded by population structure and inadequate power to map traits that explain less than 10% of the variance, typical of mouse quantitative trait loci (QTL). We report a novel strategy for association mapping that combines classic inbred strains for mapping resolution and recombinant inbred strains for mapping power. Using a mixed model algorithm to correct for population structure, we validate the approach by mapping over 2500 cis-expression QTL with a resolution an order of magnitude narrower than traditional QTL analysis. We also report the fine mapping of metabolic traits such as plasma lipids. This resource, termed the Hybrid Mouse Diversity Panel, makes possible the integration of multiple data sets and should prove useful for systems-based approaches to complex traits and studies of gene-by-environment interactions.

Абстракт Системната генетика се основава на общи генетични варианти за да намери биологичните мрежи, допринасящи за сложните фенотипи, свързани с болести. Мишките са идеални моделни организми за такива подходи, но "linkage" анализът е само донякъде успешен поради ниската проектираща разделителна способност. Асоциативният анализ при мишките има потенциал за много по-добра разделителна способност, но той е затруднен от структурата на популацията и недостатъчната мощност за картографиране на признаци, които обясняват по-малко от 10% от дисперсията, типична за локусите на количествени признаци (QTL) при мишките. Ние описваме една нова стратегия за картографиране на асоциации, която съчетава класически инбредни щамове за резолюция на картографирането и рекомбинантни инбредни щамове за мощност на картографирането. Използвайки алгоритъм на смесен модел за коригиране на популационната структура, ние валидираме подхода чрез картографиране на над 2500 локални-експресионни QTL с разделителна способност, която е на порядък по-точна от тази на традиционния QTL анализ. Също така съобщаваме за по-прецизно картографиране на метаболитни признаци, като например плазмени липиди. Този ресурс, наречен Hybrid Mouse Diversity Panel, дава възможност за интегриране на множество набори от данни и следва

да се окаже полезен за системно базирани подходи към сложни финотипове и изследвания на взаимодействията между гени и околна среда.

10. Baker, E.J., Jay, J.J., Philip, V.M., Zhang, Y., Li, Z., Kirova, R., Langston, M.A., Chesler, E.J., **Ontological discovery environment: A system for integrating gene-phenotype associations**, (2009) *Genomics*, 94 (6), pp. 377-387.

Abstract

The wealth of genomic technologies has enabled biologists to rapidly ascribe phenotypic characters to biological substrates. Central to effective biological investigation is the operational definition of the process under investigation. We propose an elucidation of categories of biological characters, including disease relevant traits, based on natural endogenous processes and experimentally observed biological networks, pathways and systems rather than on externally manifested constructs and current semantics such as disease names and processes. The Ontological Discovery Environment (ODE) is an Internet accessible resource for the storage, sharing, retrieval and analysis of phenotype-centered genomic data sets across species and experimental model systems. Any type of data set representing gene-phenotype relationships, such quantitative trait loci (QTL) positional candidates, literature reviews, microarray experiments, ontological or even meta-data, may serve as inputs. To demonstrate a use case leveraging the homology capabilities of ODE and its ability to synthesize diverse data sets, we conducted an analysis of genomic studies related to alcoholism. The core of ODE's gene set similarity, distance and hierarchical analysis is the creation of a bipartite network of gene-phenotype relations, a unique discrete graph approach to analysis that enables set-set matching of non-referential data. Gene sets are annotated with several levels of metadata, including community ontologies, while gene set translations compare models across species. Computationally derived gene sets are integrated into hierarchical trees based on gene-derived phenotype interdependencies. Automated set identifications are augmented by statistical tools which enable users to interpret the confidence of modeled results. This approach allows data integration and hypothesis discovery across multiple experimental contexts, regardless of the face similarity and semantic annotation of the experimental systems or species domain.

Абстракт Разнообразието от геномни технологии позволиха на биолозите бързо да приписват фенотипни характеристики на биологични субстрати. Централно място в ефективното биологично изследване заема оперативното определяне на изследвания процес. Ние предлагаме определяне на категориите биологични признаци, включително болестно свързани фенотипи, на базата на естествени ендогенни процеси и експериментално наблюдавани биологични мрежи, механизми и системи а не на външно проявени конструкции и текуща семантика, като например имена на болести и процеси. Ontological Discovery Environment (ODE) е ресурс достъпен на интернет за съхранение, споделяне, извличане и анализ на фенотипно центрирани набори от геномни данни за различни видове и експериментални моделни системи. Всеки тип набор от данни, представящ взаимоотношенията между гени и фенотипи, като например позиционни кандидати за локуси на количествени признаци (QTL), литературни обзори, експерименти с микрочипове, онтологични или дори метаданни, може да служи като входни данни. За да демонстрираме слу-

чай на употреба, използващ хомоложните възможности на ODE и способността му да синтезира разнообразни набори от данни, ние проведохме анализ на геномни изследвания, свързани с алкохолизма. В основата на сходството, разстоянието и йерархичния анализ на генните набори в ODE е създаването на двустранна мрежа от връзки между гени и фенотипи - уникален подход от анализ на дискретни графи, който позволява изграждане на съответствия между множества от нереперентни данни. Множествата от гени се аотират с няколко нива на метаданни, включително онтологии на общността, а трансляциите на генните множества сравняват моделите между различните видове. Компютърно изведените генни набори се интегрират в йерархични дървета въз основа на генно извлечени фенотипни взаимозависимости. Автоматизираните идентификации на набори се допълват от статистически методи, които позволяват на потребителите да интерпретират доверието в моделираните резултати. Този подход позволява интегриране на данни и откриване на хипотези в множество експериментални контексти, независимо от сходството и семантичната аотация на експерименталните системи или видовата област.

11. Song, M., Lewis, C.K., Lance, E.R., Chesler, E.J., Yordanova, R.K., Langston, M.A., Lodowski, K.H., Bergeson, S.E., Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data, (2009) Eurasip Journal on Bioinformatics and Systems Biology, 2009, art. no. 545176, DOI: 10.1155/2009/545176

Abstract

Gene expression time course data can be used not only to detect differentially expressed genes but also to find temporal associations among genes. The problem of reconstructing generalized logical networks to account for temporal dependencies among genes and environmental stimuli from transcriptomic data is addressed. A network reconstruction algorithm was developed that uses statistical significance as a criterion for network selection to avoid false-positive interactions arising from pure chance. The multinomial hypothesis testing-based network reconstruction allows for explicit specification of the false-positive rate, unique from all extant network inference algorithms. The method is superior to dynamic Bayesian network modeling in a simulation study. Temporal gene expression data from the brains of alcohol-treated mice in an analysis of the molecular response to alcohol are used for modeling. Genes from major neuronal pathways are identified as putative components of the alcohol response mechanism. Nine of these genes have associations with alcohol reported in literature. Several other potentially relevant genes, compatible with independent results from literature mining, may play a role in the response to alcohol. Additional, previously unknown gene interactions were discovered that, subject to biological verification, may offer new clues in the search for the elusive molecular mechanisms of alcoholism.

Абстракт Динамичните данни на генната експресия могат да се използват не само за откриване на диференциално експресирани гени, но и за намиране на времеви асоциации между гените. Тук е разгледан проблемът за реконструиране на обобщени логически мрежи за отчитане на времевите зависимости между гените и стимулите на околната среда базирани на транскриптомни данни. Разработен

е алгоритъм за реконструкция на мрежи, който използва статистическата значимост като критерий за избор на мрежа, за да се избегнат фалшиво положителни взаимодействия, произтичащи от чиста случайност. Реконструкцията на мрежата, базирана на многокомпонентно тестване на хипотези, позволява явно определяне на степента на фалшиво положителните резултати, което е уникално за всички съществуващи алгоритми за конструиране на мрежи. Методът превъзхожда динамичното бейсовското моделиране на мрежи в симулационно изследване. За моделиране се използват временни данни на генна експресия от мозъци на третирани с алкохол мишки при анализ на молекулярния отговор към алкохола. Гените от основните невронни пътища са идентифицирани като предполагаеми компоненти на механизма на алкохолния отговор. Девет от тези гени имат асоциации с алкохола, за които се съобщава в литературата. Няколко други потенциално значими гени, съвместими с независими резултати от литературния анализ, могат да играят роля в реакцията към алкохол. Открити са допълнителни, неизвестни досега генни взаимодействия, които, след биологична проверка, могат да предложат нови идеи в търсенето на неуловимите молекулярни механизми на алкохолизма.

Други публикации

1. Habiba, U., Sugino, H., Yordanova, R., Ise, K., Tanei, Z.-I., Ishida, Y., Tanikawa, S., Terasaka, S., Sato, K.-I., Kamoshima, Y., Katoh, M., Nagane, M., Shibahara, J., Tsuda, M., Tanaka, S., Loss of H3K27 trimethylation is frequent in IDH1-R132H but not in non-canonical IDH1/2 mutated and 1p/19q codeleted oligodendroglioma: a Japanese cohort study, (2021) Acta Neuropathologica Communications, 9 (1), art. no. 95, DOI: 10.1186/s40478-021-01194-7

Abstract

Oligodendrogliomas are defined by mutation in isocitrate dehydrogenase (NADP(+)) (IDH)1/2 genes and chromosome 1p/19q codeletion. World Health Organisation diagnosis endorses testing for 1p/19q codeletion to distinguish IDH mutant (Mut) oligodendrogliomas from astrocytomas because these gliomas require different treatments and they have different outcomes. Several methods have been used to identify 1p/19q status; however, these techniques are not routinely available and require substantial infrastructure investment. Two recent studies reported reduced immunostaining for trimethylation at lysine 27 on histone H3 (H3K27me3) in IDH Mut 1p/19q codeleted oligodendroglioma. However, the specificity of H3K27me3 immunostaining in this setting is controversial. Therefore, we developed an easy-to-implement immunohistochemical surrogate for IDH Mut glioma subclassification and evaluated a validated adult glioma cohort. We screened 145 adult glioma cases, consisting of 45 IDH Mut and 1p/19q codeleted oligodendrogliomas, 30 IDH Mut astrocytomas, 16 IDH wild-type (Wt) astrocytomas, and 54 IDH Wt glioblastomas (GBMs). We compared immunostaining with DNA sequencing and fluorescent in situ hybridization analysis and assessed differences in H3K27me3 staining between oligodendroglial and astrocytic lineages and between IDH1-R132H and non-canonical (non-R132H) IDH1/2 Mut oligodendroglioma. A loss of H3K27me3 was observed in 36/40 (90%) of IDH1-R132H Mut oligodendroglioma. In contrast, loss of H3K27me3 was never seen in IDH1-R132L or IDH2-mutated 1p/19q codeleted oligodendrogliomas.

IDH Mut astrocytoma, IDH Wt astrocytoma and GBM showed preserved nuclear staining in 87%, 94%, and 91% of cases, respectively. A high recursive partitioning model predicted probability score (0.9835) indicated that the loss of H3K27me3 is frequent to IDH1-R132H Mut oligodendroglioma. Our results demonstrate H3K27me3 immunohistochemical evaluation to be a cost-effective and reliable method for defining 1p/19q codeletion along with IDH1-R132H and ATRX immunostaining, even in the absence of 1p/19q testing.

Абстракт

Олигодендроглиомите се определят от мутация в изоцитрат дехидрогеназата (NADP(+)) (IDH)1/2 гени и хромозома 1p/19q коделиция. Диагнозът на Световната здравна организация подкрепя изследването за коделиция на 1p/19q за да се разграничат олигодендроглиомите с мутация на IDH (Mut) от астроцитомите, тъй като тези глиоми изискват различно лечение и имат различен краен резултат. Използвани са няколко метода да се определи статуса на 1p/19q; тези техники обаче не са рутинно достъпни и изискват значителни инфраструктурни инвестиции. Две скорошни проучвания съобщават за намалено имунооцветяване при триметилиране на лисин 27 на хистон H3 (H3K27me3) при олигодендроглиом с IDH Mut 1p/19q коделиция. Въпреки това, специфичността на имунооцветяването на H3K27me3 в тази среда е спорна. Затова разработихме лесен за прилагане имунохистохимичен заместител за субкласификация на IDH Mut глиома и направихме оценка на валидирана кохорта от глиоми при възрастни. Изследвахме 145 случая на глиоми при възрастни, състоящи се от 45 IDH Mut и 1p/19q коделирани олигодендроглиоми, 30 IDH Mut астроцитом, 16 IDH див тип (Wt) астроцитом и 54 IDH Wt глиобластоми (GBM). Сравнихме имунооцветяването с анализ на ДНК секвенцията и флуоресцентна *in situ* хибридизация и оценихме разликите в оцветяването на H3K27me3 между олигодендроглиалните и астроцитните линии и между IDH1-R132H и неканоничните (non-R132H) IDH1/2 Mut олигодендроглиоми. Загуба на H3K27me3 се наблюдава при 36/40 (90%) от IDH1-R132H Mut олигодендроглиомите. За разлика от тях, загуба на H3K27me3 никога не е наблюдавана при IDH1-R132L или IDH2-мутирани 1p/19q коделирани олигодендроглиоми. IDH Mut астроцитом, IDH Wt астроцитом и GBM показват запазено ядрено оцветяване съответно в 87%, 94% и 91% от случаите. Високият резултат на прогнозираната вероятност на модела за рекурсивно разделяне (0,9835) показва, че загубата на H3K27me3 е често срещана при IDH1-R132H Mut олигодендроглиома. Нашите резултати показват, че имунохистохимичната оценка на H3K27me3 е икономически ефективен и надежден метод за определяне на 1p/19q коделиция заедно с IDH1-R132H и ATRX имунооцветяване, дори при липса на 1p/19q изследване.

2. Gatti, D.M., Zhao, N., Chesler, E.J., Bradford, B.U., Shabalin, A.A., Yordanova, R., Lu, L., Rusyn, I., Sex-specific gene expression in the BXD mouse liver, (2010) Physiological Genomics, 42 (3), 456-468, DOI: 10.1152/physiolgenomics.00110.2009

Abstract

Differences in clinical phenotypes between the sexes are well documented and have their roots in differential gene expression. While sex has a major effect on gene expression,

transcription is also influenced by complex interactions between individual genetic variation and environmental stimuli. In this study, we sought to understand how genetic variation affects sex-related differences in liver gene expression by performing genetic mapping of genomewide liver mRNA expression data in a genetically defined population of naive male and female mice from C57BL/6J, DBA/2J, B6D2F1, and 37 C57BL/6J X DBA/2J (BXD) recombinant inbred strains. As expected, we found that many genes important to xenobiotic metabolism and other important pathways exhibit sexually dimorphic expression. We also performed gene expression quantitative trait locus mapping in this panel and report that the most significant loci that appear to regulate a larger number of genes than expected by chance are largely sex independent. Importantly, we found that the degree of correlation within gene expression networks differs substantially between the sexes. Finally, we compare our results to a recently released human liver gene expression data set and report on important similarities in sexually dimorphic liver gene expression between mouse and human. This study enhances our understanding of sex differences at the genome level and between species, as well as increasing our knowledge of the molecular underpinnings of sex differences in responses to xenobiotics.

Абстракт

Разликите в клиничните фенотипове между половете са добре документирани и се коренят в диференцираната генна експресия. Въпреки че полът има основен ефект върху генната експресия, транскрипцията се влияе и от сложни взаимодействия между индивидуалните генетични вариации и стимулите на околната среда. В това проучване се опитахме да разберем как генетичната вариация влияе върху свързаните с пола различия в експресията на чернодробни гени, като извършихме генетично картографиране на мРНК експресионни данни в черния дроб в генетично определена популация от наивни мъжки и женски мишки от рекомбинантни инбредни щамове C57BL/6J, DBA/2J, B6D2F1 и 37 C57BL/6J X DBA/2J (BXD). Както се очакваше, открихме, че много гени, важни за метаболизма на ксенобиотиците и други важни пътища, показват полово диморфна експресия. Извършихме също така мапиране на локуси за количествени признаци на генната експресия в този панел и открихме, че най-значимите локуси, които изглежда регулират голям брой гени, отколкото се очаква по случайност, са до голяма степен независими от пола. Съществено, ние установихме, че степента на корелация в рамките на мрежите за генна експресия се различава значително между половете. Накрая, ние сравняваме нашите резултати с наскоро публикуван набор от данни за генната експресия на човешкия черен дроб и съобщаваме за важни прилики в полово диморфната експресия на чернодробните гени между мишката и човека. Това проучване подобрява разбирането ни за половите различия на ниво геном и между видовете, както и увеличава познанията ни за молекулярните основи на половите различия в отговорите към ксенобиотици.

3.Hanumegowda, U.M., Wenke, G., Regueiro-Ren, A., Yordanova, R., Corradi, J.P., Adams, S.P., Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds, (2010) Chemical Research in Toxicology, 23 (4), pp. 749-755., DOI: 10.1021/tx9003825

Abstract

Drug-induced phospholipidosis (PLD) is an adaptive histologic alteration that is seen with various marketed drugs and often encountered during drug development. Various *in silico* and *in vitro* cell-based methods have been developed to predict the PLD-inducing potential of compounds. These methods rely on the inherent physicochemical properties of the molecule and, as such, tend to overpredict compounds as PLD inducers. Recognizing that the distribution of compounds into tissues or tissue accumulation is likely a key factor in PLD induction, in addition to key physicochemical properties, we developed a model to predict PLD *in vivo* using the measures of basicity (pKa), lipophilicity (ClogP), and volume of distribution (Vd). Using sets of PLD inducers and noninducers, we demonstrate improved concordance with this method. Furthermore, we propose a screening paradigm that includes a combination of various methods to predict the *in vivo* PLD-inducing potential of compounds, which may be especially useful in lead identification and optimization processes in drug discovery.

Абстракт Индуцираната от лекарствата фосфолипидоза (PLD) е адаптивна хистологична промяна, която се наблюдава при различни продавани лекарства и често се среща по време на разработването на лекарства. Различни *in silico* и *in vitro* клетъчни методи са разработени за прогнозиране на потенциала на съединенията да предизвикват PLD. Тези методи разчитат на присъщите физикохимични свойства на молекулата и като такива са склонни да предопределят съединенията като индуктори на PLD. Като отчитаме, че разпределението на съединенията в тъканите или тъканното натрупване вероятно е ключов фактор за индуциране на PLD, в допълнение към основните физикохимични свойства, ние разработихме модел за прогнозиране на PLD *in vivo*, използвайки мерките за основност (pKa), липофилност (ClogP) и обем на разпределение (Vd). Използвайки набори от PLD индуктори и неиндуктори, ние демонстрираме подобро съответствие с този метод. Освен това предлагаме парадигма за скрининг, която включва комбинация от различни методи за прогнозиране на потенциала на съединенията за индуциране на PLD *in vivo*, която може да бъде особено полезна в процесите на идентифициране и оптимизиране на водещи съединения при откриването на лекарства.

4. Kirova, R., Georgiev, V., Rubino, B., Sampalmieri, R., Yordanov, B., Asymptotic behavior for linear and nonlinear elastic waves in materials with memory, (2008) Journal of Non-Crystalline Solids, 354 (35-39), pp. 4126-4137.DOI: 10.1016/j.jnoncrysol.2008.06.020

Abstract

In this review, we study the Cauchy problem associated to the equation of linear and nonlinear viscoelasticity with memory. Our first point is the study of dispersive properties of the solution to the linear equation of viscoelasticity with memory. The decay estimates obtained in this first part are important to treat the corresponding nonlinear Cauchy problem. The key novelty is the fact that we admit algebraic singularities and decay at infinity for the time dependent functions in the memory kernel. This fact enables one to include models different from the classical viscoelasticity problem, where this kernel is smooth and exponentially decaying in time. © 2008 Elsevier B.V.

Абстракт В този обзор изучаваме задачата на Коши, свързана с уравнението на

линейна и нелинейна вискоеластичност с памет. Първата ни точка е изследването на дисперсионните свойства на решението на линейното уравнение на вискозната еластичност с памет. Получените в тази първа част оценки на разпадането са важни за разглеждането на съответната нелинейна задача на Коши. Основната новост е фактът, че допускаме алгебрични сингулярности и разпад на безкрайност за зависимите от времето функции в ядрото на паметта. Този факт позволява да се включат модели, различни от класическата задача за вискоеластичност, където това ядро е гладко и експоненциално разпадащо се във времето.

5. Gatti, D., Maki, A., Chesler, E.J., Kirova, R., Kosyk, O., Lu, L., Manly, K.F., Williams, R.W., Perkins, A., Langston, M.A., Threadgill, D.W., Rusyn, I., Genome-level analysis of genetic regulation of liver gene expression networks, (2007) *Hepatology*, 46 (2), pp. 548-557, DOI: 10.1002/hep.21682

Abstract

The liver is the primary site for the metabolism of nutrients, drugs, and chemical agents. Although metabolic pathways are complex and tightly regulated, genetic variation among individuals, reflected in variations in gene expression levels, introduces complexity into research on liver disease. This study dissected genetic networks that control liver gene expression through the combination of large-scale quantitative mRNA expression analysis with genetic mapping in a reference population of BXD recombinant inbred mouse strains for which extensive single-nucleotide polymorphism, haplotype, and phenotypic data are publicly available. We profiled gene expression in livers of naive mice of both sexes from C57BL/6J, DBA/2J, B6D2F1, and 37 BXD strains using Agilent oligonucleotide microarrays. These data were used to map quantitative trait loci (QTLs) responsible for variations in the expression of about 19,000 transcripts. We identified polymorphic local and distant QTLs, including several loci that control the expression of large numbers of genes in liver, by comparing the physical transcript position with the location of the controlling QTL. Conclusion: The data are available through a public web-based resource (www.genenetwork.org) that allows custom data mining, identification of coregulated transcripts and correlated phenotypes, cross-tissue, and cross-species comparisons, as well as testing of a broad array of hypotheses.

Абстракт Черният дроб е основното място, където се извършва метаболизмът на хранителните вещества, лекарствата и химичните агенти. Въпреки че метаболитните пътища са сложни и строго регулирани, генетичните различия между хората, отразени в разликите в нивата на генна експресия, внасят сложност в изследванията на чернодробните заболявания. Това проучване открива генетичните мрежи, които контролират експресията на чернодробни гени, чрез комбинация от широкомащабен количествен анализ на експресията на мРНК с генетично картографиране в референтна популация от рекомбинантни инбредни щамове мишки BXD, за които данни за еднонуклеотиден полиморфизъм, хаплотип и фенотип са публично достъпни. Профилирахме генната експресия в черния дроб на наивни мишки от двата пола от щамове C57BL/6J, DBA/2J, B6D2F1 и 37 BXD, като използвахме олигонуклеотидни микрочипове Agilent. Тези данни бяха използвани за картографиране на локуси с количествени признаци (QTL), отговорни за вариациите в експресията на около 19 000 транскрипта. Идентифицирахме полиморфни

локални и отдалечени QTL, включително няколко локуса, които контролират експресията на голям брой гени в черния дроб, като сравнихме физическата позиция на транскрипта с местоположението на контролиращия QTL. Заключение: Данните са достъпни чрез публичен уеб-базиран ресурс (www.genenetwork.org), който позволява персонализирано извличане на данни, идентифициране на корегулиращи транскрипти и корелирани фенотипи, сравнения между тъкани и видове, както и тестване на широк спектър от хипотези.