ISSN 1539–2791

Volume 1 • Number 4 • 2003

Neuroinformatics

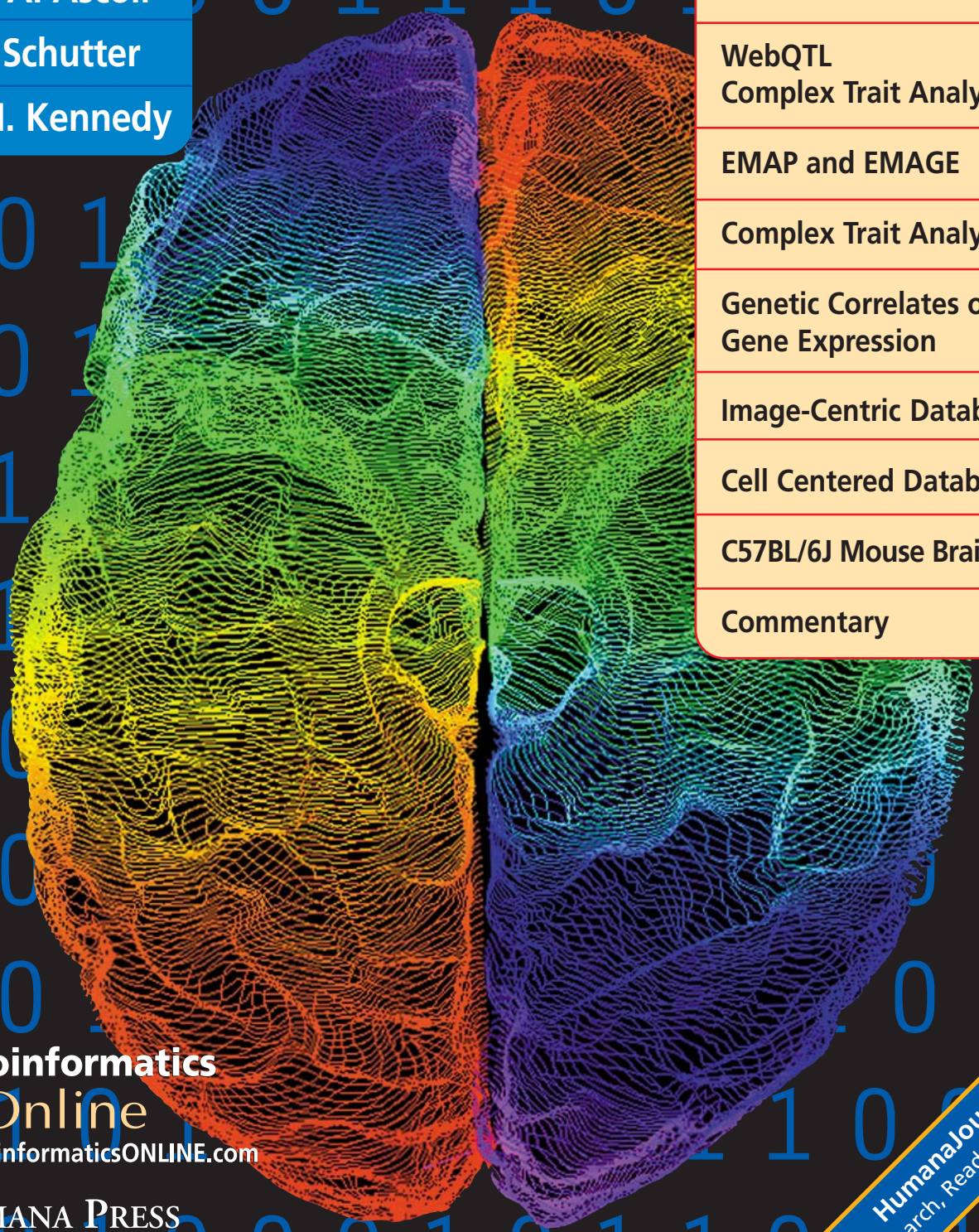
Editors

Giorgio A. Ascoli

Erik De Schutter

David N. Kennedy

IN THIS ISSUE



Neuroinformatics

ISSN 1539-2791/03/343–358/$25.00

Original Article

Genetic Correlates of Gene Expression in Recombinant Inbred Strains

A Relational Model System to Explore Neurobehavioral Phenotypes

Elissa J. Chesler,[*,1] *Jintao Wang,*[2] *Lu Lu,*[1] *Yanhua Qu,*[1] *Kenneth F. Manly,*[2] *and Robert W. Williams*[1]

[1]Department of Anatomy and Neurobiology, Center for Genomics and Bioinformatics, University of Tennessee Health Science Center, Memphis TN; [2] Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY

Abstract

Full genome sequencing, high-density genotyping, expanding sets of microarray assays, and systematic phenotyping of neuroanatomical and behavioral traits are producing a wealth of data on the mouse central nervous system (CNS). These disparate resources are still poorly integrated. One solution is to acquire these data using a common reference population of isogenic lines of mice, providing a point of integration between the data types. Recombinant inbred (RI) mice, derived through inbreeding of progeny from an inbred cross, are a powerful tool for complex trait mapping and analysis of the challenging phenotypes of neuroscientific interest. These isogenic RI lines are a retrievable genetic resource that can be repeatedly studied using a wide variety of assays. Diverse data sets can be related through fixed and known genomes, using tools such as the interactive web-based system for complex trait analysis, www.WebQTL.org. In this report, we demonstrate the use of WebQTL to explore complex interactions among a wide variety of traits—from mRNA transcripts to the impressive behavioral and pharmacological variation among RI strains. The relational approach exploiting a common set of strains facilitates study of multiple effects of single genes (pleiotropy) without *a priori* hypotheses required. Here we demonstrate the power of this technique through genetic correlation of gene expression with a database of neurobehavioral phenotypes collected in these strains of mice through more than 20 years of experimentation. By repeatedly studying the same panel of mice, early data can be re-examined in light of technological advances unforeseen at the time of their initial collection.

Index Entries: Behavioral genetics; genetic correlation analysis; relational databases; QTL mapping; recombinant inbred mice; oligonucleotide microarray.

*Address to which all correspondence and reprint requests should be sent.
E-mail: echesler@utmem.edu

Introduction

With the advent of new genomic technologies, including tremendous advances in sequence availability and mining, SNP genotyping, and gene microarray analysis, there has been a major push toward large-scale genetic studies of brain function. The potential of these emerging approaches has been embraced with excitement and a shift in focus in behavioral genetics away from the behavioral traits themselves, toward the study of biological "sub-phenotypes" has been advocated (Hamer, 2002). Ideally, a bioinformatic bridge can combine the achievements of decades of systems level research in the field with the large volumes of genome-era data for non-reductionist systems level analysis. High-throughput expression profiling, gene network analysis, functional brain imaging, and other integrative methods greatly enhance our understanding of mechanisms behind genetic mediation of behavior, but association of genetic factors with behaviorally relevant constructs must remain central. However, a platform for integration of behavioral genetics results with data from recent genomic technology is required. Studies typically carried out in isolated experimentally bred populations, or in mice of obscure/non-extant genetic backgrounds render findings difficult to incorporate with rapidly accumulating knowledge of genome structure. In contrast, data collected in recombinant inbred strains can be used in a relational fashion while retaining interesting genetic diversity and utility for genetic mapping.

Historically, recombinant inbred strains have allowed significant advances in behavioral genetics because these mice have fixed genomes. They can be repeatedly tested with no molecular analysis necessary because genotypes of known genetic markers are already stored in a database. This facilitates economically efficient genetic mapping (Belknap, 1998) and allows genetic dissection of traits that are highly variable or that require separate control groups. Using recombinant inbred mice, data can be cumulatively aggregated from different laboratories studying widely varying traits without any prior hypothesis motivating their simultaneous analysis. With a relational approach, data from all projects using these same mice can be assembled, resulting in a powerful community resource.

The mouse phenome project (Paigen and Eppig, 2000) is an excellent example of a community-wide collaborative resource for the cumulative use of data from mice of a fixed genetic background. Trait data spanning the range of morphological, physiological, life history, and behavioral traits are aggregated for multi-trait analysis in four sets of inbred strains. For behavior genetics, the inclusion of wild-derived strains has been a challenge as these mice, particularly CAST/Ei, are often not amenable to the testing conditions (Le Roy et al., 1998).

Though phenotypic means analyses of the standard inbred strains provide an ideal platform for relational studies, they are of limited utility as a mapping panel for genetic dissection of complex traits because of their small numbers, and complex, often unknown genetic backgrounds (Chesler et al., 2001; Darvasi, 2001). In typical mapping studies, the pair of inbred strains with the most extreme phenotypes are crossed, generating hundreds of unique recombinant progeny that must be individually phenotyped and genotyped. Data collected in the individuals cannot be aggregated across studies. The recombinant progeny of several of these common crosses have been inbred for many generations. In contrast to data from the unique individual progeny of the experimental cross, many genetically identical mice can be phenotyped and genotyped repeatedly. These mice have all the advantages of retrievability and relational databasing of experimental data that standard inbred strains

have, but also have sufficient statistical power (though less than that of an experimental cross) for genetic mapping.

The BXD recombinant inbred strain set has been assayed on a few hundred phenotypes, with over 250 literature reports since 1979 (PubMed, http://www.ncbi.nlm.nih.gov [1/24/03]). Since their earliest use in behavioral genetics (Crabbe et al., 1983), they have been evaluated on a vast array of neurobehavioral assays. This notably includes a large body of traits pertaining to alcoholism (Crabbe et al., 1994), leading to the identification of several quantitative trait loci determining susceptibility to these traits. Since the earliest of these mapping studies, new genotype resources have become available (Williams et al., 2001a), allowing much higher precision than was previously possible. As a result, early challenges faced by RI analysis, such as the detection of mirror loci (two loci possessing identical genotype strain distribution patterns in different regions of the genome) are now much more easily avoided.

Microarrays allow simultaneous quantification of steady-state abundance of thousands of transcript mRNAs and are often described as a massively parallel analysis of gene expression. In addition to mapping of susceptibility loci determining gene expression (*see* Wang et al., current issue), microarray assays in the RI mice can be used for genetic correlation analysis of gene expression with any other trait evaluated in these strains.

Genetic correlations determine the degree of shared genetic mediation of traits, i.e., the degree to which two traits are associated with the same biological substrates in genetically similar individuals. Using platforms for gene expression data collection that can be used cumulatively, e.g., Affymetrix oligonucleotide arrays, in a genetically retrievable resource, the genetic correlation of thousands of gene expression traits with hundreds of behavioral traits can be performed. Successful use of individual transcript correlations from our Affymetrix Gene Expression database (Hitzemann et al., 2003) and other individual transcript expression levels (Kirstein et al., 2002; Janowsky et al., 2001) with neurobehavioral phenotypes has been demonstrated.

Presently, we have assembled a database of phenotypes obtained in the BXD recombinant inbred strain set through review of the literature and personal communication with the authors (notably John C. Crabbe at Oregon Health Sciences University). Using the WebQTL (Wang et al., current issue) interface as a platform for genetic correlation analysis, users can re-map each of these traits using advanced high-density genotypes, and perform genetic correlation analysis with gene expression obtained in these same strains. This resource facilitates the integration of novel technologies for genome-wide analysis, and their deployment in this historically important strain set, allowing us to reach back into the literature to identify genetic bases for the neuroscientific phenotypes that are the ultimate target of study.

Methods

Strains, Animals, Sex, and Age

Female mice from the C57BL/6J, DBA/2J, F1(C57BL/6J × DBA/2J) and 30 RI strains were used in this study and another concurrent mapping study. In the majority of cases, we pooled tissues from sets of three animals of the same strain prior to isolation of mRNA. Mice were bred at the University of Tennessee, University of Memphis, and University of Alabama at Birmingham from stock originally obtained from the Jackson Laboratory (Bar Harbor, ME). Mice were maintained at 20–24°C on a 14/10h light-dark cycle in pathogen-free colonies. Most animals were fed a 5% fat Agway Prolab 3000 rat and mouse chow. Female mice were group housed three to six per cage in separate colonies.

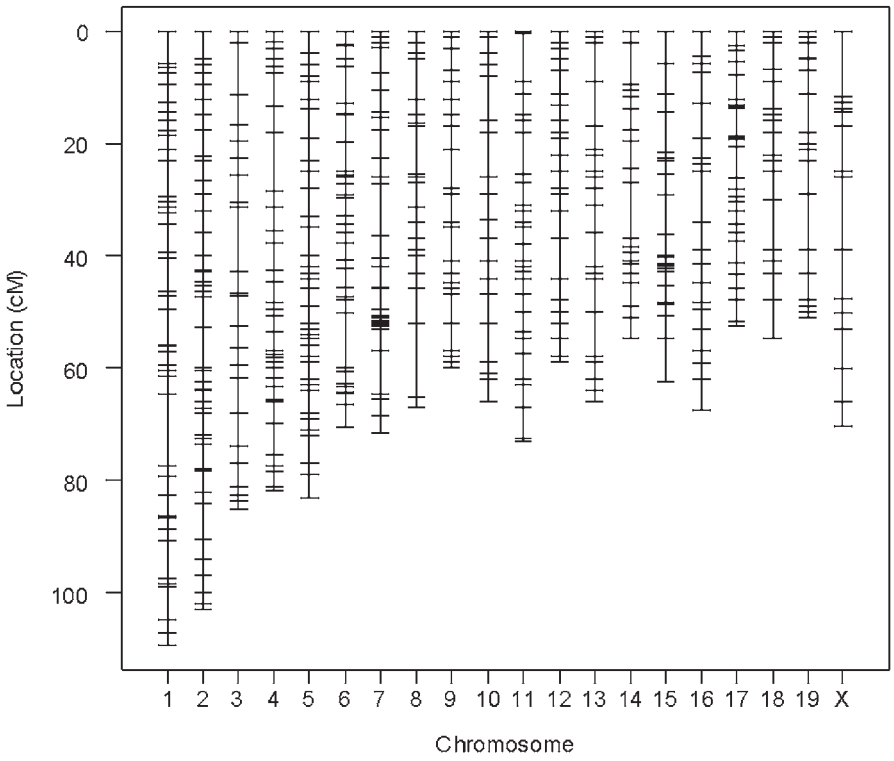


Fig. 1. Genetic map used for analysis of the BXDs. The map consists of over 570 markers polymorphic in C57BL/6J and DBA/2J mice. The resulting marker map is used for BXD RI expression QTL mapping.

Tissues

Animals were killed by cervical dislocation and brains were removed immediately. Fresh brains were immersed in RNAlater for 2–3 d at 4°C before dissection and extraction of RNA. The cerebellum and olfactory bulbs were cut free of the remainder of the brain as described in Airey et al. (2001) and Williams (2001b). The brainstem was cut at the level of the inferior colliculus and pons. The tissue sample that we will refer to as a forebrain more accurately consists of the forebrain (including subcortical structures) minus the olfactory bulb, and the midbrain and hindbrain removed. Tissues from three mice of each strain were processed together to extract total RNA using RNASTAT-60 and RNAlater according to the manufacturer's protocol. RNA quality and purity was assessed using an Agilent Bioanalyzer 2100 to assess relative quantities of 18s and 28s rRNA, the presence of DNA contamination, and evidence of RNA degradation. Using a Spectromax UV-vis spectrophotometer, the 260/280 n*M* absorbance ratio was determined to evaluate DNA:RNA ratio and to detect protein contamination. Over 95% of samples were of sufficient quality to generate cDNAs and cRNAs. cRNAs probes were prepared and hybridized to U74Av2 microarrays at Genome Explorations Inc. (Memphis, TN; www.genome-explorations.com) according to standard Affymetrix protocols. Technical variance between arrays is substantial even using the Affymetrix platform, and replication can reduce the relative impact of this variation. Thus, we used multiple (three to five) gene chips for each strain as replicates, each containing a pool of samples.

Microarray Data Generation and Preprocessing

Affymetrix DAT and CEL output files were processed using Affymetrix Microarray Analysis Suite 5.0 and dChip 1.0 (Wong et al., 2001). dChip was used to assess the quality of microarray data through model-based assessment of outliers. Microarrays that were identified as strong outliers were excluded from the data set if they did not meet two criteria: 1) The correlation of the chip data sets (MAS 5.0 signal) is greater than 0.9 with other members of the same brain tissue-type arrays; 2) The percentage of outliers detected using dChip is less than 5%; the percentage of dChip "present calls" is greater than 30%. Custom databases for all probe sets and for individual probe cells were assembled using FileMaker Pro with incorporated genome information about each gene and RNA samples.

Statistical Analysis

The probe set MAS 5.0 expression signal was transformed by $\log_2 (X+1)$ and standardized to an array mean of 8.0 and variance of 2. This transformation is a computationally efficient means of reducing the between array variance, though it assumes a constant total level of gene expression across chips. Statistical analysis, data processing, and visualization were performed using Microsoft Excel, SYSTAT (SYSTAT Software Inc., Richmond, CA), MATLAB (Mathworks), Data Desk 6.1.1 (Data Description, Inc., Ithaca NY; www.datadescription.com), and SAS (The SAS Institute, Cary, NC).

Genotype Database and Genetic Map

Genotypic data have been assembled by ongoing genotyping efforts (Williams et al., 2001). By screening genome sequence data for novel polymorphisms, many new BXD markers are being added to the BXD map (Fig. 1). See updates at http://webqtl.org/BXDGeno.html. Physical marker positions in megabases (MB) are determined by searching the most recent public assembly of the mouse genome.

Recombination position is determined by linkage analysis based on the rate of recombinations between markers. This has resulted in the construction of a high-density genetic map (Fig. 1) from which studies of genotype-phenotype associations for quantitative trait locus detection are performed.

This high-density map has been incorporated into WebQTL (Wang et al., 2003).

Literature Review

The construction of the BXD published phenotype database was largely based on a Medline search for "BXD" and "recombinant inbred mice." Unpublished results and additional sources were identified by personal communications with several of the authors of these studies who have also participated in the submission of new and revised trait data. All data are available in a Filemaker Pro database (www.nervenet.org) that contains links to PubMed for direct access to the original reference, and have also been incorporated in WebQTL. In the event that data were not tabulated but presented graphically, a vernier caliper was used to estimate trait means in each strain. The assembly of this database is an ongoing effort that will increase in value with community input.

Genetic Correlation Analysis

Pearson's product moment correlations or Spearman's rank correlations were used to obtain correlations between published BXD phenotypes and the 12,422 gene expression traits measured on the Affymetrix arrays. Normal probability plots of each phenotype allow the user to determine whether the assumptions of normality may be met by each phenotype. Scatterplots, linked from the cor-

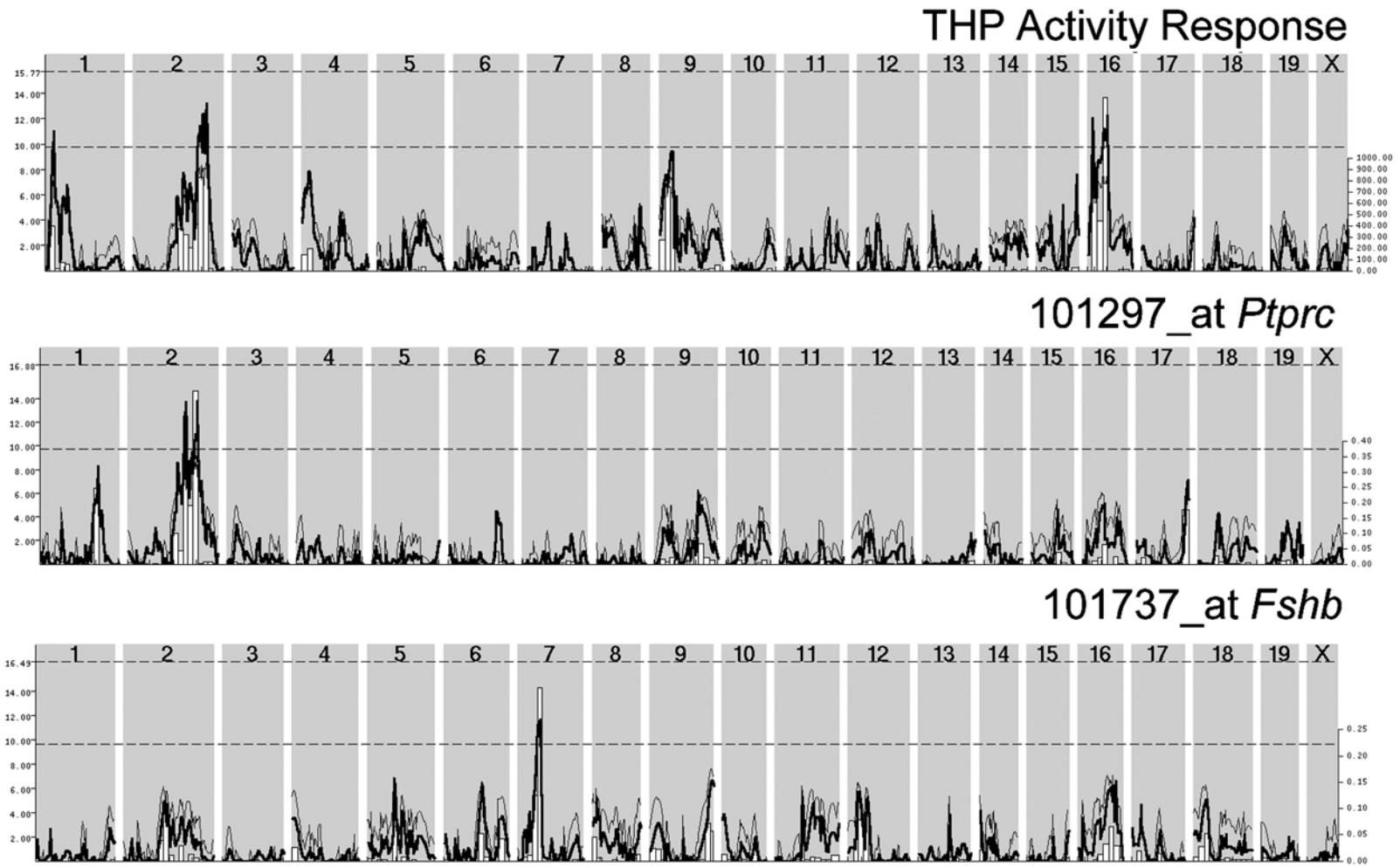


Fig. 2. Interval map showing a significant QTL on chromosome 2 regulating the activity response to a high dose of allopregnanolone, and expression levels of two transcripts highly correlated with this trait, receptor-type protein tyrosine phosphatase c (101297_at), and follicle-stimulating hormone (101737_at).

relation display, allow visual assessment of the linearity of the relationship. P-values were calculated from Fishers R-Z transformation, and control of the false discovery rate is implemented to guard against the high potential Type I error rate due to multiple testing. All analyses can be performed at www.webqtl.org.

Results

WebQTL (Wang et al., 2003) allows full text searching for traits of interest. The user may start with a search for either a transcript or behavioral trait of interest (Fig. 2), or by entering novel trait data for comparison with previously published traits.

Interval mapping of the traits can be performed in WebQTL using recently developed high-density genotype resources (Fig. 3). A few examples are illustrated in Figs. 2 and 3.

Genetic Correlations and Transcription Regulatory Quantitative Trait Loci

The activity response to a high dose of THP (allopregnanolone) is regulated by a suggestive quantitative trait locus (QTL) on chromosome 2 and two other suggestive QTLs, one on proximal chromosome 9 and proximal chromosome 16 (Fig. 2; Palmer et al., 2002). Mapping the QTL locations of several of the correlated transcripts shows that some may share a common genetic mediation, e.g., the receptor protein tyrosine phosphatase, *Ptprc* (101297_at). It should be noted that this association may be spurious, i.e., that the strain distribution pattern of the trait values are similar, and thus they map to roughly the same locations. Other highly correlated transcripts including *FSHb* (101737_at) do not have regulatory QTLs in these regions (Fig. 2). These correlations may indicate an important role for these genes in the trait, for example, that higher levels of FSH may alter plasma progesterone levels and thus affect brain neurosteroid metabolism. Indeed, the FSHb gene is located in the regulatory region for the THP-induced activity response.

Genetic Correlations by Position

Querying the Published Phenotypes for the term "seizure" results in the retrieval of several types of seizures: high pressure-induced (Plomin et al., 1991), audiogenic (Neumann and Seyfried, 1990), spontaneous and drug-withdrawal handling-induced convulsion (Buck et al., 1997), and cocaine-induced seizures (Hain et al., 2000). Each of these traits was re-mapped using an interval mapping procedure with our high-density genotype map (Fig. 3). A significant QTL was identified on chromosome 17 for high pressure-induced seizures, whereas the audiogenic seizures have common mediation by a locus on chromosome 12 (Neumann and Seyfried, 1990). Running a genetic correlation of this trait with gene expression, sorted by position, reveals that many major histocompatibilty 2 transcripts are correlated with high pressure seizures, all of which map to chromosome 17. Though these proteins have been related to seizure, the result of correlation analysis is associative, and not necessarily causal. Thus, such a finding might be explained by a genetic difference in processes resulting in elevated type I HLA levels, which also causes seizures, for example, glial scar formation (Beach et al., 1995). An alternative explanation is that this cluster of genes is regulated by a cis-acting locus on chromosome 17, which might be tightly linked to the locus regulating seizure phenotypes, such as the GABA B-1 receptor (Probe Set 102170_at), also on chromosome 17 and correlated with the seizure phenotype. Alternatively, a gene not on the Affymetrix array, linked to these regions may be involved in the phenotype. These can be explored directly through WebQTL via links from the interval map graphic display output. The human *mGLuR4* gene, linked to both

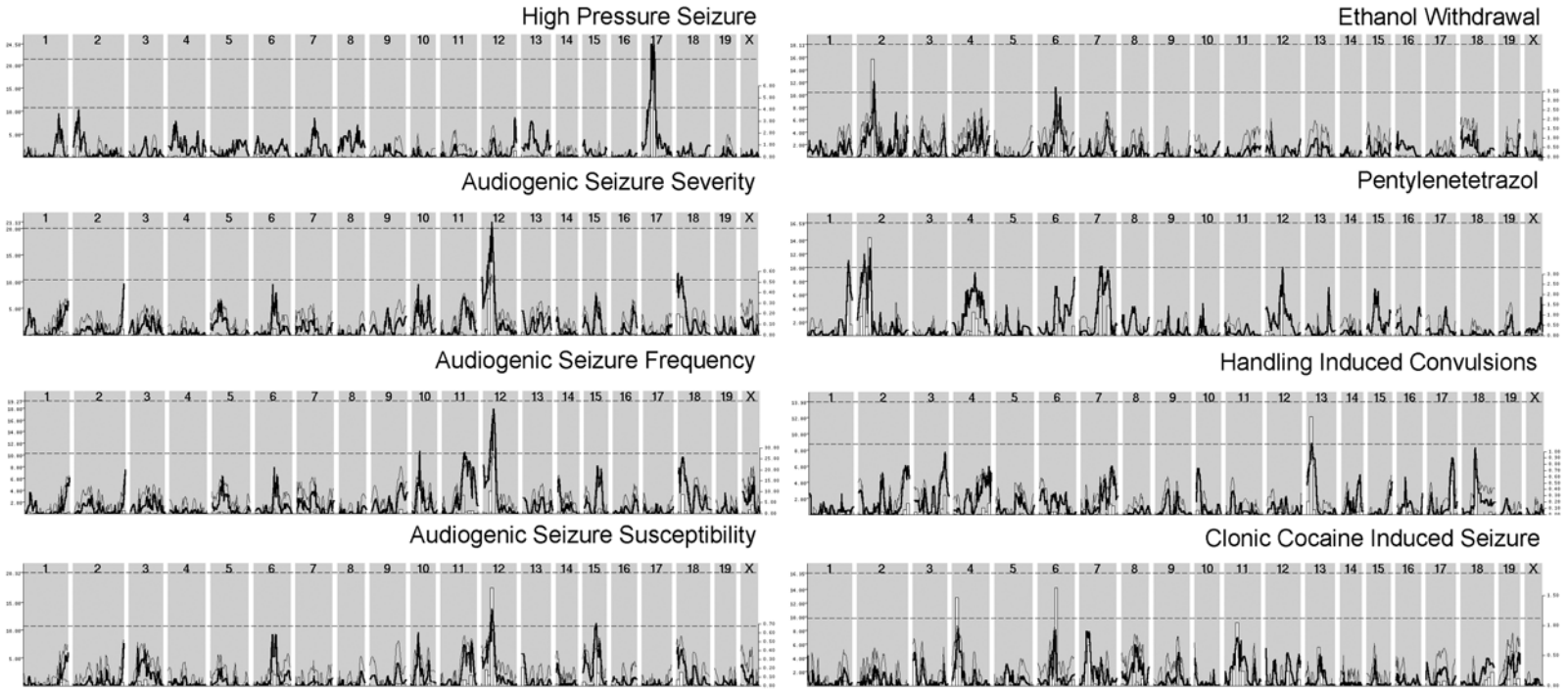
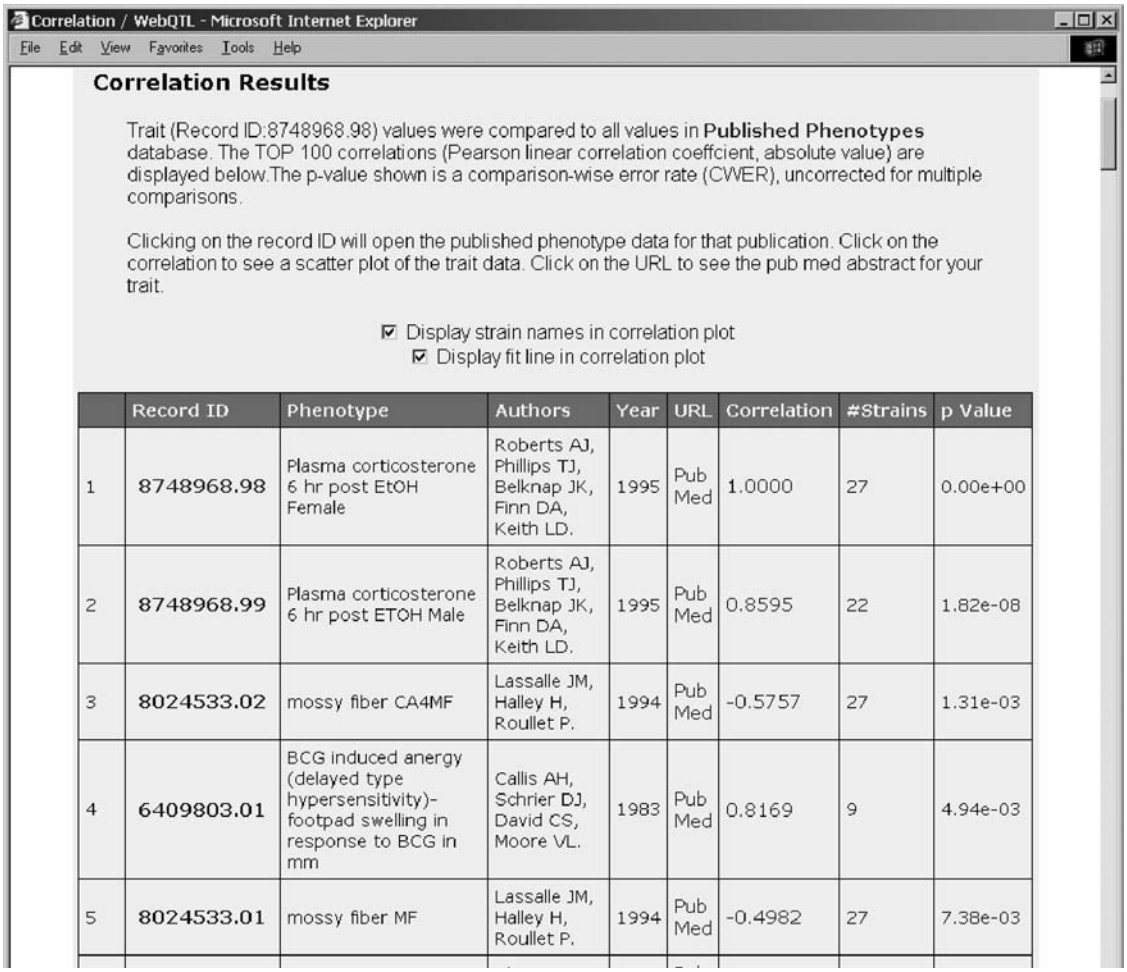


Fig. 3. Interval maps for eight published seizure phenotypes, showing different loci regulating high-pressure seizure, audiogenic seizure, handling induced convulsion, pentylenetetrazol-induced seizures, ethanol withdrawal effects on handling induced convulsions, and clonic cocaine-induced seizure.



Correlation Results

Trait (Record ID:8748968.98) values were compared to all values in **Published Phenotypes** database. The TOP 100 correlations (Pearson linear correlation coeffcient, absolute value) are displayed below.The p-value shown is a comparison-wise error rate (CWER), uncorrected for multiple comparisons.

Clicking on the record ID will open the published phenotype data for that publication. Click on the correlation to see a scatter plot of the trait data. Click on the URL to see the pub med abstract for your trait.

☑ Display strain names in correlation plot
☑ Display fit line in correlation plot

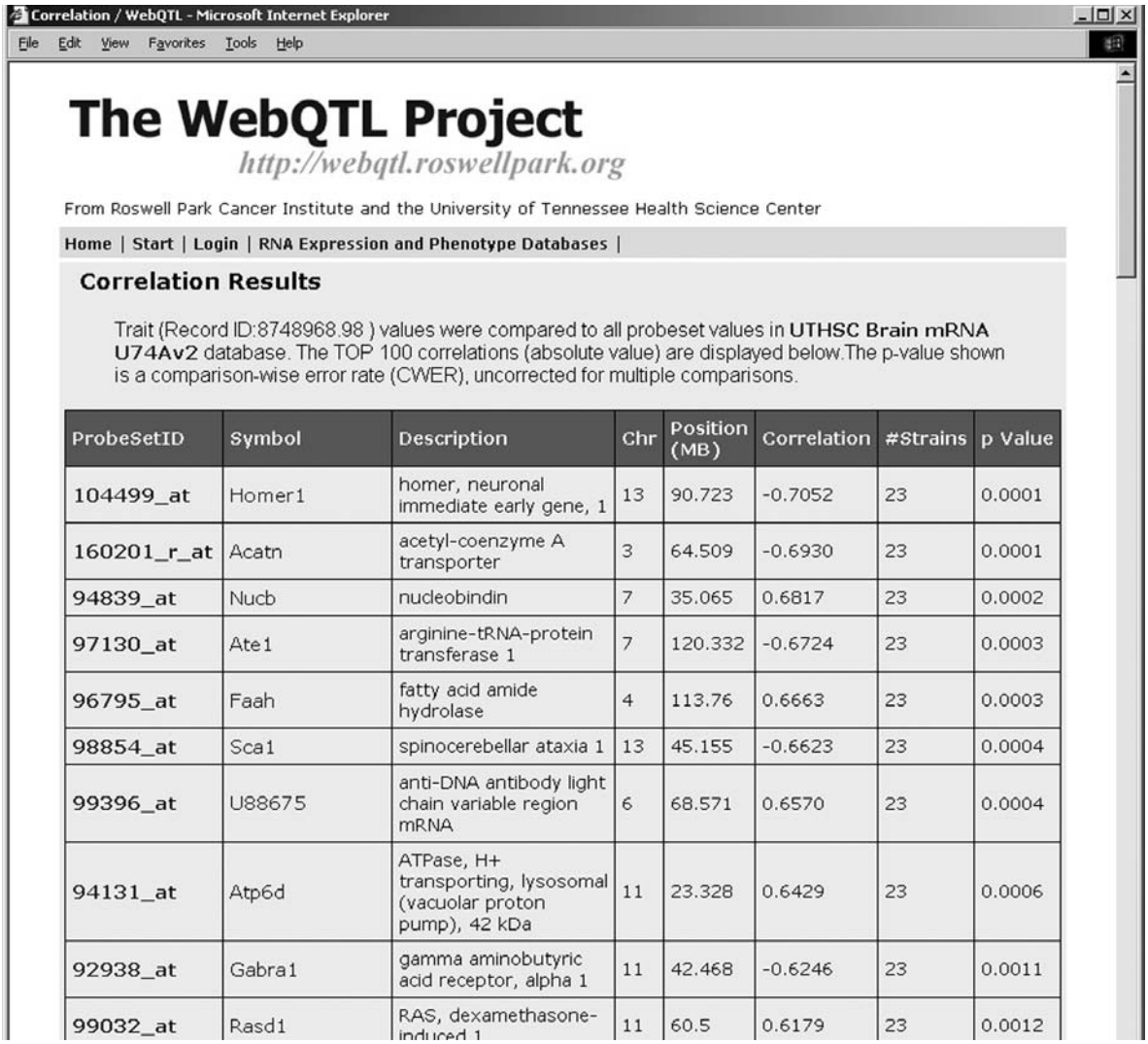
	Record ID	Phenotype	Authors	Year	URL	Correlation	#Strains	p Value
1	8748968.98	Plasma corticosterone 6 hr post EtOH Female	Roberts AJ, Phillips TJ, Belknap JK, Finn DA, Keith LD.	1995	Pub Med	1.0000	27	0.00e+00
2	8748968.99	Plasma corticosterone 6 hr post ETOH Male	Roberts AJ, Phillips TJ, Belknap JK, Finn DA, Keith LD.	1995	Pub Med	0.8595	22	1.82e-08
3	8024533.02	mossy fiber CA4MF	Lassalle JM, Halley H, Roullet P.	1994	Pub Med	-0.5757	27	1.31e-03
4	6409803.01	BCG induced anergy (delayed type hypersensitivity)-footpad swelling in response to BCG in mm	Callis AH, Schrier DJ, David CS, Moore VL.	1983	Pub Med	0.8169	9	4.94e-03
5	8024533.01	mossy fiber MF	Lassalle JM, Halley H, Roullet P.	1994	Pub Med	-0.4982	27	7.38e-03

Fig. 4. Results from correlation of published phenotypes with the corticosterone response to ethanol.

GABAbR1 and the major histocompatibility complex has also been shown to be involved in epilepsy (Wong et al., 2001). Several nonmutually exclusive hypotheses for chromosome 17 candidate genes are thus supported by mapping data, by transcription QTL analysis, by gene expression correlations, and by the literature, all of which are amenable to further testing.

Correlations Among Published Phenotypes and Gene Expression

Even in the absence of a strong QTL, genetic correlation analysis can reveal genes that have expression levels associated with a particular phenotype. The lack of a statistically significant genetic locus mediating the trait does not preclude the existence of genetic variation in the trait and association with gene expression. The corticosterone response to ethanol in females (Roberts et al., 1995) was chosen for this illustration because gene expression measurements were performed in females. This trait correlates highly with similar measurements obtained in males, and interestingly, with Bacillus Calmette-Guerin (BCG)-induced anergy and mossy fiber measurements in the hippocampus (Fig. 4).



ProbeSetID	Symbol	Description	Chr	Position (MB)	Correlation	#Strains	p Value
104499_at	Homer1	homer, neuronal immediate early gene, 1	13	90.723	-0.7052	23	0.0001
160201_r_at	Acatn	acetyl-coenzyme A transporter	3	64.509	-0.6930	23	0.0001
94839_at	Nucb	nucleobindin	7	35.065	0.6817	23	0.0002
97130_at	Ate1	arginine-tRNA-protein transferase 1	7	120.332	-0.6724	23	0.0003
96795_at	Faah	fatty acid amide hydrolase	4	113.76	0.6663	23	0.0003
98854_at	Sca1	spinocerebellar ataxia 1	13	45.155	-0.6623	23	0.0004
99396_at	U88675	anti-DNA antibody light chain variable region mRNA	6	68.571	0.6570	23	0.0004
94131_at	Atp6d	ATPase, H+ transporting, lysosomal (vacuolar proton pump), 42 kDa	11	23.328	0.6429	23	0.0006
92938_at	Gabra1	gamma aminobutyric acid receptor, alpha 1	11	42.468	-0.6246	23	0.0011
99032_at	Rasd1	RAS, dexamethasone-induced 1	11	60.5	0.6179	23	0.0012

Fig. 5. Results from correlation of gene expression with the corticosterone response to ethanol.

Both immune system anergy and hippocampal degeneration are observed in chronic elevation of corticosteroids, as might be expected in a hyperresponsive corticosteroid-releasing animal. Correlating this phenotype with gene expression reveals several interesting transcripts that are associated with the trait (Fig. 5), though not with each other (Fig. 6).

Among the transcripts that are correlated are *Homer1* (homer, neuronal immediate early gene, 1), *Acatn* (acetyl-coenzyme A transporter), *NucB* (nucleobindin), *Ate1* (arginine-tRNA-protein transferase 1), *Faah* (fatty acid amide hydrolase), *Sca1* (spinocerebellar ataxia 1), *Atp6d* (ATPase, H+ transporting, lysosomal [vacuolar proton pump], 42 kDA), *Gabra1*

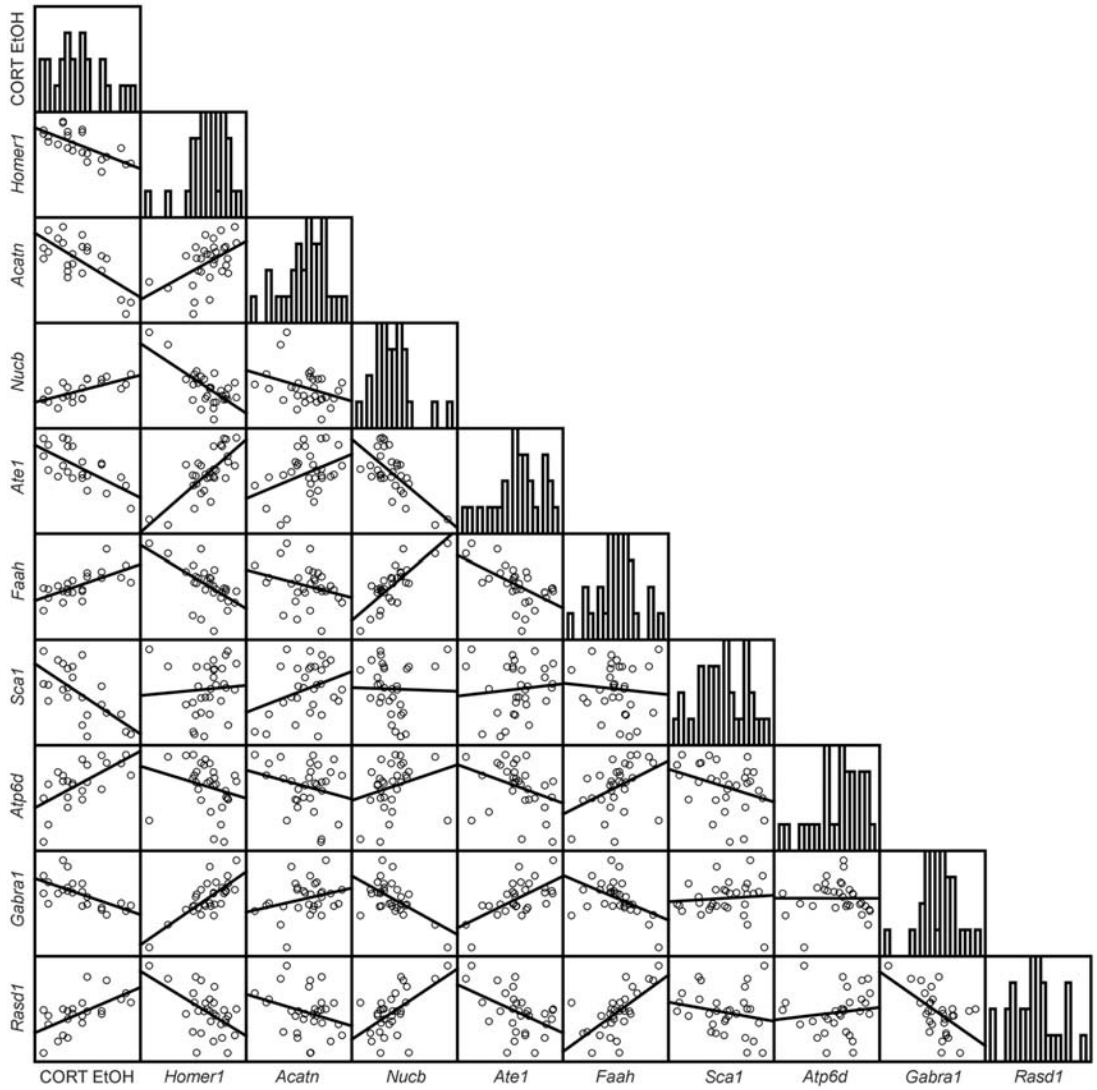


Fig. 6. Scatterplot matrix of expression levels of known genes on the Affymetrix array and the corticosterone response to ethanol. Each circle represents the x–y plotted strain means for a pair of traits. The best-fit least-squares regression line through the points is indicated. Frequency histograms of trait values are represented along the diagonal. Though each gene correlates with the phenotype (first column of plots), expression is not necessarily correlated between genes (e.g., *Sca1* and *Homer1*).

(gamma amino-butyric acid receptor, α1), and *Rasd1* (Ras, dexamethasone-induced 1).

Gene-to-Gene Correlations

Gene-to-gene correlations can be performed using WebQTL and can be quite useful in revealing pathway members. Performing a genetic correlation of the trait *Bag5* expression to the gene expression database reveals large numbers of oxidative stress and apoptosis-related transcripts. Several hundred of the correlations are greater than 0.6 and less than –0.6, and include many apoptosis-related transcripts, oxidation-related mitochondrial transcripts,

peroxidases and ionotropic receptor proteins, and ion channel proteins which suggest genetic differences in excitatory oxidative stress.

Discussion

In this report, we demonstrate a new web-based interface for the study of genetic correlations in recombinant inbred mice. The evaluation of trait-to-trait relationships in genetically identical mice allows the study of pleiotropic effects—regulation of multiple traits by single genes. Ordinary neurogenetic inquiry focuses on single-gene to single-trait hypotheses, with occasional forays into epistasis and multiple trait mapping. In contrast, this resource allows identification of novel associations between traits, including thousands of transcript expression levels. It is our hope that this becomes a true community resource, and that users submit trait data following instructions found from the database info link (http://www.webqtl.org/PublishDB.html).

The interpretations of genetic correlations are many, and this resource can be important in generation of mechanistic hypotheses from a variety of sources. The association of phenotypes using a family-means approach in genetically identical mice can indicate shared genetic mediation of phenotypes, and direct identification of some of the associated genes is possible. Using positional information about the transcripts, support for candidate genes underlying QTLs can be evaluated based on their association with traits. It should be noted that such information is not exclusionary, nor are expression differences an obligatory result of trait-relevant polymorphisms. While genetic correlations can indicate a causal relationship between gene expression and other traits, the relationship may also be due to shared association of the two correlated traits with a third variable, for example, shared linkage to a marker-strain distribution pattern, common environment, or technical influences. Despite these caveats to the interpretation of genetic correlation results, this synthesis of data, made possible only because of the fixed and definable status of recombinant inbred mice, provides much useful information about potential pleiotropic effects of segregating polymorphisms.

Genetic correlation analysis of gene expression and previously published phenotypes can now be performed using the WebQTL interface. This resource allows complex trait mapping and features links to existing bioinformatics resources, including locus link for further exploration of gene specific information, PubMed for published phenotype related information, and UCSC Genome Browser links for locus-specific information. The environment allows integration of existing molecular genetic information with trait data of a global organismic scope.

Using the relationships between genotype, gene expression, and behavior in three databases created in the same recombinant inbred strain set, advances in genome analysis technology have been applied to the reanalysis of traits that have been historically important foundations in neuroscientific research. Directly building on these early achievements is possible by using bioinformatics approaches to pull together newly developed resources and tools with the wide body of previous results in the field. As complete genome sequences in both of these strains become available, the exact locations of SNPs, which may be responsible for these phenotypic differences, will be determined. Large-scale proteomic methods can also be applied to these strain sets.

In contrast to other approaches involving outbred mice and experimental mapping crosses bred for individual experiments, mice of fixed genotypes facilitate the evaluation of traits under a variety of conditions. The influence of testing environment has been dramat-

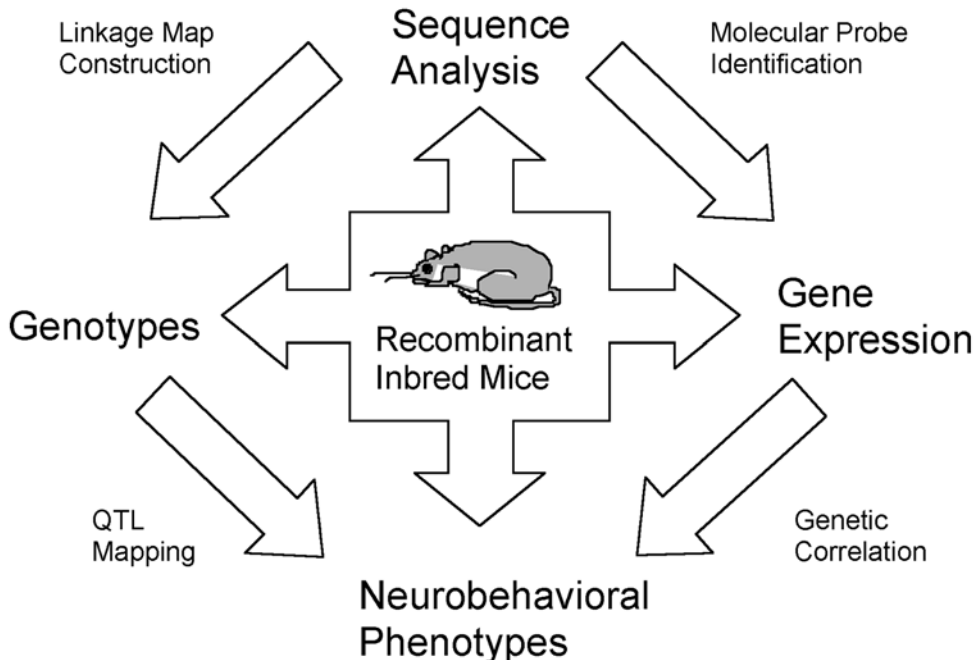


Fig. 7. The RI mouse as a model for relational genomic data analysis.

ically demonstrated by Crabbe and colleagues, testing inbred mice on several widely used behavioral assays in three different laboratories (Crabbe et al., 1999). While this has been pessimistically viewed by some to indicate the lack of generalizability of behavioral assessment in mice, it has also effectively emphasized the ability to examine the stability of traits across environments using inbred strains. Though the combination of data across laboratories might result in both false positive and false negative correlations of traits, the study of traits in multiple environments in genetically similar mice presents us with a unique ability to evaluate genes by environment interactions. Highly heritable behaviors such as locomotor activity are strongly correlated despite their study in multiple laboratories, whereas less heritable traits, such as ethanol withdrawal effects, are not as strongly correlated. By mining inbred strain data assembled through relational data models, archival data sets spanning long periods have been used together for identification of the particular laboratory influences on behavioral traits (Chesler et al., 2002 a,b).

The requisite resource for this relational approach is a genetically defined and expandable inbred strain set that allows for studies to be performed cumulatively over extensive periods of time in mice of the same genetic background. The BXD recombinant inbred strain series is an example of such a resource, and has been successfully deployed for relational genomic inquiry, linking diverse sets of experiments through a common genome (Fig. 7).

Despite the significant contributions in QTL detection made using this strain set, the mapping resolution is still too low to positively identify the genes giving rise to QTLs, even

with improved marker maps. The addition of more BXD strains is in progress, and will increase resolution approx two- to threefold. Development of an even more refined, genetically diverse, high-sample size, high-resolution strain set has been suggested (Threadgill et al., 2002). A recent proposal by the complex trait consortium for a community-wide recombinant inbred cross derived from 8-progenitors (www.complextrait.org) may result in a mapping panel with 1024 genetically retrievable inbred strains with adequate power for mapping and analysis of epistatic effects (Vogel, 2003). The power of this proposed strain set would increase the resolution of QTL mapping to levels amenable to cloning of QTLs. While the relational approach described here can be extended to newly developed or expanded recombinant inbred strain sets, the early published phenotype data itself will be more challenging to directly incorporate. However, its value in the design of future studies should not be underestimated. Historical data can serve to select a subset of these strains for analysis, chosen to maximize resolution in the regions of suggestive linkage identified in earlier studies. Transcript level data might also be a point of integration for phenotypic data from new and old strains, allowing users to generate predictive hypotheses about gene expression correlates with neurobehavioral traits.

WebQTL provides a structure for association of these diverse sources of data all tied together by a shared genome. The use of an internet-based collaborative data-mining tool will facilitate communication of research results, not just via data sharing and rapid online publicizing of results, but in the ability to engage in on-the-fly assessment of the interface of new findings with previous research. Thus, novel results can be immediately evaluated in the context of previous research in the field, leading to multiplicative integration of the growing body of findings in mouse genomics.

Acknowledgments

The authors gratefully acknowledge Dr. John Crabbe for curating the many traits studied by researchers at Oregon Health Sciences University, and Mary Kathleen Sullivan for her assistance in compiling the database. This work is a human brain project funded jointly by the NIMH, NIDA, and NSF (P20-MH62009 to KFM and RW), and is also supported by INIA grants VOIAA13499 and V24AA13513 and the William and Dorothy Dunavant Endowment in Pediatrics.

References

Airey, D. C., Lu, L., Williams, R. W. (2001) Genetic control of the mouse cerebellum: identification of quantitative trait loci modulating size and architecture. J. Neurosci. 21, 5099–5109.

Beach, T. G., Woodhurst, W. B., MacDonald, D. B., Jones, M. W. (1995) Reactive microglia in hippocampal sclerosis associated with human temporal lobe epilepsy. Neurosci. Lett. 191, 27–30.

Belknap, J. K. (1998). Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behav. Genet. 28, 29–38.

Buck, K. J., Metten, P., Belknap, J. K., Crabbe, J. C. (1997) Quantitative trait loci involved in genetic predisposition to acute alcohol withdrawal in mice. J. Neurosci. 17, 3946–3955.

Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L., Mogil, J. S. (2002a) Influences of laboratory environment on behavior. Nat. Neurosci. 5, 1101–1102.

Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L., Mogil, J. S. (2002b) Indentification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large datta archive. Neurosci. Biohav. Rev. 26, 907–923.

Chesler, E. J., Rodriguez-Zas, S. L., Mogil, J. S. (2001) In silico mapping of mouse quantitative trait loci. Science. 294, 2423.

Crabbe, J. C., Belknap, J. K., Buck, K. J., Metten, P. (1994) Use of recombinant inbred strains for studying genetic determinants of responses to alcohol. Alcohol. Suppl. 2, 67–71.

Crabbe, J. C., Kosobud, A., Young, E. R., Janowsky, J. S. (1983) Polygenic and single-gene determination of responses to ethanol in BXD/Ty recombinant inbred mouse strains. Neurobehav. Toxicol. Teratol. 5, 181–187.

Crabbe, J. C., Wahlsten, D., Dudek, B. C. (1999) Genetics of mouse behavior: interactions with laboratory environment. Science 284, 1670–1672.

Darvasi, A. (2001) In silico mapping of mouse quantitative trait loci. Science 294, 2423.

Hain, H. S., Crabbe, J. C., Bergeson, S.E., Belknap, J. K. (2000) Cocaine-induced seizure thresholds: quantitative trait loci detection and mapping in two populations derived from the C57BL/6 and DBA/2 mouse strains. J. Pharmacol. Exp. Ther. 293, 180–187.

Hamer, D. (2002) Genetics. Rethinking behavior genetics. Science 298, 71–72.

Hitzemann, R., Hitzemann, B., Rivera, S., et al. (2003) Dopamine D2 Receptor binding, Drd2 Expression and the number of dopamine neurons in the BXD recombinant inbred series: Genetic relationships to alcohol and other drug related phenotypes. Alc. Clin. Exp. Res. 27, 1–11.

Janowsky, A., Mah, C., Johnson, R. A., et al. (2001) Mapping genes that regulate density of dopamine transporters and correlated behaviors in recombinant inbred mice. J. Pharmacol. Exp. Ther. 298, 634–643.

Kirstein, S.L., Davidson, K.L., Ehringer, M.A., Sikela, J.M., Erwin, V.G., Tabakoff, B. (2002) Quantitative trait loci affecting initial sensitivity and acute functional tolerance to ethanol-induced ataxia and brain cAMP signaling in BXD recombinant inbred mice. J. Pharmacol. Exp. Ther. 302, 1238–1245.

Le Roy, I., Roubertoux, P.L., Jamot, L., et al. (1998) Neuronal and behavioral differences between Mus musculus domesticus (C57BL/6JBy) and Mus musculus castaneus (CAST/Ei). Behav. Brain Res. 95, 135–142.

Neumann, P.E., Seyfried, T. N. (1990) Mapping of two genes that influence susceptibility to audiogenic seizures in crosses of C57BL/6J and DBA/2J mice. Behav. Genet. 20, 307–323.

Paigen, K., Eppig, J. T. (2000) A mouse phenome project. Mamm Genome. 11, 715–717.

Palmer A. A., Miller M. N., McKinnon C. S., Phillips T. J. (2002) Sensitivity to the locomotor stimulant effects of ethanol and allopregnanolone is influenced by common genes. Behav. Neurosci. 116, 126–127.

Plomin, R., McClearn, G.E., Gora-Maslak, G., Neiderhiser, J.M. (1991) Use of recombinant inbred strains to detect quantitative trait loci associated with behavior. Behav. Genet. 21, 99–116.

Roberts, A. J., Phillips, T. J., Belknap, J. K., Finn, D. A., Keith, L. D. (1995) Genetic analysis of the corticosterone response to ethanol in BXD recombinant inbred mice. Behav. Neurosci. 109, 1199–1208.

Threadgill, D. W., Hunter, K. W., Williams, R. W. (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm. Gen. 13, 175–178.

Vogel, G. (2003) Genetics. Scientists dream of 1001 complex mice. Science 301, 456–457.

Wang, J., Williams, R. W., and Manly, K. F. (2003) WebQTL: Web-based complex trait analysis. Neuroinformatics 1, 299–308.

Williams, R. W., Gu, J., Qi, S., Lu, L. (2001a) The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. Genome Biol. 2, RESEARCH0046.

Williams, R. W., Airey, D. C., Kulkarni, A., Zhou, G., Lu, L. (2001b) Genetic dissection of the olfactory bulbs of mice: QTLs on four chromosomes modulate bulb size. Behav. Genet. 31, 61–77.

Wong, C. G, Scherer, S. W., Snead, O. C. 3rd, Hampson, D. R. (2001) Localization of the human mGluR4 gene within an epilepsy susceptibility locus(1). Brain Res. Mol. Brain Res. 87, 109–116.