Toxicogenetics: population-based testing of drug and chemical safety in mouse models

Ivan Rusyn[†,1,2], **Daniel M Gatti**[1], **Timothy Wiltshire**[3], **Steven R Kleeberger**[2,4], and **David W Threadgill**[2,5]

[1] Department of Environmental Sciences & Engineering, 0031 Michael Hooker Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2] Curriculum in Toxicology, University of North Carolina, Chapel Hill, NC, USA

[3] Division of Pharmacotherapy & Experimental Therapeutics, University of North Carolina, Chapel Hill, NC, USA

[4] Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

[5] Department of Genetics, North Carolina State University, Raleigh, NC, USA

Abstract

The rapid decline in the cost of dense genotyping is paving the way for new DNA sequence-based laboratory tests to move quickly into clinical practice, and to ultimately help realize the promise of 'personalized' therapies. These advances are based on the growing appreciation of genetics as an important dimension in science and the practice of investigative pharmacology and toxicology. On the clinical side, both the regulators and the pharmaceutical industry hope that the early identification of individuals prone to adverse drug effects will keep advantageous medicines on the market for the benefit of the vast majority of prospective patients. On the environmental health protection side, there is a clear need for better science to define the range and causes of susceptibility to adverse effects of chemicals in the population, so that the appropriate regulatory limits are established. In both cases, most of the research effort is focused on genome-wide association studies in humans where *de novo* genotyping of each subject is required. At the same time, the power of population-based preclinical safety testing in rodent models (e.g., mouse) remains to be fully exploited. Here, we highlight the approaches available to utilize the knowledge of DNA sequence and genetic diversity of the mouse as a species in mechanistic toxicology research. We posit that appropriate genetically defined mouse models may be combined with the limited data from human studies to not only discover the genetic determinants of susceptibility, but to also understand the molecular underpinnings of toxicity.

Keywords

collaborative cross; expression quantitative trait locus mapping; mouse genetics; population screening; toxicity testing; toxicogenetics

One of the goals of mechanistic toxicology is the discovery of the biochemical mechanisms underlying toxicity responses in humans. This research leads to the discovery of genes and polymorphisms responsible for toxicity and an understanding of how molecular pathways are altered in response to xenobiotic exposure. Recent advances in high-throughput methods of gathering biological data and in computational power to manipulate these datasets has changed the manner in which this research can be conducted. While toxicology will always rely upon systematic *in vitro* and *in vivo* investigation of model systems, the generation of new hypotheses regarding the genetic causes of toxicity has benefited greatly from the sequencing of mammalian genomes and the invention of gene-expression microarrays. The sequencing of several mammalian genomes, including mouse [1] and human [2], has opened up new methods for investigating the genetic basis of toxicity responses. Gene-expression microarrays have produced a global view of transcriptional changes in response to xenobiotic exposure and have supported the discovery of gene clusters, which may be part of the same biological pathway, that are related to injury. The combination of these two developments has opened up new approaches to the understanding of toxicity as it is affected by genetic variability [3].

Finding the associations between genes & complex traits: a preponderance of the human-centric approach

Recent advances in technologies that permit human genome-wide association studies (GWAS) with more than 1 million SNP have enabled the identification of genetic variants associated with important diseases [4]. In the past few years, more than 400 GWAS studies have been published, establishing a knowledge base that links hundreds of genetic variants to complex human diseases, as well as providing valuable insights into the complexities of their genetic architecture [5]. These studies have not been as successful as hoped for in the prediction of individual risk for developing a disease, but have been successful for identifying plausible molecular causes underlying polygenic diseases and traits. It is noteworthy that GWAS have 'found' many genes that have been known to be important in the pathogenesis of the relevant diseases [6].

With regards to adverse drug reactions, a number of important advances have also been made over the past decade with the help of GWAS approaches [7]. GWAS have identified gene targets for approved drugs, including thiazolidine diones and sulfonylureas, statins and estrogens [8,9]. The association between MHC alleles, especially *HLA-B*5701*, and susceptibility to adverse drug reactions manifesting through a diverse set of clinical phenotypes is one of the most intriguing stories in pharmacogenomics [10-12]. Genetic screening is now advised or recommended for a number of drugs on the market with known adverse drug reactions, and the introduction of new clinical tests is likely to intensify as the ongoing trials make their way to peer review [13].

Although GWAS are increasingly popular and affordable, they present a number of formidable logistical and technical challenges to the conduct of the studies and in the interpretation of the results [4]. These include the challenge of selecting a well-defined disease or trait suitable for analysis, the requirement for sufficiently large sample sizes, and the fact that most common variants, individually or in combination, confer relatively small increments in risk (up to 1.5-fold) and explain only a small proportion of heritability [5,14]. Human-only GWAS are also likely to remain expensive in terms of recruitment and characterization (both phenotyping and genotyping) of the study cohorts.

The limitations of human GWAS studies may be alleviated, at least partially, by the use of appropriate genetically-defined model systems. Cell-based models have been extensively used in preclinical drug development for years as a means to evaluate drug-induced toxicity

or to identify interactions of target compounds with drug-metabolizing enzymes and transporter proteins. Importantly, the availability of a large bank of commercially available and densely genotyped lymphoblastoid cell lines from the Centre d'Etude du Polymorphisme Humain (Paris, France) and Coriell Institute for Medical Research (NJ, USA) shows promise for *in vitro* pharmacogenetics research [15]. Rats are commonly used in drug safety evaluation experiments, and the genetically diversified inbred rat strains are being developed [16] and used for population-based toxicity testing [17]. However, among the mammalian organism-based laboratory models, the mouse offers an unparalleled wealth of genetic knowledge and resources, among which is a high-density SNP database encompassing more than 8 million polymorphic loci across hundreds of inbred strains [18,19].

Mouse genetics: a useful resource for population-based toxicology studies

Mouse genetic studies can complement many shortcomings of both *in vitro* and human-only approaches in pharmacogenomics. For example, collection of tissues from a wide variety of anatomical sites (e.g., brain and heart) or developmental stages is problematic in humans, as are many experimental interventions. A successful GWAS analysis is more likely when the phenotype of interest can be sensitively and specifically defined and measured, which is usually not a limitation in a mouse system. With regards to the environmental health sciences and toxicology, controlled exposure of people to environmental toxicants is often ethically unacceptable, which makes it challenging to interpret genetic associations produced in human cohorts exposed in the occupational or environmental setting without validation in animal studies.

In vivo toxicity screens and mechanistic studies are often carried out in a single strain of mouse [20-22]. This is done in order to fix as many variables as possible and has the benefit of standardizing the genotype across multiple chemicals. While this approach provides mechanistic information regarding toxicant activity in a single genetic background, the reality of human toxicity is more complex, including both diverse genetic backgrounds and uncontrolled environmental effects. The interpretation of the data with respect to the population-wide effects is plagued by the largely inaccurate generalizations from a single genome; inability to distinguish small and biologically important changes from background variation; ineffective exploitation of reproducible genetic variation to dissect differential response to chemical exposure; and inefficient use of defined genetic backgrounds to model particular phenotypic profiles observed in human populations. To address these important limitations, various animal models are being used by toxicologists to assess gene–environment interactions and determine genetic causes of interindividual variability in toxicity. Indeed, genetic background is an important component of toxicity responses [23,24], and a successful *in vivo* approach to modeling the effects of genetic diversity on toxicity will improve both the prediction of toxicity in humans as well as the identification of sensitive subpopulations.

Panels of genetically defined animals that provide a fixed genotype within a particular strain but encompass great genetic diversity across strains are being used in biomedical research [25]. Both standard intercrosses between inbred lines of mice, and large populations of inbred strains have been used as powerful tools for mapping quantitative trait loci (QTL). Inbred mouse strains represent fixed, renewable genotypes that are ideally suited for system biology approaches to improve our understanding of the mechanisms of toxicity and discovery of new biomarkers associated with biological responses to toxicant exposure. Panels of inbred mouse strains are also well-suited for identifying whole-genome response signatures indicative of chemical exposure because of the large knowledge base on the genetic lineage for hundreds of strains, and because the number and distribution of genetic

polymorphisms among mouse strains is equal to or exceeds that in the human population [26,27]. This approach also has the added advantage of 'repeat testing' in genetically identical individuals within a given strain, yielding important information regarding reproducibility of the response.

When a research study into the genetic basis of toxicity is initiated, the responsible genes being sought could lie anywhere in the genome. A forward genetics approach, in which the genetic basis of toxicity is investigated, is a reasonable approach to this problem. The first step is to search for evidence that the responsible genes lie in certain chromosomal regions by detecting correlations between the toxicity phenotype and genotype at loci throughout the genome. This is often carried out using QTL mapping [28,29]. This involves selecting or breeding a genetically segregating population, such as a backcross or intercross between two inbred parental strains that demonstrate quantitative variation in the toxicity phenotype of interest. The quantitative phenotype, such as liver histology score or serum alkaline aminotransferase levels, is measured in each individual, who is then genotyped at a number of genetic markers across the genome. A statistic of association, such as a likelihood ratio statistic or a linear model, is then calculated between the phenotypic values and each marker.

Using the 'mouse model of the human population' in translational clinical research

The diversity of the genotypes archived in different inbred mouse strains is ideally suited to identify and dissect genetic susceptibility in responses to toxicant exposure. To advance the understanding of health risks posed by toxicants, and the role that genetic diversity plays in determining the variability of responses between individuals and species, panels of inbred mouse strains can be used to demonstrate the benefits afforded by combining mechanistic toxicology with genetics.

A translational study, whereby candidate genes for susceptibility to toxicant-induced liver injury were discovered in a mouse population and subsequently validated in two independent human cohorts, was recently demonstrated using a well-known liver toxicant, acetaminophen [30]. A traditional 'human-only' approach to a genome-wide pharmacogenetic investigation into the genetic factors linked to liver toxicity of acetaminophen would require a much larger number of individuals to overcome statistical power limitations [31]. Conversely, a so-called 'candidate gene' analysis [10] is equally challenging owing to the complexity of the mechanism of toxicity and the inherent biases in candidate gene selection [32]. By utilizing a mouse model for acetaminophen-induced liver toxicity, whole-genome association analysis and targeted sequencing, polymorphisms in *Ly86*, *Cd44*, *Cd59a* and *Capn8* were identified that correlate strongly with liver injury [30]. Furthermore, these candidates were validated in two independent human cohorts where volunteers were exposed to the maximum recommended doses of acetaminophen. This study demonstrated that variation in the orthologous human gene, *CD44*, is associated with susceptibility to acetaminophen. Interestingly, well-characterized genes known to be essential for acetaminophen toxicity did not correlate with liver injury in the panel of mouse strains. This finding suggests that while *a priori* knowledge of the toxicant's mode of action can be useful in the selection of genes for follow-up analysis, validation of susceptibility-modulating genes in the laboratory is essential. This finding also illustrates the important difference between genes involved in mechanistic pathways leading to toxicant-induced injury and genes whose variants contribute to interindividual differences in susceptibility to toxicity, two areas that have potentially different gene sets.

It is noteworthy that the top candidate genes derived from the analysis of the mouse population were related to the immune response, and not to metabolism and detoxification of acetaminophen. Interestingly, in several cytokine knockout mouse studies of acetaminophen toxicity, the sensitivity to liver necrosis was also largely independent of glutathione depletion [32]. The traditional view on the mechanisms of toxicity, the approach widely utilized to predict individual responses to toxicants, would imply that the metabolism of acetaminophen to the reactive electrophile and/or its subsequent detoxification by glutathione conjugation should explain, at least to a considerable degree, the variability in responses. However, no apparent correlation between the levels of major metabolizing enzymes, glutathione, or acetaminophen plasma levels and liver injury was observed in the mouse population. Similarly, no correlation with sensitivity for polymorphisms in the genes encoding catalase or CYP2E1 was found, implying that variation at these key mediators of acetaminophen toxicity probably do not contribute to differential susceptibility.

It is worth noting that the variations in drug exposure and metabolism profile have been shown to be common causes of difference in adverse effects of chemicals across species and strains. For example, the well-described human interindividual variability in the metabolism of warfarin, specifically the generation of 7-hydroxywarfarin, was reproduced in a panel of inbred mouse strains [33], and it was determined that the phenotypic differences were associated with the polymorphisms in the *Cyp2c* locus. In addition, a study that used liver microsomes isolated from the panel of mouse strains demonstrated that genetic variation in *Cyp2b9* and *Ugt1a* loci played a role in the oxidative metabolism of α-hydroxytestosterone and glucuronidation of irinotecan, respectively [34]. Thus, it should be emphasized that exposure levels should always be assessed when using populations of strains for genetic biomarker identification.

Not every adverse drug effect in humans may be genotype dependent; thus, a multistrain approach may also prove useful for understanding genotype-independent toxicity responses and facilitate the identification of novel targets of therapeutic intervention that will be effective in the entire population. When liver gene-expression levels were assessed across strains exposed to acetaminophen, it was determined that the genes associated with the level of liver necrosis, independent of the genetic background, were involved in cell death pathways and form a closely linked molecular network [35]. This finding confirms a central role for cell death-inducing intracellular cell signaling in acetaminophen-induced liver toxicity [32,36].

The power of mouse-to-human translation studies that use mouse genetic tools has also been shown through the QTL analysis of pulmonary responses to the air pollutant ozone. Ozone causes highly reproducible changes in pulmonary function in humans, and significant interindividual variation in the responses have suggested that genetic background is an important determinant of susceptibility to ozone-induced toxicity [37]. Similarly, significant variation in ozone-induced pulmonary injury and inflammatory responses has been found among inbred strains of mice [38-40]. Both F_2 and backcross studies utilizing differentially responsive strains were used to discover a number of candidate QTLs for responsiveness to ozone [38,39]. These QTLs guided the selection of the candidate genes and loci for validation not only in subsequent mouse studies [41-43], but two homologs of the mouse susceptibility genes (*TNF* and *HLA-DR*) have also been associated with response to ozone in humans [44,45]. Similar QTL mapping approaches have been used to investigate many clinical phenotypes, including alcohol-related behavior [46], alcohol metabolism [47] and iron transport [48].

These studies indicate that the use of an inbred mouse strain panel may be a useful tool for understanding the mode of action of toxic agents and the identification of nodes in the

complex molecular events that may confer susceptibility to adverse events. When genotype-dependent chemical-induced toxicity in human population is identified or suspected, this approach has several benefits. First, potential genetic biomarkers may be developed to prescreen individuals prior to therapeutic drug treatment when potential adverse drug events are suspected. If the genes associated with differential susceptibility to toxicity are identified in a preclinical phase, the subsequent pharmacogenetics research may be focused on a few candidates to help overcome the challenge of small cohort size in human studies and to shorten the validation period. The data acquired with this model could therefore be influential in the analysis of individual risk to chemicals and may facilitate both drug development and human safety endeavors. At the same time, such an approach may not be fruitful in safety assessment of experimental drug candidates for which the risk in humans has yet to be determined.

Second, the genetic variation among individuals is reflected in variation in gene-expression levels [49,50], which introduces additional challenges into toxicology research on biomarkers of effect. While major research efforts are seeking genetic and genomic markers that could identify individuals susceptible to toxicity, less attention is given to the fact that genetic control of gene expression may present a challenge for finding robust population-wide expression biomarkers of toxicity responses [35]. Indeed, it is seldom appreciated in the analysis of gene-expression data that the genetic difference between individuals is by far the strongest effect on global gene expression at both basal levels and even when a considerable amount of tissue damage is present [35]. Thus, a careful evaluation of gene-expression-based biomarkers of response through multistrain experiments can avert the risk of mistakenly identifying large genotype effects in a particular strain of animals used for toxicity testing as the effects of treatment.

Translating SNP/gene associations into the mode of action paradigm

Determining which of the multitude of variants carried by an individual are responsible for a given phenotype represents a massive task, especially if the causal alleles are relatively anonymous in terms of known functional consequences. The best approaches for combining functional credibility and statistical support in the evaluation of such variants remain to be determined. GWAS tend to focus almost exclusively on statistical evidence and give lesser weight to considerations of biological plausibility, but the challenges of finding causal associations among the large number of rare variants may prompt a more careful examination of the underlying biology [51].

Toxicogenomics has been used at all stages of chemical risk assessment, and it is thought that gene-expression changes may be utilized as biomarkers of adverse effects [52]. Current approaches often attempt to classify compounds with the goals of predicting adverse responses to specific chemical classes [53], understanding the underlying biological mechanism of toxicity [54], or identifying key nodes in the toxicity pathway that may serve as effect biomarkers [55]. Extensive proprietary [56,57] and public [58] databases containing gene-expression and pathological end points derived from rodent and human tissues exposed to a variety of chemicals have been developed, thereby allowing the scientific community to mine the data for biomarkers.

Gene expression QTL (eQTL) mapping is one of the modern tools that support the evaluation of associations between transcript expression and genotype in order to find genomic locations that are likely to regulate transcript expression. The availability of gene-expression and high-density genotype data has enabled eQTL mapping in animal and human populations. These analyses have contributed significantly to our understanding of the effects that genetic polymorphisms may have on interindividual variability in normal

physiological processes, in multiple tissues, and in both animals and humans [50-61]. Furthermore, these studies have shown that genetic regulation of gene expression is a key contributor to population diversity, and is being realized not only through transcription factors and subtle variations in sequence of their response elements, but also through previously unknown mechanisms. While eQTL mapping is clearly an important new frontier in the application of 'omics technologies to biomedical research, no current approaches are available for the evaluation of the potential role of eQTLs in the response to environmental exposures or the pathogenesis of common diseases.

Early eQTL studies surveyed natural variation in crosses of model organisms such as budding yeast [62,63] and mouse [49]. In the mouse, two inbred parental stains were selected and bred in either a backcross or intercross design. All progeny mice are genotyped at a density sufficient to distinguish all recombination blocks, and microarrays were used to measure transcript expression. Previous studies reported significant numbers of eQTL (~9% of the transcripts surveyed) and demonstrated that there are genomic loci that contain more eQTL than expected by chance. These eQTL 'hotspots' are thought to regulate the expression levels of dozens of transcripts [49] and have been observed in several tissues in the mouse [49,50,59]. By examining the genes that lie beneath the eQTL hotspot, investigators can propose regulatory candidates for the transcripts with eQTL at the hotspot. eQTL studies also commonly identify both *cis*-acting eQTL [64], for which a transcript's eQTL is located near the transcript itself in the genome, and *trans*-acting eQTL, for which a transcript's eQTL is located far from the transcript. It has been hypothesized that *cis*-eQTL are caused by polymorphisms in regulatory regions close to the transcript itself, whereas *trans*-eQTL are caused by polymorphisms in distant genes that affect transcript expression, either directly or indirectly, in an allele-specific manner.

Indeed, while the study of individual genes is informative and can improve our understanding of the causes of differential toxicity in populations, a broader approach that focuses on gene networks and biological pathways may produce more interpretable results. eQTL mapping can be used to generate hypotheses regarding transcriptional regulation and can be integrated with gene coexpression data to discover gene networks or pathways that are associated with a clinical trait. For example, transcript expression data in the livers of a panel of C57BL/6J × DBA/2J F2 mice was combined with obesity data, and eQTL mapping was used to identify causal gene-expression networks [65,66]. eQTL data may also be combined with estimates of transcription factor activity to infer causal relationships between transcription factors and clusters of eQTL genes [67]. Other methods to infer causality between regulatory candidate genes under eQTL hotspots and the *trans*-regulated genes that map to the eQTL locus have also been proposed to assist in narrowing the list of candidate genes for further biological investigation [64,68].

Network-based approaches have been used in research into Type I diabetes and heart disease, and have shown the power of integrating human data with data derived from mouse models. A GWAS in a large human population proposed the receptor typrosine kinase *ERBB3* as the best candidate gene near a QTL for Type I diabetes [69]. Separate work that examined liver gene expression in a smaller cohort of human samples with and without Type I diabetes found that *ERBB3* did not have a *cis*-eQTL but that a flanking gene, *RPS26*, did. Since the disease phenotype and *RPS26* both had QTLs in the same location, this suggested the *RPS26* was a stronger candidate than *ERBB3*. The authors then used mouse liver and adipose expression data from several mouse crosses to construct causal expression networks for the *ERBB3* and *RPS26* orthologs in the mouse. They were able to show that *ERBB3* is not associated with any known Type I diabetes genes, whereas *RPS26* is associated with a network of several genes that are part of the Kyoto Encyclopedia of Genes and Genomes (KEGG) Type I diabetes pathway [61]. This type of analysis demonstrates the

power of combining human and mouse data with a network-based approach that has been proposed for use in drug discovery [70] and may prove useful in toxicology studies.

It should be noted that the accuracy of *cis*-eQTL detection has been called into question owing to the possibility of SNPs residing within the sequence queried by microarray probes [64,71,72]. Microarray probes for mice are designed based on the reference sequence of C57BL/6J. Transcripts in other strains with polymorphisms in the probe sequence will bind with lower affinity than the C57BL/6J transcript, giving the false appearance of allele-specific expression levels associated with the transcript location, which is the defining characteristic of a *cis*-eQTL. Studies in which shorter, 25-nucleotide microarray probes are used [71] appear to be more significantly affected than studies that use longer, 50- to 60-nucleotide-long probes [64,72]. This is consistent with the hypothesis that a SNP within a probe sequence will affect shorter probes more strongly than longer probes. The validity of eQTL hotspots has also been questioned [73,74] owing to the possibility that sets of highly correlated genes will naturally cluster over the same genomic marker. Furthermore, sets of highly correlated genes are likely to be part of the same gene-ontology category, and so when geneontology category enrichment is conducted on eQTL hotspots, they are likely to (falsely) appear biologically coherent. A permutation strategy, in which the sample labels are permuted and the expression labels are held fixed, has been suggested to address this problem [74,75]. False eQTL hotspots may also arise owing to intersample correlation, and a mixed-model approach has been shown to eliminate spurious eQTL hotspots in mouse data [11].

Another matter of concern in eQTL mapping studies is how to control for the massive multiple testing involved. There are two levels of multiple testing carried out in an eQTL study; multiple testing across correlated SNPs and across multiple correlated transcripts. Multiple testing across SNPs may be addressed by permuting the sample labels in the genotype data while holding the expression data fixed [75]. Multiple testing across genes may be addressed using approaches based on the false-discovery rate [76].

Conclusion & future perspective

Recognition of the challenges of currently available laboratory animal-based genetics resources led to the realization that a new general-purpose mouse population was needed to model complex human diseases, with particular emphasis on traits relevant to human health in its broadest aspects. Open discussion among members of the genetics community resulted in the conception and design of the 'Collaborative Cross' (CC) [77,78]. Establishment of this new mouse-based resource will considerably expedite gene discovery and characterization and serve as a powerful complement to ongoing studies in human genetics.

The CC provides a translational platform for systems genetics that integrates classical genetics and systems biology tools to identify genetic networks that underlie complex phenotypes. A pre-requisite for systems genetics is a realistic experimental population structure, which is essential to unravel the complex biological processes that may differ from one individual to another, such as cancer susceptibility or the response to an environmental exposure. The CC population was created through a community effort by the Complex Trait Community [101]. This resource will obviate the need for researchers to produce ephemeral backcross or F2 populations; it will need to be genotyped only once; it will archive thousands of recombination events and millions of genetic polymorphisms; and it will facilitate international and intergenerational comparisons of genetic effects.

The CC satisfies four essential criteria for an optimal experimental platform that can support systems genetic studies [79]:

- Genome-wide genetic variation;

- Randomization of the genetic variation;

- It is infinitely reproducible;

- It is sufficiently large to support statistical analysis of the data.

The overall design of CC consists of eight founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HiLtJ, CAST/Ei, PWK/PhJ and WSB/EiJ) bred through an eight-way 'funnel' breeding design established to randomly mix the variation present in the founder strains before inbreeding by brother–sister mating. The founder strains were selected from a set of over 100 strains in order to maximize genetic diversity and utility for studying traits of widespread interest. The eight CC founder strains capture approximately 90% of the known allelic diversity across all 1-Mb intervals spanning the entire mouse genome; compare this with AXB/BXA and BXD, the two most commonly used mouse recombinant inbred panels, which each capture only approximately 13% of the known variation, with much of it overlapping [27]. Furthermore, the population of emergent CC lines has a much more random distribution of genetic variation than existing panels of inbred strains such as the mouse phenome panel [80,81], which has considerable genetic linkage across chromosomes that result in high rates of false-positive associations. Since the CC strains have a population structure that randomizes existing genetic variation, this resource will provide unparalleled power to assign causality to understand the intricacies of biological networks underlying disease and toxicant response. The types, distribution and frequency of genetic polymorphisms are close to those in human populations, and the fraction of genetic diversity captured in CC lines is far superior compared with other commonly used mouse populations. Importantly, preliminary phenotypic characterization of pre-CC strains indicates that a very large variability exists within the CC population following changes in environmental conditions (e.g., diet and exercise). The recombination, inbreeding rates and statistical power of this novel cross has been examined by others and found to be optimal for systems genetics applications [82,83].

Executive summary

- Genome-wide association studies in human cohorts, large and small, provide important information on the genetic causes of complex diseases, individual susceptibility to chemical exposures and other environmental factors. Still, the validation of the 'candidate' genes and/or polymorphisms is a daunting task that is compounded by the statistical challenges in defining the candidates.

- Additional resources for gene/polymorphism discovery are available through *in vitro* experiments with densely genotyped human cell lines assembled into cohorts or familial structures, or *in vivo* approaches capitalizing on the knowledge of mouse genetics and the availability of densely genotyped reproducible panels of strains.

- Several proof-of-concept studies that utilized the 'mouse model of the human population' to advance translational clinical research have been published recently. This work points to the promise of the combination approach whereby human disease or toxicity can be modeled in animals and lead to the discovery of candidates for validation in human cohorts.

- Translation of gene/polymorphism associations between species also provides the opportunity to better understand which nodes in the complex mode of toxicity/disease pathogenesis pathways are most likely to be causes of genetic diversity in responses. Such knowledge will empower the practical use of the genetic

> information for individualized therapy and/or assessment of risk from environmental exposures.
>
> ▪ The 'Collaborative Cross' is a new mouse-based resource that will considerably expedite gene discovery and characterization, and serve as a powerful complement to ongoing studies in human genetics.

Acknowledgments

Bibliography

Papers of special note have been highlighted as:

▪ of interest

▪▪ of considerable interest

1. Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420(6915):520–562. [PubMed: 12466850]

2. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001;291(5507):1304–1351. [PubMed: 11181995]

3. Koster ES, Rodin AS, Raaijmakers JA, Maitland-van der Zee AH. Systems biology in pharmacogenomic research: the way to personalized prescribing? Pharmacogenomics 2009;10(6):971–981. [PubMed: 19530964]

4. Hardy J, Singleton A. Genomewide association studies and human disease. N. Engl. J. Med 2009;360(17):1759–1768. [PubMed: 19369657]

5. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl Acad. Sci. USA 2009;106(23):9362–9367. [PubMed: 19474294] ▪ Online catalog of SNP-trait associations from published genome-wide association studies.

6. Hirschhorn JN. Genomewide association studies – illuminating biologic pathways. N. Engl. J. Med 2009;360(17):1699–1701. [PubMed: 19369661]

7. Andrade RJ, Robles M, Ulzurrun E, Lucena MI. Drug-induced liver injury: insights from genetic studies. Pharmacogenomics 2009;10(9):1467–1487. [PubMed: 19761370]

8. Mohlke KL, Boehnke M, Abecasis GR. Metabolic and cardiovascular traits: an abundance of recently identified common genetic variants. Hum. Mol. Genet 2008;17(R2):R102–R108. [PubMed: 18852197]

9. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. N. Engl. J. Med 2008;358(22):2355–2365. [PubMed: 18445777]

10. Kindmark A, Jawaid A, Harbron CG, et al. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. Pharmacogenomics J 2008;8(3):186–195. [PubMed: 17505501]

11. Kang HM, Ye C, Eskin E. Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots. Genetics 2008;180(4):1909–1925. [PubMed: 18791227]

12. Mallal S, Phillips E, Carosi G, et al. *HLA-B*5701* screening for hypersensitivity to abacavir. N. Engl. J. Med 2008;358(6):568–579. [PubMed: 18256392]

13. Becquemont L. Pharmacogenomics of adverse drug reactions: practical applications and perspectives. Pharmacogenomics 2009;10(6):961–969. [PubMed: 19530963]

14. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461(7265):747–753. [PubMed: 19812666]

15. Welsh M, Mangravite L, Medina MW, et al. Pharmacogenomic discovery using cell-based models. Pharmacol. Rev 2009;61(4):413–429. [PubMed: 20038569]

16. Aitman TJ, Critser JK, Cuppen E, et al. Progress and prospects in rat genetics: a community view. Nat. Genet 2008;40(5):516–522. [PubMed: 18443588]

17. Nye SH, Wittenburg AL, Evans DL, O'Connor JA, Roman RJ, Jacob HJ. Rat survival to anthrax lethal toxin is likely controlled by a single gene. Pharmacogenomics J 2008;8(1):16–22. [PubMed: 17440430]

18. Frazer KA, Eskin E, Kang HM, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 2007;448(7157):1050–1053. [PubMed: 17660834]

19. Szatkiewicz JP, Beane GL, Ding Y, Hutchins L, de Villena F Pardo-Manuel, Churchill GA. An imputed genotype resource for the laboratory mouse. Mamm. Genome 2008;19(3):199–208. [PubMed: 18301946]

20. Rao GN, Lindamood C III, Heath JE, Farnell DR, Giles HD. Subchronic toxicity of human immunodeficiency virus and tuberculosis combination therapies in B6C3F1 mice. Toxicol. Sci 1998;45(1):113–127. [PubMed: 9848118]

21. Morgan DL, Mahler JF, Kirkpatrick DT, et al. Characterization of inhaled α-methylstyrene vapor toxicity for B6C3F1 mice and F344 rats. Toxicol. Sci 1999;47(2):187–194. [PubMed: 10220856]

22. Constan AA, Wong BA, Everitt JI, Butterworth BE. Chloroform inhalation exposure conditions necessary to initiate liver toxicity in female B6C3F1 mice. Toxicol. Sci 2002;66(2):201–208. [PubMed: 11896286]

23. Chorley BN, Wang X, Campbell MR, Pittman GS, Noureddine MA, Bell DA. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. Mutat. Res 2008;659(1–2):147–157. [PubMed: 18565787]

24. Lowes BD, Buttrick PM. Genetic determinants of drug response in heart failure. Curr. Cardiol. Rep 2008;10(3):176–181. [PubMed: 18489860]

25. Festing MF. Experimental approaches to the determination of genetic variability. Toxicol. Lett 2001;120(1–3):293–300. [PubMed: 11323187] ▪ One of the original papers arguing for the use of genetics in toxicity testing.

26. Ideraabdullah FY, Casa-Esperon E, Bell TA, et al. Genetic and haplotype diversity among wild-derived mouse inbred strains. Genome Res 2004;14(10A):1880–1887. [PubMed: 15466288]

27. Roberts A, de Villena F Pardo-Manuel, Wang W, McMillan L, Threadgill DW. The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. Mamm. Genome 2007;18(6–7):473–481. [PubMed: 17674098]

28. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 1989;121(1):185–199. [PubMed: 2563713]

29. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 1992;69(4):315–324. [PubMed: 16718932]

30. Harrill AH, Watkins PB, Su S, et al. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. Genome Res 2009;19:1507–1515. [PubMed: 19416960]

31. Nelson MR, Bacanu SA, Mosteller M, et al. Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. Pharmacogenomics J 2008;9(1):23–33. [PubMed: 18301416]

32. Kaplowitz N. Idiosyncratic drug hepatotoxicity. Nat. Rev. Drug Discov 2005;4(6):489–499. [PubMed: 15931258]

33. Guo Y, Weller P, Farrell E, et al. *In silico* pharmacogenetics of warfarin metabolism. Nat. Biotechnol 2006;24(5):531–536. [PubMed: 16680137]

34. Guo Y, Lu P, Farrell E, et al. *In silico* and *in vitro* pharmacogenetic analysis in mice. Proc. Natl Acad. Sci. USA 2007;104(45):17735–17740. [PubMed: 17978195]

35. Harrill AH, Ross PK, Gatti DM, Threadgill DW, Rusyn I. Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. Toxicol. Sci 2009;110(1):235–243. [PubMed: 19420014]

36. Jaeschke H, Bajt ML. Intracellular signaling mechanisms of acetaminophen-induced liver cell death. Toxicol. Sci 2006;89(1):31–41. [PubMed: 16177235]

37. Weinmann GG, Bowes SM, Gerbase MW, Kimball AW, Frank R. Response to acute ozone exposure in healthy men. Results of a screening procedure. Am. J. Respir. Crit. Care Med 1995;151(1):33–40. [PubMed: 7812569]

38. Prows DR, Shertzer HG, Daly MJ, Sidman CL, Leikauf GD. Genetic analysis of ozone-induced acute lung injury in sensitive and resistant strains of mice. Nat. Genet 1997;17(4):471–474. [PubMed: 9398853]

39. Kleeberger SR, Levitt RC, Zhang LY, et al. Linkage analysis of susceptibility to ozone-induced lung inflammation in inbred mice. Nat. Genet 1997;17(4):475–478. [PubMed: 9398854]

40. Savov JD, Whitehead GS, Wang J, et al. Ozone-induced acute pulmonary injury in inbred mouse strains. Am. J. Respir. Cell Mol. Biol 2004;31(1):69–77. [PubMed: 14975936]

41. Prows DR, Hafertepen AP, Winterberg AV, et al. Reciprocal congenic lines of mice capture the aliq1 effect on acute lung injury survival time. Am. J. Respir. Cell Mol. Biol 2008;38(1):68–77. [PubMed: 17656683]

42. Cho HY, Morgan DL, Bauer AK, Kleeberger SR. Signal transduction pathways of tumor necrosis factor – mediated lung injury induced by ozone in mice. Am. J. Respir. Crit. Care Med 2007;175(8):829–839. [PubMed: 17255564]

43. Bauer AK, Travis EL, Malhotra SS, et al. Identification of novel susceptibility genes in ozone-induced inflammation in mice. Eur. Respir. J. 2009 DOI: 10.1183/09031936.00145309. (Epub ahead of print).

44. Lay JC, Alexis NE, Kleeberger SR, et al. Ozone enhances markers of innate immunity and antigen presentation on airway monocytes in healthy individuals. J. Allergy Clin. Immunol 2007;120(3): 719–722. [PubMed: 17586033]

45. Li YF, Gauderman WJ, Avol E, Dubeau L, Gilliland FD. Associations of tumor necrosis factor G-308A with childhood asthma and wheezing. Am. J. Respir. Crit. Care Med 2006;173(9):970–976. [PubMed: 16456144]

46. Crabbe JC. Alcohol and genetics: new models. Am. J. Med. Genet 2002;114(8):969–974. [PubMed: 12457395]

47. Grisel JE, Metten P, Wenger CD, Merrill CM, Crabbe JC. Mapping of quantitative trait loci underlying ethanol metabolism in BXD recombinant inbred mouse strains. Alcohol Clin. Exp. Res 2002;26(5):610–616. [PubMed: 12045468]

48. Wang F, Paradkar PN, Custodio AO, et al. Genetic variation in MON1A affects protein trafficking and modifies macrophage iron loading in mice. Nat. Genet 2007;39(8):1025–1032. [PubMed: 17632513]

49. Schadt EE, Monks SA, Drake TA, et al. Genetics of gene expression surveyed in maize, mouse and man. Nature 2003;422(6929):297–302. [PubMed: 12646919] ▪▪ Landmark study demonstrating the power of expression quantitative trait locus (eQTL) mapping.

50. Chesler EJ, Lu L, Shou S, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat. Genet 2005;37(3):233–242. [PubMed: 15711545]

51. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature 2009;461(7261):218–223. [PubMed: 19741703] ▪ Comprehensive review of progress in QTL and eQTL mapping of complex traits.

52. Casciano DA, Woodcock J. Empowering microarrays in the regulatory setting. Nat. Biotechnol 2006;24(9):1103. [PubMed: 16964221]

53. Fostel JM. Future of toxicogenomics and safety signatures: balancing public access to data with proprietary drug discovery. Pharmacogenomics 2007;8(5):425–430. [PubMed: 17465705]

54. Dix DJ, Gallagher K, Benson WH, et al. A framework for the use of genomics data at the EPA. Nat. Biotechnol 2006;24(9):1108–1111. [PubMed: 16964223]

55. Fry RC, Navasumrit P, Valiathan C, et al. Activation of inflammation/NF-κB signaling in infants born to arsenic-exposed mothers. PLoS Genet 2007;3(11):E207. [PubMed: 18039032]

56. Castle AL, Carver MP, Mendrick DL. Toxicogenomics: a new revolution in drug safety. Drug Discov. Today 2002;7(13):728–736. [PubMed: 12110229]

57. Ganter B, Zidek N, Hewitt PR, Muller D, Vladimirova A. Pathway analysis tools and toxicogenomics reference databases for risk assessment. Pharmacogenomics 2008;9(1):35–54. [PubMed: 18154447]

58. Waters M, Stasiewicz S, Merrick BA, et al. CEBS – Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res 2008;36(Database issue):D892–D900. [PubMed: 17962311]

59. Bystrykh L, Weersing E, Dontje B, et al. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat. Genet 2005;37(3):225–232. [PubMed: 15711547]

60. Gatti D, Maki A, Chesler EJ, et al. Genome-level analysis of genetic regulation of liver gene expression networks. Hepatology 2007;46(2):548–557. [PubMed: 17542012]

61. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol 2008;6(5):E107. [PubMed: 18462017]

62. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science 2002;296(5568):752–755. [PubMed: 11923494]

63. Yvert G, Brem RB, Whittle J, et al. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat. Genet 2003;35(1):57–64. [PubMed: 12897782]

64. Doss S, Schadt EE, Drake TA, Lusis AJ. *Cis*-acting expression quantitative trait loci in mice. Genome Res 2005;15(5):681–691. [PubMed: 15837804]

65. Zhu J, Lum PY, Lamb J, et al. An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet. Genome Res 2004;105(2–4):363–374. [PubMed: 15237224]

66. Drake TA, Schadt EE, Davis RC, Lusis AJ. Integrating genetic and gene expression data to study the metabolic syndrome and diabetes in mice. Am. J. Ther 2005;12(6):503–511. [PubMed: 16280644]

67. Sun W, Yu T, Li KC. Detection of eQTL modules mediated by activity levels of transcription factors. Bioinformatics 2007;23(17):2290–2297. [PubMed: 17599927]

68. Peng J, Wang P, Tang H. Controlling for false positive findings of *trans*-hubs in expression quantitative trait loci mapping. BMC Proc 2007;1(Suppl. 1):S157. [PubMed: 18466502]

69. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–678. [PubMed: 17554300]

70. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. Nat. Rev. Drug Discov 2009;8(4):286–295. [PubMed: 19337271]

71. Alberts R, Terpstra P, Li Y, Breitling R, Nap JP, Jansen RC. Sequence polymorphisms cause many false *cis* eQTLs. PLoS One 2007;2(7):E622. [PubMed: 17637838]

72. Ciobanu DC, Lu L, Mozhui K, et al. Detection, validation, and downstream analysis of allelic variation in gene expression. Genetics 2010;184(1):119–128. [PubMed: 19884314]

73. de Koning DJ, Haley CS. Genetical genomics in humans and model organisms. Trends Genet 2005;21(7):377–381. [PubMed: 15908034]

74. Breitling R, Li Y, Tesson BM, et al. Genetical genomics: spotlight on QTL hotspots. PLoS Genet 2008;4(10):E1000232. [PubMed: 18949031]

75. Churchill GA, Doerge RW. Naive application of permutation testing leads to inflated type I error rates. Genetics 2008;178(1):609–610. [PubMed: 18202402]

76. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 2003;100(16):9440–9445. [PubMed: 12883005]

77. Churchill GA, Airey DC, Allayee H, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet 2004;36(11):1133–1137. [PubMed: 15514660]

78. Threadgill DW, Hunter KW, Williams RW. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm. Genome 2002;13(4):175–178. [PubMed: 11956758] ▪ ▪ Roadmap for the use of mouse population-based testing in biomedical research.

79. Chesler EJ, Miller DR, Branstetter LR, et al. The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm. Genome 2008;19(6): 382–389. [PubMed: 18716833]

80. Petkov PM, Ding Y, Cassell MA, et al. An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res 2004;14(9):1806–1811. [PubMed: 15342563]

81. Grubb SC, Maddatu TP, Bult CJ, Bogue MA. Mouse phenome database. Nucleic Acids Res 2009;37(Database issue):D720–D730. [PubMed: 18987003] ▪ Online resource for genotype–phenotype interactions in mouse studies.

82. Broman KW. The genomes of recombinant inbred lines. Genetics 2005;169(2):1133–1146. [PubMed: 15545647]

83. Valdar W, Flint J, Mott R. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics 2006;172(3):1783–1797. [PubMed: 16361245]

▪ **Website**

101. Complex Trait Community. www.complextrait.org