Human Molecular Genetics, 2005, Vol. 14, Review Issue 2 **R163–R169**
doi:10.1093/hmg/ddi267

Genetical genomics: combining genetics with gene expression analysis

Jun Li[1] and Margit Burmeister[2,3,*]

[1]Stanford Human Genome Center, Stanford University, Palo Alto, CA, USA, [2]Department of Psychiatry and [3]Department of Human Genetics, Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109-0720, USA

Received June 27, 2005; Revised and Accepted July 14, 2005

The biological mechanisms that link genetic variation and its phenotypic outcome stand as a central puzzle in biology. Geneticists have usually approached this problem by trying to identify genetic variants that underlie the trait in question. Ten years ago, microarray technology opened a second front by making it possible to compare expression levels for most active genes under a variety of genetic and environmental conditions. A typical study reveals up- or down-regulation of genes or pathways associated with a phenotype (case/control) or condition (treated/untreated). In the past few years, a number of groups have started to combine gene expression studies with genetic linkage analysis, leading to a new synergy between these approaches. In this strategy, expression levels are treated as quantitative phenotypes and genetic variants that influence gene expression are sought. Several studies have shown that mRNA levels for many genes are heritable, thus amenable to genetic analysis. Quantitative trait loci mapping efforts have led to the initial characterization of genetic regulation in '*cis*' probably because of variants in the gene's own regulatory regions, as well as in '*trans*', i.e. by loci elsewhere in the genome. The existence of some 'master regulators' that each affects expression levels of hundreds of genes is an important finding that will surely enrich our understanding of regulatory networks. Although this novel field is still developing, understanding the genetic basis of molecular phenotypes such as gene expression is expected to shed light on the intermediate processes that connect genotype to cellular and organismal traits and represents a critical step towards true systems biology.

INTRODUCTION

Most common human diseases, including cancer, heart disease and schizophrenia, have complex etiologies, involving the action of many genes, as well as dynamic gene–environment interactions. To elucidate the mechanisms underlying disease susceptibility and progression and to improve diagnosis and treatment, an important strategy is to use genetic methods to identify the causative DNA variants and use this knowledge as the first step towards the eventual unraveling of the complex interplay between genes and environment. A second, more recent approach, made possible by the emergence of microarray technology since the early 1990s, is to analyze gene expression patterns globally in tissues from healthy and diseased individuals and use the steady-state mRNA levels to infer the maladaptive regulatory changes accompanying the disease. Until recently, genetic and gene expression studies were largely separate endeavors, involving different study designs, biological materials and analysis tools. Occasionally, a study searching for the overlap between genes expressed in a relevant pathway and the chromosomal region identified by linkage could pinpoint a clear candidate that turned out to be the correct gene (1). A systematic integration of genetic association and linkage results with gene expression results remains, however, a significant challenge.

In the past few years, genetic and gene expression approaches have been brought together, in what has been coined 'genetical genomics' (2), to study the genetic basis of gene expression (Fig. 1). The importance of understanding the genetic basis of gene expression, and by extension of biological regulation, is predicated on the widely held view that genetic contribution to phenotypic diversity is just as

*To whom correspondence should be addressed. Tel: +1 7346472186; Fax: +1 7346474130; Email: margit@umich.edu

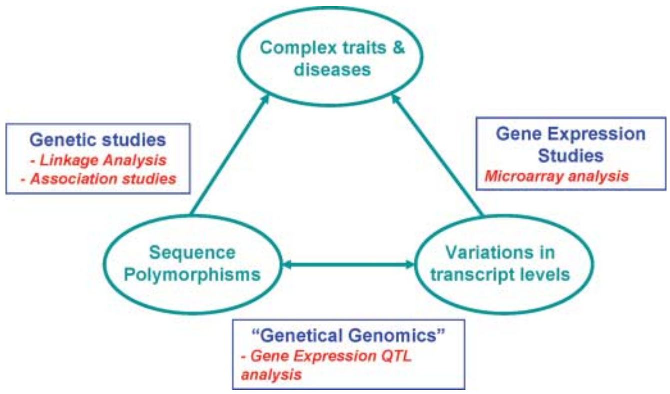

Figure 1. The 'genetical genomics' brings together the traditional genetic analysis and the gene expression studies by directly characterizing the genetic influence of gene expression.

likely to come from variations in amounts of proteins as from functional changes in them. These studies (3–11) (Table 1) follow a strategy outlined by Jansen and Nap (2): both mRNA data and DNA marker data are collected in tissue samples from genetically related individuals; the mRNA level of each of thousands of genes is treated as a separate quantitative phenotype, just like traits such as blood pressure or body weight. The chromosomal regions that affect steady-state levels of each transcript are then determined by conventional quantitative trait loci (QTLs) analysis. In the simplest terms, a significant QTL means that different genotypes at a polymorphic marker locus are associated with different trait values, in this case, expression levels. The power of this genomic strategy comes from our current ability to gather both genotype and gene expression data accurately, efficiently and on a global scale, thus enabling systems-level data mining. Although in the past, it took a significant amount of effort to show that a certain quantitative trait is variable, heritable and, furthermore, can be mapped to QTLs, with the genomic approach, each study instantly reveals hundreds of highly heritable molecular traits as well as hundreds of significant QTLs for the segregating population under study. These QTLs can lead to positional or functional candidates for downstream analysis and, especially when combined with QTL analysis of higher-level traits, may have important implication for both basic biology and medical research.

This strategy has been successfully applied to yeast (3,11,12), fly (13,14), mouse, plant and human (10). Most of these earlier studies have been reviewed elsewhere (15,16). Here, we will focus on the latest development in gene expression QTL mapping in human and rodent systems.

Table 1 summarizes the results of the two major experimental approaches discussed here. Morley *et al.* (9) and Monks *et al.* (8) both measured baseline levels of gene expression in lymphoblastoid cell lines (LCLs) from members of 14 and 15 CEPH (Center d'Etude du Polymorphisme Humain) families (17), respectively. Although differing in important details such as microarray platform, genetic markers, specific families and analysis used, both studies focussed on genes that

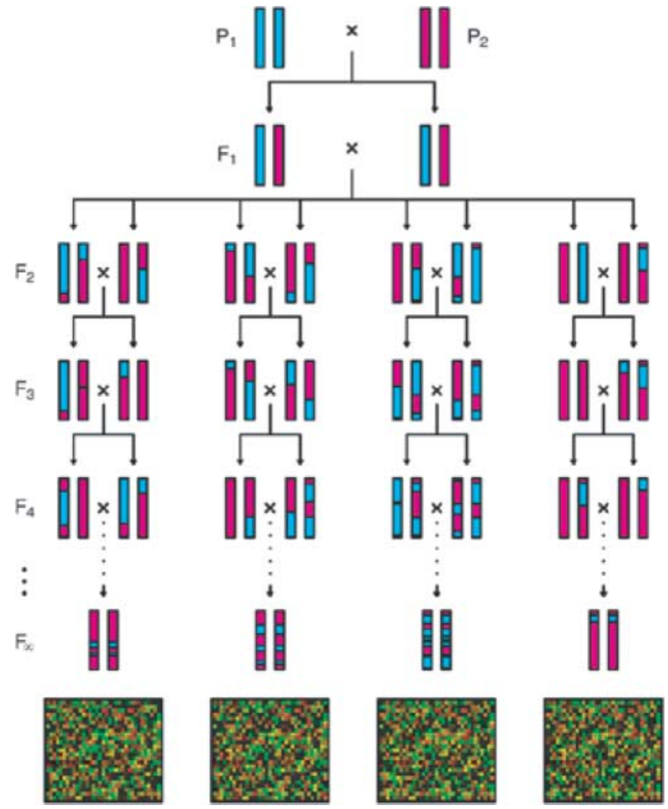

Figure 2. The construction of RI strains. Mating of two founder strains is followed by repeated sibling mating to produce many inbred lines, each of which represents a distinct genomic mosaic of the two parental genomes. One chromosome is depicted. Reprinted from Broman (33), with permission from Nature Publishing Group (http://www.nature.com).

showed high individual variation in expression and carried out genome-wide QTL analysis using genotype data that were already in the public domain, i.e. from the SNP Consortium (9) and the CEPH genotype database (8). Two more recent studies, by Chesler *et al.* (5) and Bystrykh *et al.* (4), analyzed forebrain and the hematopoietic stem cells (HSCs), respectively, in the same ~30 recombinant inbred (RI) strains of mouse. Figure 2 illustrates how these isogenic lines are formed by mating two genetically divergent parental strains, in this case the DBA/2 (D) and C57BL/6 (B) strains, followed by repeated sibling intercross of F2 progenies to produce individual inbred lines, each of which represents a distinct mosaic of the two founder genomes. The power of these strains, in contrast to standard F2 offspring, is that the recombination events are fixed, and recombinant animals of the same genotype are essentially unlimited in number. Thus, in a panel of RI lines, one can replicate experiments or characterize the consequences of perturbing environmental conditions on a diverse yet controlled genetic background (18), (Fig. 2). The fifth study, by Hubner *et al.* (7), applied the same approach to the fat and kidney tissues in the BXH/HXB rat RI strains, which were most often used as a model system for human hypertension and other metabolic syndromes.

Table 1. Five recent genetic studies of genes expression in humans and rodents

Studies	Samples	Gene expression data	Genetic data	Heritability (h^2)	QTL[a]	*Cis*-loci versus *trans*-loci	'Master regulators'	Other key findings	Ref
Morley *et al.*	LCLs from 14 CEPH families	Affymetrix arrays, focussing on 3554 most variable transcripts	~2500 SNP markers	Did not report	142 at point-wise $P < 4.3\text{E-}7$ (genome-wide $P < 0.001$); 984 at point-wise $P < 3.7\text{E-}5$	Among the top QTLs, 27 *cis*-loci, 110 *trans*-loci, five multiple QTLs	Two 'hotspots', one on chr.14 (seven QTLs), another on chr. 20 (six QTLs)	Experimentally confirmed several cis-QTLs as allelic differences in gene expression	(9)
Monks *et al.*	LCLs from 15 CEPH families	Agilent 25K oligonucleotide arrays, focussing on 2430 most variable transcripts	346 autosomal markers	762 were heritable at False discovery rate < 0.05, for a median h^2 of 0.34	33 at $P < 5\text{E-}6$; 50 at $P < 5\text{E-}5$; 132 at $P < 5\text{E-}4$ (all point-wise)	13 of the top 33, and 25 of the top 132 are *cis*-loci. *cis*-loci tend to show more significant linkage	Lack of evidence for 'hotspots'		(8)
Chesler *et al.*	Forebrain, 35 BXD RI strains of mouse	Affymetrix U74Av2 arrays, analyzed all transcripts on the array	779 non-redundant markers	Median of 11% for all transcripts, 608 with $h^2 > 33\%$	101 at $h^2 > 0.33$ and genome-wide $P < 0.05$ (FDR < 0.25)	83 of top 88 loci are *cis*-loci	Seven trans- 'bands', on chr. 1,2,6,10, 11,14,19, with up to 1650 transcripts controlled by one band	Examples of association between gene expression and behavioral traits, epistatic interaction, tissue specificity and gene networks	(5)
Bystrykh *et al.*	HSCs, 30 BXD RI strains	Same as above	Same as above	Did not report	1219 at genome-wide $P < 0.05$	At $P < 0.05$, 162 *cis*-loci, 1057 *trans*-loci. *cis*-loci tend to show more significant linkage	17 different loci controlling 10–272 transcripts	Identified 297 QTLs that were common between brain and HSCs, 222 were *trans*-QTLs, 75 were *cis*-[b]	(4)
Hubner *et al.*	Fat and kidney tissues, 30 BXH/HXB rat RI strains	Affymetrix RAE 230A arrays, all transcripts	1011 autosomal markers	Did not report	2118 (fat) and 2490 (kidney) at genome-wide $P < 0.05$	60–65% were *trans*-loci at $P < 0.05$, yet 80–100% were *cis*-loci at $P < 0.0001$	In fat, a chr.17 QTL for 42 transcripts; in kidney a chr.3 QTL for 28 transcripts	15% QTLs, mostly *cis*-QTLs, were common in fat and kidney tissues	(7)

[a]The reported statistical significance levels (*P*-values and false discovery rates) are not always suitable for direct comparisons across studies, which are often different in sample size, RNA pooling, numbers of technical replication, marker density or methods of statistical analysis. Most studies used permutation to derive genome-wide *P*-values across all genetic loci and false discovery rate to control for testing 10^3–10^4 transcripts.
[b]Although there were three times more *trans*-loci than *cis*-loci among the common QTLs between brain and HSC, the *trans*-QTLs in HSC were six times more than *cis*-QTLs.

HERITABILITY OF GENE EXPRESSION LEVELS

A prerequisite for any genetic study is to demonstrate that the trait in question is influenced by inherited factors. Several earlier studies have shown that transcript levels for many genes are indeed heritable (3,6,10). For newer examples, Chesler *et al.* (5) reported, in forebrain of BXD RI mice, a median heritability of 11% across all transcripts on the microarray, with 608 transcripts having >33% variance accounted for by strain. Monks *et al.* (8) focussed on a subset of 2430 genes, which were differentially expressed in LCLs from 15 human families, and found expression in 762 genes (31%) to be significantly heritable (at a false discovery rate $P < 0.05$), for a median heritability of 34%. These estimates are similar to those reported in yeast, where a median heritability of 27% was reported among a set of 1038 transcripts with strongest detected QTLs (12). Although the exact heritability estimates depend on factors such as sample size, tissue type, statistical model, amount of genetic diversity and environmental variabilities, these studies have revealed that hundreds to thousands of transcripts were clearly influenced by inherited factors and collectively confirmed that variations in mRNA levels are heritable traits amenable for genetic analysis, and therefore can serve as possible 'intermediate phenotypes' between genetic risk factors and grossly observable traits or diseases.

MAPPING QTLS: *CIS*- OR *TRANS*-REGULATION?

The genetic analysis of gene expression naturally leads to the classification of QTLs into *cis*-acting and *trans*-acting classes based on the relative genomic locations of the transcript and its QTL. This has provided a glimpse into some basic principles regarding the relative contributions of *cis*-acting versus *trans*-acting loci, summarized in Table 1. In human LCLs, Morley *et al.* (9) found significantly more *trans*-acting QTLs ($N = 110$) than *cis*-acting QTLs ($N = 17$). Some of the *trans*-acting QTLs were found to aggregate in genomic 'hotspots'. These hotspots presumably contain the 'master regulators', each controlling a large number of transcripts. All the three rodent RI studies found more *cis*-regulators and also strong evidence for 'master regulators'. Chesler *et al.* (5) reported that 83 of the top 88 QTLs were *cis*-acting. In addition, they found seven *trans*-acting QTLs, each of which influenced the expression of hundreds to thousands of individual transcripts. Bystrykh *et al.* (4) and Hubner *et al.* (7) also found that *cis*-QTLs tended to have more significant linkage evidence and a few *trans*-acting hotspots. In contrast, Monks *et al.* (8) found no evidence of hotspots, although they, like Morley *et al.* (9), used human CEPH LCLs, with eight of 14 families identical. This discrepancy may have many causes—these human studies, although using separately cultured, partially overlapping cell lines, differed in marker type, microarray platform, as well as in the approach in which genes were declared as differentially expressed. It is important to reconcile these differences in future studies. It should also be pointed out that with thousands of partially correlated phenotypes tested for linkage against the whole genome, the statistical problem of multiple testing brings many complications. The three rodent studies in Table 1 used a permutation test to control for testing ~1000 genetic markers and the calculations of false discovery rate to control for testing >10 000 transcripts. At a similar genome-wide *P*-value cutoff of 0.05, the Chesler *et al.* study (5) revealed 10–20-fold fewer QTLs than the other two rodent studies (4,7). While the rodent studies found generally more *cis*- than *trans*-QTLs, the two human studies reported predominantly *trans*-acting QTLs. Although part of these differences might arise from real biological differences between different tissues and between human and rodent systems (such as in genetic diversity and selective pressure), technical differences are also important to note. For example, both human studies focussed on highly variable traits. Thus, it is possible that *cis*-acting influence on expression shows smaller inter-individual variation in cultured human LCLs and that such small effects are less detectable when the environmental variability is large.

Some apparent *cis*-acting QTLs may be an experimental artifact due to hybridization to mismatched probes: if the target region of a probe contains a polymorphism, the transcript of one of the alleles will hybridize less well to the probe—effectively confounding expression differences with genotype differences. For example, some of the top *cis*-acting loci in Monks *et al.* (8) were located in the HLA region, where most genes are highly polymorphic and paralogous genes are highly similar. The authors have cautioned that these may not be regulatory QTLs but simple sequence differences. Similarly, we have demonstrated that Affymetrix chips can show an apparent *cis*-acting 'expression' difference that is completely explained by differential hybridization due to sequence differences in the probes (19,20). The impact of this sequence artifact on reported QTL results needs to be assessed. For interested readers, this additional analysis can be carried out only if the raw, probe-level data are always made available.

The fully genotyped, stable rodent RI lines allow not only a high degree of technical replication, but also the integration of data from multiple tissues, as well as comparison with organism level phenotypes. Between mouse brain tissue and blood, QTLs for 297 genes overlapped (4,5), with most of these, as one might expect, being *cis*-QTLs. Similarly, Hubner *et al.* (7) found that ~15% of the QTLs detected in rat kidney and fat tissues were common to both, with a majority of these common QTLs being *cis*-acting loci. Taken together, these data suggest that *trans*-effects are more likely to arise from tissue-specific regulation. However, *trans*-acting effects often interact with *cis*-effects and are inherently broader and more complex, reflecting the cumulative outcome of genetic, epigenetic and environmental regulations. The actual *trans*-acting polymorphisms may be coding variants in transcription factors that directly affect their binding affinities to target genes. Alternatively, indirect regulation may also come from a multitude of feedback control processes, affecting RNA stability, activities of the gene products, the state of the cell as a whole or the anatomy and cell type composition of the complex tissues. Genetic influences at all levels of biological organization, including intercellular signaling, may affect gene expression in *trans*, as has been suggested by Yvert *et al.* (11), who found that *trans*-regulators in yeast are not enriched for transcription factors *per se* but are distributed broadly across different categories of molecular function. Interestingly, Chesler *et al.* (5) pointed out that many of the hundreds of co-regulated target genes were transcription factors, pointing to a regulatory hierarchy, although the identity of the upstream 'master regulators' is still not known. This level of complexity, involving polygenic regulation (multiple regulators for one transcript), environmental input and pleiotropic effect (control of many transcripts by a master *trans*-regulator), makes it all the more challenging to delineate the underlying mechanisms.

REGULATORY NETWORKS

With the microarray data or the protein interaction data now available, one can attempt to reconstruct the associative networks: gene–gene correlation in expression levels across a large variety of perturbations, such as different growth or treatment conditions, different time points of a natural process (growth, cell cycle), disease versus health or engineered mutations, can be used to define groups of genes that are co-regulated and by inference, may serve shared functions (21), even directly interact with each other. The mapping of *trans*-acting QTLs often identifies sets of correlated transcripts as common targets of a *trans*-acting QTL, thus not only corroborating the associative networks learned from other types of perturbations, but also providing a powerful filter by reducing candidate nodes in the 'wiring diagram' of regulatory control. More importantly, the knowledge of genetic loci that influence gene expression may shed new light on such networks in terms of causality of regulatory relationships and the impact of genetic polymorphism on such networks. For example, Li *et al.* (22) described 66 QTL-derived candidate networks on the basis of 209 *trans*-QTLs from the Chesler and co-workers (4,5) data. Each network is a directed graph in which genes located in the QTL intervals are candidate regulators of the affected transcripts whereas expression levels of the regulators themselves may map to other, upstream QTLs.

In a recent study, Brem and Kruglyak (12) found that among the highly heritable transcripts in yeast, 40% had no QTL detected, 16% showed epistatic interaction and most may require more than five loci. Such amazing genetic complexity for a simple eukaryote illustrates the magnitude of the challenges lying ahead for higher organisms. The studies mentioned in Table 1 did not state whether the amount of variance explained by the discovered QTLs accounted for most of the total genetic variance as estimated from the heritabilities; nor could they systematically test interactions between pairs of loci because of the limited sample size. It is expected that, for some transcripts at least, a larger proportion of variance can be attributed to epistatic actions of multiple loci. Even with the current data set, it might be useful to rescan for interaction effects conditional on known positive QTL results. Knowledge of such statistical epistasis will be invaluable in forming testable hypotheses about actual biological interactions in gene networks.

CHALLENGES AND PROMISES

The genetic and regulatory mechanisms underlying disease etiology is one of the central challenges in today's biomedical research. One of the main difficulties lies in the inherent biological complexity. Genetic influences of gene expression occur in the contexts of the specific tissues, developmental stages and environmental inputs. For example, in human brains, from subjects with depression and bipolar disorder and controls, we found strong and widespread expression changes due to the severity of the physiological stress at the time of death (23). In this case, the impact of the condition at death was stronger than any genetic factor affecting transcription described to date. Such environmental factors distract from the factors of interest: genetic variants and effects of disorder on expression. But now, at least with LCLs and RI strains, we have the potential of separating the environmental factors from the genetic factors. Furthermore, for genome annotation, the newly identified gene expression QTLs are expected to facilitate the systematic identification of sequence elements that confer regulatory function. For disease etiology, the most likely candidate genes for future functional and association studies will be those that carry functional variants that impact either protein structure or transcriptional regulation. As an early example, Schadt *et al.* (10) found four candidate genes for obesity, which had gene expression QTLs co-localizing with clinical QTLs for obesity-related traits. One of these four, the Mup1 gene, was highly correlated in gene expression or shared common QTLs with many other genes known to be involved in the obesity trait. Several association studies in complex disorders recently identified non-coding SNPs or haplotypes associated with the disorder [e.g. G72 in schizophrenia and bipolar disorder (24,25) and GABRA2 in alcoholism (26)], usually postulating that the variants might affect expression. These claims can now be compared with the QTL results, because they predict *cis*-association between expression levels of the gene with the disease-associated variants.

The scope of genetic analysis of gene expression also presents enormous technical and analytical challenges. The genetic reference populations of rodent RI strains and the linked WebQTL (http://webqtl.org/) (27) provide an excellent example of a collaborative framework in which multiple investigators can contribute to data gathering and data mining on the vast number of possible phenotypes. In WebQTL, genetic data are stable and standardized, whereas phenotypic data at all levels, including gene expression, proteins and metabolites, anatomy, physiology and behavior, can accumulate in time and be scrutinized both for genetic influence of trait values and for correlations between different traits across multiple levels of organization. Similar multipurpose computational tools, as well as a standardized data format, will be absolutely essential for other systems such as the human LCLs. For the latter, it would be helpful if the CEPH cell lines and those lines studied in the HapMap project (28) can be integrated in a community-wide database, so that trait values such as the microarray data, collected under baseline conditions as well as during perturbation, can be routinely reanalyzed by using the genotype information that is already freely available for these lines. Although microarray expression researchers have adopted a common data management and exchange format (29,30) and a community-wide commitment to data accessibility and timely release (31,32), a similar model needs to be developed for the genetic analysis of quantitative phenotypes.

As the field of genetical genomics develops, it is expected to catalyze the formal integration of genetic and gene expression studies, which have so far been largely unrelated endeavors. A global understanding of genetic variations that affect gene expression will breathe new life into the vast amount of genetic linkage and gene expression data accumulated over previous decades for many important model systems for complex diseases. We will significantly improve our ability to dissect gene–environment interactions in light of their separable contributions to molecular phenotypes. Some common disorders will be understood as perturbations of the associated networks by both genetic and environmental factors. Normal phenotypic variation can already be integrated to some extent, at least in the case of RI strains, for which ample organism-level phenotypic data are available. In time, proteomic, physiological and functional imaging results, along with relevant interaction networks, will be integrated as well. The timely merger of gene expression with genetic analysis is just the beginning of a richer and more unified systems biology approach that the age of genomics has promised to us.

ACKNOWLEDGEMENTS

We thank Zhaohui (Steve) Qin and Viktoriya Strumba for critical reading. J.L. and M.B. are members of the Pritzker Neuropsychiatric Disorders Research Consortium, which is supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C.

Conflict of Interest statement. None declared.

REFERENCES

1. Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F. *et al.* (2003) Identification of a gene causing human cytochrome *c* oxidase deficiency by integrative genomics. *Proc. Natl Acad. Sci. USA*, **100**, 605–610.
2. Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.*, **17**, 388–391.
3. Brem, R.B., Yvert, G., Clinton, R. and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.
4. Bystrykh, L., Weersing, E., Dontje, B., Sutton, S., Pletcher, M.T., Wiltshire, T., Su, A.I., Vellenga, E., Wang, J., Manly, K.F. *et al.* (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet.*, **37**, 225–232.
5. Chesler, E.J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H.C., Mountz, J.D., Baldwin, N.E., Langston, M.A. *et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.*, **37**, 233–242.
6. Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M. and Spielman, R.S. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.*, **33**, 422–425.
7. Hubner, N., Wallace, C.A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., Mueller, M., Hummel, O., Monti, J., Zidek, V. *et al.* (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.*, **37**, 243–253.
8. Monks, S.A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J.W., Sachs, A. and Schadt, E.E. (2004) Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.*, **75**, 1094–1105.
9. Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S. and Cheung, V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
10. Schadt, E.E., Monks, S.A., Drake, T.A., Lusis, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.
11. Yvert, G., Brem, R.B., Whittle, J., Akey, J.M., Foss, E., Smith, E.N., Mackelprang, R. and Kruglyak, L. (2003) *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.*, **35**, 57–64.
12. Brem, R.B. and Kruglyak, L. (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl Acad. Sci. USA*, **102**, 1572–1577.
13. Wayne, M.L. and McIntyre, L.M. (2002) Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl Acad. Sci. USA*, **99**, 14903–14906.
14. Oleksiak, M.F., Churchill, G.A. and Crawford, D.L. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.*, **32**, 261–266.
15. Stamatoyannopoulos, J.A. (2004) The genomics of gene expression. *Genomics*, **84**, 449–457.
16. Chesler, E.J. and Williams, R.W. (2004) Brain gene expression: genomics and genetics. *Int. Rev. Neurobiol.*, **60**, 59–95.
17. Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R. (1990) Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics*, **6**, 575–577.
18. Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W. *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.*, **36**, 1133–1137.
19. Sliwerska, E., Meng, F., Speed, T., Jones, E., Bunney, W., Akil, H., Watson, S.J. and Burmeister, M. SNPs on chips: the hidden genetic code in expression arrays. *Biol. Psychiatry*, in press.
20. Dai, M., Wang, Y., Sliwerska, E., Burmeister, M., Watson, S.J., Akil, H. and Meng, F. (2005) Screening potential allele-frequency disequilibrium data from Affymetrix gene expression arrays. *Adv. Bioinform. Appl.*, in press.
21. Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N. and Davidson, G.S. (2001) A gene expression map for *Caenorhabditis elegans. Science*, **293**, 2087–2092.
22. Li, H., Lu, L., Manly, K.F., Chesler, E.J., Bao, L., Wang, J., Zhou, M., Williams, R.W. and Cui, Y. (2005) Inferring gene transcriptional modulatory relations: a genetical genomics approach. *Hum. Mol. Genet.*, **14**, 1119–1125.
23. Li, J.Z., Vawter, M.P., Walsh, D.M., Tomita, H., Evans, S.J., Choudary, P.V., Lopez, J.F., Avelar, A., Shokoohi, V., Chung, T. *et al.* (2004) Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum. Mol. Genet.*, **13**, 609–616.
24. Hattori, E., Liu, C., Badner, J.A., Bonner, T.I., Christian, S.L., Maheshwari, M., Detera-Wadleigh, S.D., Gibbs, R.A. and Gershon, E.S. (2003) Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. *Am. J. Hum. Genet.*, **72**, 1131–1140.
25. Chumakov, I., Blumenfeld, M., Guerassimenko, O., Cavarec, L., Palicio, M., Abderrahim, H., Bougueleret, L., Barry, C., Tanaka, H., La Rosa, P. *et al.* (2002) Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc. Natl Acad. Sci. USA*, **99**, 13675–13680.
26. Edenberg, H.J., Dick, D.M., Xuei, X., Tian, H., Almasy, L., Bauer, L.O., Crowe, R.R., Goate, A., Hesselbrock, V., Jones, K. *et al.* (2004) Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *Am. J. Hum. Genet.*, **74**, 705–714.

27. Chesler, E.J., Lu, L., Wang, J., Williams, R.W. and Manly, K.F. (2004) WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nat. Neurosci.*, **7**, 485–486.
28. The International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.
29. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.*, **29**, 365–371.
30. Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M. *et al.* (2002) Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.*, **3**, RESEARCH0046.
31. Ball, C.A., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M. *et al.* (2004) Submission of microarray data to public repositories. *PLoS Biol.*, **2**, E317.
32. Ball, C.A., Sherlock, G., Parkinson, H., Rocca-Sera, P., Brooksbank, C., Causton, H.C., Cavalieri, D., Gaasterland, T., Hingamp, P., Holstege, F. *et al.* (2002) Standards for microarray data. *Science*, **298**, 539.
33. Broman, K.W. (2005) Mapping expression in randomized rodent genomes. *Nat. Genet.*, **37**, 209–210.